



Date : 11/05/2006

“dCollection”; The National Network of University-Based Repository in Korea.

Park, Hong Seok
 KERIS¹
honspark@keris.or.kr

Jang, Keum Yeon
 KERIS
jkylib@keris.or.kr

Meeting:	157 Asia and Oceania
Simultaneous Interpretation:	-

WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL
 20-24 August 2006, Seoul, Korea
<http://www.ifla.org/IV/ifla72/index.htm>

Abstract

KERIS(Korea Education & Research Information Service) is erating a nation-wide system, named ‘dCollection’. The system has been developed from 2003. This system has DP (Data Provider) system and SP (Service Provider) system. DP system is to collect metadata and digital documents like thesis and journal articles from academic communities in the university. SP system is to periodically harvest the metadata from DP system and integrate these for a nation-wide service. The main functions of DP system in the university are to collect metadata and digital documents, manage the quality of data collected, expose the metadata for harvesting by SP system, and to manage copyright. The main functions of dCollection SP system in KERIS are to harvest the metadata from DP systems, manage log records, schedule and workflow of DP systems, and to integrate data from DP systems. Forty universities are using this system. To expand to universities of a small scale, a web-based dCollection hosting system will be developed within this year. And dCollection system will be connected with major open archives in other countries. In near future, dCollection will be growing to the core national research information system and to an important center for international flow of research information.

1. Introduction

Recently the open access to information, especially research information, is getting more attention. This awareness of open access would be related to recent trends in research

¹ KERIS (Korea Education & Research Information Service), a public organization under The Ministry of Education & Human Resource Development, founded for informatization of the university library and primary and secondary education.

information market; dominance by a few major online DB publishers, sharply increasing prices of research journals, limited library budget, increased awareness of open access to information by authors, and others. Open access to research information makes it easy to share research results and, as a result, to increase the visibility of author's achievements. Open access to information is a way to increase the right of authors, who usually is not paid for the publication.

One of important means for open access to information is self-archiving of research papers of an institute before or after publication. In 2004, UK Parliament Science and Technology Committee issued a report to recommend building the institutional repository for open online access of information from public-funded organizations². In Korea, KERIS(Korea Education & Research Information Service) is distributing an institutional repository system, named 'dCollection' to denote digital collection, to university libraries, and also operating a central system to integrate distributed university systems for resource sharing. The purpose of this paper is to introduce dCollection system and to report the result of 2-years' operation of the system.

2. Brief Overview of Institutional Repository

Two types of repository are found; the disciplinary repository such as arXiv.org for physics and related area and NCSTRL for computer science and related area, and the institutional repository such as DSpace by MIT and ePrint by University of Southampton. All of these are to preserve digital documents and to assure open access to information over the barrier of law, economy and technique.

Raym Crow³ defined the institutional repository in the context of the university as "digital collections capturing and preserving the intellectual output of a single or multi-university community", and pointed out two strategic aspects of it; 1) reformation of scholarly communication as it "expands access to research, reasserts control over scholarship by the academy, increases competition and reduces the monopoly power of journals, and brings economic relief and heightened relevance to the institutions and libraries that support them", and 2) development of university's competitiveness with "the potential to serve as tangible indicators of a university's quality and to demonstrate the scientific, societal, and economic relevance of its research activities, thus increasing the institution's visibility, status, and public value".

The definition and strategic points of view above are implemented in dCollection of KERIS. The objectives of dCollection is not different from those of DSpace; to strengthen the competitiveness of university libraries as for management and provision of information, to enhance the awareness of the role of the library in and outside the university, and to increase the value of the universities participating in dCollection.

