# The Pacific Rim Library: A Surprising Pearl

The Pacific Rim Library (PRL) is an initiative of the Pacific Rim Digital Library Association (PRDLA). The project began in 2006 using the OAI-PMH paradigm and now holds over 300,000 records harvested from OAI data provider libraries around the Pacific. PRL's goal is to enable the sharing of digital collections amongst PRDLA members and the world, but greater unexpected benefits have been discovered. Through mirroring their metadata, PRL increases the chance that their data will be discovered in Google and other general search engines. With its many disparate collections, PRL is not a repository for traditional information discovery and retrieval. Initially users will bounce from a Google hit, to the PRL metadata record in Hong Kong, and then begin an intensive search on the original site which hosts the full digital object, in Vancouver, Honolulu, Wuhan, Singapore, or other PRDLA member location. Serials Review 2009; xx:xxx–xxx.

*Keywords:* OAI; Pacific Rim Library; Pacific Rim Digital Library Association; PRDLA; PRL; Information discovery and retrieval; Indexing; Metadata; Deep Web

## Introduction

The Pacific Rim Digital Library Association (PRDLA) was formed in 1997 by fourteen libraries in the Pacific region. Now thirty-one libraries strong, the goal of this group is to "improve access to scholarly research materials through cooperative ventures."[1,2] Over the years the association has sponsored many such endeavors. A task force was established in 2004 to study the best means for sharing PRDLA digital collections amongst themselves and the world. In its 2005 report, the task force recommended creating a Web-based Open Archives Initiative (OAI)[3] and determined that establishing a Service Provider would be the quickest, easiest, and most effective way of accomplishing this goal.

## Open Archives Initiative Model

In the OAI paradigm, there can be many geographically or categorically disparate Data Providers hosting their own local repositories, with digital objects of full-text items or other files. Each Data Provider will expose the metadata on these items; a Service Provider can then harvest, importing the metadata (and not the digital object) into a new repository. The Service Provider will apply a search engine on this new repository. Users can then do one search across all metadata of the several Data Providers. Upon choosing one record and clicking, the user leaves the Service Provider's repository and arrives on a page in one of the local Data Provider's repositories.

## Developments and Refinements

In 2006, a new task force was funded by PRDLA to create an OAI pilot repository at The University of Hong Kong (HKU). The working title of this repository was the "PRDLA Archive," which recently became "Pacific Rim Library" or "PRL" for short. (Hereafter, "PRL" will be used, although the name change has still not been enacted.) PRDLA decided that PRL would harvest OAI metadata from at least one locally created digital collection at each of the PRDLA member libraries. HKU became the OAI Service Provider, harvesting data from the many OAI Data Providers. Several PRDLA libraries then implemented the OAI protocol for metadata harvesting (OAI-PMH) for the first time. The PRL OAI data providers created XML compatible, UTF-8 compliant metadata using the simplest Dublin Core (DC)[4] schema: "oai_dc" consisting of fifteen unqualified data elements. At that time, HKU had begun to use DSpace for its institutional repository, and thus, created a second instance of DSpace to be the OAI Service Provider for harvesting and hosting the metadata from the many PRL OAI Data Providers. Each PRDLA member became a "community" in the DSpace paradigm (Fig. 1). Q2

