

---

# Building an institutional repository: sharing experiences at the HKUST Library

---

Ki-Tat LAM and Diana L. H. CHAN

## The author

Ki-Tat Lam is the Head of Library Systems at HKUST Library. Diana L. H. Chan was the Head of Reference at HKUST Library and is now the Associate Librarian of Public Services at the City University of Hong Kong.

## Keywords

Institutional repositories, open access, self-archiving rights, content recruitment, DSpace, digital libraries, HKUST

## Abstract

**Purpose** - To document HKUST's experiences in developing its Institutional Repository and to highlight its programming developments in full-text linking and indexing, and cross institutional searching.

**Design/methodology/approach** - This paper describes how HKUST Library planned and set up its Institutional Repository, how it acquired and processed the scholarly output, and what procedures and guidelines were established. It also discusses some new developments in systems, including the implementation of OpenURL linking from the pre-published version in the Repository to the published sources; the partnership with Scirus to enable full-text searching; and the development of a cross-searching platform for institutional repositories in Hong Kong.

**Findings** - It illustrates what and why some policy issues should be adopted, including paper versioning, authority control, and withdrawal of items. It discusses what proactive approaches should be adopted to harvest research output. It also shows how programming work can be done to provide usage data, facilitate searching and publicize the repository so that scholarly output can be more accessible to the research community.

**Practical implications** - Provides a very useful case study for other academic libraries who want to develop their own institutional repositories.

**What is originality/value of paper** - HKUST is an early implementer of institutional repositories in Asia and its unique experience in policy issues, harvesting contents, standardization, software customization, and measures adopted in enhancing global access will be useful to similar institutions.

## Introduction

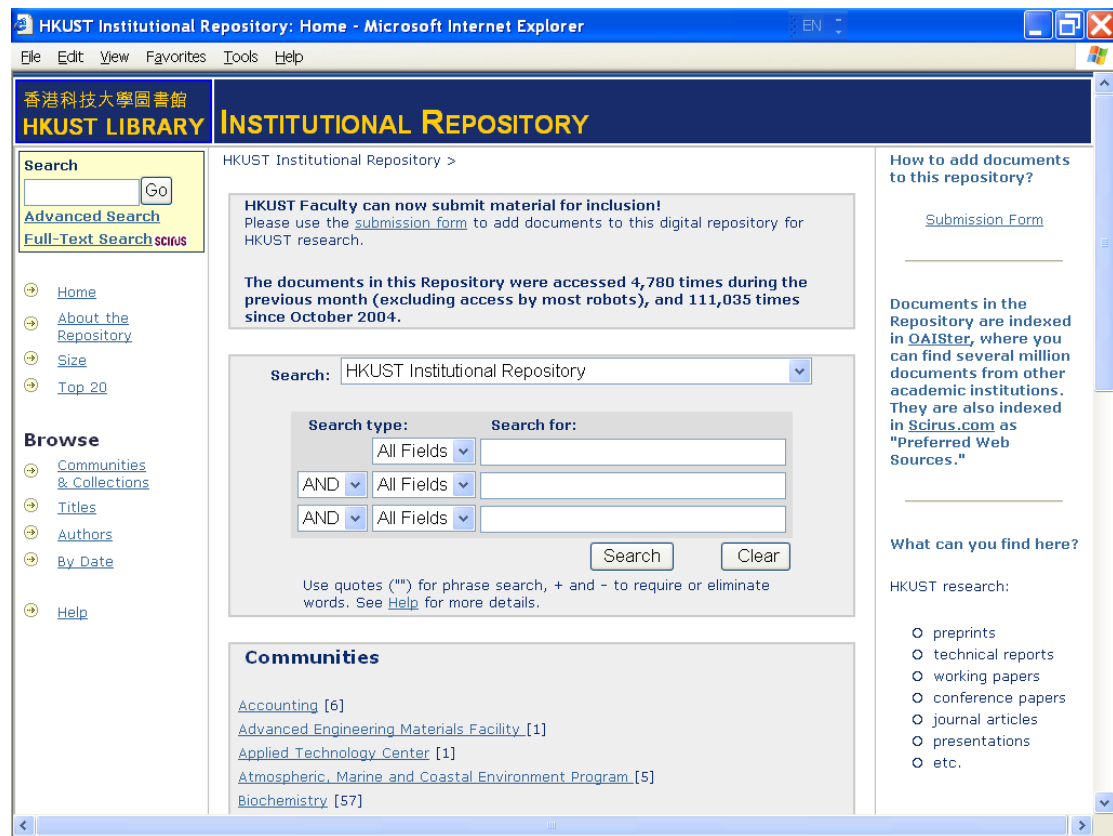
The Hong Kong University of Science and Technology (HKUST) is a young institution opened in October 1991. It offers taught and research programs in science, engineering, business, humanities and social science, with 430 full-time faculty members, 5,600 undergraduates and 3,200 postgraduates. Despite its short history, HKUST has rapidly evolved into a world class institution, and was ranked number 43 in the world by The Times Higher Education Supplement in 2005.

