

RLG Cultural Materials: The Potential to Change Research and Scholarship

Steve Hensen, Duke University, hensen@duke.edu

An open forum at the American Library Association conference in June 2001 gave attendees the most in-depth view to date of the coming RLG Cultural Materials online resource, including a comprehensive system demonstration. RLG staff discussed the reasons the membership organization undertook this initiative. Steve Hensen, director of planning and project development in the Rare Book, Manuscript, and Special Collections Library of Duke University, was a featured speaker, whose comments are excerpted here. Duke has been a significant and early participant in the members' alliance to create RLG Cultural Materials. Among the first collections in the new resource are Duke's Historic American Sheet Music and photographs by William Gedney.

For scholars, researchers, and the general public, perhaps one of the most important aspects of the RLG Cultural Materials alliance is the fact that the focus is principally on material that is often by definition unique or fugitive and not easily accessible. Manuscripts, archival materials, photographs, art and museum objects, film and dance collections—in short, the unpublished side of what constitutes “cultural materials”—typically exist in only one copy and place. While the Internet has made it possible for scholars now to use such materials “virtually,” there has been no organized, integrated way to do this—until now.

A pooled resource

We at Duke are proud of the various collection materials we have made available digitally. But they are “stand-alone” exhibits. Apart from links to high-level MARC records and the degree to which they may share some

cohesion through Encoded Archival Description, these are not integrated in any systematic way with other materials in our library—let alone with other libraries. Unfortunately, this is the case with most such projects in other libraries.

The RLG Cultural Materials initiative addresses this problem by creating a

union database of cultural resource materials from a wide variety of institutions. This may sound simple on the face of it, but it raises many interesting and complicated questions. Fortunately, RLG's principal focus has been on addressing and trying to answer these questions.

(continued on page 2)

Mission Accomplished: RLG Union Catalog Converted from Wade-Giles to Pinyin

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

RLG has finished converting all Chinese-language records in the RLG Union Catalog from Wade-Giles to Pinyin romanization.

This marks the end of a project begun in 1997, when the Library of Congress announced that the ALA/LC romanization for all new Chinese bibliographic and authority records would be Pinyin, subsequently determined to begin October 1, 2000. This milestone meant that American libraries joined the international community in using Pinyin, facilitating access to Chinese materials by scholars everywhere. (See the October 2000 joint LC-OCLC-RLG press release, www.rlg.org/pr/pr2000-pinyin.html.)

In preparation, OCLC converted all LC name authority records with a

Wade-Giles heading to Pinyin, which LC then redistributed and RLG loaded into RLG's copy of the LC Name Authority File in September 2000. Changes were made to RLG's RLIN® system so that its users had to enter new Chinese-language records in Pinyin from October 2, 2000.

Since November 2000, RLG has been converting the over 2.5 million Chinese-language records in the RLG Union Catalog from Wade-Giles to Pinyin. After first converting clusters with Library of Congress records and those from several large RLG member's Chinese collections, RLG converted remaining records alphabetically by Library Identifier. RLG sent notifications to each library when

(continued on page 11)

(RLG Cultural Materials...continued from page 1)

Content integration

1. How? Through a clever combination of site spidering, indexing, local storage of "microsurrogates," and remote access to original digital source files, RLG and the alliance have found a workable model to collect and integrate cultural resources from a wide variety of locations and repository types.

2. What? Back in the early days of integrating archival and mixed collections (AMC) into the RLIN processing system and union catalog, we knew that the forces of interdisciplinarity were moving us towards a database for research with no artificial boundaries of subject or material type. This is certainly the case with RLG Cultural Materials. While an alliance subcommittee is looking at developing coherent "selection criteria," the initial thrust has been towards richness, variety, and potential research value. The beauty

and challenge of scholarly interdisciplinary research is that it is nearly impossible to anticipate where and how connections will be made.

Metadata integration

Perhaps not surprisingly for those who know me, metadata integration interests me most in RLG Cultural Materials. The primary barrier to fuller integration of cultural material types in the past has been the differing perspectives and requirements of various communities about how their materials are described and cataloged. With the plethora of new metadata schemes for the World Wide Web, it has become even more difficult to find common ground.

I firmly believe that we can and must eventually achieve agreement on basic metadata standards, perhaps aided by the sheer accumulation of what is perceived to be "best practice." Until then, RLG has neatly finessed the issue by providing an underlying database structure that will translate the various schemes into consistent and common elements—all the while losing nothing that is "particular to the originating discipline."

Usability

An early metaphor for Web use was trying to sip water from a fire hose. The "dot.com" revolution made it even worse. The mass of scholarly and research-oriented resources available online has increased exponentially over the past five years, thoroughly stirred and blended into the petabytes of other "stuff" that make up the great stew of the Web. A scholarly search using some of the best Web searching tools, such as Google, iLor, Yahoo, and AltaVista, will certainly yield useful results. Unfortunately, they must be culled from often hundreds of thousands of useless hits. More to the point, this information—even once selected—is hard, if not impossible, to consolidate and collocate in any useful way.

RLG Cultural Materials addresses these problems. First, it concentrates the available resources through a careful process of digital collection development, intended to make the universe

RLG Cultural Materials is the result of an unparalleled effort from an international alliance of distinguished institutions. It enables users to explore and study historical documents, artifacts, photos, works of art, film clips, music, and much more.

Licensed for educational use, the growing online resource will be available at an introductory subscription price of \$5,000 (significantly less for RLG members), covering January 1 through August 31, 2002. Free trials will be available before January.

For more information about access to RLG Cultural Materials, send e-mail to Pamela.Dewey@notes.rlg.org or call +1-650-691-2208 (fax: +1.650.964.1461). Visit www.rlg.org/culturales for more about the RLG Cultural Materials initiative and alliance.