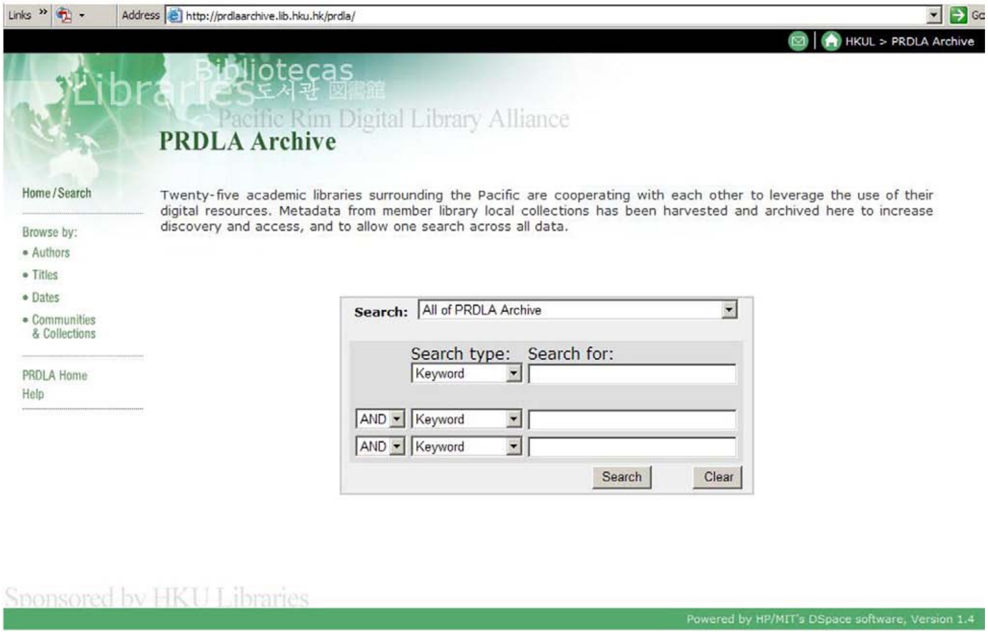
DSpace worked well when there were few records. After more PRDLA member repositories were harvested, however, and record numbers reached 300,000, the response time slowed considerably. We also found problems in OAI harvesting and updating. We then found a replacement in CDS Invenio,[5] another open source software, developed by CERN. Invenio promises fast searching across a repository of up to 1.5 M records at one time. CDS Invenio was easier to customize for OAI harvesting and provided several out-of-the-box interfaces in different languages and scripts. Our developer also contributed to the code base by making new translations for traditional and simplified Chinese as shown in the following illustration (Fig. 2).[6] Q3

In order to increase the value of PRL, we changed the oai_dc schema used among the libraries, to include one more qualification on Identifier, that of an "identifier thumbnail." This identifier allows for the harvesting and storage of thumbnail images along with the usual bibliographic details.

**Palmer** is Scholarly Communications Head, HKU Libraries, The University of Hong Kong, Hong Kong; e-mail: dtpalmer@hku.hk. Q1

0098-7913/$ – see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.serrev.2009.04.004

Please cite this article as: David Palmer, The Pacific Rim Library: A Surprising Pearl, Serials Review (2009), doi:10.1016/j.serrev.2009.04.004


Links
Address http://prdlaarchive.lib.hku.hk/prdla/
HKUL > PRDLA Archive
Bibliotecas
Libraries 도서관 図書館
Pacific Rim Digital Library Alliance
PRDLA Archive
Home/Search
Browse by:
• Authors
• Titles
• Dates
• Communities & Collections
PRDLA Home
Help
Search:
Search type:
Search for:
AND
Search
Clear
Sponsored by HKU Libraries
Powered by HP/MIT's DSpace software, Version 1.4

Twenty-five academic libraries surrounding the Pacific are cooperating with each other to leverage the use of their digital resources. Metadata from member library local collections has been harvested and archived here to increase discovery and access, and to allow one search across all data.

