An OAI-based Approach to Build and Maintain Union Catalogue of OPACs

Francis Jayakanth, Sharada B, and Filbert Minj

National Centre for Science Information

Indian Institute of Science

Bangalore - 560 012

{franc, sharadab, filbert}@ncsi.iisc.ernet.in

Abstract: Libraries world-over are facing an economic crisis due to the imbalance in skyrocketing of subscription prices and shrinking or at best stable library budgets. Librarians have long ago realized the fact that however rich their collections are, they will not be able to meet the ever growing demands of their clientele. And, for this very reason, the inter library loan (ILL) service was started by the librarians so that it could, to some extent, supplement their collections. However, for the ILL service to be effective, the participating libraries in the ILL service should have their online public access catalogues (OPACs) in place. Also, having a union catalogue (a centralized database) of the participating libraries will make the service more efficient. Past experiences in creating and maintaining union catalogues clearly reveal that maintaining the currency of the union catalogues is not an easy task, especially, if the processes of obtaining the updates of the OPACs from the member libraries and updating the union catalogue involve manual procedures. In this paper, we propose an OAI-based approach to build and maintain union catalogues. In this approach the individual OPACs (data providers) expose their metadata for harvesting. The OAI-based service providers harvest the exposed metadata automatically. The metadata thus harvested is ingested into a centralized database, which will serve as an online union catalogue, supporting search, browse, and other value-added functionalities for the end-users. The OAI-based approach will eliminate the manual work involved in getting the OAPC updates and also in updating the union catalogues.

Keywords: OPAC; Union Catalogue; OAI-PMH; Interoperability; Metadata Harvesting; Static Repository

1 Introduction

A well-maintained union catalogue will serve as a discovery tool for the end-users. Therefore, maintaining the currency of union catalogues is very essential. The currency of the union catalogues is, however, dependent on the promptness with which the updates are received from the libraries that are part of a union catalogue. Libraries being growing organisms (Ranganathan, 1963), new documents are added constantly to a library collection. The visibility of such collections will improve if OPACs are built and made

available for the end-users rather than just having a printed card catalogues. To further improve the visibility of the individual, distributed OPACs, an online union catalogue is a necessity. So, an Online Union Catalogue is a system that stores the geographically distributed catalogue information from heterogeneous OPACs at single location. All the catalogue information stored in a union catalog will have a field pointing to its original location. This enables users to find the details of the resources they are looking for at a central repository and therefore do not have to search the OPACs individually.

In a well-maintained union catalogue, getting the updates from the individual OPACs and updating the union catalogue itself are on-going processes. The developments in the information and communication technologies (ICT) do facilitate in sending / getting the updates of the individual OPACs over the computer networks. However, the process may involve some amount of manual work with respect to exporting of new / updated records from the OPACs and making them available on a computer server accessible over the network or in transferring them to a remote computer server. Also, the process of updating the union catalogue itself may involve some manual work.

Another means of providing virtual union catalogue service is through the usage of Z39.50 protocol. Z39.50 is a client server protocol for searching and retrieving information from remote computer databases. It supports a number of actions, including search, retrieval, sort, and browse. In practice, however, the functional complexity is limited by the uneven implementations by developers and commercial vendors (http://en.wikipedia.org/wiki/Z39.50).

Yet another means of maintaining an online union catalogue of OPACs is by harvesting metadata from the individual, distributed OPACs facilitated by the open archives initiative protocol for metadata harvesting (OAI-PMH) (Lagoze, et al, 2002), established by the open archives initiative (OAI).

OAI-PMH provides an application-independent interoperability framework to share metadata among information systems. The framework is at present predominantly used for the world-wide consolidation of eprint archives. We are proposing the same concept for building and maintaining of union catalogue of OPACs, wherein, the participating libraries expose their metadata, which are harvested and ingested into the union catalogues thus eliminating human intervention in maintaining the currency of union catalogues.