3. Main Features of dCollection System of KERIS

This project started in 2003. The first version of the system had been developed with 20 university libraries, 4 libraries participating in the pilot project in 2003 and 16 libraries to be joined in the project in the following year. Until 2005, 3.5 million US dollars were invested

² Albanese, A. 2004 UK Report Calls for Publicly Available STM Research. *Library Journal*, 129(13): 16-17

³ Crow, R. 2002. *The Case for Institutional Repositories: A SPARC Position Paper*. Washington, DC: SPARC. <http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf>

for system development, system distribution, data migration, data correction, and other training and promoting programs for 40 universities.

dCollection system has four particular features; data provider(DP) system for the university library to collect and manage metadata and digital documents; service provider(SP) system in KERIS to collect and integrate metadata from DP systems; OAI-PMH(Open Archive Initiative Protocol for Metadata Harvesting) for communication between DP system and SP system; and copyright management for open access to information and protection from misuse of information.

The overall workflow of dCollection operation is linear. The university library, with dCollection system installed or its own system, collects metadata and digital documents from the author. The librarian proofread and corrects the collected metadata and digital documents. If necessary, the librarian requests re-submission of these. The digital document is converted into PDF format automatically. The document is available for the within-university service. SP system in KERIS harvests the metadata from DP systems in the universities and integrates them. All of the collected metadata is transferred to RISS, the national integrated research information service provided by KERIS

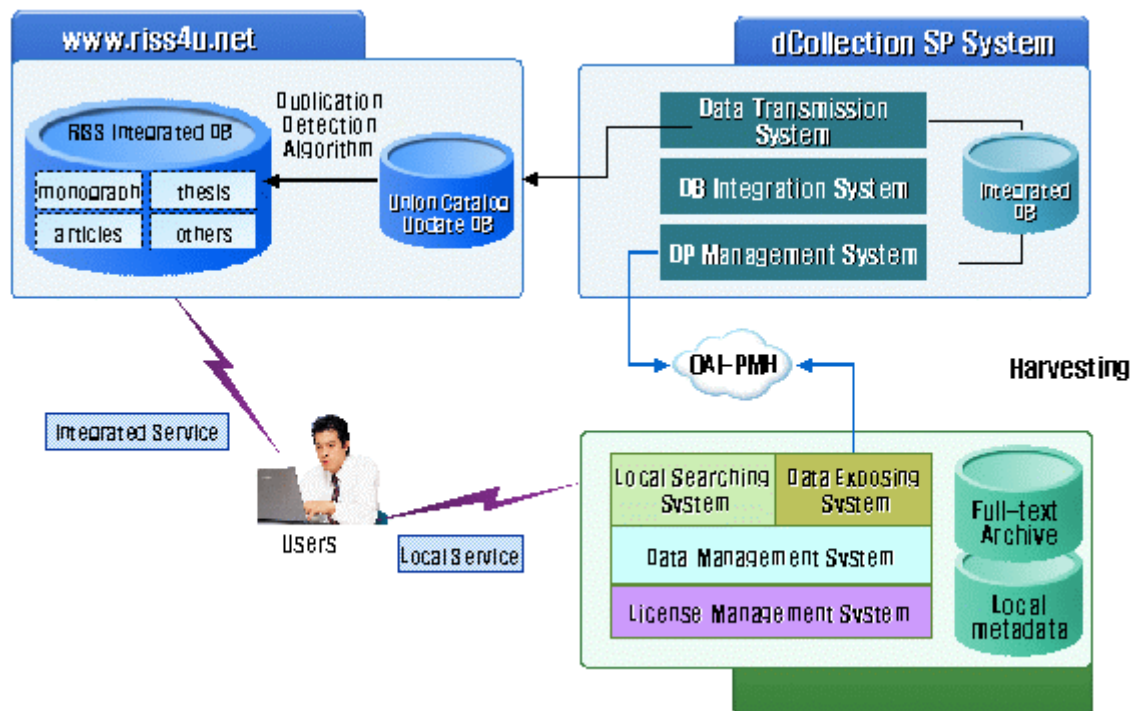
4. dCollection System Architecture

dCollection, which is named by KERIS to denote digital collection, is following the SP/DP model of OAI⁴. The overall architecture is shown in Figure 1. The participating libraries are DP(Data Provider) and KERIS is SP(Service Provider). The system is a total system of collecting, managing, preserving, and sharing digital contents. OAI-PMH (OAI Protocol for Metadata Harvesting) uses HTTP for web applications and XML and SOAP(Simple Object Access Protocol) for data exchange.

