

Opening Yale's Treasure House

Yale University Library Annual Report to Arcadia
Year 3

SEPTEMBER 2012

CONTENTS

4	I. CATALOGING EFFORTS
	A. <i>Cataloging Non-Spanish Iberian Languages</i>
	B. <i>Cataloging African Languages</i>
	C. <i>Cataloging Cambodian Newspapers/Khmer Language</i>
	D. <i>Cataloging Vietnamese Language Collection</i>
	E. <i>South Asia Collection</i>
	F. <i>OCLC Reclamation Project</i>
16	II. DIGITIZATION PROJECTS
	A. <i>New Initiative in Year 3</i>
	1. South Africa Now
	B. <i>Continued Year 2 Projects</i>
	1. Preserving Cuban Exile Broadcast Tapes
	2. Making Searchable Islamic Fundamentalist Tapes
	3. Digitizing Joel Sumner Smith's Collection of Slavic Literature
21	III. YEAR 4 PROJECTS
	A. <i>Continuing Projects</i>
	1. Digitizing Joel Sumner Smith's Collection of Slavic Literature
	B. <i>New Projects</i>
	1. Digitization of Hán Nôm Handwritten and Woodblock Manuscripts Held in the Maurice Durand Collection
	2. Enriching, Enhancing, and Connecting Yale's Digital Collections: Yale Persian Collection
	3. Arabic and Persian Medicine: Middle Eastern Texts in a Western Collection
	4. Cataloging Indigenous African Language Pamphlets
	5. Processing the North African Jewish Collection
24	IV. BUDGET SUMMARY
25	APPENDICES

Opening Yale's Treasure House

THE ARCADIA GRANT first came to Yale University Library through interactions led by Dr. Alice Prochaska, Yale's 15th University Librarian. The Arcadia projects began in September 2009, and this report summarizes the activities of the grant's third year, September 2011 through August 2012.

I began my tenure as University Librarian in July 2011, and therefore have had the pleasure of working on the Arcadia Year 3 projects from start to finish. I have also witnessed the powerful impact that the grant has had on the access and discovery of our collections. Through this generous grant, Yale University Library (YUL) has advanced considerably the efforts to make its unique collections more accessible to a wider audience through the cataloging of "hidden collections" and the digitization of many at-risk materials. YUL, in turn, has shared broadly the lessons learned, processes designed or enhanced, and the expertise developed with the cultural heritage community at large.

What follows is a summary of the third year of activities sponsored by Arcadia funding at YUL. Arcadia projects focus on: 1) cataloging projects aimed to increase access to Yale's holdings; and 2) digitization initiatives to make unique or fragile materials available for scholarship worldwide. We also note the projects planned for Year 4, should the Arcadia Fund decide to continue its support of activities at YUL.

YUL staff members contributing to this report include Kathleen Bauer, Carolyn Caizzi, Sarah Calhoun, Kevin Glick, Robert Klingenberg, Graziano Krätli, Ellen Jaramillo, Rich Richie, Charles Riley, DaJin Sun, Joan Swanekamp, Roberta Pilette, Dorothy Woodson, Debbie McGraw, Chris Edwards, Elizabeth Beaudin, and Amanda Patrick.

Respectfully submitted by Susan Gibbons

I. CATALOGING EFFORTS

A. *Cataloging Non-Spanish Iberian Languages* Europe's Iberian Peninsula – an area consisting primarily of Spain, Portugal, Andorra, and adjacent islands – is home to a variety of languages beyond Spanish and Portuguese which are insufficiently addressed by libraries within the United States, and indeed, worldwide. Yale, however, has a long tradition of collecting library materials in these languages, and consequently had a sizeable collection of uncataloged items because of language difficulties. The majority of these uncataloged materials were written in Basque, Catalán, Galician (Gallegan), Bable (Asturian), Aragonese, Occitan, and followed by a number of dialects such as Navarrese, Montañés or Cantabrian, Leonese, Extremeño, and Andalusian.