RLG Focus is a bimonthly publication of the Research Libraries Group, 1200 Villa Street, Mountain View, CA 94041-1100 USA.

Editor: Karen Smith-Yoshimura
Contributing Editors: Tony Gill, Ed Glazier, Sue Marsh, and Sharon Vaughn-Lahman
Graphic Production Artist: Monique Funnie
Advisory Board: Kristine Brancolini, Indiana University (Bloomington); Sherman Clarke, New York University; Carla Kupitz, Brigham Young University; Jeffrey Luttrell, Princeton University; Chris Ward, New York State Archives and Record Administration.

To subscribe to RLG Focus, for address changes, or for additional copies, contact ldp@notes.rlg.org. Please send all comments and contributions to the editor at kss@notes.rlg.org.

Articles in this issue are available at www.rlg.org/r-focus. For additional information on services mentioned, contact the RLG Information Center at 800-537-7546 or bl.ric@rlg.org.

To join the RLG services news and discussion mail group, send the message **subscribe rlg-services-forum <your name>** to lyris@lists2.rlg.org.

CJK®, Eureka®, Marcadia®, and RLIN® are trademarks of the Research Libraries Group. Ariel® is a registered trademark of Ariel Corporation, used by RLG under license. Windows® is a registered trademark of Microsoft Corporation.

Correction: Page 7 of the print version of RLG Focus issue 51 gives an incorrect web address for an article on Yale's uses of RLG Archival Resources. The URL for the earlier article is www.rlg.org/r-focus/i43.yale.html.

of available resources more meaningful and to the point. (I must say that it is the responsibility of the participating institutions to ensure the database reaches its full critical mass and potential.) Second, it provides a congenial and intuitively designed workspace, in which users can not only search across a wide variety of institutions for relevant materials, but also bring these materials together to "compare and contrast" on their desktops.

It is no wild-eyed hyperbole to declare that efforts such as this have the potential to change the nature of research and scholarship forever. The Web and the Internet have completely altered the information and communications universe. This initiative, in seeking to more sharply and intelligently focus this information, will make it more useful.

When the Web burst upon the scene nearly ten years ago, there were many who imagined that the dream of universal access to information would finally be realized. We now understand that the essence of this dream can only be fully attained through the kind of cooperation and collaborative mediation achieved in efforts such as RLG Cultural Materials. ■

Building the RLG Cultural Materials Infrastructure

David Richards, RLG, david.richards@notes.rlg.org

The RLG Cultural Materials system is designed to provide integrated access to digital surrogates (e.g. images, audio, video) of cultural artifacts of all kinds, along with extensive descriptive and contextual information. The system can be divided into three functional parts: surrogate storage, search and retrieval of descriptive information, and a Web user interface.

Hardware platform

The system is spread across three groups of servers: digital surrogate servers, a database server, and appli-

cation servers. These are all 64-bit Sun servers running Solaris 8.0, Sun's version of UNIX. The database server is an UltraSparc 420R, while the digital surrogate and application servers are UltraSparc 220Rs. An additional server is used for processing images; JPEG and MrSID images in various resolutions are derived from the high-resolution source image (supplied by a contributor) before being placed on an appropriate digital surrogate server.

The digital surrogate files and databases are stored on a RAID-5 IBM 2105 Enterprise Storage System (ESS). This is the same highly reliable integrated disk storage that RLG uses for its bibliographic and citation databases.

This flexible architecture makes it easy to incorporate additional digital surrogate and application servers as demand for the service grows.

Software architecture

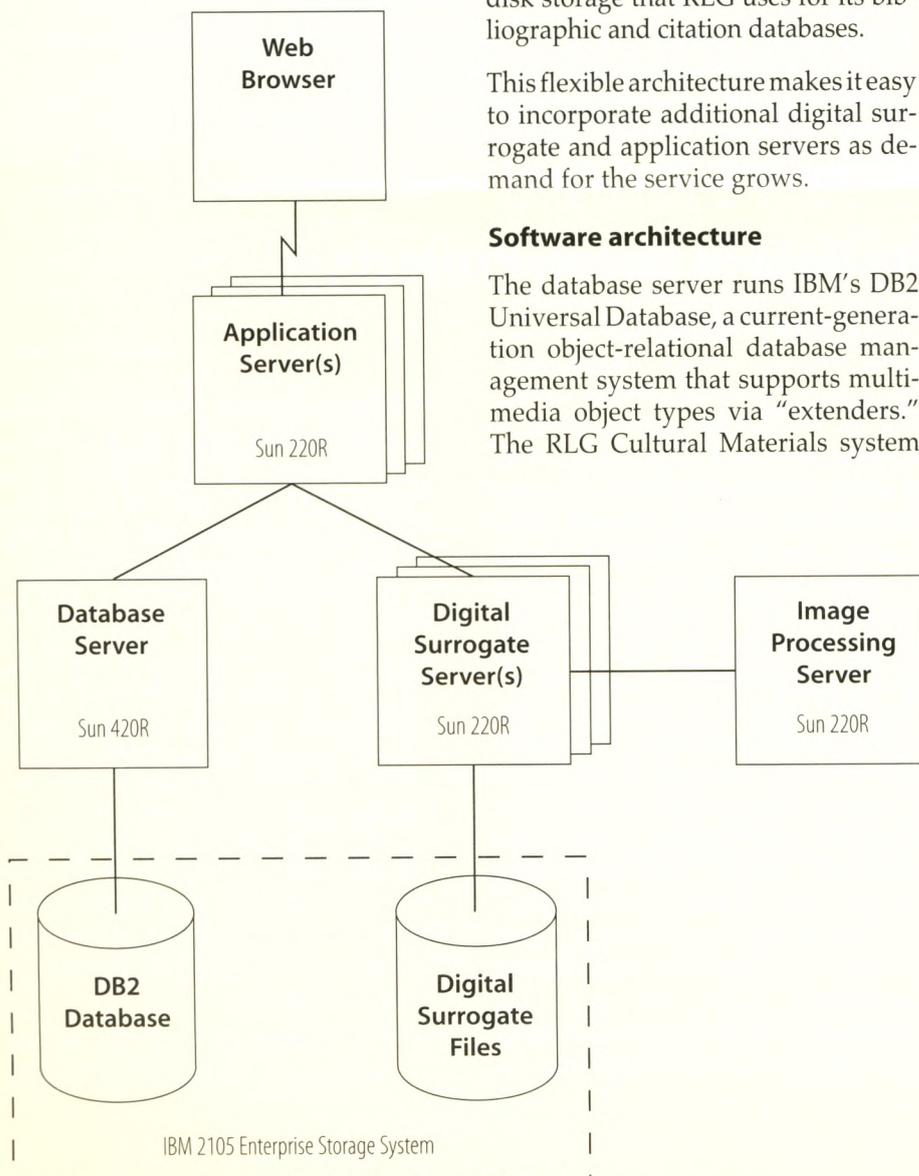
The database server runs IBM's DB2 Universal Database, a current-generation object-relational database management system that supports multimedia object types via "extenders." The RLG Cultural Materials system

