Installation and Customization Experience of Metadata Harvester System: Case from University of Ruhuna

Fernando, I.D.K.L[1]and Hettiarachchi, N[1]

Abstract

University of Ruhuna is a thriving modern Sri Lankan university with strengths in both research and teaching. Library system of University of Ruhuna provides an enormous contribution to the university in order to provide access to scholarly information through journals (online and printed), books, e-repositories and other means. Library network of the university spend substantial amount of money for subscribing to printed and electronic journals annually. Access to paid scholarly literature is limited depending on the subscription type to the database or journal. There is a renewed trend of searching for open-access (OA) knowledge which break the price boundaries and provide better access to the current knowledge among academic community. Open access repositories provide searching and downloading of scholarly information materials for free of charge. Spending extra time on sorting and filtering of OA-materials by information seekers out of the search results obtained has been a major demerit of using OA knowledge. Finding and searching on different open access repositories separately is also been a difficult and time consuming task. A unified interface for searching on different OA repositories worldwide can facilitates readers by reducing the searching time and effort of different scholarly materials and OA repositories in the world. Harvester System of University of Ruhuna (HaSURu) deployed in order to fulfil the gap of a unified interface that can connect number of OA archives together and search over the internet for OA information. HaSURu keeps OA information of OAI-registered open-access archives worldwide. OAI is an organization that promotes the better dissemination of knowledge among researchers. There are 1944 OA repositories worldwide has registered under OAI as data providers. Present study aimed at acquiring all the OA repository information from OAI and provide access through a unified interface using OAI-PMH driven metadata harvester. This paper discusses the process of installing and customizing PKP OHS Harvester2 as OA information finder carried out at University of Ruhuna. All the Information seekers

---

[1] Corresponding author:

Assistant Librarian, Main Library, University of Ruhuna, Sri Lanka.

E mail: kusala@lib.ruh.ac.lk

[1] Senior Assistant Librarian, Main Library, University of Ruhuna, Sri Lanka.

E mail: nimal@lib.ruh.ac.lk

of Sri Lanka will find this new service as a better source of access to OA repositories worldwide for their future research work.

Key words: Metadata harvester, Open-access repositories, OAI-PMH, PKP, OHS

Introduction

University, as defined by the Oxford Dictionary (Oxford University press, 2007, p. 3445) is,

"A corporation of teachers and students formed for the purpose of giving and receiving instructions in a fixed range of subjects at a level beyond that provided at a school. Later, and institution of higher education, offering courses and research facilities in mainly non-vocational subjects and having acknowledged powers and privileges, esp. that of conferring degrees. "

In other words, Universities are 'high-level educational institution in which students study for degrees and academic research is done' (Oxford University press, 2007, p. 1415). Therefore academic institutes conducts massive amount of research in various disciplines. Scholars who engaged in research need to get access to relevant academic knowledge (literature) to conduct their research studies successfully.

Prime responsibility of an academic library is to facilitate researchers (Brandt, 2007, pp. 365-396) with effective, efficient and reliable sources of scholarly literature from a wide spectrum of subject array. Libraries get subscription to electronic and printed journals from various subject disciplines in order to access to the current knowledge. Library system of University of Ruhuna (UOR) has currently subscribed to an array of printed and electronic journals. In addition to this, university also receiving journals as gifts and there are avenues to exchange journals to fulfill the literature requirements of academics. Journals are primary means of sharing scholarly knowledge among academics (Chan, Gray, & Kahn, 2012, p. 7). According to the Library statistics, university of Ruhuna has spends substantial amount of money from the acquisition budget to get the subscription to online and printed journals annually. Depending on the amount of budget allocations, the university can only access to a limited collection of journal articles under the paid

subscription to databases and journals. Paid subscription to databases is very expensive and there are also hidden limitations behind the technical walls and license restrictions (Charles & Bailey, 2006). Even though there are a number of means to access to current knowledge through journals. The present paid subscription to databases and individual journals is not sufficient to cover all subject disciplines of the University and therefore it is essential to find supplementary means to fill out the lacuna of the information supply in the university.

Open Access to Research

There is a renewed focus on browsing for Open Access (OA) scholarly literature among academics (Chan, Gray, & Kahn, 2012, p. 5) in universites. Access limitations resulted from the copyright, patent, and licensing of commercial publications (Chan, Gray, & Kahn, 2012, pp. 5-10) is also tend readers shifting towards OA materials.A study conducted by Swan & Brown (2005) shows that most of the scholars tend to use Google Scholar for searching desired academic information. According to European Commission, Open-access is "free access to the publicly funded academic knowledge that includes actual research publications, research data, and variety of other digital media and objects" (2008).Open-access journals are free to download and use their contents with proper acknowledgements (Charles & Bailey, 2006) to the original authors.