Search: All of PRDLA Archive

Search type: Keyword

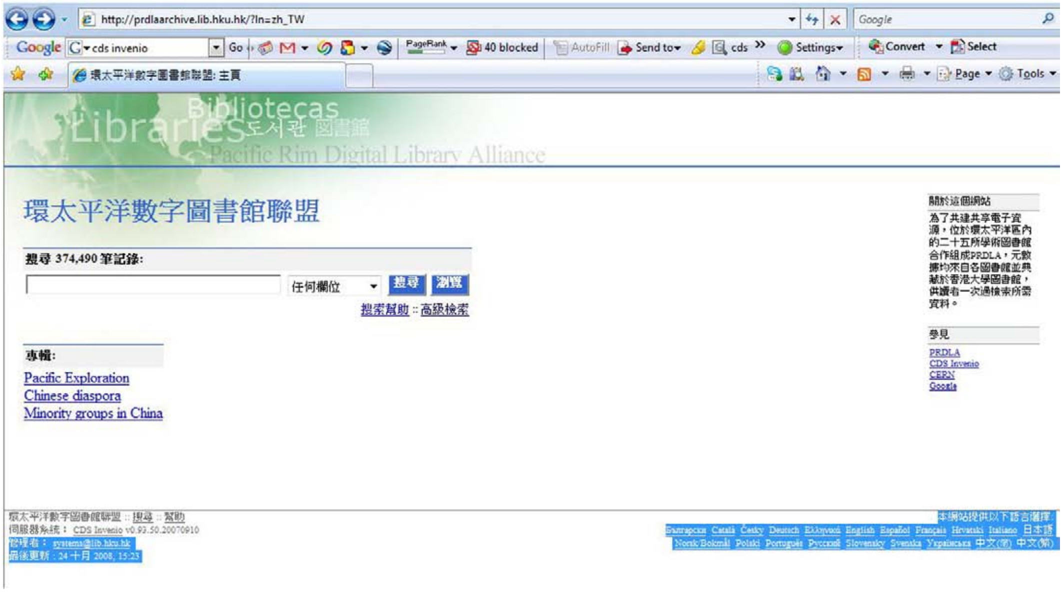
**Figure 1.** PRDLA archive in DSpace (2006).

The 2007 Berkeley PRDLA meeting pronounced this repository fit for purpose and asked that it go into "production." We removed the password access and specifically invited external robots and crawlers inside the system by creating sitemaps for the major search engines to index content.[7] We created records for PRL in the following major OAI-PMH registries and repositories:

- Open Archives list of registered OAI repositories: http://www.openarchives.org/Register/BrowseSites
- OAI registry at University of Illinois at Urbana-Champaign: http://gita.grainger.uiuc.edu/registry/
- Celestial OAI registry: http://celestial.eprints.org
- Registry of Open Access Repositories: http://roar.eprints.org/
- Directory of Open Access Repositories — OpenDOAR: http://www.opendoar.org/

## The Purpose of PRL

Although the membership of PRDLA was generally pleased with the progress of PRL, at the 2007 meeting some members expressed uneasiness at the disparate nature of the collections hosted in PRL. Traditionally, databases are created and chosen for searching because of a selection process in the creation of these databases that strives to include as much relevant content in a given subject as possible, and which, therefore, excludes material beyond the scope of that subject. Although there are many collections in PRL that are unquestionably relevant to the Pacific area, such as the "Sea of Korea Map Collection" and the "Hawaiian Photo Album," there are also several that do not have direct relevance to the Pacific area, such as "The Automobile Club of Southern California."


http://prdlaarchive.lib.hku.hk/?ln=zh_TW
Google
cds invenio
Go
PageRank
40 blocked
AutoFill
Send to
cds
Settings
Convert
Select
環太平洋數字圖書館聯盟: 主頁
Page
Tools
Bibliotecas
Libraries 도서관 図書館
Pacific Rim Digital Library Alliance
任何欄位
搜尋
瀏覽
搜索幫助 :: 高級檢索
關於這個網站
參見
PRDLA
CDS Invenio
CERN
Google
本網站提供以下語言選擇:
Български Català Česky Deutsch Ελληνικά English Español Français Hrvatski Italiano 日本語 Norsk/Bokmål Polski Português Русский Slovensky Svenska Українська 中文(简) 中文(繁)

環太平洋數字圖書館聯盟

搜尋 374,490 筆記錄:

專輯:
Pacific Exploration
Chinese diaspora
Minority groups in China

為了共建共享電子資源，位於環太平洋區內的二十五所學術圖書館合作組成PRDLA，元數據均來自各圖書館並典藏於香港大學圖書館，供讀者一次過檢索所需資料。

環太平洋數字圖書館聯盟 :: 搜尋 :: 幫助
伺服器系統： CDS Invenio v0.93.50.20070910
管理者： systems@lib.hku.hk
最後更新：24 十月 2008, 15:23