uses this capability in several ways. The descriptive data is stored as XML (Extensible Markup Language), which can be searched using IBM's XML Extender and formatted for display using XSLT (Extensible Stylesheet Language: Transformations). The IBM Text Information Extender supports flexible searching of textual data by offering features such as wildcards, proximity searching, fuzzy searching, and relevance ranking.

The application servers run IBM's WebSphere and an Apache Webserver. WebSphere provides the framework and "plumbing" for a Web application; for example network connectivity, session management for multiple users, and access to the database. The cultural materials application is written in the Java language and follows the "Model-View-Controller" design pattern: the core business logic or *model* of the system is implemented with Enterprise Java Beans (EJBs). Some EJB software modules are standard components provided by IBM—for example database access modules—while others were developed by RLG and are unique to the cultural materials application. The user's *view* of the system is implemented with "Java server pages" (JSPs), which consist of HTML and optional embedded logic expressed in Java, and determine what the user sees on the computer screen. The system's *control* functions (e.g., user authentication, logging requests, managing cookies, and managing the connection) are implemented with "Servlets," programs written in Java that run on the application server and are invoked by browser access.

From a software standpoint, the digital surrogate servers are relatively simple. Digital surrogates are stored as operating-system files and retrieved via an Apache Web server. The image processing server uses NFS (network file system) to transfer the service images to a digital surrogate server after they are created.

(continued on page 4)



(Building...continued from page 3)

Managing digital collections

Because we expect the RLG Cultural Materials database to grow rapidly, we devoted considerable effort to designing an efficient and scalable digital surrogate and data collection process. To this end, we developed the Cultural Materials Alliance Content Manager, which both contributors and RLG staff use to enter collection descriptions, file locations, and other administrative information about the digital collections that are contributed to the RLG Cultural Materials service. When all Content Manager information is complete and the digital collection is ready for processing, RLG obtains the descriptive data for analysis, conversion, and loading into the DB2 database. As part of that process, information in the content manager and the descriptive data is used to drive the automated gathering and processing of the corresponding digital surrogates.

Flexibility for the future

This system represents the first of a new generation of RLG services that will all use the same underlying technology. After carefully evaluating available database and application server software, we chose IBM's DB2 and WebSphere products because

they offer the best combination of capabilities and price for our intended applications. As we gain experience with these new tools, we are beginning to see how they could be used to reengineer our "legacy" services, to reunify and simplify RLG's operational environment. ■

For more information on RLG Cultural Materials and the various technologies mentioned in this article, see:

RLG's Cultural Materials Initiative: www.rlg.org/culturalres/

RLG Services: www.rlg.org/r-focus/i50.html#technology

Sun UltraSparc: www.sun.com/microelectronics/UltraSPARC/

Sun Solaris: www.sun.com/solaris/

MrSID: www.lizardtech.com/products/mrsid/

IBM WebSphere: www-4.ibm.com/software/webservers/appserv/

IBM UDB DB2: www-4.ibm.com/software/data/db2/

IBM Enterprise Storage: www.storage.ibm.com/hardsoft/products/ess/news.htm

XML: www.xml.com/

XSLT: www.xslt.com/

MARC Link Expands its Role in Recon Projects

by Sharon Cline McKay, MARC Link Corporation, smckay@marclink.com

MARC Who

You may know that MARC Link Retrospective Conversion provides high-quality retrospective conversion services to libraries around the world, but can you name all the companies from which it has descended? Here's a brief lesson in MARC Link's genealogy:

In 1998, epixtech (previously Dynix, then Ameritech) decided to focus on its core automation business. At that time, it began to outsource the business of Retro Link Associates (a subsidiary) to MARC Link Corporation, a company established by former managers and employees of Retro Link. By 1999, MARC Link had absorbed all the people, projects, and software tools from Retro Link, including Marcadia® development work, maintenance, and processing.

MARC What

Retrospective conversion (recon) is the process of converting manual catalog records into machine-readable cataloging records. In traditional recon projects a data technician works from catalog cards, searching the item online and verifying each possible match against the card. Marcadia, on the other hand, is an automated cataloging service developed in partnership with RLG. Marcadia matches brief electronic records with records from the RLG Union Catalog. Using either approach, MARC Link customizes results based on parameters outlined with the library. Normal match points include title, publisher, and date of publication (within one year), plus verification on author, form of item, ISBN, LCCN, and edition (if known).

For recon projects, data technicians can also work from photocopies of title pages and versos, or even from the actual publications. Whenever an acceptable match isn't found, information supplied by the library is keyed into an original entry in MARC21 format according to AACR2. This last step in the recon process enables MARC Link to return 100% of a library's records in machine-readable format.

