Touring the Information Landscape: The Open Archives Initiative

Lennie Stovel, RLG, lennie.stovel@notes.rlg.org

The Open Archives Initiative (OAI) is one effort to improve access to resources in the "deep Web"—Web-based and Web-accessible resources that are not easily discovered because they are hidden in databases or dynamic Web pages that cannot be "crawled" by the familiar search engines such as Alta Vista and Google. By some estimates these resources contain 400 to 500 times more information than the "surface Web." RLG's databases are a good example of resources in the deep Web.

Originally, the Open Archive Initiative grew out of the need to enhance access

National Library of Russia Contributes File to Hand Press Book Database

Curtis Lavery, RLG, crl@notes.rlg.org

RLG recently added more than 4,000 records from the National Library of Russia to the Hand Press Book Database. These new records are the initial load of a collection that contains more than 70,000 volumes. The collection is particularly rich in incunabula: over 6,000 15th-century works provide a valuable resource for scholars interested in the history of European printing. When the file is complete, it will include the remarkable collection produced by Elseviers, the celebrated 17th-century Dutch family firm that published mainly utilitarian books and works by the great humanist printer of the Renaissance, Aldus Manutius, and his heirs. Early Russian printed books are well represented in the

---

SHOW ROMANIZATION ONLY

— Record 1 — SAVE REQUEST

Author: Voltaire, F.M.A. de (François Marie Arouet de)
Вольтер, Ф.М.А. де (Франсуа Мари Аруэ де), 1694-1778.

Title: General'noe mnenie o Taktike g. Giberta / Socinennoe g. Volterom, ; A s francuskago âzyka perevedennoe na rossijskoj Lejb-gvardii Preobrazenskago polku podpraporsikom Fedorom Levcenkovym.
Генеральное мнение о Тактике г. Гиберта / Сочиненное г. Волтером, ; А с францускаго языка переведенное на российской Лейб-гвардии Преображенскаго полку подпрапорщиком Федором Левченковым.

Published: Sankt-Peterburg : [Tip. Art. i inz. kad. korpusa], 1777.
Санкт-Петербург : [Тип. Арт. и инж. кад. корпуса], 1777.

Physical Details: 14 с. ; 8o.

Location: National Library of Russia
Shelfmark: 18.316.4.161

Location: National Library of Russia
Shelfmark: 137/124

Notes: Прозаический перевод "La Tactique" Вольтера.
References: СК XVIII 1094.

Other Authors: Levcenkov, F. (Fedor)
Левченков, Ф. (Федор)
Artillerijskij i inzenernyj kadetskij korpus (Sankt-Peterburg.) Tipografiâ.
Артиллерийский и инженерный кадетский корпус (Санкт-Петербург.) Типография.
Volter.
Волтер.

Other Entries: Sankt-Peterburg.
Санкт-Петербург.

Subjects: Giber, Z.A.I., 1743-1790 (Zak Antuan Ippolit). Taktika.
Гибер, Ж.А.И., 1743-1790 (Жак Антуан Ипполит). Тактика.

Source: National Library of Russia.

Record ID: RUNC2378-E

Sample Hand Press Book Record

RLG Forum: Issues in Pinyin Transition

Karen Smith-Yoshimura, RLG, kss@notes.rlg.org

On October 1, 2000, the North American library community adopted Pinyin romanization for all Chinese characters in bibliographic records and in name and subject authority file headings. (See *RLG Focus*, 46:5, October 2000). Libraries are now working in a mixed Wade-Giles and Pinyin environment.

The RLG Forum during the American Library Association's midwinter conference in Washington, DC updated participants with the current status of LC, OCLC, and RLG's coordinated conversion of authority and bibliographic records to Pinyin romanization. A panel of representatives from the RLG Pinyin Working Group provided perspectives on how their libraries are addressing issues arising from the transition to Pinyin.

- Philip Melzer, Chair of the Library of Congress Pinyin Task Force, outlined the cleanup tasks now

underway and new documentation being added to the LC Pinyin Project Web site or prepared for the library community's review. LC gave special thanks to the 14 NACO participants who volunteered to modify more than 8,400 "undifferentiated" personal name authority records.

- Glenn Patton, OCLC Director, Metadata Standards and Quality, reported on OCLC's final testing of its bibliographic conversion software and reviewed options for libraries to receive Pinyin-converted bibliographic records from OCLC.