**Figure 2.** PRDLA archive in CDS Invenio; simplified Chinese interface.

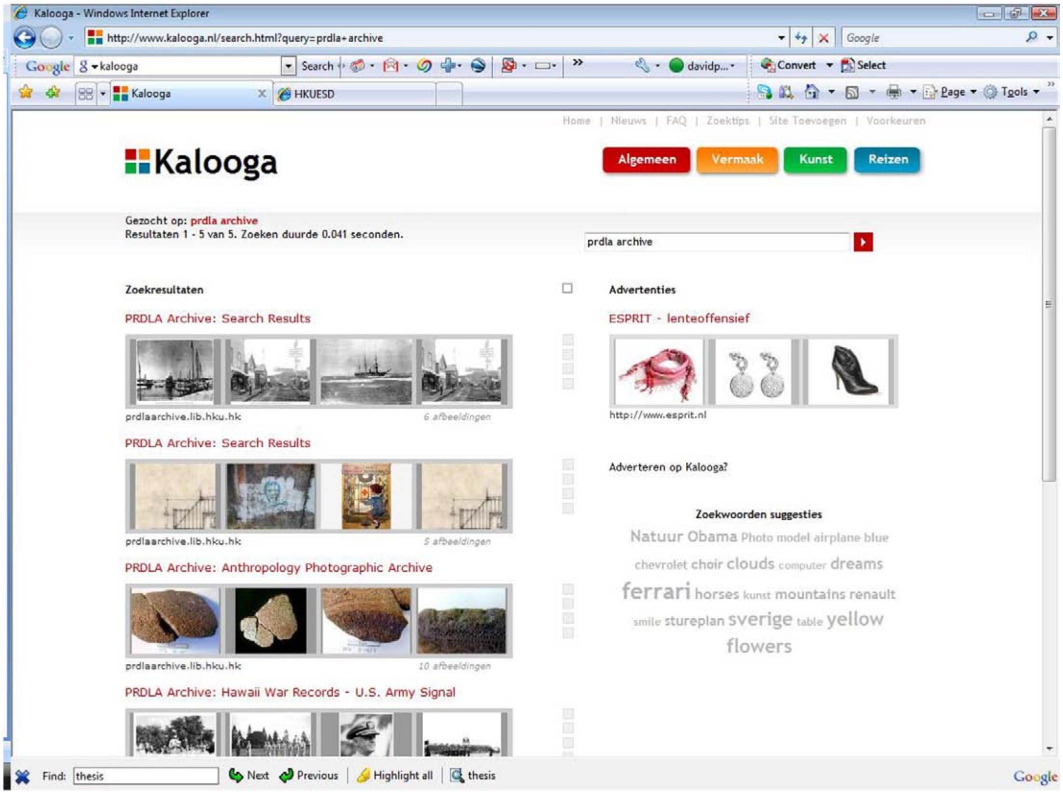

Kalooga - Windows Internet Explorer
http://www.kalooga.nl/search.html?query=prdla+archive
Google
Kalooga
HKUESD
Home | Nieuws | FAQ | Zoektips | Site Toevoegen | Voorkeuren
Kalooga
Algemeen
Vermaak
Kunst
Reizen
Gezocht op: prdla archive
Resultaten 1 - 5 van 5. Zoeken duurde 0.041 seconden.
prdla archive
Zoekresultaten
PRDLA Archive: Search Results
prdlaarchive.lib.hku.hk
6 afbeeldingen
PRDLA Archive: Search Results
prdlaarchive.lib.hku.hk
5 afbeeldingen
PRDLA Archive: Anthropology Photographic Archive
prdlaarchive.lib.hku.hk
10 afbeeldingen
PRDLA Archive: Hawaii War Records - U.S. Army Signal
Advertenties
ESPRIT - lenteoffensief
http://www.esprit.nl
Adverteren op Kalooga?
Zoekwoorden suggesties
Natuur Obama Photo model airplane blue
chevrolet choir clouds computer dreams
ferrari horses kunst mountains renault
smile stureplan sverige table yellow
flowers
Find: thesis
Next
Previous
Highlight all
thesis