However, to implement an OAI-based system, the data providers (OPACs) should be OAI-compliant. The OAI-based approach for interoperability is relatively a new concept. Many of the library OPACs, which have been in

existence for several decades now and are being maintained by a variety of library automation software packages, can not be expected to be OAI-compliant.

There are several commercial library automation packages available in the market today. Several more are available as freeware and a few are available as open-source packages. Koha, an integrated library system, developed initially in New Zealand by Katipo Communications Ltd. (http://www.koha.org/) is one of the best examples for an open-source library automation package. In India, many of the research libraries and universities are using commercial library automation packages like LibSys (http://www.libsys.co.in/home.html), SLIM (http://www.slimpp.com/SlimPPSITE/why_slim.asp), NewGenLib (http://www.kiikm.org/newgenlib.html), SOUL (http://www.inflibnet.ac.in/soul/soul.htm) and host of other packages. Several other libraries use CDS/ISIS, a specialized tool meant for handling textual data, developed by UNESCO since 1988, for maintaining their OPACs. A quick review of features of these packages indicates that except NewGenLib, none of the other packages mentioned above are based on open standards hence harvesting of metadata from such OPACs is not feasible. We therefore propose two possible approaches to make OPACs OAI-complaint so that metadata from OAI-compliant OPACs can be harvested by the OAI-based service providers.

2 Union Catalogue Models

There are two possible approaches to build union catalogues – Virtual and centralized. In a virtual union catalogue, the catalogue is not maintained at a single location but would be created in real-time by searching each of the member libraries' OPACs through the Z39.50 protocol. This would eliminate the redundancy of record storage as well as the expense of loading and maintaining access to the central catalog (Coyle, 2000). This approach may suffer from high network latency and uncertainty about the availability of servers. An alternative solution is typified by the OCLC WorldCat project (http://www.oclc.org/worldcat/) that collects bibliographic data from libraries all over the world into the OCLC Online Union Catalog (Suleman, Atkins, Gonçalves, France & Fox, 2001).

Our proposal of building and maintaining an OAI-based union catalogue is a centralized approach, wherein a single centralized database will hold all the bibliographic data from the heterogeneous, distributed OPACs. However, in this approach, the metadata from the distributed OPACs will be harvested automatically by OAI-based service providers. Harvested data will be used to

build a central database, which will serve as an online union catalogue providing search, browse, and other functionalities. The important requirement in this approach is that the participating OPACs should be OAI-compliant. As many of the existing OPACs are not OAI-compliant, we suggest two approaches to make the OPACs OAI-compliant – static and dynamic.

3 Static and Dynamic approaches to make OPACs OAI-compliant

OAI-PMH is becoming a de facto standard for the interoperability of heterogeneous, distributed, OAI-compliant digital libraries (DLs). It is a low-barrier, non-intrusive, and HTTP-based request-response transaction framework for communication between an OAI-based service provider (harvester) and OAI-compliant data providers. An OAI-compliant data provider exposes metadata, which can be harvested by OAI-based service providers. In a typical OAI-compliant system (such as the software DSpace (htttp://www.dspace.org) or GNU EPrints.org (http://www.eprints.org) for institutional repositories), the OAI interface of the system act as an intermediary between an OAI-compliant data provider and an OAI-based service provider (such as OAIster (http://oaister.umdl.umich.edu/o/oaister/), ARC (http://arc.cs.odu.edu/)).

Library OPACs are being maintained using various software packages. Most of these packages do not comply to open standards. Hence interoperability amongst such OPACs is not feasible. However, by way of static and dynamic approaches, it is possible to make the OPACs OAI-compliant so that they become interoperable.