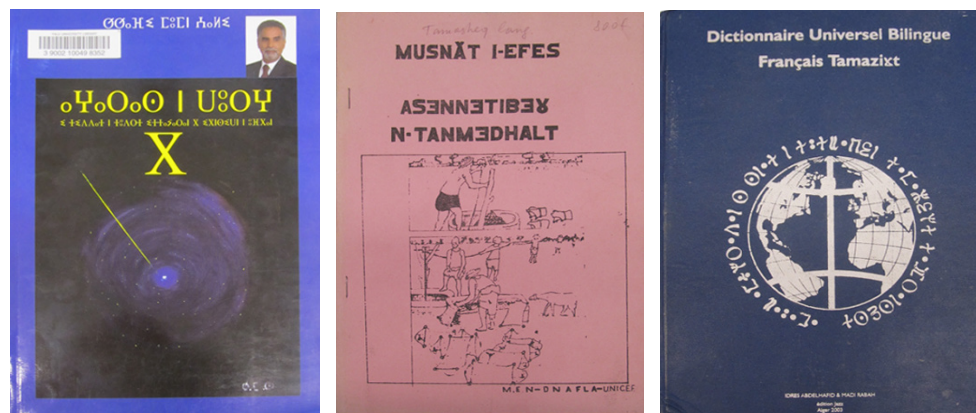
These languages were selected for attention because Spain ranks among the top publishing countries in the world with Barcelona, center of Catalan culture, an active and growing center of literary publications. In 1983, Spain legislated the use of its regional languages as the primary language of education within those regions. With the end of the Franco dictatorship (1939–1975) and the end of Spanish censorship in all languages during the last century, the government elevated the regional languages to co-official status in tandem with Spanish, thus promoting a continuous increase in the number of people able to speak, read, and write their regional tongues. Additionally, the Spanish government provides subsidies to publishers that produce titles in the regional languages, so that production of materials in those minority languages continues to grow annually.

The project goals included producing full-level catalog entries and creating NACO (Library of Congress' Name Authority Cooperative Program) authorities to support the access points in the bibliographic records created. Along with normalizing the language codes for the Iberian dialects, each entry received a language note (variable field 546 in the metadata record) in all of the bibliographic records cataloged during this project. The catalog librarian also qualified the country of publication in the imprint field and supplied the locations in Spanish, when the place names in the vernacular are not well known. For example, the city name *Irun*, sometimes also called *Iruña*, appearing as the place of publication on an item would be transcribed as: 260 \a Iruña [i.e. Pamplona, Spain] since Pamplona is the official and more widely known form of the city's name. This cataloging treatment will facilitate patron searching, and will be helpful for librarians who may want to update the records in the future.

Cataloging the Basque language materials presented several challenges, not the least of which was identifying and finding the materials within YUL's uncataloged collections. Moreover, Basque grammatical structure adds difficulties since word order in a sentence can indicate different meanings, pronouns are often dropped, and modern literary grammar differs from the non-literary. As stated in the Arcadia Year 2 final report, the only other significant Basque collection in the United States is held at the University of Nevada in Reno. Because of this cataloging effort supported with Arcadia funding, we can confirm that the vast majority of the titles owned by Yale are not held in the Basque Library in Reno.

In summary, more than 92% of the items needed completely original cataloging. Approximately 40% of items cataloged fall into 'belle lettres' or fiction titles. The remaining 60% cover a variety of subject areas (internet information security, dubbing for films and television, ethnic assimilation, folklore, architecture, linguistics, politics and economics, mountain climbing, social change, education reform, plastic arts, theater, handmade wooden toys, and ancient stone crucifixes used as road markers). By the project's end, the cataloging librarian had completed 3,519 pieces, surpassing the projected number by 27.25%. All of these materials are now discoverable by scholars worldwide through Yale University Library's online catalog (ORBIS) and the OCLC database (WorldCat). As expected, once the collection was cataloged researchers began requesting access. From the start of 2010 to the 2012 spring academic term, there were 528 circulations of items from the non-Spanish Iberian languages collection (starting from a baseline of zero circulations because the collection was undiscoverable). Just over 57% of those circulations were to individuals outside of the Yale community.

B. Cataloging African Languages



A representative sample of the kind of African language materials that were cataloged in year 3. On the left and far right are books in the Tamazight language using Tifinagh script, for which a Romanization table was created. In the middle is a booklet in the Tamashek language using extended Latin orthography. Other materials in Tamashek make use of Tifinagh script, but in a way that differs from Tamazight usage.

The Africana collection at YUL dates to the 1910s and 1920s when Howell Wright, a Yale graduate, began to collect southern Africana materials. He was both interested in the exploits of Cecil Rhodes and befriended many others who were more directly associated with Rhodes. His collection contained numerous items in indigenous southern African languages, and in particular grammars and dictionaries, which are located throughout the stacks of YUL.