**Figure 3.** PRL thumbnails in Kalooga.

PRDLA charged a new task force to survey the membership and propose scope and direction for PRL. This task force reported its findings at the 2008 Singapore meeting and made the following recommendations:

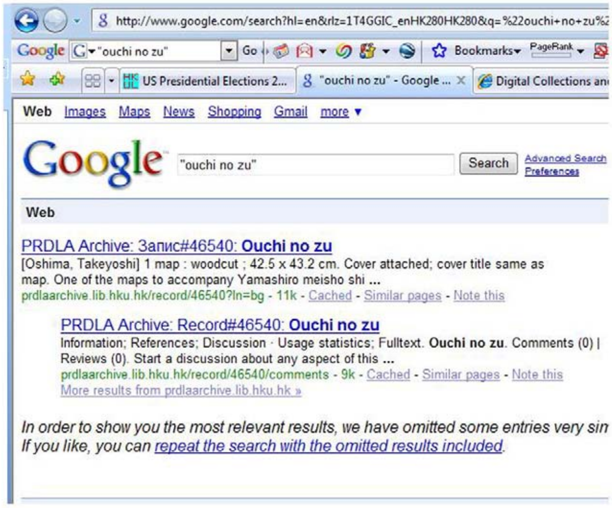
1. Change the name to "Pacific Rim Library" with the acronym "PRL," pronounced as "Pearl" (as in "pearl of the orient") to more accurately represent the content of this repository.
2. That the contents of PRL be about the Pacific Rim areas and that the purpose of PRL is to provide people with a way of finding material about the peoples, cultures, history, etc. of the Pacific Rim.
3. That PRDLA should invite others with Pacific Rim-related content to allow harvesting into PRL, and that this invitation be extended to anyone with relevant content, within or external to the PRDLA membership.

## Unexpected Pearls

The PRL Technical Committee also presented a report at this meeting that described its findings on the usage statistics of PRL. These statistics showed high activity by all of the major search bots, including Googlebot, Yahoo Slurp, Cuil's Twiceler, BaiduSpider, etc. One top referring site was a Wikipedia entry about James Wong (黃湛森), a recently deceased Cantopop singer and songwriter from Hong Kong. At that time, the entry hyperlinked to the record in PRL, which, in turn, hyperlinked to his full-text PhD thesis in the HKU Libraries. One top target identified in PRL was the War Poster Collection of the University of Washington. Thumbnail images provided by this collection in PRL are very attractive to Kalooga, and other major projects seeking thumbnail images (Fig. 3).

## Search Engine Results

Comparing searches in Google and Yahoo on a known PRL title, with the original host providing the metadata, gives interesting results. A Google search in the HKU network on "Lise's Lunchwagon" shows only two hosts, the originating collection at the University of Hawaii-Manoa and PRL (the UH-M record appears above PRL, as expected, because they are the original provider); however, a search on "Hoo Hoo House," a record from the University of Washington,


http://www.google.com/search?hl=en&rlz=1T4GGIC_enHK280HK280&q=%22ouchi+no+zu%2
Google
"ouchi no zu"
Go
Bookmarks
PageRank
US Presidential Elections 2...
"ouchi no zu" - Google ...
Digital Collections an
Web Images Maps News Shopping Gmail more ▾
Google
Search
Advanced Search
Preferences
Web