3.1 The Static (repository) Approach

In the static (repository) approach (Jayakanth, et al., 2005), the records from an OPAC are exported to a textual file. This file is then converted into the format that conforms to the static repository xml file (http://www.openarchives.org/news/news2.html#StaticRepo). Once this has been done, the xml file can be ingested into the Kepler system (http://dlib.cs.odu.edu/#kepler), a light-weight, self-contained, OAI-compliant tool developed for individuals to archive their documents. Alternatively, the xml file can be made OAI-compliant through the intermediation of a static repository gateway (SRG) (Hochstenbach P, et al, 2003) (http://srepod.sourceforge.net/). The SRG approach has a limitation with respect to number of records. As of now, the static repository file cannot have more than 5000 records. Initially, when an OPAC needs to be made OAI-compliant, all the records are to be exported from an OPAC and the exported records are converted into a static repository (an XML file). This file could be imported into the Kepler system thus making the records contained in an

OPAC, OAI-compliant. Alternatively, depending on the number of records contained in an OPAC, number of static repository files (each file containing 5000 records) can be created. Each of these static repository files can be made OAI-compliant through the intermediation of SRG. These records will then be harvested by the OAI service providers upon registration. Figure 1 illustrates the system diagram for the static approach.

Subsequent harvesting will just be incremental and may not involve in the creation of multiple static repository files. The static (repository) approach does involve manual work, contrary to the notion that OAI-compliance facilitates transparent interoperability. This contradiction is due to fact that the OPACs being maintained using variety of library automation packages, do not adhere to open standards, which are publicly available and implementable.

An alternative way to make the OPACs OAI-compliant is the dynamic approach, wherein the OAI service providers will be able to interact with the OPAC data in real-time.

