

Checking your Typeset Proof

Multi-Authored Papers

In the case of multi-authored papers, authors are advised to collaborate when checking the typeset proof. One author should be nominated to either accept or submit corrections on behalf of all of the authors of the paper.

We can only accept one set of revisions, or one acceptance of the typeset proof, from the nominated author. Once an author approves the typeset proof further revisions may not be requested.

Replying to us

After you review the typeset proof, you need to click on the ‘Author Verify Typeset Proof’ button (available at the link you downloaded the typeset proof from). You will then need to select the appropriate option to proceed.

Option 1: Accept Typeset Proof

To be selected when your paper is ready for publication

- Please thoroughly check the typeset proof before accepting it. You will not have further opportunities to make additional changes after the typeset proof has been accepted.
- Once you have accepted the typeset proof of your paper it will be ready to be published. You will be notified when your paper has been published and given instructions on how to access the published version.

Option 2: Request Resubmission of Typeset Proof

To be selected when your paper requires corrections

- Please see section on ‘Documenting your Corrections’.
- The typesetter will receive notification of your requested corrections. Once the corrections have been completed you will be notified of the availability of a revised typeset proof for your approval.

Bibliographical Details

Please note that full bibliographical details (issue and page numbers) will not be available until final publication of your paper. Once your paper has been published you will be able to obtain these details. We will notify you as soon as your paper is published.

Checklist for Reviewing the Typeset Proof

We recommend that you print the typeset proof and proofread it slowly and with great care. Request that a colleague also proofread your paper as they may notice errors that you may miss due to your familiarity with the content.

Remember to check your typeset proof for:

- Completeness: inclusion of all text, figures, illustrations and tables
- Correct title and subtitle
- Correct authorship and order of authors
- Current affiliation details
- Heading levels
- Position and size of illustrations and figures
- Matching of captions to illustrations and figures
- Position of tables
- Presentation of quotes
- Presentation of equations
- Footnotes and footnote numbering
- Inclusion of acknowledgements
- References and reference style
- Typesetting or conversion errors

Please check the Journal Standard Style prior to requesting changes to style as we adhere to standard presentation requirements for all papers to ensure consistency throughout the Journal.

It is important that all of your corrections (and those of your co-authors if applicable) are submitted to us in one communication.

Please note that careful proofreading is solely your responsibility.

Journal Standard Style

Order of the Paper:

1. Cover page
2. Copyright/imprint page
3. Paper: title/subtitle; author names with affiliation; abstract; keywords; body of paper; acknowledgement (if applicable); reference list; appendix (if any); about the author section
4. Journal colophon

Journal Standard Style:

- Paper title/subtitle and all headings appear in Title Case whereby only definite and indefinite articles (e.g. 'the' and 'a'), conjunctions (e.g. 'and'), and prepositions (e.g. 'in', 'of' etc.) appear in lower case.
- No italics in titles and subtitles.
- Affiliation of the author will include only the name of the author, university or organization name and country. Honorifics are not included.
- Abstract will appear in italics as a single paragraph.
- No italics included in the keyword list.
- No footnotes attached to title/subtitle, authors or the abstract.
- The first paragraph of the paper will appear in floating style - first three words appear in capital case and bold.
- Footnotes within tables have separate numbering to that of the footnotes within the paper.
- Hyphenation cannot be altered.
- No underline will be included.
- Figure captions are centred below the figure. The figure number and caption appear on the same line.
- Table titles appear above the table, left justified, in bold. The table number and table title appear on the same line.
- Flow of columns: If a figure or table appears in the middle of the page then the flow of the text will be from top left column to top right column, followed by table or figure. The remaining text will begin in the left column under the figure/table and will continue in the bottom right column.
- About the Author section: The honorific will reflect in this section. Contact details such as email addresses will not be included.

Documenting your Corrections

Changes to the Abstract

If you wish to make changes to the abstract of your paper please provide the revised abstract either as a Word document (if there are also changes to the text), or by entering it in the text box provided when you select Option 2.

Additional Authors

If you need to add a co-author we require the following information for each additional author to be added:

1. Name of the co-author
2. Affiliation details
3. Email address of the co-author (Mandatory)
4. Short Biography (limit of 30 words)
5. Long Biography (limit of 200 words one paragraph only)

Corrections to Text

If you have changes to the text please complete these in the Word version of your paper available at the link where you downloaded this PDF (or an existing word version). You can then upload the revised document for typesetting by selecting Option 2.