Starting in 2007 with assistance first from a Rausing Grant and then with Arcadia funding, the Catalog Librarian for African Languages, Charles Riley, has increased and enhanced the cataloging for African language materials in the YUL collection. During the current grant year, approximately 1,000 titles were cataloged, including materials in several languages which were new to the project: Bamun, Bangolan, Benga, Boma, Bubi, Cross-River Mbembe, Dime, Epie, Fang, Guosa, Hunde, Isekiri, Kaansa, Kabola, Kalanga, Kpan, Kuwaataay, Lukpa, Mandingo, Mandjak, Manyanka, Meroitic, Nande, Nilamba, Njebi, Nyamwanga, Nyature, Saafi-Saafi, Safaliba, Saho, Shambala, Vengo, Xaasongaxango, and Yao. New countries of publication for this period include Canada and Pakistan. For roughly 50 of the languages, this project marks the first time that a text in that language has been professionally cataloged by any library, anywhere in the world.

One of the challenges associated with these materials has been the lack of support for the non-Roman alphabet characters in OCLC and other bibliographic utilities. Consequently, OCLC's decision to "add additional scripts during fiscal year 2013, including support for some African scripts" is important progress. During this past year, a decision was made at Yale to begin cataloging directly on the OCLC-supplied platform, *Connexion*. As a result, cataloging procedures and workflows that have been developed over the course of this project locally can be formalized and shared more broadly between institutions.

Language identification of materials in this project has become more efficient over time as language familiarity increased. However, some items have continued to require a fair amount of research. This was the case especially for items in Punu, Nupe, and Sebat Bet Gurage. In most other cases, there were at least a few clues regarding the language of publication, but these cases required collaborating with a linguist of Gabonese languages in Lyon, a missionary in Nigeria, and a native speaker familiar with other Gurage languages, to narrow down and confirm the language identification for the work in hand. Other items required the assistance of the author or publisher to provide more detailed subject analysis, or confirmation as to the specific genre of a work. This happened in the case of a Somali book on stand-up comedy and with what we believe is one of a few copies of the first Amharic science fiction novel to appear in print.

Looking back over the course of the work on Africana supported at YUL by the Raising Trust and Arcadia Fund, one can easily see the results. Before this work started, YUL had only a general sense of the quantity of materials in its uncataloged collections. Very little could be said about the diversity of languages in the collections, aside from those in Afrikaans, Amharic, Hausa, Kinyarwanda and Swahili. The true, astonishing scope of YUL's Africana collection is now known. While there are likely to be many records where the language identification is not perfect, in the process of cataloging these extraordinary materials, those involved have come to have a greater appreciation for the knowledge they contain and for what it requires to make broad user access to them easier, faster, and more accurate.

The use of materials of this kind has seen a marked increase in attention in recent years. Since the start of the project in 2007 through August 2012, there have been 902 circulations of the newly cataloged material, starting from a baseline circulation of zero. Approximately 31% of those circulations were to students and scholars outside of the Yale community. In addition, in October 2011, at the African Language and Literary Studies Theme Group lecture series hosted at Yale, Wendy Belcher of Princeton discussed the influence of African thought on Samuel Johnson as he was writing *The History of Rasselas, Prince of Abissinia*. Yale's record of the 1964 Amharic translation of this work is accessible in Ethiopic script, which is of most use to native speakers, but also to students learning the classical language of Ge'ez. Students in classes on language policy in Africa, epidemiology and public health, linguistics, anthropology, art history, and environmental studies are becoming more aware of the availability of materials directly relevant to their work as it relates to preparation for fieldwork. Two examples of recent requests were for materials in Saafi-Saafi, a Senegalese language, and the other in Bakossi, spoken in Cameroon. Before this project, the discovery and location of materials in these languages would have been virtually impossible.

The impact of this African languages cataloging project has been global as well. For example, this year, Mr. Riley assisted in the creation of a Romanization table for transcribing the Tifinagh script in Moroccan languages (appendix A) which has been approved by the ALA and the Library of Congress, and is now posted and available.¹ Its accompanying International Standards Organization (ISO) bibliographic code is expected to be approved shortly. This is the second Romanization table for an African language contributed by Arcadia-funded work at YUL, with the first being for Vai (appendix B).² In addition, Mr. Riley also edited a Romanization table for Cherokee (appendix C), which while a Native American language, has an intriguing relationship

1 <http://www.loc.gov/catdir/cpsol/romanization/tamazight.pdf>

2 <http://www.loc.gov/catdir/cpsol/romanization/vai.pdf>

with Vai.³ Two further African language Romanization tables are in process: N’ko and Tifinagh as used in Tamshek languages. Moreover, Mr. Riley worked closely with the Microsoft Local Language Program to assist in the development of a new font, Ebrima, which contains the specialized characters for N’Ko, Tifinagh, Vai and Osmanya (appendix D). With Romanization tables, fonts, and system-level support for diacritics, scholars of African languages move closer to more direct, intuitive access to full text indigenous language materials and the ability to better source data for interaction with transliteration and translation tools.