The HKUST Library has been engaged in library digitization projects since its foundation 15 years ago, including the early project on Course Reserve Imaging in 1993 and the CJK (Chinese, Japanese, Korean) capable systems for Digital University Archives and Electronic Theses in 1997. The experiences gained through these projects have facilitated a smooth creation of its Institutional Repository.

The Library showed its early support to the open access concept by joining SPARC in 2001. And in November 2002, Kimberly Douglas, the University Librarian of the California Institute of Technology, was invited to the Library to give a staff development workshop on *E-prints, OAI (Open Access Initiatives) and institutional repository*. The Library decided to build the HKUST Institutional Repository after the workshop, aiming to create a permanent record of the institution's scholarly output in digital format, and to make the Repository globally and openly accessible.

The HKUST Institutional Repository (see Figure 1) was launched in February 2003 with its first batch of 105 computer science technical reports. It has grown to 2,369 documents from 42 academic departments in September 2006, holding preprints, technical reports, working papers, conference papers, journal articles, presentations, book chapters, patents, and PhD theses. They are mainly PDF files with some PowerPoint and program files. These scholarly works were accessed 69,000 times excluding robots in the period from September 2005 to August 2006. A research study on how Hong Kong Chinese students learned English was downloaded 800 times in just a month.

Figure 1 Home page of the HKUST Institutional Repository (<http://library.ust.hk/repository/>)



The HKUST Institutional Repository was reported in a number of conference presentations (Chan 2004a; Chan 2004b; Lam 2004; Lam 2006) and journal articles

(Chan, Kwok, Yip 2005; Kwok, Chan, Wong 2006). This paper will summarize and update the issues discussed in these reports, including how to plan and set up the Repository, how to acquire and process the scholarly output, and what procedures and guidelines were established. It will also discuss some of the new developments, including the implementation of OpenURL linking from the pre-published version in the Repository to the published sources; the partnership with Scirus to enable full-text searching; and the development of the HKIR, the cross-searching platform for institutional repositories in Hong Kong.

## **Planning Stage**

The Library adopted a bottom-up approach for building its Institutional Repository. As the concept of institutional repository was quite new in Hong Kong in the year of 2002, it would be easier to begin the project in a small scale but with gradual expansion in scope and institutional participation. With a tangible repository at hand, the Library could approach the faculty and university administration, explained and demonstrated what IR was and how they could benefit from it. Another advantage of starting small was that the investment of resources would be relatively small as compared to a large-scale project and the Library could have more freedom to test the water before moving on.

### *Establishing a Task Force*

The project began with the establishment of a task force in December 2002, consisting of four librarians from the Reference and Systems departments and the Associate University Librarian. The task force's charge was to identify the issues involved in creating the IR, to evaluate and select the software for hosting the Repository, and to develop action plans. Findings and recommendations were reported to the Library Administration Committee, the main decision-making body, for approval.

A number of policy issues were resolved during the early stage of planning. For example:

- Make the IR totally open and accessible to the world. If a faculty member wishes to restrict access, then the document will not be accepted.
- The IR is a deposit of research documents, not merely an index with links to external sources; if the Library does not have the right to deposit the full-text papers, they cannot be included in the IR.
- Undertake retrospective work to include documents previously published in addition to the current ones.
- Do not include ephemeral materials such as faculty-produced course notes, popular works or feature columns from newspapers, but would limit the coverage to published material and grey literature only.
- Allow authors to submit documents online and they will sign a permission agreement, granting the Library non-exclusive distribution rights.
- Adopt Adobe's PDF format as the default document format.
- Build a single database and not to have multiple databases such as the model adopted by California Institute of Technology.

### *Selecting IR Software*

While there are many options for selecting IR software and hosting services today, the choices were extremely limited in the year of 2002. Like many of its digital library projects, the Library decided to use open source software for its IR. The main advantage of open source software was that it provided flexibility for local

customization and feature enhancements. The significant software cost savings was also a consideration, as the Library did not receive extra funding for the IR project.

The task force decided to focus on open source software that supported OAI-PMH (Open Access Initiative - Protocol for Metadata Harvesting). Two such IR software programs were evaluated, namely EPrints and DSpace. EPrints was developed by the University of Southampton and was widely used by IR implementers in 2002. DSpace was jointly developed by MIT Libraries and Hewlett-Packard Company, and began its first release on Sourceforge at the time of the Library's evaluation. DSpace was developed with experience gained from EPrints, but with a clever move from the Perl programming language to Java and Servlet. And at that time, it also had better Unicode support, which was essential to the Repository that would contain Chinese materials. With the above consideration, the Library decided to adopt DSpace.

