

Formalization of the Paper Registration Process to Institutional Repositories

Sakaguchi, Koji
Hitotsubashi University

Nakayama, Satoshi
University of Tsukuba

Baba, Kensuke
Research and Development Division, Kyushu University Library

<https://hdl.handle.net/2324/25179>

出版情報 : Proceedings of 2012 IIAI International Conference on Advanced Applied Informatics,
pp.14-18, 2012-09. IEEE

バージョン :

権利関係 :

Formalization of the Paper Registration Process to Institutional Repositories*

Koji Sakaguchi

Satoshi Nakayama

Kensuke Baba[†]

Abstract

An institutional repository is a system for archiving research outputs such as scholarly papers. In order to increase the number of archived papers in institutional repositories, the registration process controlled by repository managers should be improved. The authors of this paper conducted an investigation on the efforts involved in paper registration with institutional repositories in several universities. The paper shows the result as a formalized workflow of the process. The investigation showed that one of the hardest tasks for repository managers is confirmation of the copyright policy of publishers. The authors developed a system that automatically searches the copyright policy on external databases.

Keywords: archiving; institutional repository; copyright policy of publishers.

1 Introduction

Scholarly communication has been described as comprising four essential components: registration, certification, awareness, and archiving [17]. The institutional repository (IR) is a system for archiving scholarly papers produced by researchers. In Japan, IRs have become very popular; there are 170 IRs as of May 2012 [5]. Most IRs are developed by standard searching and harvesting protocols such as OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).

The aim of our research is to increase the number of archived papers in IRs. Especially, archiving (the authors' version of) journal articles is expected to indirectly solve the subscription problem of some electronic journals. In all Japanese IRs, as of May 2012, 42.1% (569,549/1,351,410) of the total content consists of departmental bulletin papers, and 22.0% (297,541/1,351,410) consists of journal articles [4].

The low percentage of archived journal articles seems to be caused by the great efforts required of

authors for paper registration. According to a study about several well-known IR by Xia and Sun [20], most papers were registered by repository managers (librarians) rather than by authors. In order to encourage authors to register their papers, we developed a system that reduces the efforts of authors for paper registration with the IR at Kyushu University (QIR) [6] by connecting with the researcher database of the university [13]. However, the problem of the confirmation process of the copyright policy of publishers still remains. In most institutions, the process is conducted manually by repository managers. It is very difficult and complicated for repository managers to check copyright policy. Therefore, this process can be a bottleneck of paper registration with an IR.

In order to clarify the efforts of repository managers for paper registration, we made an investigation in several universities in Japan. Repository managers at 20 universities were interviewed and a questionnaire survey was conducted in 168 institutions. This kind of comprehensive investigation into the efforts of repository managers for paper registration with IR was conducted for the first time, at least in Japan. In this paper, we show the results of the investigation and formalize the processes of paper registration as a workflow.

As a result of the investigation, we confirmed that the process of copyright policy confirmation is one of the hardest tasks for repository managers. Additionally, according to the results, we developed a system to confirm the copyright policy of publishers using copyright policy databases [9, 10]. The system is part of a combined system for paper registration with IR that includes another system to offer researchers paper registration by e-mail [12]. Fig. 1 shows the overview of the combined system. The automated system was realized by the formalization of workflow, and hence the system is applicable to general IRs.

The rest of this paper is constructed as follows. Section 2 describes the current situation of the IRs in Kyushu University and Hitotsubashi University in order to clarify the problem. Section 3 shows the results of our investigation concerning the efforts of repository managers for paper registration with IR. Section 4 introduces the copyright policy confirmation system developed based on the results of the investigation.

*An edited version of this report was published in: *Proc. 2012 IIAI International Conference on Advanced Applied Informatics*, pp. 14–18, IEEE, Sep, 2012.