*C. Cataloging
Cambodian Newspapers/
Khmer Language*

YUL’s collection of Cambodian newspapers spans 1986 through 1996 and consists of approximately 97,000 pages of newsprint from 65 individual newspaper titles. Because of its historical value and increasing fragility, the Cambodian newspaper collection was selected as an Arcadia project, and a visiting Fellow from Cambodia joined the Arcadia team in March 2011 to assist. His efforts focused on the identification, organization, and description of the primarily Khmer language titles. By using online library catalogs and WorldCat to collect data, the Fellow created an online guide (LibGuide) to describe YUL’s United Nations Transitional Authority Period (UNTAC) Cambodian newspaper collection and to provide a listing of Yale’s holdings along with those of other research libraries in the United States (appendix E).⁴ Examples of some of Cambodian newspaper profile and holdings summaries created through this project are included as appendix F.

At the 2012 spring meeting of the Association for Asian Studies Southeast Asia Microforms Project,⁵ the Yale Southeast Asia curator presented the Cambodian Newspaper Project Guide to the other participants as a first step towards consolidating North American library holdings of these newspapers for a reformatting effort through digitization. SEAM members expressed a strong interest in funding such a project with YUL’s Southeast Asia Collection as the coordinating body.

By the completion of the project in March 2012, the visiting Fellow had cataloged 112 individual monograph titles, created brief (searchable by author or title) records for an additional 377 titles, 19 DVDs and CDs in the Khmer language, and wrote minimum level catalog records for all 81 of the Cambodian newspaper titles, including full holdings information for each newspaper. The newspapers were placed in acid free boxes for interim storage in the Library Storage Facility under optimum climatic conditions

3 Konrad Tuchscherer and P. E. H. Hair, “Cherokee and West Africa: Examining the Origins of the Vai Script” *History in Africa*, Vol. 29, (2002), pp. 427–486. <http://www.loc.gov/catdir/cpso/romanization/cherokee.pdf>

4 <http://guides.library.yale.edu/southeastasiakh>

5 <http://www.crl.edu/area-studies/seam>

for future digital collaborative efforts. Along with his cataloging efforts, the visiting Fellow was a key contributor to the newly approved revisions to the Library of Congress Khmer language Romanization tables (appendix G).⁶

Because of Arcadia funding, Khmer newspaper researchers at Yale and around the world may locate and identify these newspapers in ORBIS, WorldCat, and web search engines. With an expected growth in interest in these newspapers, the collection's curator hopes to collaborate with other libraries on a future proposal for the digitization of these newspapers.

*D. Cataloging Vietnamese
Language Collection*

Identified as an important resource which was largely inaccessible due to a lack of cataloging, YUL's Vietnamese language collection was selected as an Arcadia grant recipient starting in Arcadia Year 2. Through this funding, a native Vietnamese cataloging/acquisitions assistant focused on the cataloging of the many unique and rare Vietnamese holdings in YUL's collection. During its continuation to Arcadia Year 3, close to 1,700 fully cataloged monograph titles have been added to ORBIS and sent to WorldCat.

In June 2012, the South and Southeast Asia Collections curators hosted the first of several collaborative collection development meetings with other Borrow Direct South and Southeast Asian studies librarians. During this workshop, the curator of Yale's Southeast Asia Collection began initial discussions with Cornell University's Southeast Asian studies librarians to reduce duplication of collecting efforts for the same titles from Vietnam, and to collaborate on collecting for a broader Vietnamese studies collection. Such cooperative collection development efforts would not have been possible without first cataloging the existing collections in order to make cross-institution collection assessments. Yale can now fully participate in an inter-institutional Vietnamese language cooperative collection development program and can help to build a broader collection of Vietnamese materials across the U.S.

⁶ <http://www.loc.gov/catdir/cpsd/romanization/khmer.pdf>



December 2011: Geshe Lobsang Tenzin and six Tibetan monks from Gaden Jangtse Tsawa in South India visited Yale and read through parts of the Tibetan Buddhist manuscript collection housed in the Beinecke. As part of this, they also taught a group of Yale students about its contents and importance. The event came about through a collaboration between Sarah Calhoun, Bruce Blair (Yale Buddhist Chaplain), George Joseph (Office of International Affairs), and the Beinecke staff.

With the support of Arcadia funding, YUL hired an Assistant Curator for South Asia for a one-year term (July 2011–June 2012) to develop a plan to address the rapid growth of South Asian studies and the resulting addition of twelve new faculty members since 2008. The curator’s impact at Yale was so profound that YUL and the Yale South Asian Studies Council formed a partnership to extend the funding for the curator position indefinitely when the Arcadia phase ended in June 2012. The primary goals for the curator during the Arcadia project included: 1) evaluating the historical collection, 2) identifying preservation and reformatting needs, and 3) ensuring that the collection is available and identifiable through ORBIS.