Once DSpace was selected, the Library began to develop an initial prototype, using the 105 working papers from the Department of Computer Science, which were freely available on their website in postscript format. During this prototyping, a number of design issues were resolved, including how to organize the documents by departments and by document types, and what fields were required in the metadata.

### ***Staffing***

As there was no extra funding and manpower provided for the project, the Library relied on existing library staff to create and maintain the Repository. In addition to the initial planning and system setup done by the task force, an IR Team of eight reference librarians and five data entry staff was established to handle on-going work such as faculty liaison, document acquisition and processing, and the actual data input tasks. It was estimated that about 350 man-days for librarians and another 350 man-days for support staff were spent in the first three years of the project. It was found that the time and efforts taken to acquire and process the content so as to achieve a critical mass were quite substantial. Institutions that are interested in creating IRs should be aware of the staffing implications and should request sufficient funding for the project.

### **Organizing the Repository**

In DSpace, a repository is made up of a hierarchy of communities, collections, items, metadata and bitstreams. A document is represented by an *item*, which contains *metadata*, i.e., a description of the document and a bundle of *bitstreams*, such as PDF and PowerPoint files that hold the actual content of the document. Items are held in *collections*, which are further grouped under *communities*.

Documents in the HKUST Institutional Repository are organized by academic departments (communities), and within a department, they are further grouped according to the document types (collections), such as conference papers and journal articles. As of September 2006, the Repository has 2,369 documents in 42 communities and 139 collections. As expected, disciplines which have an established tradition of sharing preprints and working papers, such as computer science and engineering, are ranked as top contributors to the Repository (see Table 1). Conference papers, journal articles, preprints, working papers and doctoral theses constitute the major document types held in the Repository (see Table 2).

Table 1 Top 10 contributing departments (September 2006)

Academic Departments and Centers	Size	Percentage
Computer Science	478	20.2%
Electrical and Electronic Engineering	324	13.7%
Mechanical Engineering	164	6.9%
Marketing	154	6.5%
Mathematics	130	5.5%
Physics	126	5.3%
Chemistry	106	4.5%
Social Science	106	4.5%
Biology	84	3.5%
Language Center	84	3.5%
Others	613	25.9%
<b>Total</b>	<b>2369</b>	<b>100.0%</b>

Table 2 Total number of documents by document types (September 2006)

Document Types	Size	Percentage
Conference papers	636	26.8%
Working papers, technical papers, research reports, preprints	549	23.2%
Journal articles	537	22.7%
Doctoral theses	473	20.0%
Presentations	70	3.0%
Patents	58	2.4%
Book chapters	38	1.6%
Others	8	0.3%
<b>Total</b>	<b>2369</b>	<b>100.0%</b>

The metadata of the documents is encoded in qualified Dublin Core schema. DSpace's default DC registry was followed, except for a locally defined qualifier `openurl` for the element `identifier`. The purpose of defining this local field will be discussed later in this paper.

### *Document Submission and Processing*

To make the document submission as simple and effortless as possible, the Library decided to develop its own Faculty Submission Form, a web-based interface outside of the DSpace workflow. Faculty members are only required to input minimal bibliographic data, such as title, author and citation source when submitting the actual files to the server. The Form contains a Non-Exclusive Distribution License (Figure 2). They need to check the "I Agree" box to grant permission to the Library.

The IR Team then verify and enhance the metadata, ascertain publishers' policies to avoid depositing the wrong versions, harvest and convert the files to PDF format as needed, and add the documents to the Repository. A web-based Add Item program was also developed for the IR Team so that they can integrate these document submission and processing tasks seamlessly with DSpace.

Figure 2. Non-Exclusive Distribution License in the Faculty Submission Form

---

### HKUST Institutional Repository

#### NON-EXCLUSIVE DISTRIBUTION LICENSE

I am submitting this collection of files and associated bibliographic metadata for inclusion in the HKUST Institutional Repository.

I hereby grant to The Hong Kong University of Science and Technology (HKUST) the irrevocable, non-exclusive royalty free right to reproduce, distribute, display, and perform this work in any format including electronic formats throughout the world for educational, research and scientific non-profit uses during the full term of copyright including renewals and extensions via the HKUST Institutional Repository mechanisms maintained by the HKUST Library. I also hereby grant to HKUST the non-exclusive right to sub-license these rights to others should the University forego the ability to maintain distribution. I warrant that I have the copyright to make this grant to HKUST unencumbered and complete.

\*  I agree to the above Distribution License

Submit	Cancel
--------	--------

### *Harvesting Research Output*

While some faculty members and researchers had the initiative to submit their documents via the submission form, most of them were not responsive at all. Therefore, the Library had to take a more proactive approach to discover and harvest research output for the Repository. For example, the IR Team had:

- Visited faculty members' personal and departmental websites as well as the websites of the research centers and institutes on campus to harvest full-text research papers and publications posted on the web.
- Surveyed academic departments to harvest collections of working papers and technical reports.
- Searched the library catalog to identify proceedings of conferences held at HKUST.
- Scanned through boxes of pre-published research papers held in the University Archives.
- Searched electronic databases and open access sources such as Web of Science and DOAJ to identify papers published by the HKUST researchers.
- Contacted individual faculty members to ask for their complete publication lists and their full-text documents.