⁴ “The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication.”

<http://www.openarchives.org/organization/index.html>

[Figure 1.] dCollection System Architecture



Based on this system, digital documents of the university are available to the general public as well as users in the university. The typical overall service flow is; 1) registration of a community (department, laboratory etc.) and collection (thesis, article, etc.) into the local dCollection system, 2) submission of documents and metadata, 3) agreement by the author for open access to the document, 4) checking and correcting errors by the librarian, 5) requesting the author (or community) to correct or re-submit the metadata or document, 6) approval for local service by the librarian, 7) harvesting the local metadata by SP system, 8) quality verification and improvement by KERIS, 9) integration with the old data, and 10) integrated service through RISS (Research Information Service System) of KERIS dCollection DP system is installed in the university library. The main function of DP system is to collect metadata and digital documents, manage the quality of collected metadata and digital documents, expose the metadata for harvesting by SP system, and to manage copyright.

The system has 4 main system modules;

- Data Management System: to collect and re-collect metadata and digital documents from authors; to check and correct errors in metadata and digital documents; to inform errors and request re-submission to the author.
- Data Exposing System: to response to the request from SP system and to expose the collected metadata for harvesting by SP system.
- Local Searching System: for the librarian and institutional user to search the local data
- License Management System: to inform the user the conditions and options of uses of the document; and to protect the document from copyright infringement.
- The main function of dCollection SP system in KERIS is to harvest and to integrate the metadata from DP systems. The system has 3 main system modules;
- DP Management System: to manage the log record, schedule, and workflow related to DP system and data harvesting.

- DB Integration System: to remove duplication of and integrate data from DP systems; and to manage the integrated dCollection metadata.
- Data Transmission System: to correct errors in the integrated metadata; and to transfer the data to the service system.

5. Copyright Issues

Copyright is one of the most important issues to consider for information service. Copyright is managed in three ways. First is to get an agreement from the author for open access to the document. Most of documents in dCollection are open to the public with the agreement by the author. The general agreement form was developed by KERIS and shared by the member libraries with minor modifications. The second one is to set the condition of use according to CCL⁵. The conditions of use are “Attribution”, “Noncommercial”, “No Derivative Works”, and “Share Alike”. The last one is document protection. To protect a document from misuse, the library can limit the use of document to displaying, printing, or saving.

6. Current Status of dCollection Service

Forty universities are participating in dCollection project. They provided about 74 thousand items in total. Most of them (about 93%) are doctoral and master’s thesis as shown in Table 1. Many universities require graduate students to submit the thesis for preservation and information service. Graduate students may or not agree for open access to their thesis. Most of students agree for the open access. For journal articles, the library needs to contact individual university laboratory or professor. Few universities have a regulation for submission of the research paper of the professor.

In most universities, dCollection system is used mainly for preservation and service of thesis. Besides thesis, there are several kinds of materials which need to be collected and shared by universities. To facilitate and expand the use of dCollection system, KERIS operates a model university library program to develop a best practice model for dCollection operation. In 2005, Pusan National University and Sungkyunkwan University were designated as a model university. The major products developed by these universities through this program are preparation of university regulation for digital submission of thesis and research paper, contract with laboratories in the university to participate in dCollection, and development of a manual for proper submission and management of research information in cooperation with the university R&D center. These products were distributed to all the university to share the experience of model university libraries.

To assist the university library and to shorten the time from submission to national service, central quality management was done by KERIS in 2005. This is to reduce the job load of the university librarian for dCollection system operation. In case of theses, they are submitted mostly within 3 months in a year. Most of the library has less than one professional staff member for dCollection system operation. The main activity of central quality management is the correction of metadata and digital documents the author submitted.