As mentioned in the Year 3 mid-year report, the curator has added LibGuides (online subject guides) to the YUL web site to increase the visibility of the collection and to provide a helpful resource for research. The release of these new LibGuides coincided with an initiative to track usage for later assessment review of resources made available with Arcadia funding, the results of which are in the table below. The work on preparing a LibGuide for statistics and census data uncovered a need on campus for such information from South Asia, and India in particular. The South Asia curator, funded by Yale, will continue collaborations with colleagues at the Yale Center for Science and Social Science Information and the YUL Maps Department to compile detailed information about both new digital resources and older data sources from South Asia that were originally purchased by the defunct Economic Growth Center.

Name of LibGuide	Launched	Page hits	Top pages of interest	%of total
South Asia: Forestry and Environmental Studies	Sept 2011	2,380	Anthropology/Forestry class, "Environment, History and Society in India"	85%
South Asia: Himalayan Studies	Dec 2011	64,885	Collection Highlights Sanskrit/Pali/Prakrit Tibetan Nepali/Newari	13% 15% 10% 5%
South Asian Stats and Census Data	June 2012	76		

A new web page for the South Asia collection was launched in May 2012.⁷ While the YUL's Information Technology (IT) staff is renovating the entire YUL web site, this simplified web page serves as a placeholder until the website renovation is complete in 2013 while providing basic collection information and helpful links to other resources. At the same time, the curator created a Facebook page for the South Asia Collection to push information out to patrons about hidden collections.⁸ Additionally, during Arcadia Year 3, the South Asia curator has assisted with the cataloging of South Asian material throughout the library system and has supported the Beinecke Rare Book and Manuscript Library as a consultant to assess their "Indic" collection and define its scope and language composition. This small but wide-ranging collection contains items such as palm leaf manuscripts in Pali/Sinhala and the personal papers of a prominent Sanskrit scholar, Dwight Whitney, from the mid-19th century. The curator has also evaluated the South Asia portions of the old Yale Classification system material, specifically the Urdu and Persian sections, covering the condition of the books and their level of cataloging. For this particular effort, the curator has collaborated with other YUL departments about best practices. Since there are faculty at Yale who work on less commonly taught South Asian languages (Oriya, Pali, and Prakrit), the curator has also added titles to the collection to support this need.

Ongoing collaborations include work with other South Asia librarians to enhance the national South Asia collection through collaborative efforts. This conversation started at a national level with CONSALD and has been continued by a Borrow Direct (Ivy League) South and Southeast Asian cooperative collection development initiative co-organized by the Southeast Asia and the South Asia curators.⁹

⁷ <http://www.library.yale.edu/southasia>

⁸ <http://www.facebook.com/YULSouthAsia>

⁹ For more on CONSALD – the North American organization of South Asia librarians – see: <http://consald.org/newsite/node/4>

*F. OCLC
Reclamation Project*

The OCLC Reclamation project has been a singular endeavor of YUL to upload its holdings in OCLC's global WorldCat database so that users worldwide can discover Yale's rich collections widely and easily through the WorldCat and Google search engines. The project reflects a close and extensive collaboration among various departments of the library and has gone through several phases of development, including extracting records from ORBIS, adding holdings in OCLC, loading OCLC Master Records and Institutional Record control numbers back into ORBIS, and post-reclamation record cleanup.

Applying criteria established by the OCLC Reclamation Workgroup, and based on a careful and thorough analysis of the catalog, YUL IT staff extracted approximately 6,000,000 records, batched them by collections, and sent them to OCLC for processing in the winter of 2010. While OCLC worked on these datasets, Yale librarians developed an interim cataloging workflow and placed a moratorium on creating and editing non-Roman catalog records directly in OCLC to ensure the library's records were still coordinated in both databases. In processing Yale's records, OCLC applied different routines to map Yale's records to appropriate Master Records and create related Institution Records at the same time.

By June 2011, OCLC had processed and returned successful records to Yale and had rejected approximately 41,000 records, which became candidates for the cleanup project funded by the Arcadia grant. Newly loaded Master Record and Institutional Record control numbers were reviewed for accuracy and searching. Communication continued between OCLC and Yale to resolve data issues promptly. New workflows were implemented smoothly for general and non-Roman cataloging while non-Roman cataloging directly into OCLC was resumed.