Requirement of a Unified Searching Interface

Searching result produced by Google or other online search engine may contain both free and fee based information sources in the search result (Jacsó, 2005). Information seekers have to spend extra time and resources to sort out and filter the desired OA information among miscellaneous records with a noise. To avoid this problem, information seekers have to find OA information source and search on desired OA information separately (Liu, Kurt, Zubair, & Nelson, 2001). Finding information about different types of OA resources and searching on them separately is a difficult and time consuming task (Donaldson & Nelson, 2011). A unified searching interface can address this issue by amalgamating different OA sources together and providing a unified interface to perform search queries

(Liu, Kurt, Zubair, & Nelson, 2001). It will provide information seekers to reach to their desired information in an effective and efficient way.

Aim of the Present project is to design a unified interface for searching and accessing to open-access repositories registered in OAI (Open archves initiatives, 2002). OAI provides a comprihensive list of Open access archives in the world.

Open Archives Initiative (OAI)

OAI is an organization that promotes Open access of reseach in the world. They keep a record of all the open access archives who registerd with OAI in the world. These repositories contain information from a broad array of subjectes and with multidisciplinary knowledge. According to the (Open Archives Initiative, n.d) 'Open Archives Initiative (OAI) develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content' and increasing the availability of scholarly communication. The fundamental technological framework and standards that are developing to support this work are independent and opening up access to a range of digital materials. OAI maintains a list of 1944 data providers who registered and confirmed by OAI in order to disseminate scholarly knowledge among researchers (Open archves initiatives, 2002). Community can get use of these OA information sources with a proper acknowledgment.

There are various projects maintained by OAI. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a protocol that design to facilitate interoperability of information repositories through metadata exchange (Open Archives Initiative, n.d).Users can access to data providers through the OAI-PMH to download (harvest) its metadata information. Data providers are administer systems that supports OAI-PMH as a means of exposing metadata to harvesters (Lagoze & Sompel, 2008). Lagoze & Sompel (2008) explained OAI-PMH as an "application-independent interoperability framework based on metadata harvesting from metadata providers". OAI-PMH is a low-barrier mechanism for repository interoperability that provide a set of six verbs, requests or services that are invoked within Hyper Text Transfer Protocol (HTTP) (Open Archives Initiative, n.d).

Harvesters use these verbs to fetch and request metadata information form data providers. Metadata Harvesters are designed to facilitate this requirement of incorporating different archives into a single database and let users to search all of these archives in the same time through a unified searching interface(Liu, Kurt, Zubair, & Nelson, 2001; Public knowledge project, 2012). Lagoze & Sompel (2008) explains Harvester as a 'client application that issues OAI-PMH requests and collects metadata from other online repositories worldwide'.

Metadata

Metadata are 'structured data about data' or descriptive information about an object or resource whether it is physical or electronic (Greenberg, 2010, p. 3611). Metadata can be available in various standards (Greenberg, 2010, p. 3612). Most of the Digital libraries and repositories use Dublin Core (DC) as the major format of metadata (NISO, 2004). DC is a simple form of metadata which organized under 15 headings (NISO, 2004). Contents hosted in the archives are a complex of the digital material and metadata. Digital materials can be born digital or digitized and exists in different formats (documents, music, image, video, etc.). Metadata are carrying all the bibliographic information of the digital materials hosted in the archive (William, 2000, p. 14). Each digital material that hosted in data provider (repositories) are incorporated with metadata (Greenberg, 2010, p. 3610). Harvester application can access to external data providers (data sources) using the OAI-PMH and download only the metadata information related with the digital materials hosted in the archive.

Universities, research institutes and other database vendors around the world produce their own Digital Libraries (DL), Institutional repositories (IR) or Digital Archives (DA), Open Journal systems (OJS) etc. to collect, store and disseminate scholarly materials of different file formats (William, 2000, p. 70). Most of these DLs, IRs and DA are publicly available through internet (McCray & Gallagher, 2001, p. 51) and provide their content to share by researchers (Chan, Gray, & Kahn, 2012).