Corrections to Style:

You will need to clearly indicate all corrections in the following manner:

1. Page Number - paragraph number - line number - correction to be made

eg:

1. Page 4 - last paragraph, line 4, please put a comma after Tom in the sentence Mary, Tom, Jane and her friends...

The page number is the actual page of the PDF. As the paper has not been paginated yet, no numbers appear on the pages.

Submitting Corrections

Click the 'Author Verify Typeset Proof' button (available at the link you downloaded the typeset proof from) and select Option 2.

Option 2: Request Resubmission of Typeset Proof

- Please upload the corrected Word document, or add your instructions for corrections in the text box provided
- Note that you can only upload one document, and this document must contain all of the corrections (and those of your co-authors if applicable).

The typesetter will receive notification of your requested corrections. Once the corrections have been completed you will be notified of the availability of a revised typeset proof for your approval.

The International
JOURNAL

of

KNOWLEDGE, CULTURE
& CHANGE MANAGEMENT

Volume 8

An Architecture for Efficient Resource Discovery
with Metadata Harvesting in a Multidisciplinary
Distributed Repository

Jayan C Kurian, M.G. Sreekumar, Dion Hoe-Lian Goh,
Diljit Singh, Abrizah Abdullah and Joy Lynn Wheeler

THE INTERNATIONAL JOURNAL OF KNOWLEDGE, CULTURE AND CHANGE MANAGEMENT
<http://www.Management-Journal.com>

First published in 2008 in Melbourne, Australia by Common Ground Publishing Pty Ltd
www.CommonGroundPublishing.com.

© 2008 (individual papers), the author(s)
© 2008 (selection and editorial matter) Common Ground

Authors are responsible for the accuracy of citations, quotations, diagrams, tables and maps.

All rights reserved. Apart from fair use for the purposes of study, research, criticism or review as permitted under the Copyright Act (Australia), no part of this work may be reproduced without written permission from the publisher. For permissions and other inquiries, please contact [<cg-support@commongroundpublishing.com>](mailto:cg-support@commongroundpublishing.com).

ISSN: 1447-9524
Publisher Site: <http://www.Management-Journal.com>

THE INTERNATIONAL JOURNAL OF KNOWLEDGE, CULTURE AND CHANGE MANAGEMENT is a peer refereed journal. Full papers submitted for publication are refereed by Associate Editors through anonymous referee processes.

Typeset in Common Ground Markup Language using CGCreator multichannel typesetting system
<http://www.CommonGroundSoftware.com>.

An Architecture for Efficient Resource Discovery with Metadata Harvesting in a Multidisciplinary Distributed Repository

Jayan C Kurian, Nanyang Technological University, Singapore, SINGAPORE

M.G. Sreekumar, University of Malaya, Kuala Lumpur, MALAYSIA

Dion Hoe-Lian Goh, Nanyang Technological University, Singapore, SINGAPORE

Diljit Singh, University of Malaya, Kuala Lumpur, MALAYSIA

Abrizah Abdullah, University of Malaya, Malaysia, Kuala Lumpur, MALAYSIA

Joy Lynn Wheeler, Nanyang Technological University, SINGAPORE

Abstract: The profusion of non-relevant information for a given query on the Web explains the pressing need for formulating ebullient strategies for pertinent Web resource discovery and retrieval. One of the major requirements for effective document retrieval is its diligently encoded metadata. At the same time metadata standards to be followed for annotating documents from large collections are pretty complex. This is because the standardized global metadata cannot represent all the elusive forms of document metadata for improved retrieval ranking. In this context, we propose an approach to facilitate document retrieval from multidisciplinary domains where each belonging to discrete domains would be indexed in a segregated instance of a repository. This would facilitate document metadata customization for each specific discipline by adding specific metadata themes. Since the approach retains the standard metadata schema in addition to the customized metadata schema, it would result in enhanced resource discovery. The metadata retrieval process will be supported by an extended protocol for metadata harvesting (X-PMH) [1] and will be implemented in each repository. The extended metadata harvesting approach has been used to tie together the metadata customization components made at various repository instances. The proposed framework could be integrated into Open Digital Libraries (ODLs) [2] and shall serve as an intrinsic model that adds value in the context of multidisciplinary metadata simplicity, maintenance, and descriptive metadata availability in the event of repository instance failures. Our approach is to implement this cost-effective architecture using the PKP-OAI (Public Knowledge Project – Open Archive Initiative) [3,4] harvester on DSpace [5], an open source digital repository platform that supports metadata harvesting in its innate form. Once this is fully achieved, a federated search build upon such repository instances using open source technologies [6] would yield promising results in the context of information retrieval.