MARC How

MARC Link uses both the RLG Union Catalog and our own internal database for retrospective conversion projects. The MARC Link internal database combines:

- The Library of Congress MARC database—Over five million biblio-

graphic records distributed by the Library of Congress since 1968 for books, maps, music, serials, and visual materials. This database includes LC foreign records.

- The Quickcat database—Approximately three million bibliographic records converted from various academic and government sources. Many of these records predate the LC MARC database.
- MARC Link's original entry and audiovisual databases—Over 800,000 records converted from various academic, special, public, and school libraries. These records were originally created by MARC Link's data technicians and the database is continually updated as projects are completed.

Access to these two major resources gives clients the option of using RLG's Union Catalog as the primary source, supplemented by MARC Link's in-house database ("RLIN® Premium Service")—or vice versa ("RLIN® Minimal Service"). Typically, samples are processed to help clients decide which service to use. Results show the relative match rates and actual records that are retrieved from the two databases.

MARC Twain

After using our unique system of scanning cards to create images on the computer, MARC Link uses specially developed software, called MARC Twain, to search the RLG Union Catalog directly. MARC Twain enables data technicians to link the scanned images to records derived from RLIN in one step, quickly searching for matches and then editing records for clients. These linked records provide for more efficient processing and easier quality-control checking, and they eliminate the need to handle catalog

cards more than once. Technicians can manipulate the display of multiple records to show specific fields and can use macros to speed up data entry of frequently used input.

MARC Original Scripts

MARC Link has in-house expertise with Japanese, Chinese, Korean, Arabic, Hebrew, and Cyrillic scripts, which are particularly prone to inconsistency and ambiguity when romanized. Data technicians use keyboard maps printed out from the RLIN® Terminal for Windows® software as a guide. MARC Link uses the RLG Union Catalog for projects that require original scripts. Our new Hebrew team has completed Hebrew-script projects for New York University's Bobst Library and for Hebrew Union College.

MARC Recon Projects

MARC Link projects are varied in size and content and require extensive attention to detail. For the Huntington Library (235,000 records) and the Historical Society of Pennsylvania (320,000 records), matches found on the MARC Link databases were edited to fit the requirements of the libraries. The University of Pennsylvania (more than 380,000 records) already had a "brief record" database, which provided holdings information for their new records. The existing BobCat records (17,000) were edited online for the New-York Historical Society (130,988 records).

Special collection projects require a variety of approaches. For the Library Company of Philadelphia Print Department (7,500 prints and maps), entries were keyed from photocopies of multiple cards. The Kurth Memorial Library Genealogy collection (8,000 records) had records for their microfilm collection keyed from ledgers and

copies of the film boxes. For New York City Municipal Reference and Research Center (24,000), MARC Link entered subject tracings from cards and standardized them to the Center's own name and subject authority files. For the Victoria and Albert Museum (48,000), technicians searched the Auction House Catalogs in RLG's SCIPIO database.

MARC International

MARC Link is expanding internationally with recently completed projects for the Literary and Philosophical Society of Newcastle Upon Tyne (the "Lit and Phil"), the Victoria and Albert Museum (where a new project is now underway), and the University of London. MARC Link was recently selected as the preferred supplier for the retrospective conversion of the Guardbook catalogue for Edinburgh University Library in Scotland. The Guardbook catalogue is a union catalog of pre-1985 acquisitions. This covers the majority of the printed material added to Edinburgh University Library since its foundation in 1580. In addition, MARC Link has recently completed its acquisition of the assets and goodwill of Saztec Europe, Ltd. Saztec had provided retrospective conversion services to libraries worldwide, primarily in the U.K. and Western Europe, from offices in Ardrossen, North Ayrshire, Scotland.

MARC Where

For more information, please contact MARC Link directly at:

MARC Link Retrospective Conversion
175 North Freedom Blvd., Suite 100
Provo, UT 84601
800-288-1265
801-356-1852
Fax: 801.356.8220 ■

Resource Enrichment

Over 1,500 Citations Added to the Chicano Database

Sharon Vaughn-Lahman, RLG, svl@notes.rlg.org

Author: Valle, Maria Eva.

Title: Las chicanas : alcances y retos hacia el futuro.

In: VICEVERSA [Mexico City, Mexico] No. 80 (January, 2000), p. 26-31.

UC Berkeley Ethnic Studies Library, Chicano Studies Collection

Call No.: AP63.1.V539 Serials

Subjects: Chicanas.

Chicano Studies.

Feminism.

Chicano Movement.

Stereotypes.

Identity.

Genre Or Form: Journal Article.

Record ID: XCDB57471

Chicano Database Record

In August 2001, RLG added over 1,500 bibliographic citations to the Chicano Database, enhancing the currency of this important research tool. The database received high praise in *Library Journal* *netConnect's* winter 2001 issue in "Accessing America's Minorities," a review of the most notable ethnic studies files. Cheryl LaGuardia and Christine Oka described the Chicano Database as "indispensable for college and university students and researchers in Chicano and Hispanic Studies," stating it was "the comprehensive source for all types of materials (books, journals, and newspaper articles) about Mexican Americans (Chicanos)." As the US continues to see dramatic shifts in its population, with Hispanic-Americans forming an increasingly significant portion, interest in Chicano studies remains high.

Compiled by the Ethnic Studies Library, University of California, Berkeley, the Chicano Database is recognized as the single best source for information on Mexican-American topics and the only specialized database for Chicano refer-

ence. The broader Latino experience of Puerto Ricans, Cuban Americans, and Central American immigrants is also included in this index.

The Chicano Database provides a unique way to view changing sociocultural attitudes throughout history. The database cites materials, generally not available elsewhere, written by and about Chicanos. Researchers will discover extensive coverage of materials dating back to the 1960s, with selective coverage back to the early 1900s.