This cleanup became a major part of the reclamation project intended to resolve the wide variety of problems from the unresolved records returned from OCLC as well as other related catalog maintenance issues. Two new staff members were hired for this portion of the project; after thorough training by the Catalog Training and Documentation Librarian, the new staff began to work on OCLC unresolved records under the supervision of the Metadata and Catalog Management Team Leader. To ensure that the new staff received appropriate support, the team leader also set up time to work with each staff member on a weekly basis and developed the workflow and procedures that would advance the cleanup process.

Efficient workflow plus staff dedication led to desirable productivity for the reclamation cleanup project, which has successfully processed a total of 30,435 ORBIS records from the OCLC reports and resolved a total of 169,378 errors associated with the

Home Search Create lists, bibliographies and reviews: [Sign in](#) or [create a free account](#)

WorldCat®

Advanced Search Find a Library

<< Return to Search Results Cite/Export Print E-mail Share Permalink

Add to list Add tags Write a review Rate this item: ☆☆☆☆

Yale daily news

Publisher: New Haven, Conn. : Yale Daily News Pub. Co.
 Edition/Format: eJournal/eMagazine : Document : Periodical : English
[View all editions and formats](#)
 Rating: ☆☆☆☆ (not yet rated) 0 with reviews - Be the first

[Borrow / obtain a copy](#)

You are connected to the **YALE UNIV LIBR** network
[Hide local services for this item](#)
[Check for electronic versions](#) [Connect to the catalog at Yale University Library](#) [OCLC FirstSearch](#)

Orbis YALE UNIVERSITY LIBRARY CATALOG [Log in to your account](#)
[Ask Yale Library](#) | [Help](#) | [Feedback](#)

New Search My Searches My List My Account Morris (Law Library) Borrow Direct / ILL

Quick Search: Search [Search History](#)

Titles 1 of 2 Next

Title: [Yale daily news](#) [electronic resource].
ISSN: 0890-2240 0890-2240
Published/Created: New Haven, Conn. : [Yale Daily News](#) Pub. Co.,
Links: [Full text](#)

Yale Holdings

Location: [Yale Internet Resource](#)
[Information about Off Campus Access](#)

Call Number: None
Status: No information available
Notes: Online resource

Frequency: [Daily](#) (except Saturday and Sunday) during school year

Actions
 · Print
 · Export
 · E-mail
 · Add to My List

This item
Full View
 · Staff (MARC) View

Screen shots from OCLC and Orbis, showing an example of a search for the Yale Daily News that was added to the OCLC holdings.

catalog records. To date, 30 OCLC reports (approximately 900 records) and 30 ORBIS loading error reports (approximately 2,600 records) remain for review by December 2012.

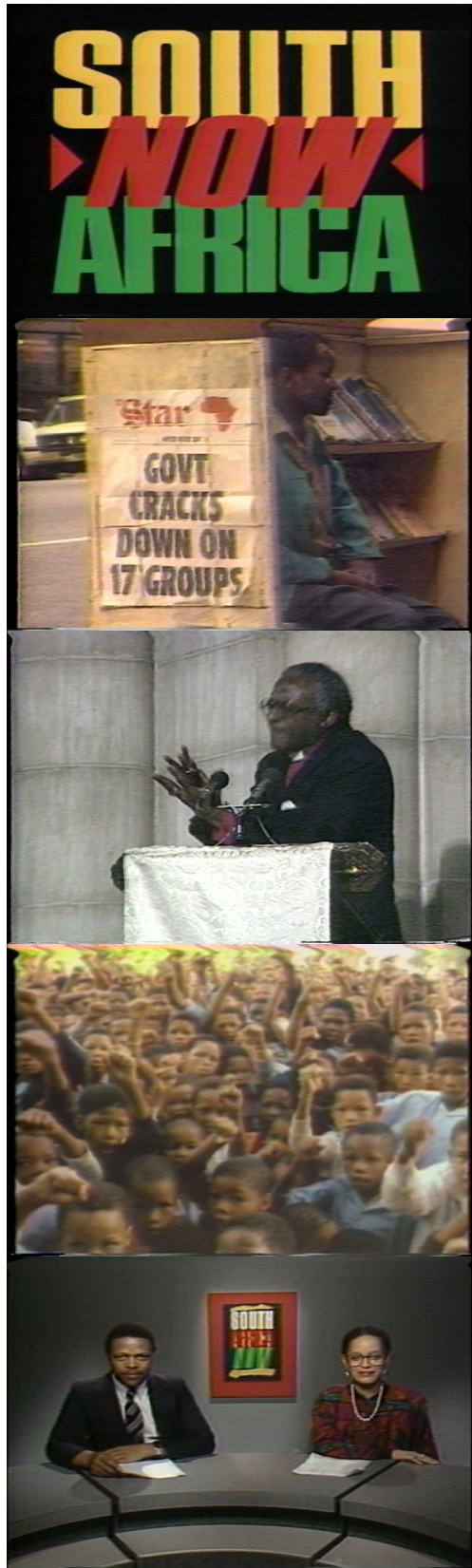
Two major data issues were discovered regarding the accuracy and quality of Yale holdings in WorldCat, thanks to the OCLC reclamation process. First, OCLC and Yale identified “shadow” holdings records in the hundreds of thousands, likely from the earlier merger of the RLIN database (or other record migrations) and the lack of holdings maintenance over the years.¹⁰ Yale officially notified OCLC to delete this dataset of false Yale holdings records from OCLC in August 2012.