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

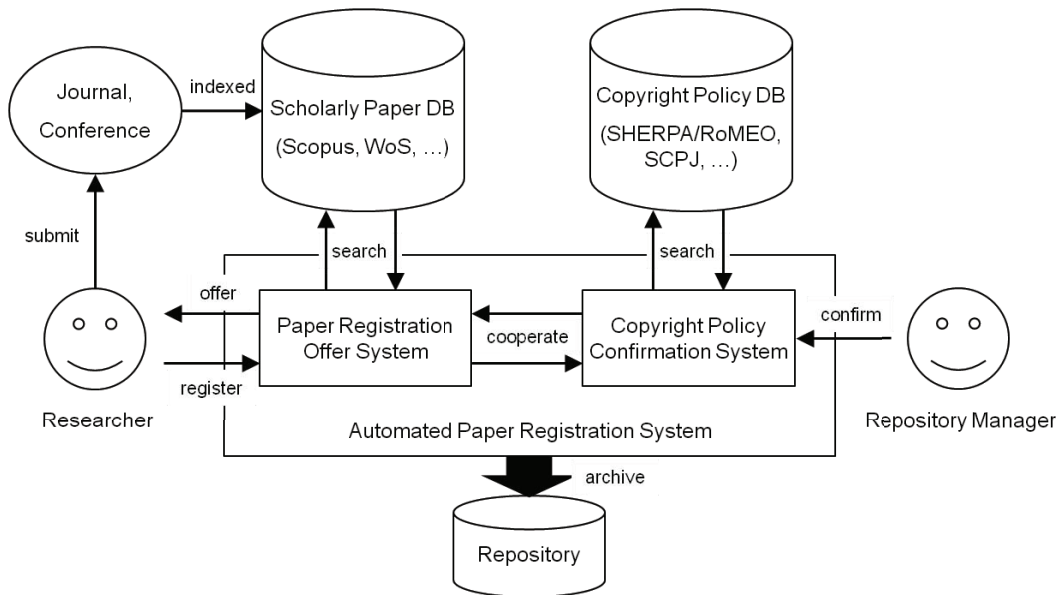


Figure 1: Overview of the automated paper registration system.

2 Current Situation

The IR of Kyushu University (QIR) [6] is based on Dspace [2]. The number of the archived papers (full-text) in QIR is about 19,000 as of April 2012. The number of the metadata of the papers registered in the researcher database of Kyushu University (DHJS) [1] is about 90,000, and the number of distinct papers is at least 72,000 even if we consider duplication by co-authors [14]. This implies that a large number of papers produced in Kyushu University are not archived in the IR. Most institutions are estimated to be in the same situation. According to Ranking Web of World Repositories [7], the number of papers in QIR ranks 68th among about 2,000 institutions as of April 2012. An institution ranked under 68 should have many non-archived papers if the number of produced papers is not extremely small compared with Kyushu University. Harnad et al. [18] pointed out that only about 15% of articles are being spontaneously self-archived. Björk et al. [16] reported that the rate of deposited papers in IRs for about 1,350,000 peer-reviewed journal articles published in 2006 is estimated to be about 5%.

We considered that an effective means to increase the number of archived papers would be to encourage researchers to register the buried papers with IR. As a previous work, we developed a system which connects QIR with DHJS [13]. The system links the metadata of a paper in DHJS to the full text in QIR and leads the author to the registration form if the corresponding full text is not archived in QIR. However, as the result of verification of the link system, the number of paper registrations by the system from June

2008 to October 2010 was only 8% of the total number 7,411 [15].

The main idea of this work is to encourage researchers more directly. The IR of Hitotsubashi University (HERMES-IR) [3] is also based on Dspace [2]. The number of archived papers (full text) in HERMES-IR is about 13,000 as of April 2012. We requested the researchers in Hitotsubashi University to register their papers in the researcher database of Hitotsubashi University to HERMES-IR by e-mail in 2011. The number of the actual paper registrations was 70% of our requests. Therefore, we found that registration requests to researchers are effective.

On the basis of the previous idea, we developed a system that can offer paper registration to authors by e-mail based on the results of a search of external databases [12]. There exist some databases of published scholarly papers that are widely used on the internet and produce API [8, 11]. While paper registration with an IR is currently done by researchers voluntarily in Kyushu University, by searching published papers written by researchers in our own institution, we can encourage researchers to register their papers with the IR. Actually, in Scopus [8] the number of articles written by researchers in Kyushu University (papers for which the affiliation of at least one of the authors is Kyushu University) was about 4,000 in 2009. Since the number of articles in QIR is about 400 per year, a significant effect is expected if the system works well.