The multidisciplinary nature of Chicano studies is reflected in the wide-ranging topical coverage. Bibliographic citations encompass art, bilingual education, economics, film, health, history, labor, law, literature, mental health, music, psychology, public policy, religion, sociology, and much more.

For information about trial and subscription access to the Chicano Database, or any of RLG's Citation Resources, contact the RLG Information Center, bl.ric@rlg.org. ■

Weinberg Contributor

Joan Biella, Library of Congress, jbie@l...

Imagine you're a cataloger, pulling a book written in Hebrew from a box that just arrived from Jerusalem. While you search for the item in the RLG Union Catalog, you think to yourself that the book is so new you won't find copy for it. But there is copy—a brief record in MARC21 format with ISBN and alternative title access and even the appropriate parallel field in Hebrew script! You've found a vendor record supplied by the A.I. Weinberg Book Agency, a popular provider of current Israeli monographs and serials to U.S. libraries.

Technical services staff working with French, Italian, Japanese, Spanish and German language materials have long benefited from the presence of vendor records in the RLG Union Catalog. Now timely dual-script records for Hebrew books are available. In the spring of 2001, RLG added vendor records representing Israeli monographs into the RLG Union Catalog. [Editor's note: see www.rlg.org/global.htm for descriptions of other record contributors to the RLG Union Catalog.]

Bibliographic information direct from the vendor benefits users in many ways. For bibliographers these records provide easily accessible information on books available for purchase. Acquisitions and cataloging staff can derive their own records from the vendor's brief records, saving considerable keying. In the RLG Union Catalog, records provide just-in-time information on available books to individual libraries. Reference staff and end-users can get the books immediately or request them through interlibrary loan.

RLG and MARC21 standards require vernacular fields to be paired with romanized fields. When it was initially proposed that A.I. Weinberg contribute records to the RLG Union Catalog, however, their records contained vernacular data only. Weinberg volunteered to provide romanized fields for personal names (100 field), titles/statements of responsibility (245 field), and imprints (260 field). The Library of Congress (LC) volunteered to reformat the data.

Hebrew Records into the RLG Union Catalog

Author: Katz, Ktzia.

כ"ץ, קציעה.

Title: Simlah dakah, levanah / Ketsi'ah Kats.

שמלה דקה, לבנה / קציעה כ"ץ.

Published: Yerushalayim : Karmel, 2000.

ירושלים : כרמל, 2000.

Physical Details: 170 p. ; 21 cm.

Location: A. I. Weinberg Book Agency

Other Title: Verso t.p.: Thin, White Dress.

Other Entries: Israel--Jerusalem.

ISBN: 9654073153

Record ID: XWEI2001088191-B

Example of an A.I. Weinberg Book Agency Record

office at LC's Network Development and MARC Standards Office, the records pass through MARC-maker to restructure the data in MARC21 format. Then the records are transmitted to RLG and become accessible to users of the RLG Union Catalog worldwide. Throughout the process the Library of Congress, in consultation with other RLG member libraries, advises A.I. Weinberg of ways to enhance the usefulness of the romanized fields.

native title in English—usually from the title page verso (246 field)—appear when appropriate. A Weinberg record can be identified by the RLG library identifier "XWEI", or the NUC symbol "IsJeAIW". In Eureka®, an item is identifiable by the presence of "A.I. Weinberg Book Agency" in the Location field.

This project was initiated by Linda Lerman, then RLG's member services officer for Jewish studies. It was made possible through the hard work of the Jerusalem bookseller Arodi Weinberg, Gail Shirazi of LC's African and Middle Eastern Acquisitions Section, Randall Barry of LC, RLG, and RLG member libraries.

Library staff and users everywhere welcome dual-script vendor records in the RLG Union Catalog—a remarkable example of technical achievement and international cooperation. ■

A truly cooperative process makes these records available within the RLG Union Catalog. At A.I. Weinberg, a cataloger creates the records and native speakers of Hebrew enter the vernacular fields before they are sent to LC. In Randall Barry's

The romanization of the author's name (100 field) is from the verso of the title page and every record contains the country and city of publication in English (752 field). The ISBN (020 field), and the alter-

Pierpont Morgan Library Adds Literary and Historical Manuscripts Records to the RLG Union Catalog

Elizabeth O'Keefe, Pierpont Morgan Library,
eokeefe@morganlibrary.org

About 70% of the Pierpont Morgan Library's collection of literary and historical manuscripts is now represented in the RLG Union Catalog. In June 2001, RLG loaded 23,317 records describing collection items from the 12th to the 20th century, including literary, historical, scientific, and biographical works, correspondence, documents of all types, and some ephemera and graphic material. About 10,000 more records documenting recent acquisitions and several large collections will be added to the RLG Union Catalog soon. In addition, Pierpont Morgan staff will be preparing finding aids for two recently acquired collections, the Archives of *The Paris Review* and of the Pierre Matisse Gallery.

Author: Wilde, Oscar, 1854-1900.

Title: The picture of Dorian Gray : [n.p.] : autograph manuscript signed. [1890].

Physical Details: 1 item (264 [i.e., 262] leaves) ; 34 cm.

Brown morocco.

Location: Pierpont Morgan Library

Pierpont Morgan Library Dept. of Literary and Historical Manuscripts.

LHMS

Call No.: Loc Unassigned

Notes: Original version in thirteen chapters as published in Lippincott's Monthly Magazine for July, 1890.

The William Andrews Clark Library, UCLA, holds the typescript copied from this manuscript, which was then revised by Wilde before being sent to the printer.

From the Louis I. Haber (sale, 8 Dec., 1909, lot 1776) and Eugene Meyer, Jr. collections. Said to have been bought by J. Sabin for \$1000 at the Haber sale; acquired from Meyer, 30 June, 1913.