PKP Metadata Harvester

PKP Harvester2 is an open source metadata harvester and aggregator that have been developed by the Public Knowledge Project (Public Knowledge project, 2013) which aimed at expanding and improve access to global research. Harvester2 (Public knowledge project, 2010) has designed as a flexible tool for fetching, storing, indexing and searching data from different types of information sources (Liu, Kurt, Zubair, & Nelson, 2001). Harvester2 supports multiple harvesting protocols versions (OAI version 1 and 2), metadata standards (Dublin core, MODS, MARC), and languages with an emphasis on performance and simplicity of use (Public knowledge project, 2010). Among different types of harvesters, PKP Harvester2 provides easy management and installation of the base system. It can be further customized by designing new plugins, patches to the base system (Public knowledge project, 2010).

Materials and Methods

An open source software platform named Public Open Harvester System (OHS) that developed by Knowledge Project (PKP) was used to deploy the Harvester System for the University of Ruhuna (HaSURu). Following steps were accomplished in the process of deployment.

Installation of the Harvester System

Present study used the latest version (OHS-2.3.2) of the PKP open harvester system that installed on Ubuntu Linux 12.04.1amd64 LTS. OHS system uses MySQL 5.5.28 as the relational database, PHP5 5.3.10, and Apache2 2.2.22 as the HTTP server.

HP PortLiant ML350 G6 server was used to install the base system of the harvester. HP PortLiant ML350 G6 server is configured with redundant power and Network supplies. Redundant power supply was used to install two separate Un-interrupting Power Supply (UPS) systems. Redundant network connections were configured for the LAN/WAN access and local backup management separately (Figure 01). Having separate connections for LAN and remote backup storage access will be increased effective through put of the

cables. Since the read/write (I/O) ratio is high in the Harvester, total storage of 1TB was configured as RAID 0. RAID 0 provides only the striping with zero redundancy. Mirroring of storage space was not configured since separate backups were maintained in another local server. Scheduled updating and indexing scripts will keep the system up-to-date with the frequent changes in the data providers. This will enhance the searching efficiency. The files and directories were write-enabled (including config.inc.php,public, cache, cache/t_cache, cache/t_config, cache/t_compile, cache/_db). Client, connection and database character sets were set to Unicode (UTF-8) to get better encoding functionality in different locales and MD5 was used as the password encryption algorithm.

Customization

Since the OHS is a common platform, it should have to be modified according to the requirement of the institute. The initial interface, security and access policies were considered in the customization step. The browsing content of the system was organized under four access points: Author, Title, Subject and Date.
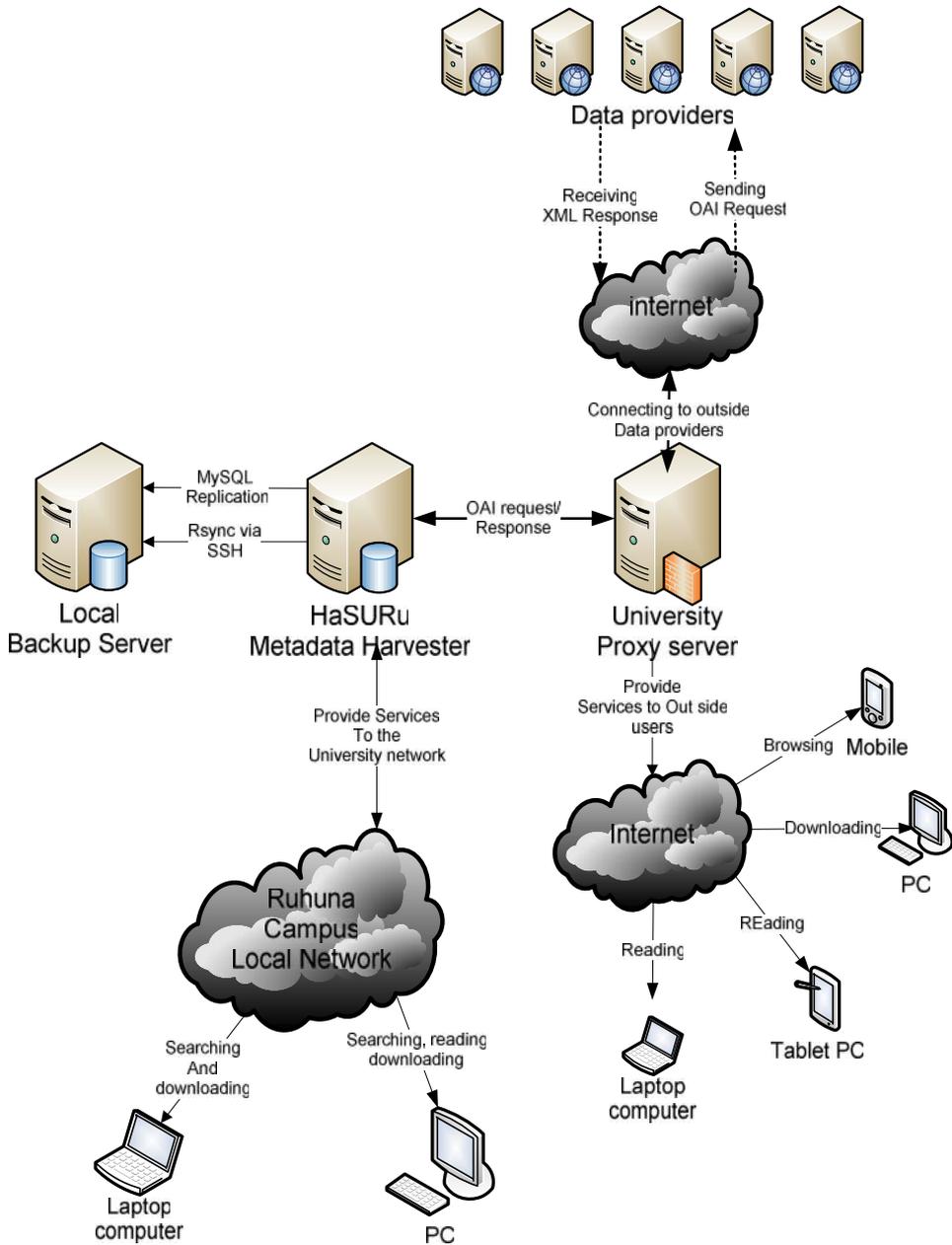
Figure 01. An overview of Harvester System-University of Ruhuna