3 Efforts of the Repository Manager

In order to increase the number of archived papers, the throughput of the efforts for paper registration by repository managers should be improved in addition to increasing the number of permissions for paper registration from researchers. In the current process of paper registration with QIR and HERMES-IR, repository managers have to confirm the reply status of permission from the authors and the copyright policy of the publisher. This process may become a bottleneck if requests for permission to researchers increase. This type of situation is concerning especially for small institutions that cannot spare enough workers for IR. Therefore, it is necessary to reconsider the paper registration process especially from the viewpoint of repository managers.

We conducted an investigation into the tasks of repository managers in several universities and formalized the process. We distributed a questionnaire about the management of information and status in the paper registration process to 168 universities in Japan and received answers from 133 universities (79%). We also conducted interview surveys with 20 universities.

From the results of the questionnaire, we found that confirmation of the copyright policy is one of the hardest processes for repository managers. Kim [19] reported that one of the obstacles to self-archiving for researchers is concern about the copyright. The results imply that the matter of copyright is also a problem for repository managers. Some publishers and academic societies that publish scholarly papers have copyright policies concerning whether a paper can be deposited in an IR. Usually, 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. Additionally, some publishers ask for grace to open the paper in an IR (a so-called “embargo”). At most universities, first, the confirmation process is conducted by searching external databases such as SCPJ [9] and SHERPA/RoMEO [10] for copyright policy. Some publishers’ policies cannot be found in the databases. In such cases, repository managers have to make inquiries with the publishers. Making such inquiries and managing the status of the inquiries for a number of papers occupy the greater part of the work of repository managers.

Additionally, we investigated the following:

- The information that should be managed for paper registration with IR for each paper,
- The status in the paper registration process with

IR.

As a result, we obtained (a) a list of the information that should be managed and (b) a list of statuses in the paper registration process.

(a) Information

- Bibliographic information
- Address of the contact author
- How to request the author’s permission to open the paper
- Answer from the author to a request of permission to open the paper
- Copyright policy of the publisher
- (Author/publisher version, terms of embargo, and so on)
- Address of the publisher
- How to request the publisher’s permission to open the paper
- Answer from the publisher to a request of permission to open the paper
- Remarks

(b) Status

- Not yet dealt
- Confirming the copyright policy of the publisher
- Searching for the address of the contact author
- Searching for the address of the publisher
- Asking for permission from the author to open the paper
- Asking for permission from the publisher to open the paper
- Waiting for submission of the full text
- Waiting for digitization of the full text
- Opened
- Quit
- Remarks

According to the results of the investigation, we formalized the processes of paper registration with IR as the workflow shown in Fig. 2. The flow of information and status were described as dotted and solid lines, respectively.