⁵ CCL(creative commons license, <http://creativecommons.org>)

7. Future Plan

This project began in 2003 and it is in a developing stage. For further development, the system needs to expand the business domestically and internationally. First of all, within this year, a web-based hosting system will be developed for small universities. A university library will be free from the maintenance of the system, which were mentioned as a big burden for the university library of a small scale. With this hosting system, dCollection system will be distributed to all the university, and it will be the most important means of the university to connect all the digital university libraries in Korea.

Secondly, under development is the fee management system for the document without the agreement from the author for open access. According to the copyright law in Korea, a published material can be transmitted in digital format between libraries 5 years after publication with the payment of copyright fee. With this fee management system, digital resource sharing between university libraries will be expanded.

Thirdly, dCollection will be connected with major open archives in other countries. Beginning with university repositories of a large scale, the interconnection will be expanded to electronic journals and various repositories other than the university. A mid-term plan will be set up for the international network of research information open archive.

Lastly, all the participating universities will be evaluated. For a university repository to be successful, there would be several factors to be managed. For example, up-to-dateness and diversity of digital collections are important for user satisfaction on the content. Also, for authors submitting metadata and digital documents, the system needs to provide an interface of easy use. If a university has a regulation for dCollection, the system can be easily spread into the university community. The evaluation criteria are the important success factors for dCollection. All of the participating universities will be evaluated with these factors to find strength and weakness of each university, and to suggest customized solutions for the university. Evaluation criteria are listed in Table 2.

8. Closing Remark

This system, what we call 'dCollection', is the result of cooperation of KERIS and 40 universities during last 3 years. There is one dominant success factor; active participation of users (university libraries) in the system development process. The voice of customers is assimilated in the system. The active participation from the university library should be mentioned first. From the next year, the system will be expanding fast with the web-based hosting system. dCollection will be growing to the core national research information system and to an import center for international flow of research information.

<Table 1> The Number of Contents in Service (As of March 31, 2006)

	Thesis	Journal Article	Total
Daegu University	3,913	2,555	6,468
Pusan National University	4,660	509	5,169
Kyungpook National University	2,399	-	2,399
Sungkyunkwan University	2,353	2	2,355
Ewha womans University	2,378	401	2,779
Chungnam National University	2,259	-	2,259
Gyeongsang National University	1,128	760	1,888
Hongik University	953	343	1,296
University of Seoul	463	452	915
Keimyung University	878	-	878
others	47,659	161	47,820
Total	69,043	5,183	74,226

<Table 2> dCollection Evaluation Criteria

Evaluation Area	Evaluation Criteria
Digital Collection and Metadata	Diversity of digital collection, Up-to-dateness of contents, Proportion of documents submitted to dCollection, # of access points (# of metadata elements), Quality of metadata
System/Network	Integration of service with OPAC, Linkage with related administration system, Quality of homepage, Additional services for community, # of linkage with other open archives
Policy and Management	University committee for dCollection, Flexibility of document submission, dCollection advertisement in the university, Portion of budget for dCollection, Portion of dCollection staff member, Provision of copyright information, Formalized dCollection manual, Incentive for self-archiving, Diversity of digital collection development, A library organization for dCollection, Formalized preservation policy, Awareness of librarians/head of library/student/professor on dCollection.
Author/ User	Satisfaction of the user, Satisfaction of submitting author, Assistance for submitting author, # of use of dCollection documents, Proportion of successful download of digital documents.

Park, Hong Seok

Email: honspark@keris.or.kr

Phone: 82-2-2118-1331



Bachelor in Department of Chemistry, Dankook University (1985)

Master in Department of Library & Information Studies, Emporia State University (1989)

Ph.D. in School of Communication, Information & Library Studies, Rutgers University (1995)

Researcher in Korea Research Information Center (1998-1999)

Senior Researcher in Korea Education & Research Information Service (1999 -)

Jang, Keum Yeon

Email: jkylib@keris.or.kr

Phone: 82-2-2118-1231



Bachelor in Department of Library, Archive & Information Studies, Pusan National University, (1995)

Master & Ph.D. program in Department of Library & Information Science, Sungkyunkwan University (2000 -)

Assistant Researcher in Korea Education & Research Information Service (1999 -)