In most of the above cases, the IR Team had to contact the original authors to obtain permissions before loading the harvested documents to the Repository. And if the electronic version was unavailable, the paper document would be digitized.

The HKUST Electronic Theses database was built a few years earlier than the IR. Not all these theses are open access. The Library decided to include only those PhD theses with author permission in the IR so that all of them would be openly accessible. Instead of depositing a second copy to the IR, only metadata was created, together with a link to retrieve the full-text from the Electronic Theses database.

HKUST faculty members need to report annually to the Research Output Collection System. The Library asked the office in-charge of this system to include a checkbox in the submission page to indicate the reporters' willingness to deposit their publications into the Repository. If the box is checked, an email containing the

citations will be sent to the IR Team for follow up actions. This automatic alert mechanism has enabled us to harvest research output on an annual basis.

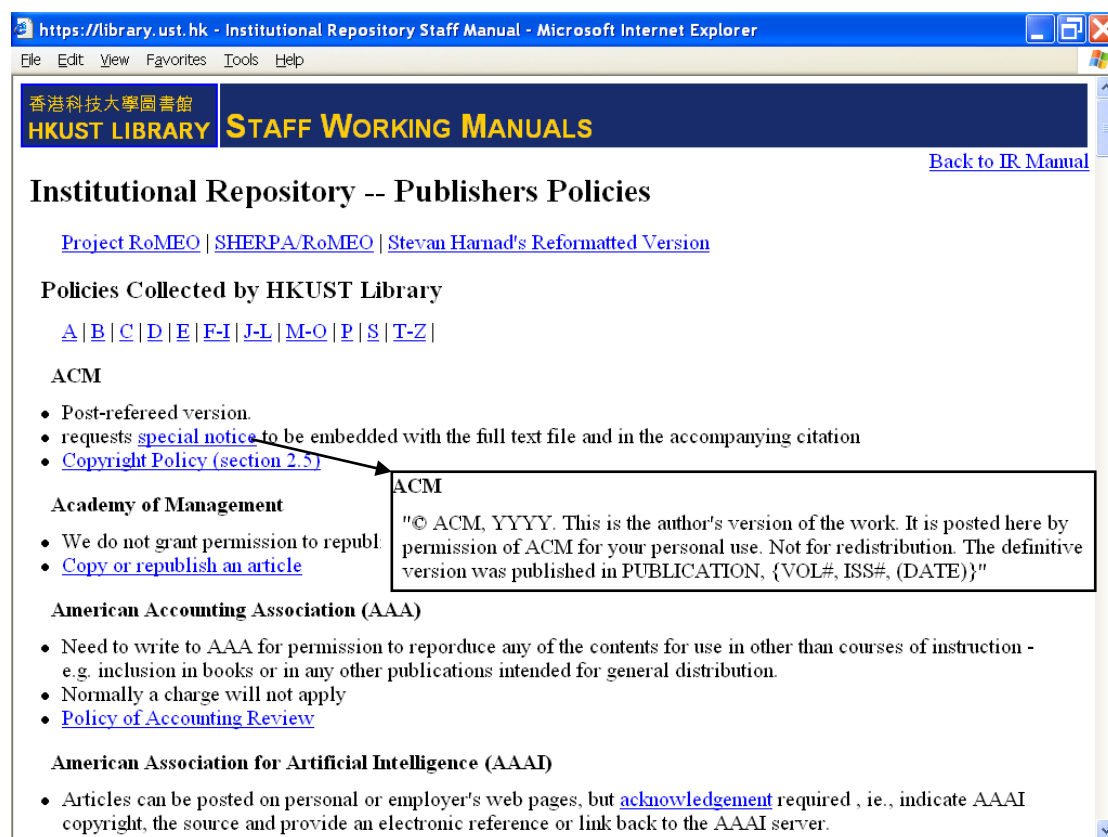
### ***Publishers' Policies and Deposit Guidelines***

Verifying and selecting a version of the document for depositing to the Repository is far from a trivial task. While more and more publishers nowadays have their self-archiving policies clearly spelt out on their websites, this was not the case in 2003. Project RoMEO, which provides a directory of publisher self-archiving policies, was just launched in 2002.