Keywords: Metadata Harvesting, Federated Searching, Information Retrieval, Multidisciplinary Distributed Repository, Public Knowledge Project, Open Archives Initiative, Open Digital Libraries

Introduction

THE ABUNDANCE OF heterogeneous information among multidisciplinary organizations has made the pervasive task of managing information a complex process. Hence in an information management context that involves aggregating relevant information and disseminating to the right audience, repositories play a magnanimous role due to their open-source nature. Thus information management and repositories are highly interrelated in a library context and cannot be seen as segregated and indolent in nature. To support this cohesive attainment towards scholarly information processing and dissemination, institutional repositories play a significant role. Institutional repositories, being developed as a major source of scholarly dissemination based on open-access policies, have become an integral component of almost all libraries especially in a multidisciplinary educational environ-

ment. We refer to a multidisciplinary repository (Open Access Archive) as a collection of peer-reviewed post-prints of scholarly articles, image and video objects, learning objects, student theses, and industrial attachment reports. In fact, the present digital libraries depend on repository collections to garner, process, and disseminate scholarly information, which otherwise would not have achieved its full potential in terms of scholarly dissemination. The crux of any institutional repository includes support for the dissemination of materials that include research papers, technical articles, theses, and working papers. In addition to that, the basic principles that govern the outreach of institutional repositories include support for learning objects, corporate assets, and granular content [7]. Learning objects includes organizational manifested materials (e.g. lecture notes), corporate assets include organizational records (e.g. annual reports), and granular content describes organizational content to its minute details



(e.g. file level metadata description rather than compound object level description). Thus repositories play a significant role by supporting the scholarly communication lifecycle model of Roosendaal and Guertz that includes registration, certification, awareness, archiving, and rewarding [7].

The popular open source repository packages include Archimede, CDSware, DSpace, EPrints.org, Fedora, and OPUS. Archimede [8] was designed for the preservation and dissemination of scholarly content of university level research communities where as CDSware [9] initially started with the high energy physics research document collection and later traversed into a collection of general library resources. EPrints [10] also was based on the concept of open-policy dissemination of research content and the Fedora repository [11] brought in the concept of a scalable digital library system with an object model supporting multiple representations of unique digital object [7]. OPUS [12] was developed keeping in mind the concept of an information system for publishing staff and student electronic documents. DSpace [5] evolved as a collaborated project between Hewlett Packard Labs and MIT Libraries and has a wide community base with several volunteered committers for its evolving development. We base our further study and proposal of a multidisciplinary distributed architecture on this widely accepted open-source software due to our familiarity with the software and its implementation in our affiliate institutes.

Two of the major value adding features of OA archives, among many others, are their Internet presence (omnipresence) and interoperability. The interoperability feature keeps all OA archives virtually a single digital library system wherein they share their metadata through some common services called metadata harvesters or service providers using the OAI-PMH protocol. A number of tools are now available for starting such services and the PKP archive harvester is the appreciated and simple one used by many.

Research Issue

There is ample evidence of increasing acceptance of institutional repositories world wide. For instance, the Directory of Open Access Repositories (Open DOAR) maintained by the University of Nottingham, UK has recently crossed the 1100 listings [13]. At the same time the growth in the number of repositories, their distributed nature, size and diversity also make the quality of metadata an increasingly import-

ant issue [14]. Bruce and Hillmann acknowledge the difficulties in defining metadata quality [15]. They have identified seven metadata quality criteria: completeness, accuracy, provenance, conformance to expectations, logical consistency and coherence, timeliness, and accessibility. And it requires both human expertise and appropriate tool support for accessing these criteria in diverse collections [16,17]. Metadata quality is of paramount importance to a repository and the metadata completeness is a mark of the record quality of the repository.