- Karen Smith-Yoshimura, RLG's Pinyin project manager, reviewed the most recent RLG Pinyin conversion milestones and new documents added to the RLG Web site. She acknowledged the significant contributions made by the RLG Pinyin Working Group members and staff at the Cleveland Museum of Art, University of Toronto, and University of Washington Gallagher Law Library. They did intensive reviews of test conversion runs on samples of their own records and discovered gaps in LC's specifications. The entire library community benefited from these libraries' efforts.

The panelists' presentations and the discussion afterward comprised the main portion of the program.

- Robert Wolven, director of bibliographic control and processing at Columbia University, noted that the Pinyin conversion fit into Columbia's existing plans to have its entire three-million record catalog processed by an authority control vendor. The lack of Wade-Giles cross-references in some of the Pinyin authority records would mean that some non-Chinese bibliographic records would retain their Wade-Giles headings.

- Don Thornbury, head of cataloging at Princeton University, focused on the complexities of Chinese-language holdings dispersed among catalogs and issues for staff and users in searching for materials in both Wade-Giles and Pinyin.

- Phyllis Valentine, head of library management systems at the University of Michigan, outlined the impact on workflow and systems processing and the staffing needed. Michigan's OPAC contains records for their entire collection, requiring authority work for non-Chinese records as well as the newly converted Chinese records.

- Joan Swanekamp, chief cataloging librarian at Yale University, described the Pinyin conversion in the context of other changes at Yale, and the local decisions Yale had made on cataloging Chinese-language materials during the transition period.

About 80 participants from a range of libraries attended the forum and participated in a lively discussion. They focused on the manual review all libraries can expect to do after their bibliographic records have been converted and the need for Wade-Giles cross-references on *all* name authority records with Pinyin headings. As a result, LC subsequently agreed to let OCLC add Wade-Giles cross-references to the Pinyin-converted Name Authority Records that do not currently have them. These updated records will be part of OCLC's normal contributions to the LC Name Authority File. In addition, LC's Cataloging Distribution Service has agreed to do a separate distribution of Pinyin-converted Name Authority records, so that any library needing these records can receive them directly from LC.

See the full report on the January 14, 2001 RLG Forum at www.rlg.org/eas/rlgforum3.html. ∎

Marcadia® Projects at the Harvard College Library
How the automated cataloging service from RLG and MARC Link is helping Harvard College Library reduce its cataloging backlog

Lynda S. Kresge, Harvard College Library Technical Services, lkresge@fas.harvard.edu

Harvard College Library (HCL) Technical Services adds 115,000 new English and Western European language titles annually to the cataloged collections of Widener Library, the research library for Harvard's Faculty of Arts and Sciences. Over the past decade, HCL Technical Services reduced its backlog of uncataloged materials from 325,000 to 100,000 volumes, all of which are represented by brief bibliographic records in HOLLIS, Harvard's integrated library system.

Every week the Harvard Office for Information Systems runs these brief records against the Harvard Books File, a separate, local database of full bibliographic records from the Library of Congress. The Books File includes LC cataloging records from the last five years and is updated weekly. Matches from this standing search automatically overlay the brief HOLLIS records. In addition, HCL Technical Services regularly compiles and runs batch search files through the OCLC Cataloging MicroEnhancer program to identify and download copy from WorldCat. As full records are found and loaded into HOLLIS, the corresponding items are pulled from the backlog and processed for the library stacks.

In 2000, HCL Technical Services contracted with MARC Link Corporation to use the Marcadia service, which searches for matches against the RLG union catalog, as a new third step in batch searching for cataloging copy for the backlog. The first three files included titles in French and Italian, titles in Germanic languages, and titles published in New Zealand. These were chosen because they represent areas of the backlog with the lowest percentage of copy cataloging available in WorldCat. The files were sorted by HOLLIS record number within publication date ranges (dates of publication

ranged from 1906 to 2000). The arrangement of these search files reflected the order of these materials in the backlog; fortunately MARC Link was able to retain this sort order when returning matches.

The results were successful beyond expectation. Of the 27,950 titles searched in French and Italian, Marcadia returned 12,200 matches, for a match rate of 44%. A search of 45,147 records in Germanic languages yielded 9,043 matches, a 20% match rate, and the 1,128 New Zealand records returned a whopping 79% match rate, with 889 matches. These are high returns for these categories, especially for titles that had already been searched regularly against both an LC file and OCLC's WorldCat.