The IR Team surveyed the publisher's policies for the journal articles via SHERPA/RoMEO and publisher websites. Findings were recorded in the IR Staff Working Manual for easy future reference. The list currently contains more than 60 publisher policies (see Figure 3). Publishers' policies can be summarized as follows:

- no archiving allowed
- allow pre-refereed version only
- allow post-referred version only
- allow pre- and post- refereed versions
- allow publisher's version
- allow all versions
- not specified

Figure 3. List of publishers' policies collected by the HKUST Library, showing links to special notes and acknowledgment text, and links to publishers' website



If the policy is unknown, the IR Team will write to the publishers for clarification and ask them for the archiving permission. And if the Library's version is not usable, the IR Team will contact the authors to ask for an acceptable version.

The Library also encourages authors to negotiate with publishers so as to retain their self-archiving rights and the rights for personal educational use. They should also avoid granting an exclusive long-term license that extends beyond first publication.

### ***Versioning***

It is essential that users know whether the version deposited is the published version or not. To avoid confusion, a watermark “This is the Pre-Published version” is stamped on the first page of the document if it is a pre-refereed or post-refereed version. A note “pre-published version” is also displayed in the “Files in This Item” box of the item record display page (Figure 4).

A piece of scholarly work may undergo several rounds of revisions. Authors are encouraged to submit revised versions as separate documents. They can also replace the previous versions as long as they are not published items. By doing so, the revised version will bear the same identifier (handle) number as the previous one.

### ***Author Name Authority Control***

Authors may have their works published under different names. It is essential to perform some form of authority control for consistency. Policies for entering the names of HKUST researchers were established. Name authority records for some authors are readily available from the library catalog. For those without record, several university publications such as the *Academic Calendar*, *Faculty’s Profile* and *Communications Directory* are consulted. In a few cases, emails were sent to the authors to seek their preferences on the names used, especially in the case of maiden and married names, names in Chinese, and the inclusion of Christian names. For Chinese documents, the English name of a HKUST author will be taken from the title page, should it appear in English or bilingually in English and Chinese. If HKUST authors do not provide their English names in their Chinese documents, the IR Team will look up their English names and add them to the metadata. For bilingual names of the same author, the Chinese name will be entered in parentheses after the English one, e.g., “Chan, Diana L. H. (陳麗霞)”. Some documents in the Repository were jointly written with non-HKUST authors. Since it is difficult to identify these non-HKUST affiliated authors, the Library decided not to perform authority check on them.

### ***Subject Keywords***

When authors submit records, they can supply three to eight keywords or phrases for indexing. If the subject field is not filled, the IR Team will extract the keywords from the abstracts. English keywords are used for Chinese documents as well. The Library decided not to use thesauri or LC subject headings in assigning subjects.

### ***Withdrawal of Items from the Repository***

At the specific request from authors, documents in the Repository can be permanently removed. To retain the historical record, such transactions will be noted in the metadata record. Since these documents may have been cited by others, the system will supply a “tombstone” when they are requested. A withdrawal statement will be displayed in place of the view document link.



## Programming Efforts

The advantages of using open source software such as DSpace as the platform for the Institutional Repository became more apparent when the needs and requests to customize the software flooded in. Apart from the Faculty Submission Form and the Add Item Form as mentioned in the previous section, the following are other customizations that are worth mentioning.

### *Linking to the Published Version*

Some publishers only allow institutions to archive the pre-published version. From time to time, the Library receives authors' feedback that they prefer users to read the published version rather than the pre-published version archived in the IR. One can easily resolve this problem by adding the direct URL link to point users to the published version residing on an aggregator's or the publisher's website to which the institution has a subscription. The Library objected to this approach because such links would become broken due to subscription changes. After much study, the Library decided to implement an OpenURL linking mechanism on DSpace so that users can be dynamically redirected to library-subscribed resources that host the published version.

To enable this linking feature, the metadata (item record) must contain the OpenURL string. This is made possible by using a locally defined Dublin Core field `identifier.openurl`. DSpace's item record display page was modified to enable a link-resolver button when the item contains an OpenURL (see Figure 4). Program was developed to query OCLC's OpenURL Resolver Registry on-the-fly while displaying the button. By doing so, users will see their own institution's link-resolver, such as WebBridge for HKUST. When users click on the button, the link-resolver will try its best to identify external sources that contain the published version.

Constructing OpenURL manually is extremely painful. To automate this process, a web-based program was developed. It can intelligently parse the title, contributor.author and the identifier.citation fields to obtain most of the required key-value pairs for constructing the OpenURL. As the journal's ISSN is not available in the Repository's metadata, the program searches the library catalog to obtain the numbers. With this program, the OpenURL string can be quickly created within a few mouse clicks. Figure 5 shows the web interface for this highly user-friendly OpenURL Builder program.