Adding data Providers' to the Harvester

Harvester works with the metadata information downloaded by another open access archive called data providers. These data providers are identified and access by an OAI Base-URL and URL. These URL information should be configured into the Metadata Harvester. PKP Harvester comprise with a special interface namely 'Add archive' for inserting URL details of data providers (Figure 02). Harvester will automatically updating other information related with the desired data provider and writing information to the database.

Add Archive

Title*	
Description	
Archive Image	Choose File No file chosen Upload
URL*	e.g. http://www.yourarchive.com ☑ Enabled
Public ID	This unique identifier can be used in URL-based searches to identify this archive.
Type*	OAI ▾
OAI Base URL*	Fetch archive metadata e.g. http://www.yourarchive.com/oai/index.php
Admin Email	
Options	☐ This is an OAI Static Repository
Index Method*	ListRecords ▾
Metadata Format*	Click Refresh to update supported formats list for this archive. Dublin Core ▾ Refresh

Save   Cancel

* Denotes required field

Figure 02: Add archive interface

Number of data provider information can be added through the 'add archive' interface. Since the harvesting process is a time consuming task, should be initiated after defining

several data providers in the harvester. Harvesting of all addedd archives can be initiated by providing the following command at linux shell (root terminal).

#php[dir]/tools/harvest.php all useLastSets

New addition of data providers can be proceeding after harvesting metadata from initially defined archives. After defining different data providers in each update, the harvesting process can be initiated by using the following command.

#[dir]/tools/phpharvest.php all from=last skipExistingEntries

This command will update newly added data provider information only. If the first command used, it will update all the records from the beginning including the existing entries. This will take a long time to complete the metadata base. Since it using long-lived transactions, can cause internet traffic under low bandwidths.

Indexed list of all the data providers added to the harvester (Figure 03) can be obtained by,

#php[dir]/tools/phpharvest.php list

```
70: arXiv (0 records)
71: ASDM Sonoran Desert Digital Library (6113 records)
72: Ask  Biologist (21 records)
73: Aston University Research Archive (4713 records)
74: Athenea Digital - Revista de pensamiento e investigaciÃ³n social (650 recor
ds)
75: AtÄ±lÄ±m Ãniversitesi AÃ§Ä±k EriÅim Sistemi (554 records)
76: AUP publications (1977 records)
77: AUT Scholarly Commons (2504 records)
78: AÃ§Ã£o Educativa (2170 records)
79: B3Kat Repository - Library Union Catalog of Bavaria, Berlin and Brandenburg
 (234857 records)
80: BahÃ§eÅehir Ãniversitesi AÃ§Ä±k EriÅim Sistemi (4 records)
81: Ball State Digital Media Repository (125777 records)
82: Baltic Marine Environment Bibliography 1970- (0 records)
83: Barrel (1 records)
84: Base de publications de recherche - UniversitÃ©Â® Paris-Dauphine (0 records)
85: BDJur no STJ (0 records)
86: Banques d'images et de portraits de la BibliothÃ¨que interuniversitaire de
santÃ© (Paris) (0 records)
87: Berkeley Law Scholarship Repository (0 records)
88: Berlin-Brandenburgische Akademie der Wissenschaften, edoc (2000 records)
89: Biblioteca Virtual Sierra-Pambley (84 records)
90: Biblioteca de la Universidad de La Laguna (1791 records)
91: Biblioteca Digital - Facultad de Ciencias Exactas y Naturales - Universidad
 de Buenos Aires (2852 records)
```