Second, during the initial loading of Yale archival records in WorldCat, neither OCLC nor Yale staff was aware that the Yale archival records being processed had been already created in WorldCat through different processes many years ago. As a result, several thousand duplicate Yale archival records were created in WorldCat. With the full support from OCLC, Yale was able to determine the scope of duplication and develop an effective solution to remove these duplicate records from OCLC in April 2012.

Additionally, in view of the distinctive content and nature of Yale archival and special collections, Yale has made a special arrangement with OCLC to allow Yale’s archival records to be added in WorldCat as unique records. This was possible because each title from Yale archival collections is unique. This special processing routine has not only increased the efficiency of batch loading Yale’s archival holdings but also has kept Yale’s unique records up-to-date.

Because of these efforts, there is tremendous improvement in the quality of the YUL online catalog and a much more comprehensive and accurate representation of Yale’s holdings in WorldCat.

¹⁰ The former Research Libraries Information Network was absorbed by OCLC in 2007.



II. DIGITIZATION PROJECTS

A. New Initiative in Year 3 *South Africa Now*

This collection consists of 2,418 videotapes and 49 printed transcripts from the television program *South Africa Now* including a nearly complete run of the approximately 150 broadcast episodes produced by Globalvision, Inc. from 1988 to 1991 as well as file footage. Because of its rare position operating just outside the control of the South African authorities, *South Africa Now* was able to capture a unique view and in doing so documented the cultural richness of the region.

The digitization of the broadcast episodes has been completed as projected. The digitization project was undertaken with both Yale staff and Broadway Video Digital Media in New York. Three different digital video files were produced according to Yale specifications, including:

- preservation master using MXF with 10 bit lossless Motion JPEG 2000 format for video and 16 bit PCM format for audio;
- mezzanine copy using MPEG-4 QuickTime with lossy JPEG video and lossy AAC audio for creation of future access copy files;
- access copy using MPEG-4 QuickTime with 8 bit AVC for video and 128 Kbps AAC audio.

The YUL Preservation Department completed a rigorous set of quality assurance and quality control procedures on every video, which consisted of a review of file parameters (bit depth, file type, etc.) as well as a visual and audio inspection of the digital video files, including picture, sound, intelligibility, completeness, etc. In collaboration with the Preservation team members, Library IT has checked the digital video and metadata files for proper formatting and file-naming conventions and has ensured data integrity had not been compromised during transfer.

This project has provided a proving ground for a workflow and quality control (QC) database-tracking tool for use in all future video digitization projects. The database-tracking tool facilitated the documentation of success and failure of each digitized video based on completeness and appropriateness of metadata as well as picture and sound quality for the preservation master and mezzanine copies. As part of the process, any type of QC problem is recorded for each tape. For an unsuccessful ingestion, the tool records the failure and the reason for it and alerts the Preservation Department.

The remaining categories of QC are still performed even if ingestion has failed. Thus, the process allows digitization vendors to learn of all the problems with a set of files rather than just those revealed during ingestion.

The project of digitizing all the videotapes and the printed transcripts has been completed. The rate of failure showed a steady decline as the vendor put to use the information conveyed in earlier QC reports. Inefficiencies in the ingestion process by Yale's Library IT were also discovered and corrected. For example, success rates for files climbed from 45% in April to 72% by the end of June. The rate of QC also increased from an average of 17 per day to nearly 20 per day during the same 3-month period.

The project has allowed us to preserve the contents of the tapes. Of the videotapes in this collection, the vast majority (96%) are in formats that are becoming rapidly obsolete. Digital migration and preservation is the only approach presently available to safeguard this unique collection. This preservation effort has safeguarded materials documenting an important period in South African history.