Figure 4. A pre-published document record, showing versioning information, with WebBridge link to the published sources

Please use this identifier to cite or link to this item:  
<http://hdl.handle.net/1783.1/2329>

**Title:** Experimental study on RC beams with FRP strips bonded with rubber modified resins

**Authors:** Gao, Bo; Kim, Jang-Kyo; Leung, Christopher K. Y.

**Keywords:** Fiber reinforced plastic (FRP); Epoxy adhesives; Rubber modified resins; Load carrying capacity; Ductility

**Issue Date:** Dec-2004

**Citation:** Composites Science and Technology, 64(16), 2557-2564

**Abstract:** This paper presents the effects of adhesive properties on structural performance of reinforced concrete (RC) beams strengthened with carbon fiber reinforced plastic (CFRP) strips. The epoxy adhesives modified with rubber modified resins were used to bond the CFRP strips to the RC beams. The experimental results show that the use of rubber modified resins significantly improved the load carrying capacity and ductility of the RC beams. The experimental results also show that the use of rubber modified resins significantly improved the bond strength between the CFRP strips and the RC beams.

**Files in This Item:**

File	Description	Size	Format
Paper03CST07.pdf	pre-published version	254Kb	Adobe PDF

View published version via WebBridge

**Watermark:** This is the Pre-Published Version

**Click on this image to launch an OpenURL link-resolver to locate the published version**

**Document deposited in the Repository is a pre-published version**

Figure 5. OpenURL Builder - automating the construction of the OpenURL

香港科技大學圖書館  
**HKUST LIBRARY INSTITUTIONAL REPOSITORY**

**OpenURL Builder**

Journal title: Composites science and technology

Print ISSN: 0266-3538 e.g. 9876-5432

Online ISSN: 0266-3538 e.g. 1234-5678

Volume: 64 Issue: 16 Year: 2004

Start page: 2557 End page: 2564

Title of article: Experimental study on RC beams with FRP strips bonded with rubber modified resin

Author's lastname: Gao firstname: Bo

**Check Library catalog and auto-insert the ISSNs to the form**

bib number:

**Click this link to test the OpenURL**

<http://ustlib.ust.hk:4550/resserv?sid=HKUST:IR&eissn=0266-3538&issn=0266-3538&volume=64&issue=16&spage=2557&epage=2564&date=2004&aufirst=Bo&aulast=Gao&atitle=Experimental+study+on+RC+beams+with+FRP+strips+bonded+with+rubber+modified+resins&title=Composites+science+and+technology>

**Click this button to create this OpenURL fragment**

© HKUST Library

### Usage Statistics and Top 20 Most Accessed Documents

In-house usage analyzing programs were developed to supplement the usage reports that come with DSpace. They are based on the web access logs captured by the server

when users issue requests to download the bitstreams, e.g., the PDF files. The Repository is open to the world and allows visits from search engines, robots and OAI harvesters. As a result, it receives tens of thousands of web requests per day. Program was developed to enable the Library to know how many times the IR documents were downloaded by “real” users, excluding robot accesses. This figure was updated monthly to the Repository home page.

Another customized program was the monthly listings of the Top 20 most accessed documents. It is interesting to analyze these Top 20 lists as they give a good account of documents, topics and authors that users are most interested in. Such information is useful for IR promotion. For example, the Library wrote to the authors in the lists to inform them about the high usage of their papers. The IR Team also showed the lists to the faculty members during departmental visits. While the majority of the documents are from the academic departments, it is worth mentioning that a number of documents authored by the HKUST Language Center made their way into the lists, together with the ones HKUST Library wrote on institutional repository and virtual reference.

### ***CJK Search and Display***

In the early versions of DSpace, there were problems on searching and displaying Chinese characters. The authors managed to fix these problems by revising and replacing some of the DSpace source codes. While some of these problems were eventually fixed in DSpace’s later versions, the timing of fixing them was critical to the Library’s IR software selection. Had they not been fixed during software evaluation, the Library would not have selected DSpace.

Thanks to open source, one could dig into the source codes and fix problems quickly. The main CJK problem was attributed to the use of the CJK-illegible string tokenizer. DSpace is Unicode capable, meaning that it supports data and strings in multiple scripts, including CJK. However, like many other non-Roman scripts, the way Chinese strings are sorted, indexed and searched can be quite different from that for English. Global software developers should be aware of these differences in order to avoid problems similar to the ones encountered with DSpace.

### **Enhancing Global Access**

It is essential to publicize an institutional repository so that the research output can be made known to the world. In addition to making the Repository readily available and openly accessible, the Library has implemented the following measures to allow search engines, agents and harvesters around the world to discover documents in the Repository.

#### ***OAI-PMH Compliance***

OAI-PMH (Open Access Initiative – Protocol for Metadata Harvesting) is a protocol that allows metadata to be easily harvested by computer programs. Like other IR systems, DSpace is OAI-PMH compliant. It is useful to register the OAI Base Path of the IR to various OAI registries, such as the ROAR (Registry of Open Access Repositories) maintained by EPrints. OAI harvesters can then follow this registered path and retrieve the metadata for their own searching and indexing services. At least two well known services, namely OAIster and Scirus, are constantly harvesting HKUST’s IR metadata via this protocol.