In a multidisciplinary repository context effective representation of digital objects would be facilitated by various metadata schemas. This leads to a lack of universal representation schema that supports diverse digital objects. In this context we propose a harvesting-cum-indexing approach supported by the 2.0 version of the protocol for metadata harvesting (OAI-PMH 2.0) [18] and facilitated by the PKP-OAI (Public Knowledge Project – Open Archives Initiative) harvester [19]. The advantage of using this approach is extensible metadata schema and that can be used to represent diverse objects in a multidisciplinary repository. This model adds value in the context of multidisciplinary metadata simplicity, maintenance, and descriptive metadata availability in the event of repository instance failures. This also relieves from defining a single metadata schema encompassing the requirements for effectively describing variety of resources as well as large-scale resource intensiveness of central repositories. An additional research issue would to provide an interface by which a user can search harvested metadata based on encoded schema representation (e.g. Dublin core, Learning Object Model etc). This would stress on adaptive retrieval of digital resources

Proposed Case Study

The two affiliate institutes in this project are University of Malaya, Malaysia and the Nanyang Technological University, Singapore. Each university has built its own individual repository based on specific collection themes. Metadata schema may be tuned according to individual collection requirement. Harvesting can be applied between local repositories that act as data providers. Each university can also act as a service provider of its local collections. Since both universities can act as service providers metadata harvesting can be done between individual repositories. The proposed architecture is shown in figure 1.

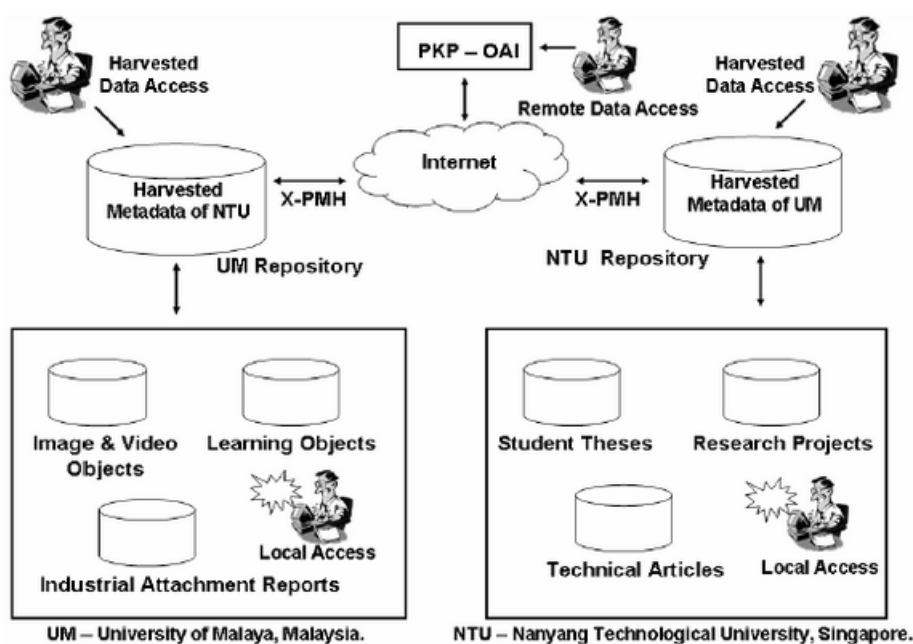


Figure 1

As an initial step we would be investigating how different metadata schemas are interoperable by performing a migration between distributed repositories. To perform this we utilize a repository script implemented in the DSpace software and schedule the script that would automatically export digital objects from one instance in a package format to be easily imported into another instance. Before importing into a second instance some amount of data

cleansing has to be done that would take into account the schema representation of the second instance. At a later stage this may be automated to represent complex objects (consisting of more than one object) in different schema representations in different local repositories. The sample scheduler script that would be used in conjunction with the main DSpace Export feature is given below.

```
set DayDateMonthYear=%date:~0,3%%date:~7,2%%date:~4,2%%date:~10,4%
set TimeinHours=%time:~0,2%
set TimeinMinutes=%time:~3,2%
set FolderNameWithDayTime=%DayDateMonthYear%-%TimeinHours%hr%TimeinMinutes%mts
nd E:\%FolderNameWithDayTime%
cd E:\dSPACE\bin
dsrun org.dspace.app.itemexport.ItemExport -t
COLLECTION -i 123456789/135 -d E:\%FolderNameWithDayTime% -n 1
```

Once this process is over, the open source PKP archive software will be installed, configured and customized to harvest the metadata records from both the UM and NTU repositories. The PKP harvester shall act as a common index having search and browse facilities over the Internet. The index service will be tested for its functionality and correctness over a period of time and subsequently more repositories will be added to the index.