The matching records returned by MARC Link have been loaded into the Harvard Books File, where catalogers retrieve them for overlay in HOLLIS as the materials are pulled from the backlog. Macros developed for the HOLLIS technical services

module facilitate quick and easy search and overlay between the two databases. Catalogers are evaluating the quality of the Marcadia records returned as matches, to determine if direct overlay without review is advisable. The MARC Link reports of "non-hits" and "non-matches" are used to identify materials from the backlog for in-house original cataloging and for outsourced original cataloging projects.

Building on the success of these initial projects, HCL Technical Services is planning another Marcadia project to search Spanish, Portuguese, and selected English language materials, and is considering a follow-up project for French, Italian, and Germanic language materials.

[Further information about Marcadia is available from RLG's Web site at www.rlg.org/marcadia.html. Lynda Kresge has also kindly agreed to answer any follow-up questions readers may have via e-mail.] ∎

---

collection, with Slavonic incunabula; editions issued by Slavonic print shops in the Balkans, Venice and Ugro-Wallachia; and the greatest collection of books from the era of Peter the Great. About half of the items are printed in Cyrillic script. The majority of the records recently loaded represent Russian-language materials. Other languages include German, Latin, French, modern Greek, Italian, and Polish. Though light on topical headings, the file contains many geographic and place-of-publication entries. The National Library of Russia will send regular updates for addition to the database.

The National Library of Russia in Saint Petersburg, which joined the Consortium of Research Libraries (CERL) in 1993, is Russia's oldest state library. One of the world's largest libraries, it possesses the most complete collection of publications in Russian. With this recent addition, the Hand Press Book Database has grown to more than 875,000 records. The Hand Press Book Database is created and contributed to by CERL.

For more information on the Hand Press Book Database, see its fact sheet at www.rlg.org/hpb.html. ∎

Macro Express® Works with RLIN® Terminal for Windows® for Productivity Gains

Paul Eugene Maher,
Library of Congress, pmah@loc.gov

Users of RLIN Terminal for Windows know that they can make their searching and cataloging operations more productive by storing frequently used commands on chosen keyboard buttons, for convenience and speed. (For examples, see *RLIN Focus* 23:5, December 1996: www.rlg.org/r-focus/i23.buttons.html.) RLIN Terminal for Windows also allows for saving and reusing strings of frequently needed text (with the store/copy keys). Several programs that enable users to create "macros" for use in their technical processing offer many more productivity gains.

A macro is a string of commands that "can be played back at will to perform a given task" (Macro Express® home page overview). Since delays can be built into a macro, it is possible to execute a command, wait for a system response, and continue with additional commands—all by pressing one key. A real advantage of creating and using macros is that they can tab across fixed fields and field boundaries, or even press the Send Message key.

Macro Express is one such macro utility program that performs complex repetitive actions. (Other programs that can be used to create macros include Microsoft Visual Basic.) By configuring Macro Express to run from a pop-up menu, one can use it within the RLIN system. Macro Express can also take advantage of the text saved with RLG Terminal for Windows' store/copy keys—including Arabic, Cyrillic, Chinese, Hebrew, Japanese, and Korean characters.

Here, for example, is a customized template that the Library of Congress uses for preliminary cataloging of Hebrew-script materials. The template is invoked with a single click from a macro created in Macro Express:

```
ID:             RTYP:c  ST:p  FRN:  MS:  EL:5  AD:02-13-01
CC:9554 BLT:am   DCF:a  CSC:  MOD:  SNR:  ATC:  UD:
CP:is   L:heb   INT:?  GPC:?  BIO:?  FIC:?   CON:????  TOC:
PC:s    PD:2001/   REP:  CPI:?  FSI:?    ILC:????   II:?
007 MICROFORM (COM:h)
SMD: OR: POL: DM: RR:    COL: EML: GEN: BSE:
020 __
040    DLC-R$cDLC-R
042    pcc
LDB    $bye01 to RCCD 02/13/01
LDF    0$bibc$corignew$d3$encip$f19$gn-rlinjack
LDG 0  acquire$b1 shelf copy$xdefault policy
```

And here is what kind of coding went into this reusable macro. I've included extra spaces and carriage returns for readability and—in square brackets—interspersed the template lines and some comments.