### ***Indexed by OAIster***

OAIster is a project of the University of Michigan Digital Library Production Service. By using OAI protocol, it has collected almost 10 million of metadata records of academically-oriented digital resources from 680 institutions around the world. The Library contacted OAIster in June 2003 and since then the Repository records are included in OAIster. HKUST research output is therefore available to all OAIster users via its one-stop searching interface.

### ***Full-text Searching on Scirus***

Scirus is Elsevier's free search engine for scientific information. In addition to web pages, it also harvests content from selective institutional repositories. In November 2005, Scirus proposed to index the HKUST Institutional Repository. The project involved building the mechanism to harvest the content of the Repository, indexing both the metadata and full-text of the documents, making them searchable on the Scirus platform, and integrating the Scirus search form within the Repository home page. This feature was rolled out in May 2006. Thanks to Scirus, the Library is able to offer full-text searching external to DSpace as well as to open up the content to a larger scientific research community.

### ***Crawling by Google and Yahoo***

Robots from search engines are allowed to visit and crawl the web pages of the Repository. By enabling robot access, HKUST's research output is readily available via popular search engines such as Google and Yahoo, as well as their subsets, such as Google Scholar. The following story, as told by the Library's reference librarians, shows the effectiveness of using these search engines to discover documents in the Repository: "Once, we received an email from someone in the U.K. who wanted to contact the author of a PhD thesis. It turned out that the requestor was the father of a son suffering from a type of cancer called Ewing Sarcoma. He discovered the thesis on the IR via the web. We acted as the intermediary and passed his enquiry to the author concerned." (Kwok, Chan, Wong 2006)

### ***Searching with SRW/U***

SRW/U (Search and Retrieval for the Web, or by URL) is a protocol for searching heterogeneous databases using XML and HTTP. It retains the core functionality of Z39.50 but in the form of web services. With SRW/U, search service providers can broadcast a search to various institutional repositories and deliver the search results in their own GUI interface. To allow such federated searching, the Library implemented the SRW/U layer to the Repository in October 2004, based on OCLC's SRW/U open source software.