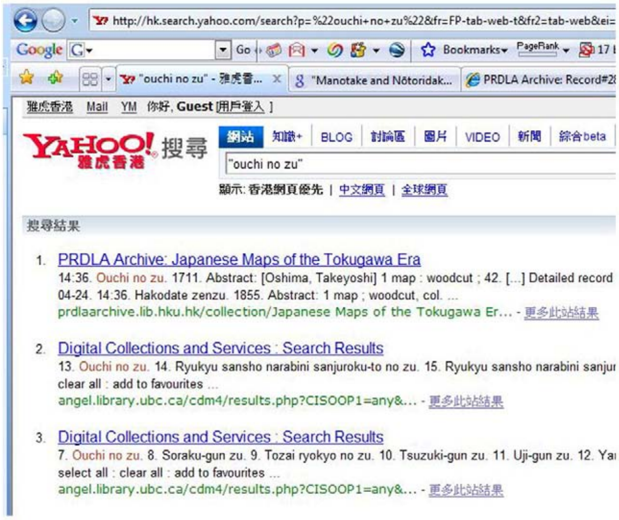
PRDLA Archive: Запис#46540: **Ouchi no zu**
[Oshima, Takeyoshi] 1 map : woodcut ; 42.5 x 43.2 cm. Cover attached; cover title same as map. One of the maps to accompany Yamashiro meisho shi ...
prdlaarchive.lib.hku.hk/record/46540?ln=bg - 11k - Cached - Similar pages - Note this

PRDLA Archive: Record#46540: **Ouchi no zu**
Information; References; Discussion · Usage statistics; Fulltext. **Ouchi no zu**. Comments (0) | Reviews (0). Start a discussion about any aspect of this ...
prdlaarchive.lib.hku.hk/record/46540/comments - 9k - Cached - Similar pages - Note this
More results from prdlaarchive.lib.hku.hk »

*In order to show you the most relevant results, we have omitted some entries very sin*
*If you like, you can repeat the search with the omitted results included.*

**Figure 4.** Google: "Ouchi no zu."

Please cite this article as: David Palmer, The Pacific Rim Library: A Surprising Pearl, Serials Review (2009), doi:10.1016/j.serrev.2009.04.004


http://hk.search.yahoo.com/search?p=%22ouchi+no+zu%22&fr=FP-tab-web-t&fr2=tab-web&ei=
Google G
Go
Bookmarks
PageRank
17
"ouchi no zu" - 雅虎香...
"Manotake and Nōtoridak...
PRDLA Archive: Record#2
雅虎香港
Mail
YM
你好, Guest [用戶登入 ]
YAHOO! 雅虎香港 搜尋
網站
知識+
BLOG
討論區
圖片
VIDEO
新聞
綜合beta
"ouchi no zu"
顯示: 香港網頁優先 | 中文網頁 | 全球網頁
搜尋結果

1. PRDLA Archive: Japanese Maps of the Tokugawa Era
14:36. Ouchi no zu. 1711. Abstract: [Oshima, Takeyoshi] 1 map : woodcut ; 42. [...] Detailed record 04-24. 14:36. Hakodate zenzu. 1855. Abstract: 1 map ; woodcut, col. ...
prdlaarchive.lib.hku.hk/collection/Japanese Maps of the Tokugawa Er... - 更多此站結果

2. Digital Collections and Services : Search Results
13. Ouchi no zu. 14. Ryukyu sansho narabini sanjuroku-to no zu. 15. Ryukyu sansho narabini sanjur
clear all : add to favourites ...
angel.library.ubc.ca/cdm4/results.php?CISOOP1=any&... - 更多此站結果

3. Digital Collections and Services : Search Results
7. Ouchi no zu. 8. Soraku-gun zu. 9. Tozai ryokyo no zu. 10. Tsuzuki-gun zu. 11. Uji-gun zu. 12. Ya
select all : clear all : add to favourites ...
angel.library.ubc.ca/cdm4/results.php?CISOOP1=any&... - 更多此站結果

**Figure 5.** Yahoo: "Ouchi no zu."

shows PRL first and UW second. A search on "Ouchi no zu," a record from University of British Columbia, only shows the PRL entry in Google. Yahoo search results, however, show PRL first, followed by UBC, as shown in Figs. 4 and 5.

During the same week that HKU performed these searches, colleagues in Honolulu and London confirmed receiving the same results on the same searches.

These results demonstrate that PRL is linking users to those resources that may not have been discovered otherwise. In this manner, PRL is adding value to the original locally hosted databases that comprise PRL.