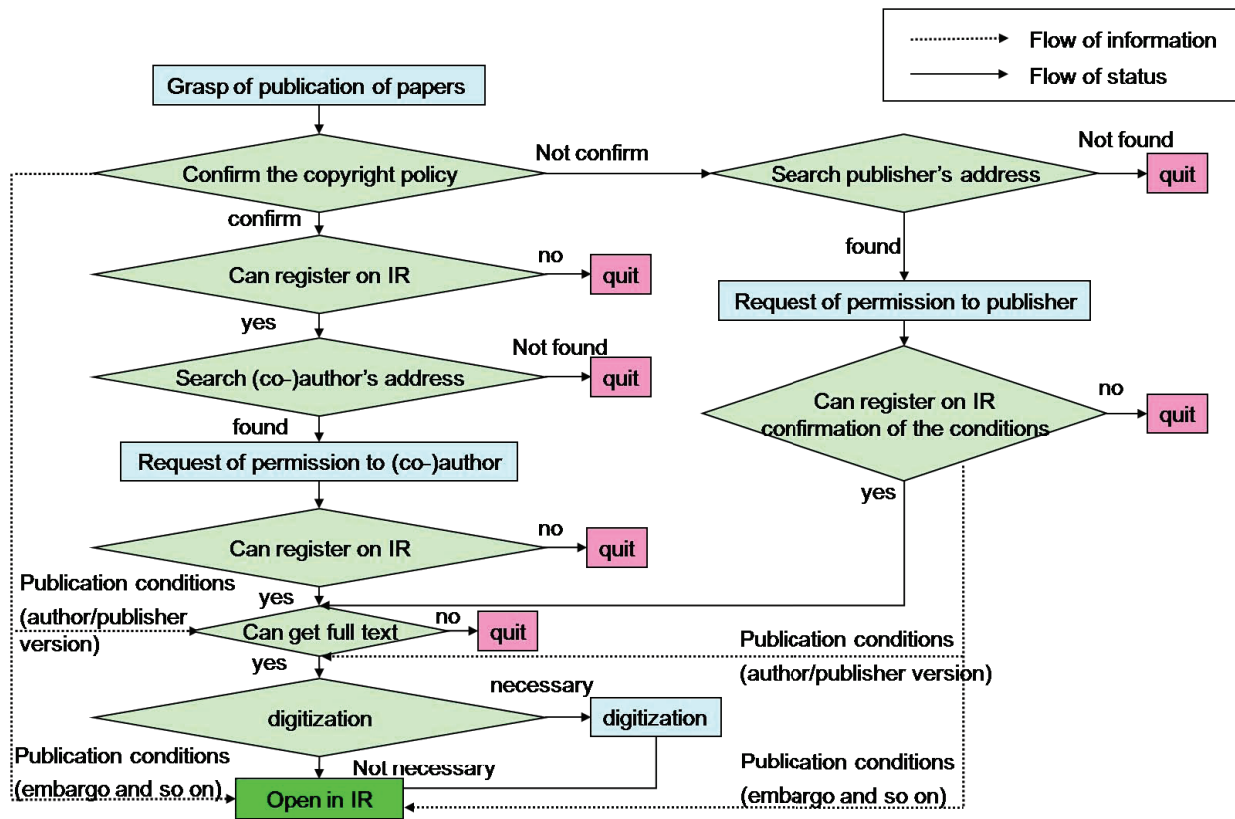


Figure 2: The workflow of the processes of paper registration with IR. The solid line and dotted line express the flows of status and information, respectively.

4 Copyright Policy Confirmation System

As mentioned in the previous section, one of the hardest processes of repository managers for paper registration is confirmation of the copyright policy of publishers, and this confirmation is done by the following two processes:

- Searching external databases of copyright policy,
- Making inquiries with publishers.

Currently, both processes of confirmation are conducted by repository managers, and the efficiency of this process depends on case-by-case improvements based on individual experience.

We developed a system that

- Manages the status of the registration process for each paper,
- Operates the first process of the copyright policy confirmation automatically.

These functions were implemented on the basis of the formalization in Section 3. Fig. 3 is an example of the interface of the developed system. In the interface, the

column “Status” was defined according to the formalization, and each value corresponds to the status in the workflow in Section 3. The status “Unsent” corresponds to the node “Search (co-)author’s address” in the workflow, and means that the address has been retrieved but no request of permission has been sent to the address yet.

The column “Color” corresponds to the policy of the publisher of each paper. For each paper stacked in the system, the copyright policy is confirmed by searching external databases automatically (the trigger is made by a repository manager). The policy can be categorized into four types on the basis of the versions of a paper. The categories of the policy are generally expressed by colors as follows [10]:

- Green: IR can open pre-print and post-print versions (or publisher version),
 - Blue: IR can open the post-print version (or publisher version),
 - Yellow: IR can open the pre-print version,
 - White: IR cannot open any version,
- and we also use another category [9]:
- Gray: the policy is not clear.