Other Authors: Haber, Louis I., former owner.

Meyer, Eugene, 1875-1959, former owner.

Other Entries: England.

MA 883

Record ID: NYPR120386-A

Example of a Pierpont Morgan Library Record for a Manuscript

(Pierpont Morgan...continued from page 7)

Literary and historical manuscripts in the Morgan

The library's collection of literary and historical manuscripts is one of the finest in this country. It includes complete manuscripts and working drafts of poetry and prose as well as correspondence, journals, and other documents of important British, American, and European authors, artists, scientists, and historical figures from the 15th to the 20th century. Highlights include the sole surviving manuscript of John Milton's *Paradise Lost*, John Locke's *Essay Concerning Human Understanding*, Alexander Pope's *Essay on Man*, Charlotte Brontë's *Professor*, Lord Byron's *Don Juan*, Wilkie Collins's *Moonstone* and *The Woman in White*, Charles Dickens's *Christmas Carol*, Mark Twain's *Pudd'nhead Wilson*, and Antoine de Saint-Exupéry's *Le Petit Prince*. There are also fine groups of correspondence of Jane Austen, Thomas Jefferson, Voltaire, and John Ruskin, and journals of Henry David Thoreau, Nathaniel Hawthorne and his family, and John Steinbeck.

The collection also includes many lesser-known treasures, such as the manuscript of Anita Loos's *Gentlemen Prefer Blondes*, a childhood notebook of Sylvia Plath, a collection related to the film *Viva Zapata!*, and Marianne Moore's correspondence with the Ford Motor Company concerning the naming of the Edsel. Historical documents range from 12th-century papal bulls to a manuscript record of colonial Peru. Artists' letters range from a 1454 letter from Leon Battista Alberti concerning his Tempio Malatestiano to receipts illustrated by Picasso in the 1920s. Items of scientific interest include letters of Galileo and the manuscript of an article by Einstein on the theory of relativity.

Records for Morgan Manuscripts in the RLG Union Catalog

About 90% of the records for Morgan manuscripts consist of item-level descriptions. All records give the author, title or description of item, recipient of the letter or signer of the document,

and place and date of writing (if known). Many records also include the names of former owners and donors, summaries of content, biographical notes, citations of published sources, descriptions of accompanying material, and information on bindings or cases. In addition, the names of subcollections or extra-illustrated works are traced as related titles.

You can retrieve these records with searches by author, title, place of writing, date of writing, associated name (recipients, signers of documents, and former owners and donors, if traced), accession number, and name of subcollection, if the item belongs to one. To retrieve the Pierpont Morgan's holdings using Eureka®, RLG's public search interface, limit your results by **RLIN File AMC** and by **Location Code NYPR**. (Limiting by **Material Type Archival or Mixed Collections** is not recommended, as only a small percentage of the library's holdings is coded as under archival control.)

A few sample searches that illustrate the Morgan's holdings (using Eureka's Advanced Search):

Author Word *byron* and **Author Word** *italy*

This search will retrieve material written in Italy by or to Byron. (The **Author** index also indexes the 752 field, where place of writing is traced.)

Author Word *bronte*

Choose **All**, then limit by **Other**; choose **Note** and enter the **value** *microscopic*.

This search will find Brontë manuscripts written in a microscopic hand.

Related Title *mrs fields album*

This will find items in Mrs. Fields's album, which contains poems and sentiments written at Mrs. Fields's request by the literary lions and lionesses of her day.

The data conversion project

The library creates and updates records for its collection using RLIN. These records for literary and historical manuscripts represent the library's first batch

contribution to the RLG Union Catalog. They are based on a card file maintained for about 75 years by the Department of Literary and Historical Manuscripts. Many of the cards date back to the era when library cards were handwritten; more recently, cards were typed or generated from word-processing files. Before the retrospective conversion project was launched, researchers were obliged to visit the library to search the catalog.

In 1997, The Andrew W. Mellon Foundation awarded \$850,000 to the Morgan Library to implement a major project to make the collections more accessible to the scholarly community. Funding from the Mellon Foundation, along with grants from the Arcana Foundation, Booth Ferris Foundation, the E. Rhodes & Leona B. Carpenter Foundation, the Mary Flagler Cary Charitable Trust, The Arthur Vining Davis Foundations, The Florence Gould Foundation, The Alice Tully Foundation, and others, enabled the library to acquire an integrated online library system and initiate a series of data conversion projects to ensure that all its collections were represented by MARC records. Many of these records were entered using RLIN. The RLG Union Catalog now contains records for more than 60% of the library's printed book collection, for all its printed music, and for the library's collection of medieval and Renaissance manuscripts. Data creation or conversion for other collections, such as literary and historical manuscripts, music manuscripts and musicians' letters, drawings and prints, portions of the printed book collection, and ancient Near Eastern cylinder seals, was outsourced, or done directly into the library's database.

The task of converting the literary and historical manuscripts card file into MARC records was entrusted to Electronic Scriptorium, a data conversion firm that employs cloistered orders. The nuns were delighted to work with material of such humanistic value (an additional inducement was the presence in the collection of numerous manuscripts by popes, ecclesiastics, and even a few saints). Their multilin-

gual skills and broad general knowledge proved extremely useful in resolving ambiguities on the cards or supplying additional clarifying information, such as place names.

Future dataloads to the RLG Union Catalog will include records for the rest of the literary and historical manuscripts, as well as records for music manuscripts, musicians' letters, and rare books that were not cata-

loged using RLIN. In addition, the library's OPAC, CORSAIR, will become accessible on the library's Web site (www.morganlibrary.org) by the end of 2001. ■

Touring the Information Landscape: RLG Backs METS

Merrilee Proffitt, RLG, merrilee.proffitt@notes.rlg.org

Take me out to the ballpark?