Figure 03: Output of the List command which displaying the archive ID, name
and number of records harvested from the data provider

Harvesting metadata

Information provided in the list can be used to harvest information from individual data providers separately. Depending on the archive ID, harvesting process can be initiated with following command.

#[dir]/tools/phpharvest.php [archive ID](archive ID = 81, 92 …)

Since the output (Figure 03) showing the list of data providers with the archive ID and number of records that harvested from that particular site, administrator can easily refer the newly defined data provider out of the list and execute further update process on it.

Updating harvester base system

Harvester system can be updated time to time with the release of new versions by the PKP. There are various ways to perform this task. Following commands can be used in linux command line interface to check and update the existing version to a new harvester version (Figure 04) .

#[dir]/tools/phpupgrade.php check
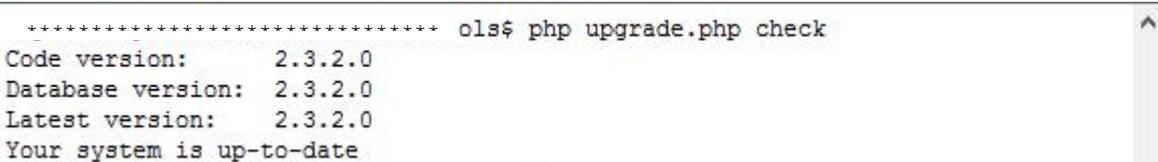

```
****************************** ols$ php upgrade.php check
Code version:       2.3.2.0
Database version:   2.3.2.0
Latest version:     2.3.2.0
Your system is up-to-date
```

Figure 04: output of the check command showing the existing and available product versions

Administrator can use following command to check for latest versions of harvester

#[dir] /tools/phpupgrade.php latest

After confirmed about a new release of the harvester base system following command can be used to download and apply the patch over the existing system

#php[dir]/tools/upgrade.php patch

Rebuilding the Search Index

Rebuilding of the search index can perform by issuing the follwoing command at the Linux command line interface.

#[dir]/tools/phprebuildSearchIndex.php

Rebuilding of the search Indexing need to be perform after creating new sort orders. Sort orders can be created depending on various metadata fields i.e. Author, Title, Date, etc. Re-indexing of the contents can increase the searching efficiency of the system.

Backups

MySQL replication was configured between the master server and a PC in the Local area network. Backup server will keep a mirror copy of the master database with same configurations at any given time. Replication system will result a fail-safe backup of the HaSURu system. Additionally, a pre-scheduled crontab script compress (tar.gz) and copy (rsync) the installed directory to the remote PC via SSH. Manual backup of the installation is taking after any updates to the base system. These updates are copied via secure shell (SSH) to the same local server that keeps the MySQL replication (Figure. 01). This combined system can be used as a fail-safe backup since it can used to restore the original database in case of a failure. All of these configurations will ensure the high availability, accessibility, and speed of the service.