The actual code would not include any of the square-bracketed material, and one would include spaces only where they constitute part of the data.

```
[The following code fills in the fixed fields:]
<ACTIVATE:RLIN> <HOME> <TAB> <TAB> <TAB> <TAB> <TAB> 5 <TAB>
[ID:             RTYP:c  ST:p  FRN:  MS:  EL:5  AD:02-13-01]
9554 <TAB> am <TAB> a <TAB> <DELETE> <TAB> <DELETE> <TAB> <TAB> <TAB>
[CC:9554 BLT:am  DCF:a  CSC:  MOD:  SNR:   ATC:    UD:]
is <TAB> heb <TAB> <TAB> <TAB> <TAB> <TAB> <TAB> <TAB>
[CP:is   L:heb   INT:? GPC:? BIO:? FIC:?   CON:???? TOC:]
s <TAB> 2001 <TAB> <TAB> <DELETE> <DOWN ARROW> <TAB>
[PC:s    PD:2001/      REP:  CPI:? FSI:?    ILC:????  II:?]


[The following code creates the 040 and 042 fields:]
<SHIFT> <INSERT> 040 <TAB> <TAB> DLC-R <KEYP-> c DLC-R <TAB>
[040    DLC-R$cDLC-R]
<SHIFT> <INSERT> 042 <TAB> <TAB> pcc <TAB>
[042    pcc]

<SHIFT> <INSERT> LDB <TAB> <TAB> <KEYP-> b ye01 to RCCD <MM/DD/
YY{YF000}>
[LDB    $bye01 to RCCD 02/13/01]
[ "<KEYP->" refers to the "minus" key on the keypad, i.e. a
delimiter; "<MM/DD/YY{YF000}>" inserts the current date taken from
the computer's clock.]


[LC uses the following fields for internal "housekeeping"]
<TAB> <SHIFT> <INSERT> LDF <TAB> <TAB> 0 <KEYP-> b ibc <KEYP-> corignew
<KEYP-> d 3 <KEYP-> e ncip <KEYP-> f 19 <KEYP-> g n-rlinjack <TAB>
[LDF    0$bibc$corignew$d3$encip$f19$gn-rlinjack]
<SHIFT> <INSERT> LDG <TAB> 0 <TAB> acquire <KEYP-> b 1 shelf copy <KEYP->
x default policy <TAB>
[LDG 0  acquire$b1 shelf copy$xdefault policy]
```

A number of Macro Express features make it valuable in other ways. For example, Macro Express can read an ID from a file, search that ID in the RLG union catalog, wait a specified period for a system response, and then print the result, issue the Put command, start the update process, and more. This has proven very useful for file maintenance projects and the like.

You can find more information about Macro Express at www.macros.com. ∎

to scholarly e-print archives, but the participants soon realized that the technology framework they were developing could be applied to a broader range of applications and have embodied this expansive vision within the initiative's mission statement:

> The fundamental technological framework and standards that are developing to support this work are, however, independent of both the type of content offered and the economic mechanisms surrounding that content, and promise to have much broader relevance in opening up access to a range of digital materials. As a result, the Open Archives Initiative is currently an organization and an effort explicitly in transition, and is committed to exploring and enabling this new and broader range of applications.

The OAI is sensitive to interpretations of the word "archives." Its Web site states,

> Members of the archiving profession have justifiably noted the strict definition of an 'archive' within their domain; with connotations of preservation of long-term value, statutory authorization and institutional policy. The OAI uses the term 'archive' in a broader sense: as a repository for stored information.

The first piece of the technological framework to be promulgated is the Open Archives Initiative Protocol for Metadata Harvesting, which collects metadata from data providers. The protocol utilizes the HyperText Transfer Protocol (HTTP) to communicate and consists of a set of standard requests and responses for determining what data is available from a supplier, and for transferring it in XML. Dublin Core is mandated as a required metadata format and other formats are likely to result from devel-

Preview of RLG Prices, 2001–2002

This article recaps RLG president James Michalko's March 9 preview memo to service liaisons. Page references are to the current (2000–2001) *RLG Price Lists for Information Resources and Related Library Services*. The new *RLG Price Lists*, effective from September 1, 2001 through August 31, 2002, will be published in June.