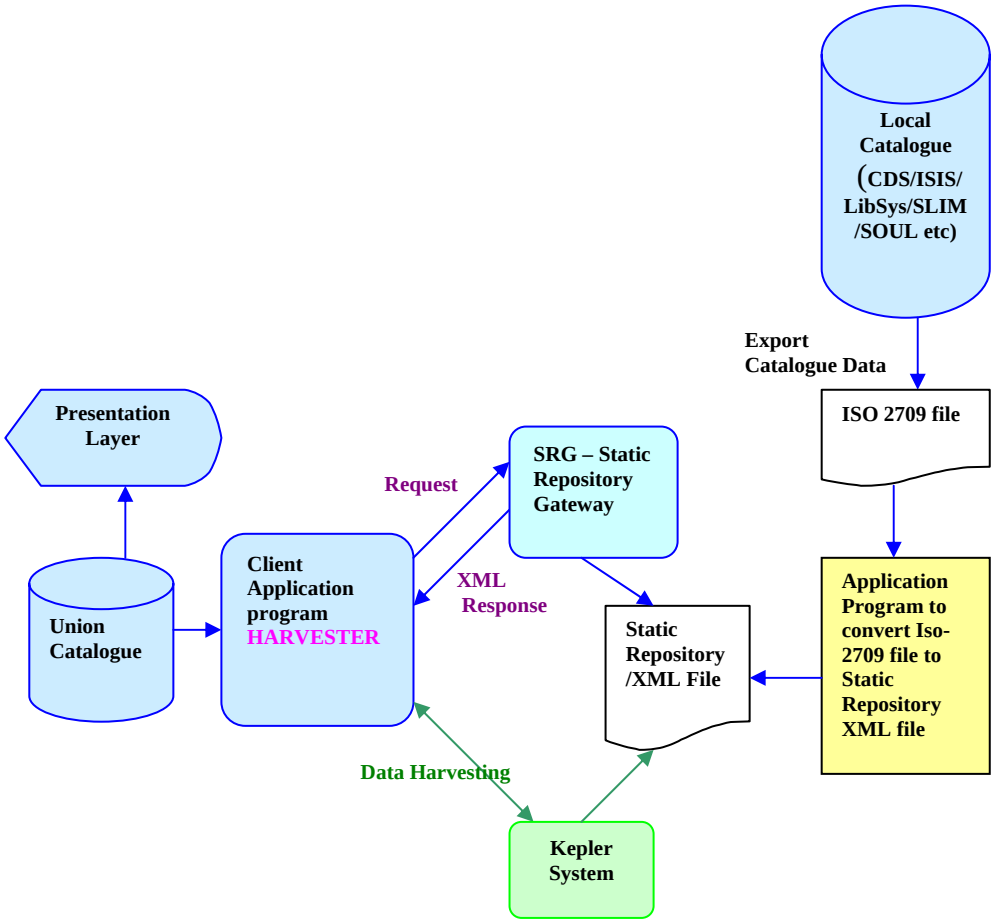

Figure 1: Union Catalogue – Static Approach

3.2 The Dynamic Approach

In the dynamic approach (Jayakanth, et al., 2006), the OAI service providers will be able to interact with the OPACs in real-time through the OAI interfaces of data providers available at specific URLs. In such of those OPACs, wherein the metadata about the resources are stored in relational database management systems (RDBMS), it is straightforward to make such databases OAI-compliant provided such RDBMS have ODBC drivers for database interaction, programmatically. Several open-source tools http://www.openarchives.org/tools/tools.html) are available, which facilitate the data providers (OPACs) to be OAI-compliant.

Once the individual OPACs are made OAI-compliant, the base URLs of such OPACs are to be registered with an OAI-based service provider. The service provider will then periodically harvests metadata from the registered OPACs and update their central index. There are several open-source software (such as ARC (http://arc.cs.odu.edu) and Public Knowledge Project (PKP) harvester (http://pkp.sfu.ca/harvester/)), which facilitate federated search service based on OAI-PMH. Figure 2 illustrates the system diagram for dynamic approach.