Conclusion

In this on-going research work that is in its preliminary stage we have been successful in scheduling

semi-automatic migration between local repository instances. Further work needs to be done in investigating metadata interoperability between repository instances. Another area to be investigated is metadata harvesting among local repositories to act as a data and service provider among affiliate institutes. We are hopeful that the PKP archives harvester shall be able to meet this efficiently and effectively. Once this model evolves as a promising initiative we plan to extend this with other institutes having similar interest in open source movement.

References

- [1] Suleman, H., Fox, E.A. "Leveraging OAI Harvesting to Disseminate Theses". *Library Hi Tech*, 21(2), pp. 219-227, 2003.
- [2] Fox, E.A., Suleman, H., Luo, M. "Building digital libraries made easy: toward open digital libraries", In *Proceedings of the 5th International Conference of Asian Digital Libraries*, 2002.
- [3] The Public Knowledge Project, accessible at: <http://pkp.sfu.ca/>
- [4] Suleman, H and Fox, E.A. "Designing protocols in support of digital library componentization". In *Proceedings of the European Conference on Digital Libraries*, pages 568–582, Rome, Italy, 2002
- [5] Dspace, accessible at: <http://www.dspace.org/>
- [6] Library Find Project, accessible at: <http://www.libraryfind.org/>
- [7] Richard Jones, Theo Andrew, John MacColl, "The Institutional Repository", Chandos Publishing, 2006
- [8] Archimede, available at: <http://www.bibl.ulaval.ca/archimede/index.en.html>
- [9] CDSware, available at: <http://cdsweb.cern.ch/>
- [10] EPrints.org, available at: <http://www.eprints.org/>
- [11] Fedora, available at: <http://www.fedora.info/>
- [12] OPUS, available at: <http://elib.uni-stuttgart.de/opus/index.php?la=en>
- [13] OpenDOAR available at <http://www.opendoar.org>
- [14] Nicholas, David M. et al. "A tool for metadata analysis". Working Paper 2/2008, Department of Computer Science, University of Waikato. 2008.
- [15] Bruce, T.R. and Hillmann, D.I. "The continuum of metadata quality: defining, expressing, exploiting". In: *Metadata in Practice*, American Library Association, Chicago, IL. 238-256. 2004.
- [16] Beall, J. "Metadata and data quality problems in the digital library". *Journal of Digital Information*. 6(3) 2005.
- [17] Ochoa, X and Duval, E. "Towards automatic evaluation of learning object metadata quality". In: *Advances in Conceptual Modeling – Theory and Practice*, ER 2006 Workshops BP-UML, CoMoGIS, COSS, ECDM, OIS, QoIS, SemWAT, Springer, 372-381. 2006.
- [18] OAI-PMH 2.0, available at <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [19] PKP Harvester 2.x, available at <http://pkp.sfu.ca/?q=harvester>

About the Authors

Jayan C Kurian

Nanyang Technological University, SINGAPORE

Dr. M.G. Sreekumar

University of Malaya, MALAYSIA

Dr. Dion Hoe-Lian Goh

Nanyang Technological University, SINGAPORE

Dr. Diljit Singh

University of Malaya, MALAYSIA

Dr. Abrizah Abdullah

University of Malaya, Malaysia, MALAYSIA

Joy Lynn Wheeler

Joy Wheeler has worked at the NTU Library in the Library Technology Systems Division for the past 2 years. She splits her time as a subject librarian for the Art Design & Media Library.