Unified interface of HaSURu

Metadata harvester for university of Ruhuna is formed with the name of "HaSURu" - Harvester System University of Ruhuna. System can be accessed through the URL: http://dilru.lib.ruh.ac.lk/hasuru (figure 04). Currently HaSURu harvesting from 186 data providers out of 1944 OAI-registered data providers in the world. Harvesting process is currently running real time and have resulted 737700 of metadata records from various OA data providers. These records consist of metadata related with journals, articles, scholarships, thesis, dissertations and other scholarly materials.
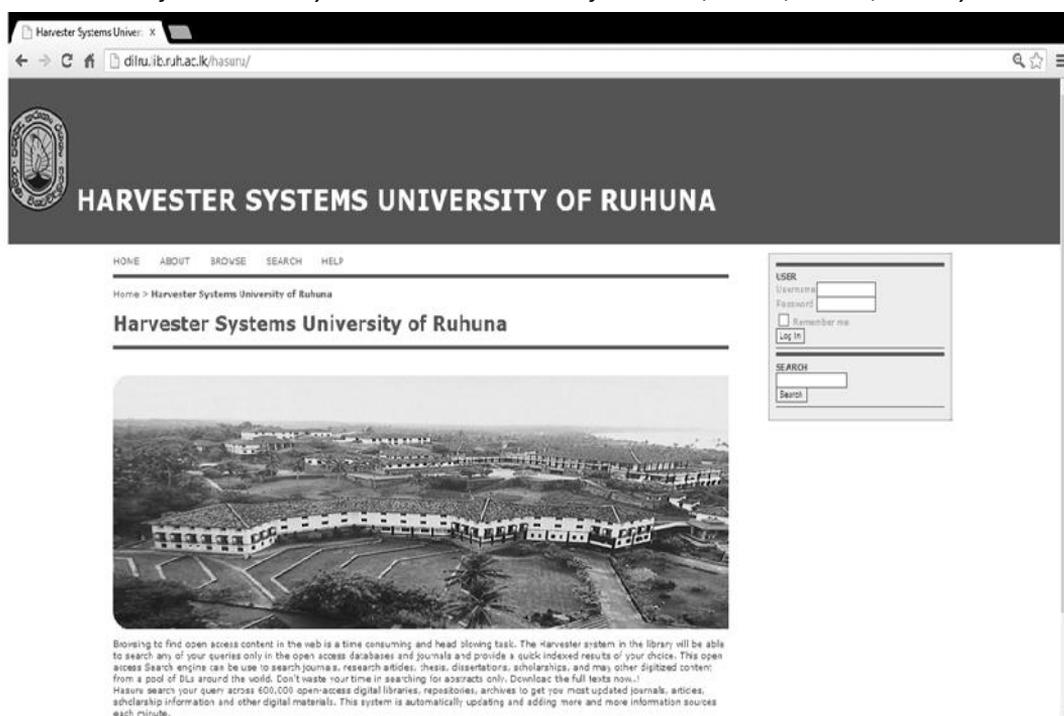
Figure 05: Homepage of Harvester system – University of Ruhuna (HaSURu)

Number of harvested metadata records increases daily with addition of new data providers to the harvester. Self-harvesting process is maintained by automated PHP script. Different user friendly interfaces provide easy access to information through the harvester.

Searching Interface

Harvester provides a search box (Figure 06) to perform search and advanced search using different searching techniques and operators. Search terms are case-insensitive and it ignores common words are ignored. It can generate search quires using simple Boolean operators (AND, OR, NOT) and more complex combination of quarries using parentheses e.g., archive ((journal OR conference) NOT theses). In addition to this harvester can search;

- for an exact phrase by putting it in quotes; e.g., *"open access publishing"*

- Exclude a word by prefixing it with **-** or *NOT*; e.g. *online -politics* or *online NOT*

- Use ***** in a term as a wildcard to match any sequence of characters; e.g., *soci* morality* would match documents containing "sociological" or "societal"

SEARCH

All	Library science
Archives	**All Archives**
	Universität Stuttgart, Fakultät 5, Germany, Computer Science Archive
	@rchiveSIC - @HAL
	A+BE | Architecture and the Built Environment
	A-plus

Search

Figure 06: Searching interface of the Harvester System

Information Organization of the Harvester

Information gathered in the harvester can be viewed in different forms according to the need of the user. As shown in Figure 07, information can be presented as an indexed list of records. This interface provides all the records indexed alphabetical order to the name of the archive (data provider). It also provides a connection to the original source through "view original" link. Through this link the user can get access to the original data providers archive and get the information required. Some data providers required to get registered into their archive through the archive administrator. All the facilities and relevant information can be found in the pages of the original data provider.

In other view it provides information of individual archives (Figure 08). In this interface user can find the name of the desired archive (data provider) all the contents harvested from that data provider, and the link to access to the original record.

Harvester also provides summarized details about all the data providers in the harvester (Figure 09). OAI-registered archives in the Harvester. This interface provide details about archive name, number of materials hosted, total number of data providers, and number of pages ahead.

Defining New access Points

There are different access points to the information indexed in the harvester system, e.g. Author, Title, Date, and Subject. Users are allowed to define their own indexes using different access points comply with available Dublin Core elements in the harvester.

Database Management System Use in HaSURu

HaSURu system uses MySQL database to store and manage harvested data. MySQL database management system has a specific feature of keeping mirror copy of a database in a remote location, which is named as MySQL replication. MySQL replication system keeping a fail-safe backups of the Harvester system. Together with backups, redundant power supplies, and network connections will ensure the data security and continuous uptime of the service.