Information Resources
(Price lists 1 and 2, pages 4–9)

Library Resources and Archival Resources. All stay at 2000–2001 prices. (Library Resources are the RLG union catalog, including authority files and access to the Deutsche Bibliothek Database, CURL Union Catalogue, and National Library of Australia Catalogue; the English Short Title Catalogue; the Hand Press Book Database; and the SCIPIO database.) *New:* In the coming service year, you will have a choice of subscribing or paying by the search for access to the SCIPIO database and the Hand Press Book Database. The SCIPIO per-search charge will be the same as for the RLG union catalog; the Hand Press Book per-search charge will be the same as for the English Short Title Catalogue.

Citation Resources. Annual subscription prices will increase an average of 3% across all databases. Increases reflect the growth of these resources and of costs to maintain them. *New:* Anthropological Index from the Royal Anthropological Institute, at $1,700 (nonmembers) or $1,610 (members) for five simultaneous users.

Cataloging Services
(Price list 3, pages 10–12)

As already communicated directly to customers for catalog cards and various printed reports, cards and reports have been discontinued. Institutions needing catalog cards are encouraged to contact MARC Link

Corporation (see www.marclink.com). Record transfers and Marcadia® stay at 2000–2001 prices.

Customer Support—Unchanged
(Price list 4, page 13)

PC Software
(Price list 5, pages 14–15, plus www.rlg.org/ariel/ari3reqpri.html)

Ariel® version 3 pricing was added to our Web site in January and is good through 2001–2002. With the release of this new software, RLG will no longer offer any earlier version. Upgrade pricing from version 2.2 includes a significant discount in the first 90 days after version 3.0's release.

Network Connections
(Price list 6, page 16)

Connections via the Internet and via UUNET dedicated IP network or private asynchronous connection stay at 2000–2001 prices. Service via direct dial will be discontinued after August 31, 2001. Given diminished use, we are still evaluating whether to discontinue the UUNET reserved dial and network dial options. We encourage users not now connecting over the Internet to contact us about that option.

Prepayment Deposit Accounts
(Service Terms, page 19)

A number of customers have prepayment deposit accounts with us. Effective September 1, 2001, the minimum deposit to establish or maintain a prepayment deposit account will increase from $125 to $400. Monthly interest will be credited only on balances of $1,000 or more. A service fee may be applied whenever a deposit account drops below $200. Late-payment fees applied when an account falls to a debit balance are unchanged. ∎

opments within specific communities of interest. The requester is assumed to be a service provider who will use the metadata to build value-added services, such as notification (selective dissemination of information) or content federation (consolidating data from a variety of sources around a topical theme). The first version of the protocol was the subject of standing-room-only open meetings in Washington and Berlin earlier this year.

The OAI is based at Cornell University and is supported by the Coalition for Networked Information (CNI) and the Digital Library Federation (DLF). A steering committee sets policy. The OAI maintains a registry of data providers, although registration is not required in order to use the protocols. The OAI provides a number of tools to help protocol implementers, including a repository "Explorer" that can be used to test data providers' implementations.

The OAI protocol is quite simple, especially compared to the more familiar library-oriented protocols, Z39.50 (information retrieval) and ISO 10162 (Interlibrary Loan). This makes it easy to implement, but may eventually limit its applications. As yet, the nature of the services that will be based on metadata harvested using this protocol is not clear. Additional information is available from the Open Archives Initiative Web site, www.openarchives.org. ∎

Contents

1200 villa st., mountain view, ca 94041-1100 usa

RLG

Some Upcoming Events

See also RLG's Web site at www.rlg.org/events.html

June 14–20, 2001. San Francisco, CA. American Library Association (ALA) Annual Meeting. RLG exhibiting. For more information, contact lynn.pazar @notes.rlg.org or 650-691-2278.

June 24–27, 2001. La Jolla, CA. 36th Annual Conference of the Association of Jewish Libraries (AJL). RLG meeting. For more information, contact kss@notes.rlg.org or 650-691-2270.

August 27–September 2, 2001. Washington, DC: The Society of American Archivists. RLG exhibiting. For more information, contact kris.tague@notes.rlg.org or 650-691-2264, or anne.van.camp @notes.rlg.org or 650-691-2237.