What could be timelier than an article about baseball in October? Alas, while many of us would love to be virtually whisked away to Shea Stadium, this is not an article about New York's other baseball team. While the Mets enthrall many baseball fans, METS (Metadata Encoding and Transmission Standard) will capture the attention of data owners and users alike.

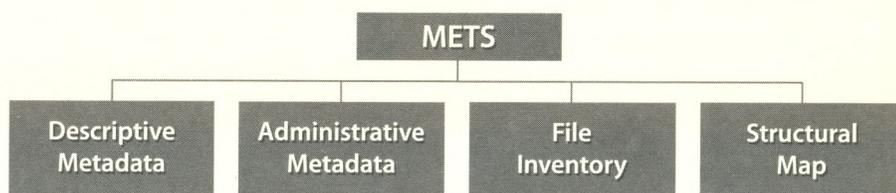
What is METS and how can it help?

METS was developed to encode the descriptive, administrative, and structural metadata for objects within a digital library. METS provides for the responsible management and transfer of digital library objects by bundling and storing appropriate metadata along with the digital objects. METS is expressed using XML, which means that METS data is stored as ASCII text, allowing for platform and software independence. METS has been conceptualized as an application of the Open Archival Information System (OAIS) reference model and as such can function as a Submission Information Package (SIP) for use as a transfer syntax; a Dissemination Information Package (DIP) for display or other applications; or an Archival Information Package (AIP) for storing information internally.

METS had its beginnings in 1996 as the brainchild of the Digital Library Federation (DLF) Architecture Group, which identified metadata and architecture problems as an area of critical need for digital libraries. Thus was born the Making of America 2 (MOA2) project, sponsored by the DLF in the early stages and funded by the National

Endowment for the Humanities. New York Public Library and the libraries of Cornell, Penn State, and Stanford collaborated under the leadership of the University of California, Berkeley Library, contributing images and data towards an investigation of structural and administrative metadata for digital objects. As more and more institutions created digital images in the course of projects, there was growing concern about sensible storage for the digital objects, defined as digital files plus associated metadata. It was the beginning of a serious discussion, tying together many important aspects of digital library research. The MOA2 Document Type Definition (DTD), which was the direct predecessor of METS, was developed for the MOA2 project to encapsulate what were then seen as the required metadata elements.

The MOA2 project was completed in early 2000, the Council on Library and Information Resources (CLIR) published the group's findings, and the MOA2 DTD was circulated for assessment and discussion. While MOA2 aroused considerable interest within the library community, the MOA2 DTD was too restrictive in some respects and lacked some basic functionality, especially for time-based media such as audio and video. Again, the DLF stepped up to the plate and hosted a meeting in February 2001 for the various parties interested in advancing the MOA2 DTD to the next stage. Following this meeting, METS was born.



A Simple View of METS

METS has a very simple structure with just four major components: descriptive metadata, administrative metadata, file inventory, and structural map. Only the file inventory and structural map are required by the schema.

- The descriptive metadata, used for discovery and identification, is optional. A METS object can contain a Metadata Reference or a Metadata Wrapper. A Metadata Reference is a link to external descriptive metadata. A Metadata Wrapper packages descriptive metadata associated with the object, as either Base64 encoded binary data or XML. METS does not require a particular scheme for description, so the implementer can choose the most appropriate descriptive scheme.
- The administrative metadata, also optional, has four optional subcomponents for technical metadata, rights metadata, source metadata, and preservation metadata. Each of these subsections acts like the descriptive section in that the

(continued on page 10)

(Touring...continued from page 9)

metadata can be encoded ("wrapped") within the METS document or pointed to in an external location ("referenced").

- The file inventory allows for listing all the files associated with a digital object. Files can be grouped; some groupings might include master files, thumbnails, etc. The files may be pointed to or can be contained internally as Base64 encoded binary data.
- The structural map forms a simple or complex tree structure that describes the digital object. The map outlines a hierarchical structure linking the content files and metadata to the digital object.

Current METS implementations

Though only released in beta, METS has already attracted a great deal of attention:

- National Gallery of the Spoken Word, a large-scale audio archive at Michigan State University, plans to use METS to store all of their data. The project's principal investigator Professor Mark Kornbluh says "METS appealed to us as a comprehensive XML metadata schema that can contain all of our metadata needs.... The fact that the [Library of Congress] audio-visual librarians

are also working with METS adds to its attractiveness for our project."

- MacKenzie Smith, Harvard University, reports that Harvard is "using MOA2 to encode structural metadata for reformatted print material, including simple items like letters and articles, and complex items like multi-year report and journal runs." The MOA2 objects may be converted to METS later. METS is also under consideration as "the underlying metadata model for a new e-journal archive under development."
- The METAe project, which is developing an automated metadata extraction system for books and journals, plans to use METS to encode their system output (descriptive, administrative, and structural metadata). Alexander Egger states, "The Metadata Engine will use METS as its output format, thus for distribution. For demonstration purposes the project will also develop a ...Web Application, which will store METS documents and make them accessible over the Internet. The related Austrian literature online project will use METS for archiving electronic books."
- "The current MOA2 DTD is the cornerstone of UCB's digital library projects," writes Rick Beaubien from

UC Berkeley. "Some of UCB's digital library initiatives...are strictly local. For these, the library produces digital objects encoded according to the MOA2 DTD, which are stored in an online repository, and served up by a servlet-based, MOA2-aware viewer." The library also participates in digital library initiatives administered by the California Digital Library. The CDL has also adopted MOA2 as an encoding standard; for CDL projects, UC Berkeley submits its contributions in MOA2. "Once METS is out of beta, it is expected to supersede MOA2 both for UCB and CDL.... Objects currently encoded according to the MOA2 DTD will be converted to METS format; and all new initiatives will produce METS objects. UCB will adapt the viewer that currently presents MOA2 objects to handle METS compliant encodings."