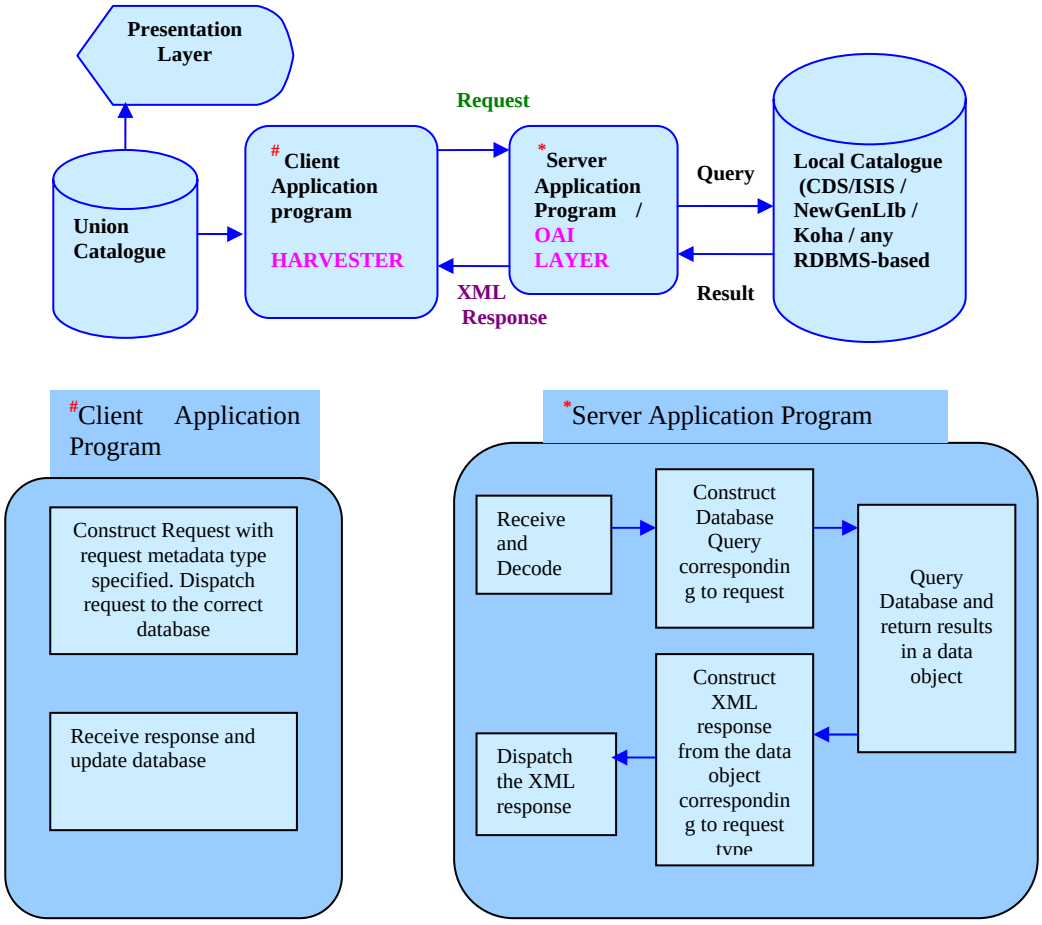

Figure 2: Union Catalogue - Dynamic Approach

4 Conclusions

Library OPACs have been in existence for several decades now. Hence the heterogeneousness amongst the OPACs is to be expected. Asking the librarians to switch over to OAI-compliant software for maintaining their OPACs is not practical. Instead, finding a solution to make their existing systems OAI-compliant is a much better option so that such OPACs become interoperable. For such of those OPACs, which are being maintained using proprietary software and access to OPAC data is not possible in the real-time; static (repository) is the only possible solution. This approach does involve some amount of manual work in periodically exporting the newly added records and also the ones that have been modified since the previous export. In fact this process of exporting the new and modified records and converting them to static repository (XML file) can be automated by writing a simple script by the OPAC maintainers. The dynamic approach is the preferred way to make data providers OAI-compliant. Depending on the nature of the software being used for maintaining the OPACs, either an open-source tool could be used to make such OPACs OAI-compliant or OPAC specific tools may need to be developed.

5. Acknowledgement

The work reported in this paper is part of on-going project, Development of OAI-based Institutional Research Repository Services in India. The project is sponsored by department of scientific and industrial research (DSIR), Govt. of India. We take this opportunity to thank DSIR for funding this project.

6 References

1. Coyle, K.. (2000). The Virtual Union Catalogue. D-Lib Magazine, 6, 3.. http://www.dlib.org/dlib/march00/coyle/03coyle.html

2. Hochstenbach, P., Jerez, H., & Van, De, Sompel, H. (2003). The OAI-PMH Static Repository and Static Repository Gateway. *Joint Conference on Digital Libraries 2003, Houston, Texas*. http://public.lanl.gov/herbertv//papers/jcdl2003-submitted-draft.pdf

3. Jayakanth, F., Maly, K., Zubair, M. & Aswath, L. (2005). Approches to make CDS/ISIS databases interoperable with OAI-compliant digital libraries. Program: electronic library and information systems, 39, 3, 269-278. http://eprints.iisc.ernet.in/archive/00003643/
4. Jayakanth, F., Maly, K., Zubair, M. & Aswath, L. (2005). A dynamic Approach to make CDS/ISIS databases interoperable over the Internet using the OAI protocol. Program: electronic library and information systems, 40, 3, 269-278. http://eprints.iisc.ernet.in/archive/00008252/

5. Lagoze, C., Van, De, Sompel, H., Nelson, M. & Warner, S. (2002) The Open Archives Initiative Protocol for Metadata Harvesting, Version 2.0. June 2002. http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

6. Ranganathan, S., R. (1963). The Five Laws of Library Science. Ranganathan Series in Library Science (12). Madras. India. Asia Publishing House.

7. Suleman, H., Atkins, A., Goncalves, M.A., France, R.K. & Fox, E.A. (2001). Networked Digital Library of Theses and Dissertations: Bridging the Gap for Global Access – Part 2: Service and Research. D-Lib Magazine, 7, 9. http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html