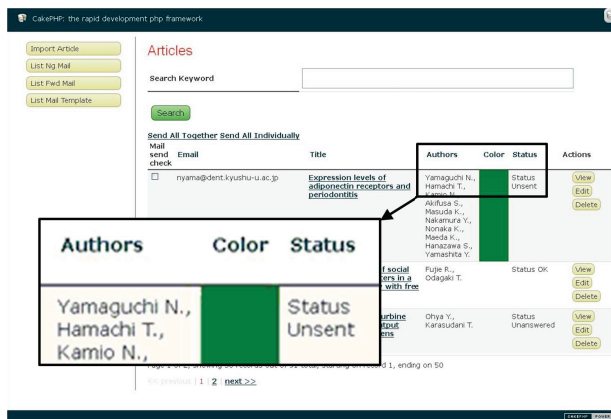


Figure 3: An example of the interface of the developed system.

In the system, the result of a search of copyright policy is shown as one of the colors. Repository managers have to change their approach to paper registration according to the color. For some publishers, they have to manage the embargo. Additionally, for some publishers, the copyright policy cannot be found and the repository manager must thus make inquiries with the publisher as a second process.

5 Conclusion

We investigated efforts for paper registration with IRs in several universities from the viewpoint of repository managers. The information they should manage and the status of the process were clarified, and thereby the workflow of paper registration was formalized. Additionally, we found that one of the most difficult tasks of repository managers for paper registration is confirmation of the copyright policy of publishers. Therefore, we developed a system to confirm the policy automatically by utilizing a number of external databases for the policy.

The developed system is applicable to general IRs by the formalization. Especially for small institutions that cannot spare enough workers for the IR, it is expected that the system can help them operate their own IRs. The developed system also can manage the status of the paper registration process, which is helpful for repository managers because it reduces their workloads.

We will install this system in several universities on a trial basis and examine the change in the number of paper registrations with IRs in order to evaluate the developed system. Additionally, we will improve the system on the basis of the feedback obtained.

Acknowledgments

This work was partially supported by the Institutional Repositories Program of the National Institute of Informatics (NII) from 2010 to 2012.

References

- [1] DHJS: Kyushu University Academic Staff Educational and Research Activities Database. Retrieved April 30, 2012, from <http://hyoka.ofc.kyushu-u.ac.jp/search/index.e.html>.
- [2] DSpace. Retrieved April 30, 2012, from <http://www.dspace.org/>.
- [3] HERMES-IR: Hitotsubashi University Repository. Retrieved June 25, 2012, from <http://hermes-ir.lib.hit-u.ac.jp/ir/index-e.html>.
- [4] JAIRO: Japanese Institutional Repositories Online. Retrieved May 9, 2012, from <http://jairo.nii.ac.jp/en/>.
- [5] NII Institutional Repositories Program. Retrieved April 30, 2012, from <http://www.nii.ac.jp/irp/en/list/>.
- [6] QIR: Kyushu University Institutional Repository. Retrieved April 30, 2012, from <https://qir.kyushu-u.ac.jp/dspace/>.
- [7] Ranking Web of World Repositories. Retrieved April 30, 2012, from <http://repositories.webometrics.info/>.
- [8] Scopus. Elsevier. Retrieved April 30, 2012, from <http://www.scopus.com/>.
- [9] SCPJ. Retrieved April 30, 2012, from <http://scpj.tulips.tsukuba.ac.jp/>.
- [10] SHERPA/RoMEO. Retrieved April 30, 2012, from <http://www.sherpa.ac.uk/romeo/>.
- [11] Web of Science. Thomson Reuters. Retrieved April 30, 2012, from <http://science.thomsonreuters.jp/products/wos/>.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] B.-C. Björk, A. Roos, and M. Lauri. Scientific journal publishing: yearly volume and open access availability. *Information Research*, 14(1):391, 2009.
- [17] R. Crow. The case for institutional repositories: A SPARC position paper. Discussion Paper, 2006. Scholarly Publication and Academic Resources Coalition, Washington, D.C.
- [18] 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.
- [19] J. Kim. Faculty self-archiving: Motivations and barriers. *Journal of the American Society for Information Science and Technology*, 61(9):1909–1922, 2010.
- [20] J. Xia and L. Sun. Assessment of self-archiving in institutional repositories: Depositorship and full-text availability. *Serials Review*, 33(1):14–21, 2007.