- Carl Fleischhauer of the Library of Congress says the Audio-Visual Prototyping Project "is developing approaches for the digital reformatting and preservation of historical sound recordings and moving image collections. The project is partly motivated by the decreasing viability of analog copying (the accepted reformatting approach) and the establishment of the National Audio-Visual Conservation Center, scheduled to open in Culpeper, Virginia, in 2004. The Prototyping Project plans to use the emerging Metadata Encoding and Transmission Standard (METS) together with some extension schemas."

RLG's role in the future of METS

The standard is currently maintained in the Network Development and MARC Standards Office of the Library of Congress. Having played a key role in moving this initiative forward and serving as the work coordinator, the DLF has helped to bring the METS work to the forefront. Now RLG will take over as the new coordinator. This is a natural step for a number of reasons. The METS standard will be applicable to RLG's member community of libraries, museums, archives, and historical societies. METS fits in

More information on the Web:

METS: www.loc.gov/standards/mets/

Digital Library Federation: www.diglib.org

MOA2: sunsite.berkeley.edu/MOA2/

National Gallery of the Spoken Word: www.ngsw.org

METAe project: meta-e.uibk.ac.at

The Library of Congress Audio-Visual Prototyping Project: lcweb.loc.gov/rr/mopic/avprot/metsmenu.html

OAIS: ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html

OCLC/RLG Preservation Metadata Working Group

White paper: www.oclc.org/digitalpreservation/presmeta_wp.pdf

New York Mets: www.mets.com

nicely with much of RLG's ongoing work in digital preservation. The joint OCLC and RLG work on digital preservation metadata has identified the OAIS reference model as a potential starting point for developing the preservation metadata framework. RLG has actively supported the NISO standards work, such as Technical Metadata for Digital Still Images, which RLG cosponsors. The work of this group will fit neatly into the administrative section of METS. Finally, RLG has always advocated community standards such as EAD and Z39.50, and METS is viewed as an emerging standard. For the next six to eight months, RLG will continue the process of education, information dissemination, and gathering of feedback on METS. Then RLG, like the DLF, will review METS based on use. For more information, contact Merrilee Proffitt. ■

(Mission Accomplished...continued from page 1)

all that library's records were converted, with the number of records affected, so each could order snapshot files of its Pinyin records and load them into its local system. RLG members received a discount rate on these "Pinyin snapshots." (If you need a copy of the notification sent to your library, please contact the RLG Information Center.)

Libraries now face a cleanup task to manually review the Pinyin records that included one or more fields the conversion program flagged as ambiguous. The RLG Pinyin Working Group, comprising representatives from Columbia, the Library of Congress, Princeton, University of Michigan, and Yale, collaborated on "Pinyin Manual Review: Sharing Experiences" (www.rlg.org/eas/pinyinreview.html) to guide other libraries. We welcome your experiences to add to this page.

The RLG Pinyin Working Group continues to hold conference calls to address any outstanding questions. If you have a question that is not addressed in "Frequently Asked Questions for RLIN Users" (www.rlg.org/eas/pinyinfaq.html), please feel free to

submit it. To understand what was converted and what was not, check the Library of Congress's Pinyin Conversion Project Web site at lcweb.loc.gov/catdir/pinyin/pinyin.html.

With the project's completion, you can now use just Pinyin romanization when searching the RLG Union Catalog to retrieve Chinese-language records. For example, the following searches will retrieve records originally created with Wade-Giles romanization:

author mao, zedong
title yi jing
title word jingji fazhan
subject chinese poetry—qing dynasty

This project benefited from close collaboration among Library of Congress, OCLC, and RLG staff. RLG is also grateful to staff from the RLG Pinyin Working Group and additional RLG-member Pinyin testers at the Cleveland Museum of Art, University of Toronto, and University of Washington Gallagher Law Library who joined LC in testing the RLG conversion program on samples of their own records. These testers helped improve the LC specification for everyone. ■

Some Upcoming Events

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

October 1-5, 2001. Ljubljana, Slovenia: 17th Interlending and Document Supply International Conference.

October 5, 9 a.m.-12 noon. RLG's Dennis Massie speaking: "Consortial Resource Sharing: a New Marketplace."

October 24-28, 2001. New Orleans LA: Medical Library Association Triple Chapter Annual Meeting. RLG's John Eilts exhibiting Ariel® and ILL Manager and conducting ILL Manager workshop.

November 8-9, 2001. Ann Arbor, MI: Shaping Interlibrary Loan/Document Delivery in the 21st Century. RLG's Linda West speaking.

November 15-18, 2001. San Diego, CA: Theater Library Association/American Society for Theatre Research Annual Conference.

November 16, 1 p.m. Panel: "The Impact of New Technologies on the Study of the Performing Arts." RLG's Anne Van Camp presenting RLG Cultural Materials.

November 16-20, 2001. San Francisco, CA: Middle East Librarians Association (MELA) and Middle East Studies Association (MESA) Annual Meetings.

Contents

RLG Cultural Materials: The Potential to Change Research and Scholarship	1
Mission Accomplished: RLG Union Catalog Converted from Wade-Giles to Pinyin	1
Building the RLG Cultural Materials Infrastructure	3
MARC Link Expands its Role in Recon Projects	4
Over 1,500 Citations Added to the Chicano Database	6
Weinberg Contributes Hebrew Records into the RLG Union Catalog	6
Pierpont Morgan Library Adds Literary and Historical Manuscripts Records to the RLG Union Catalog	7
Touring the Information Landscape: RLG Backs METS	9

1200 Villa St., Mountain View, CA 94041-1100 USA



Non-Profit Organization
U.S. Postage
Paid
Permit No. 19
Mountain View, California