Implementation Eexpenditure of the HaSURu system

Overall setup was initiated with a minimum requirement of Capital expenditure (CapEx) which include hardware cost for the server, backup server and UPS units. Operating system (Ubuntu Linux), PKP Harvester2 and other dependent software (MySQL, PHP5, and Apache2) are excluded from the CapEx since they are available under open source licenses. HaSURu with multidisciplinary academic knowledge can be accessed through both web and emerging mobile platforms.

Figure 07: Index of total content in HaSURu

1.	Link to the original data source
2.	Indexed list of data providers
3.	Number of total articles in the data base

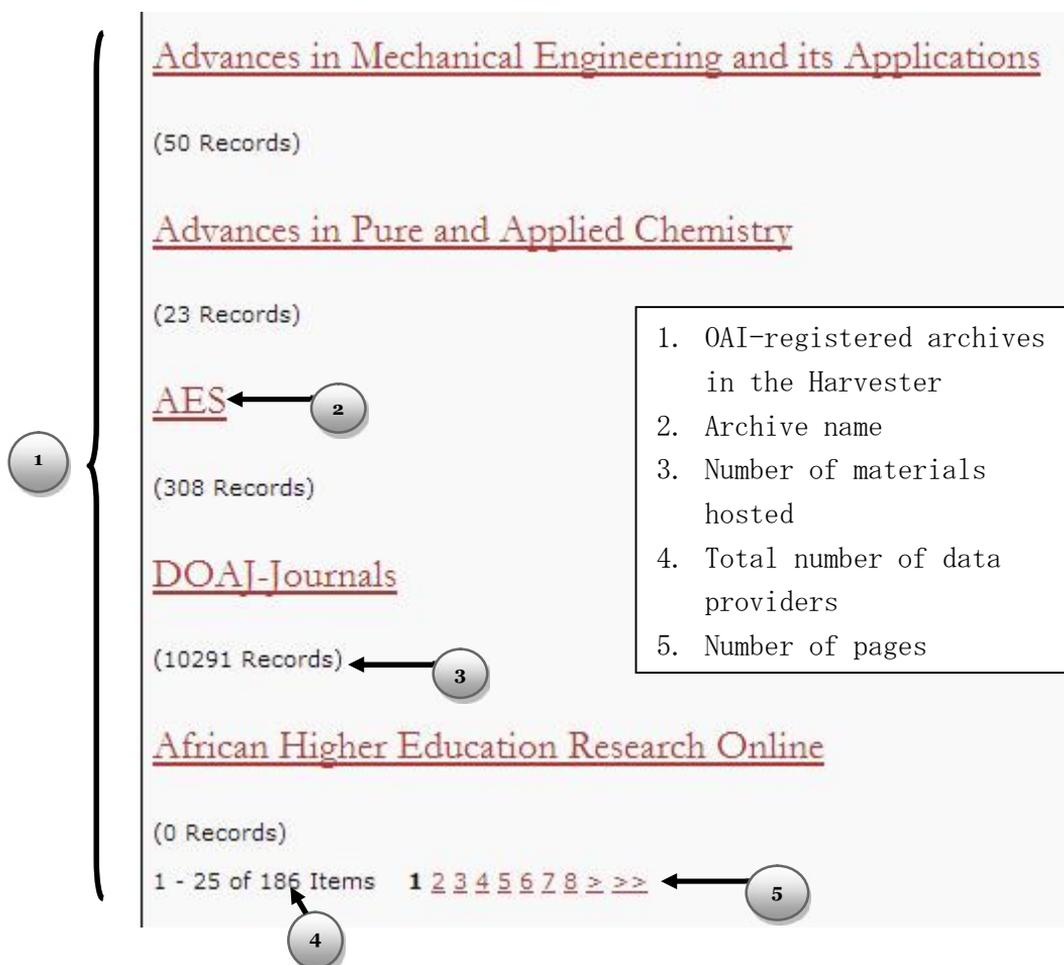**Figure 08: Information provided by a single archive**

Figure 09: List of archives that linked via HaSURu

Conclusion

Library system of University of Ruhuna successfully managed the challenge of providing access to free scholarly information by implementing the harvester system (HaSURu). The significant of HaSURu is it enables the searching through OAI-registered open access archives through unified interface. Since HaSURu covers most of the open-access archives in the world, it consists of multidisciplinary information from a broad subject array. Implementation of a metadata harvester is a low-cost solution for the journal requirement of a university.