EDITORS**Mary Kalantzis**, University of Illinois, Urbana-Champaign, USA.**Bill Cope**, University of Illinois, Urbana-Champaign, USA.**EDITORIAL ADVISORY BOARD****Verna Allee**, Verna Allee Associates, California, USA.**Zainal Ariffin**, Universiti Sains Malaysia, Penang, Malaysia.**Robert Brooks**, Monash University, Melbourne, Australia.**Bruce Cronin**, University of Greenwich, UK.**Rod Dilnutt**, William Bethway and Associates, Melbourne, Australia.**Judith Ellis**, Enterprise Knowledge, Melbourne, Australia.**Andrea Fried**, Chemnitz University of Technology, Germany.**David Gurteen**, Gurteen Knowledge, UK.**David Hakken**, University of Indiana, Bloomington, Indiana, USA.**Sabine Hoffmann**, Macquarie University, Australia.**Stavros Ioannides**, Pantion University, Athens, Greece.**Margaret Jackson**, RMIT University, Melbourne, Australia.**Paul James**, RMIT University, Melbourne, Australia.**Leslie Johnson**, University of Greenwich, UK.**Eleni Karantzola**, University of the Aegean, Rhodes, Greece.**Gerasimos Kouzelis**, University of Athens, Greece.**Krishan Kumar**, University of Virginia, USA.**Martyn Laycock**, University of Greenwich and managingtransitions.net, UK.**David Lyon**, Queens University, Ontario, Canada.**Bill Martin**, RMIT University, Melbourne, Australia.**Pumela Msweli-Mbanga**, University of Kwazulu-Natal, South Africa.**Claudia Schmitz**, Cenandu Learning Agency, Germany.**Kirpal Singh**, Singapore Management University, Singapore.**Dave Snowden**, Cynefin Centre for Organisational Complexity, UK.**Chryssi Vitsilakis-Soroniatis**, University of the Aegean, Rhodes, Greece.

THE UNIVERSITY PRESS JOURNALS

International Journal of the Arts in Society

Creates a space for dialogue on innovative theories and practices in the arts, and their inter-relationships with society.

ISSN: 1833-1866

<http://www.Arts-Journal.com>

International Journal of the Book

Explores the past, present and future of books, publishing, libraries, information, literacy and learning in the information society. ISSN: 1447-9567

<http://www.Book-Journal.com>

Design Principles and Practices: An International Journal

Examines the meaning and purpose of 'design' while also speaking in grounded ways about the task of design and the use of designed artefacts and processes. ISSN: 1833-1874

<http://www.Design-Journal.com>

International Journal of Diversity in Organisations, Communities and Nations

Provides a forum for discussion and builds a body of knowledge on the forms and dynamics of difference and diversity.

ISSN: 1447-9583

<http://www.Diversity-Journal.com>

International Journal of Environmental, Cultural, Economic and Social Sustainability

Draws from the various fields and perspectives through which we can address fundamental questions of sustainability.

ISSN: 1832-2077

<http://www.Sustainability-Journal.com>

Global Studies Journal

Maps and interprets new trends and patterns in globalization. ISSN 1835-4432

<http://www.GlobalStudiesJournal.com>

International Journal of the Humanities

Discusses the role of the humanities in contemplating the future and the human, in an era otherwise dominated by scientific, technical and economic rationalisms. ISSN: 1447-9559

<http://www.Humanities-Journal.com>

International Journal of the Inclusive Museum

Addresses the key question: How can the institution of the museum become more inclusive? ISSN 1835-2014

<http://www.Museum-Journal.com>

International Journal of Interdisciplinary Social Sciences

Discusses disciplinary and interdisciplinary approaches to knowledge creation within and across the various social sciences and between the social, natural and applied sciences.

ISSN: 1833-1882

<http://www.Socialsciences-Journal.com>

International Journal of Knowledge, Culture and Change Management

Creates a space for discussion of the nature and future of organisations, in all their forms and manifestations.

ISSN: 1447-9575

<http://www.Management-Journal.com>

International Journal of Learning

Sets out to foster inquiry, invite dialogue and build a body of knowledge on the nature and future of learning.

ISSN: 1447-9540

<http://www.Learning-Journal.com>

International Journal of Technology, Knowledge and Society

Focuses on a range of critically important themes in the various fields that address the complex and subtle relationships between technology, knowledge and society. ISSN: 1832-3669

<http://www.Technology-Journal.com>

Journal of the World Universities Forum

Explores the meaning and purpose of the academy in times of striking social transformation.

ISSN 1835-2030

<http://www.Universities-Journal.com>

FOR SUBSCRIPTION INFORMATION, PLEASE CONTACT
subscriptions@commonground.com.au