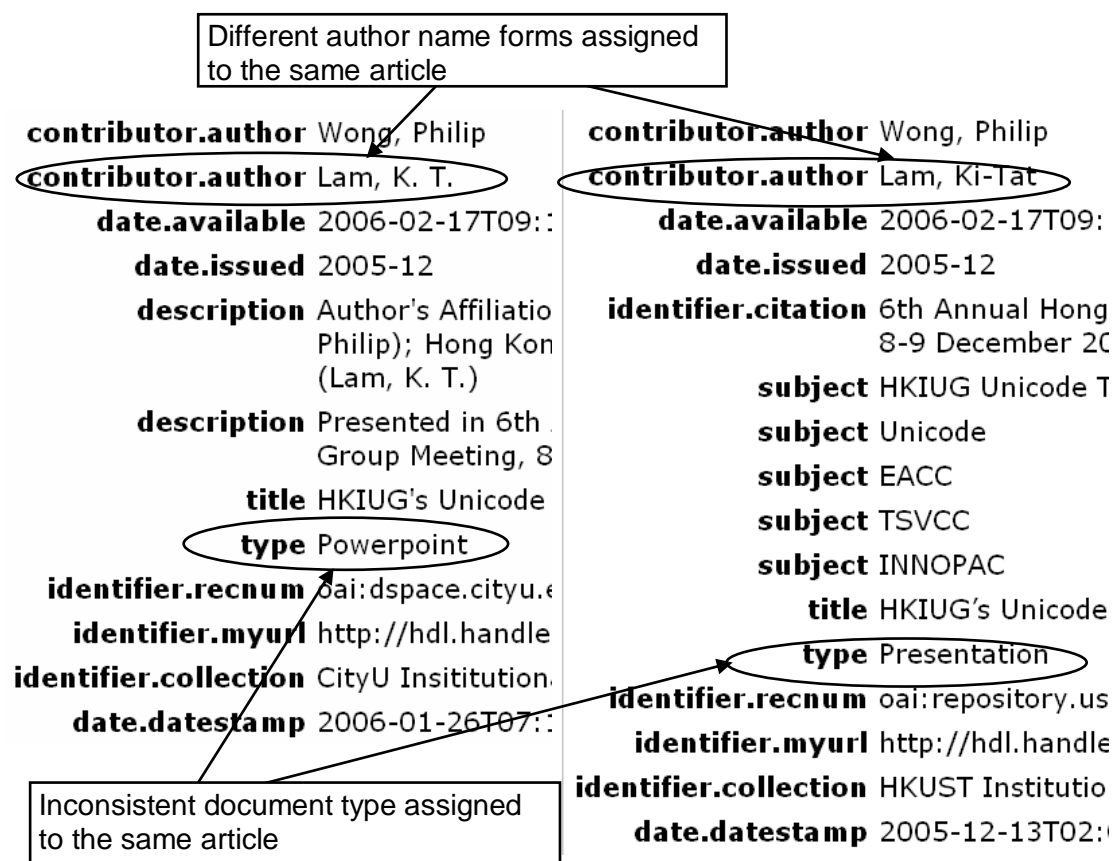
### ***The HKIR Experiment***

Other universities in Hong Kong have also started to build their institutional repositories. There is an emerging need to share IR experiences among them and to collaborate. One of the possibilities is to develop a union repository for scholarly output in Hong Kong. To demonstrate the feasibility of such collaboration and to study the issues involved, the Library developed an experimental system called HKIR (Hong Kong Institutional Repositories) in February 2006. The system is powered by the DSpace software, with OCLC's OAIHarvester2 software for harvesting OAI metadata. As of September 2006, six collections of ETD and institutional repositories

from five Hong Kong universities were created, allowing cross-searching of local scholarly output.

A number of issues were identified during the study. Many of them are related to the standardization of metadata description among institutions. These include standardization in author names, subject analysis, document types and metadata schema (Figure 6).

Figure 6. Two records of the same article in HKIR (<http://lbapps.ust.hk/hkir/>), showing different metadata description



Another problem related to OAI harvesting was also identified during the study. While DSpace uses qualified Dublin Core as the metadata schema, OAI's default metadata format `oai_dc` requires unqualified Dublin Core. As a result, metadata that contains qualifiers, such as `identifier.citation`, would become `identifier` after the OAI harvesting. Unless there is revision from the `oai_dc` schema, local institutions will be required to use a HKIR defined metadata format.

## Conclusions

The open access movement began with the establishment of SPARC to address market dysfunctions in scholarly publishing, followed by the formation of the OAI to promote author self-archiving and interoperable standards. After almost a decade of hard work, the authors see a number of good converging signs: publishers are more supportive of open access, the number of open access journals continues to grow, research funding bodies understand and better embrace open access, and more importantly, the flourishing of institutional repositories to preserve scholarly output and to make it openly accessible.

Installing IR software such as DSpace is straightforward, but tailoring the software and setting up policies and procedures to make it work effectively in one's institutional environment are uphill tasks. Even more difficult is the effort needed to recruit content. IR providers need to continue to educate researchers about the IR and encourage them to deposit their research to the Repository. They also need to campaign for government support.

While more and more institutions in Asia are beginning to develop their own repositories, the authors see the need of experience sharing, collaboration and standardization. HKUST is an early implementer of institutional repositories in Asia and its unique experience will be useful to similar institutions in this region.

## References

- Chan, D. (2004a). "Managing the challenges : acquiring content for the HKUST Institutional Repository", *International Conference on Developing Digital Institutional Repositories: Experiences and Challenges*, Hong Kong, December 9-10, 2004, California Institute of Technology Libraries, Pasadena, CA, and the Hong Kong University of Science and Technology Library, Hong Kong, available at <http://hdl.handle.net/1783.1/1973> (accessed September 28, 2006).
- Chan, D. (2004b). "Strategies for acquiring content : experiences at HKUST", *International Conference on Developing Digital Institutional Repositories: Experiences and Challenges*, Hong Kong, December 9-10, 2004, California Institute of Technology Libraries, Pasadena, CA, and the Hong Kong University of Science and Technology Library, Hong Kong, available at <http://hdl.handle.net/1783.1/1974> (accessed September 28, 2006).
- Chan, D., Kwok, C. and Yip, S. (2005). "Changing roles of reference librarians : the case of HKUST Institutional Repository", *Reference services review*, v. 33, no. 3, pp. 268-282, available at <http://hdl.handle.net/1783.1/2039> (accessed September 28, 2006).
- Kwok, C., Chan, D. and Wong, G. (2006). "From idea to reality: building the HKUST Institutional Repository", *University library journal*, v. 10, no. 1, March, available at <http://hdl.handle.net/1783.1/2528> (accessed September 28, 2006).
- Lam, K.T. (2004). "DSpace in action : implementing the HKUST Institutional Repository system", *International Conference on Developing Digital Institutional Repositories: Experiences and Challenges*, Hong Kong, December 9-10, 2004, California Institute of Technology Libraries, Pasadena, CA, and the Hong Kong University of Science and Technology Library, Hong Kong, available at <http://hdl.handle.net/1783.1/2023> (accessed September 28, 2006).
- Lam, K.T. (2006). "Exploring IR technologies", *Workshop on Managing Scholarly Assets in Institutional Repositories: Sharing Experiences Among JULAC Libraries*, Hong Kong, February 24, 2006, the Hong Kong University of Science and Technology Library, Hong Kong, available at <http://hdl.handle.net/1783.1/2501> (accessed September 28, 2006).