Intended Expansions to the System

HaSURu system can be hosted in the centralized server in university of Ruhuna and all other university libraries can link to this service via the WWW in the form of Software as a service (SaaS) of cloud computing (library) concept. Data providers that registered on OpenDOAR will be able to add into the harvester as a second step of the implementation. Repositories hosted in 2227 (on 1st May, 2013) OpenDOAR consists of open-access repositories from Asia, Africa, Australia, Caribbean, Central America, Europe, North America, Oceania and South America to the HaSURu database.

References

Brandt, D. S. (2007). Librarians as partners in e-research Purdue University Libraries promote collaboration. *College & Research Libraries News,, 68*(6), 365-396.

Chan, L., Gray, E., & Kahn, R. (2012). Open Access and development: journals and beyond. Retrieved March 5, 2013, from http://www.ids.ac.uk/files/dmfile/OpenAccessandDevelopmentJournalsandBeyond_Nov2012_Gray_et_al.pdf

Charles, W., & Bailey, J. (2006). What is open access? Retrieved December 10, 2012, from http://digital-scholarship.org/cwb/WhatIsOA.pdf

Donaldson, R. L., & Nelson, D. W. (2011). The 2012 Promise of Open Access Textbooks:. Florida Virtual Campus. Retrieved February 13, 2013, from http://www.openaccesstextbooks.org/pdf/ModelDraft.pdf

European Commission. (2008). Open access: opportunities and challenges. Hand Book. Retrieved December 10, 2012, from http://ec.europa.eu/research/science society/document_library/pdf_06/open-access-handbook_en.pdf

Greenberg, J. 3. (2010). Metadata and digital information. In M. J. Bates, & M. N. Maack (Eds.), *Encyclopedia of Library and Information Sciences* (3rd ed., Vol. V, pp. 3610-3623). CRC press.

Jacsó, P. (2005). Google Scholar: the pros and the cons. *Online Information Review, 29*(2), 208-214. Retrieved March 2 , 2013, from http://www.emeraldinsight.com/journals.htm?articleid=1464936&show=abstract

Lagoze, C., & Sompel, H. V. (Eds.). (2008). *The Open Archives Initiative Protocol for Metadata Harvesting*. Retrieved December 10, 2012, from Open Archives: http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

Liu, X., Kurt, M., Zubair, M., & Nelson, M. L. (2001). Arc: an OAI service provider for cross-archive searching. *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*, (pp. 65-66). Retrieved March 12, 2013, from http://dl.acm.org/citation.cfm?id=379451

McCray , A. T., & Gallagher, M. E. (2001, May). Principles for digital library development. *Communications of the ACM, 44*(5), 49-54. Retrieved February 15, 2013, from http://lhncbc.nlm.nih.gov/dlb/pubs/200105_cacm_mccray.pdf

NISO. (2004). Understanding metadata. Retrieved December 11, 2012, from www.niso.org/standards/resources/UnderstandingMetadata.pdf

Open Archives Initiative. (n.d). *About OAI*. Retrieved Novenmber 21, 2012, from Open Archives Initiative: http://www.openarchives.org/OAI/OAI-organization.php

Open archves initiatives. (2002). *Registered data providers*. Retrieved November 22, 2012, from Open archves initiatives:http://www.openarchives.org/Register/BrowseSites

Oxford University press. (2007). *Shorter Oxford English Dictionary on historical principles* (6th ed.). Oxford: Oxford University press.

Oxford University press. (2007). *The Oxford dictionary, thesaurus, and wordpower guide.* Oxford : Oxford University press.

Pasqui, V. (n.d.). Digital preservation and open access archives: persistence access to open access digital assets. A briefing paper on Digital preservationeurope. Retrieved December 12, 2012, from http://www.digitalpreservationeurope.eu/publications/briefs/open_archives_pasqui.pdf

Public knowledge project. (2010). Harvester2: technical reference. London: Simon fracer university library. Retrieved December 15, 2012, from http://pkp.sfu.ca/harvester2/TechnicalReference.pdf

Public knowledge project. (2012). *Open harvester system*. Retrieved December 20, 2012, from http://pkp.sfu.ca/?q=harvester

Public Knowledge project. (2013). *Public Knowledge project*. Retrieved from Public Knowledge project: http://pkp.sfu.ca/

Swan, A., & Brown, S. (2005). Open access self-archiving: An author study. Retrieved December 10, 2012, from http://eprints.soton.ac.uk/260999/1/jisc2.pdf

William, A. Y. (2000). *Digital libraries.* Massachusetts institute of technology.