## The Deep Web

Indexing or revealing the many hidden Web objects in the deep Web has long been a goal toward which many commercial and academic search engines continue to strive. Several reasons explain why these Web objects remain hidden. For example, to prevent undue traffic, some repositories use protocols to specifically exclude robots. Also, unless there are other links that point to these hidden Web objects, robots and crawlers will not find them.[8]

Most major search engines now provide procedures for repositories to follow in order to expose more of the deep Web to their robots. At one time, Google and Yahoo included OAI-PMH in these procedures; however, Google recently announced that they were retiring support for OAI-PMH in their sitemaps.[9] This lack of support, therefore, shows the further value of PRL as a discovery source. PRL harvests OAI-PMH data from Web objects that may be hidden, in repositories that may not be completely visible to these search engines. PRL then exposes this harvested metadata to these search engines for their indexing. A recent study by the OAIster project indicated similar results.[10]

These search results, combined with the dissimilar nature of the PRL collections, suggest that PRL's true value is as an indirect tool of discovery. PRL has enabled the discovery of this hidden Web to users of Google and Yahoo, and presumably pushed ranking of these pages higher within these and other search engines. The end-user will presumably find an item in the search engines, bounce to PRL in Hong Kong, and then link to the originating database in Hawaii, Wuhan, Singapore, or other PRDLA member location, to begin a more focused search.

## The Future Value of PRL

PRL has become a flagship project of PRDLA and serves to rally member support and to promote their existence to the world. In recognition, PRDLA has begun two new projects focused on PRL. PRLDA is redesigning its logo and PRL graphics to enhance its image. PRDLA has also created a collaboration for a new collection, Oceania Digital Libraries (ODiL), expected for release in 2009.[11]

In summary, the PRDLA goal for PRL is the creation of a tool for searching content about the Pacific Rim. However, the way that PRL does this is not as a traditional destination database, but one that acts as an intermediary between the search engines and the originating database. Although this is beyond the original goal, the PRDLA membership understands the value that PRL provides and will continue to use this project and its Web pages to showcase their collections.

## Notes

[1] "About [PRDLA]," http://www.prdla.ucmercedlibrary.info/?page_id=2, (accessed March 30, 2009).

[2] "Membership [of PRDLA]," http://prdla.ucmercedlibrary.info/about/membership/, (accessed March 30, 2009).

[3] "OAI for Beginners," http://www.oaforum.org/tutorial/, (accessed March 30, 2009).

[4] "Dublin Core Metadata Initiative," http://dublincore.org/usage/terms/dc/current-elements/, (accessed March 30, 2009).

[5] "CDS Invenio Overview," http://cdsware.cern.ch/invenio/, (accessed March 30, 2009).

[6] A "thank you" on this page to Dr. Ku Kam-ming of HKU, "CDS Invenio Overview," http://cdsware.cern.ch/invenio/index.html, (accessed March 30, 2009).

[7] "Sitemaps," http://en.wikipedia.org/wiki/Sitemaps, (accessed March 30, 2009).

[8] Frank McCown, Michael L. Nelson, and Mohammad Zubair, "Search Engine Coverage of the OAI-PMH Corpus," IEEE Internet Computing (March/April 2006), http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01607990.

[9] John Mueller, "Retiring support for OAI-PMH in Sitemaps," Google Webmaster Central Blog (April 23, 2008), http://googlewebmastercentral.blogspot.com/2008/04/retiring-support-for-oai-pmh-in.html, (accessed March 30, 2009).

[10] Kat Hagedorn and Joshua Santelli, "Google Still Not Indexing Hidden Web URLs," D-Lib Magazine 13, no.7/8 (2008), http://www.dlib.org/dlib/july08/hagedorn/07hagedorn.html, (accessed March 30, 2009).

[11] "PRDLA," http://prdla.ucmercedlibrary.info/2008/10/oceania-digital-libraries-odil-collaborative-digitization-project/, (accessed March 30, 2009).

Please cite this article as: David Palmer, The Pacific Rim Library: A Surprising Pearl, Serials Review (2009), doi:10.1016/j.serrev.2009.04.004