B. Continued Year 2 Projects 1. *Preserving Cuban Exile Broadcast Tapes*

YUL holds 602 unique audio tapes of radio programs, *Al cantío de un gallo*, from the 1980s and 1990s, which were broadcasted by Cuban exiles from various locations outside the island back to Cuba. They offer audiences an uncensored report on world developments very different from what was available through conventional Cuban media. The digitized audio files from this completed project can now be accessed via workstations located in YUL's Manuscripts and Archives department. Researchers and scholars worldwide have open access to a 418-page Finding Aid, which provides summaries in Spanish and English of the content of all 602 audiotapes (sample pages in appendix H). With this information, the patron can then listen to particular files in the Manuscripts and Archives reading room or request copies of the digital files for research purposes.¹¹ This restriction of access is because Yale University General Counsel ruled early in 2012 that the Library could not provide streaming audio or video files digitized by the Library if the Library did own copyright to the originals. YUL is currently engaged with University General Counsel to explore the concept of a "digital reading room" which could create controlled access via the Internet to scholars who first register with YUL Manuscripts & Archives department and agree to terms of acceptable use.

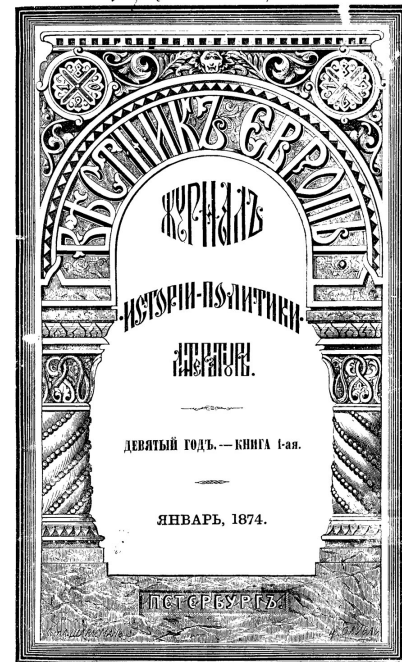
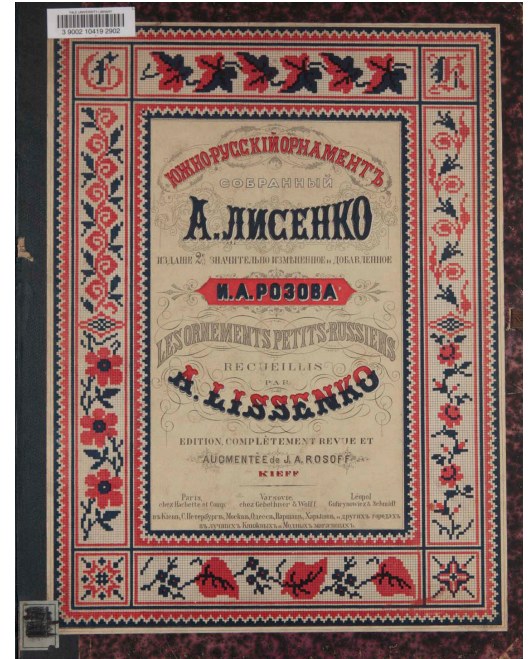
¹¹ <http://hdl.handle.net/10079/fa/mssa.ms.1960>

2. Making Searchable Islamic Fundamentalist Tapes

Through a complicated set of circumstances, 1,400 audio tapes belonging to Islamic fundamentalists that were found in a residential compound in Kandahar, Afghanistan, made their way to YUL. Most of the tapes are in Arabic, dating from the late 1960s through 2000, and feature over 200 speakers from more than a dozen countries in the Middle East, Indian subcontinent, and Africa. Included are speeches by clerics; the voice of Osama bin Laden (estimated 20 tapes); bin Laden's predecessor as head of the Arab volunteers in Afghanistan; and various other influential persons.

While digitization of the tapes was funded by YUL, the Arcadia grant provided for the employment of two temporary library staff members who are compiling descriptions of each tape in English and Arabic. The project was completed on time and the digitized audio files can now be accessed via workstations located in YUL's Manuscripts and Archives department. Similar to the Cuban Exile Broadcast tapes, this restriction of access is due to an earlier decision by the Yale University General Counsel. A 460-paged Finding Aid, providing summaries in Arabic and English of the content of the 1400 audiotapes, was created and can be searched via the Internet (sample pages in appendix I).¹² Scholars on Yale's campus may listen to the audio files in the Manuscripts and Archives reading room and may also request copies of specific digital files. This type of remote inquiry was tested when a Norwegian scholar who is writing a book on the Arabs in Afghanistan in the 1980s contacted the Library with a list of search terms ("Muslim Brotherhood," "Muslim World League," "Individual duty," "America" or "USA" or "Americans," "Bin Ladin" or "Bin Laden"), and received a corresponding list of tapes in which these terms appear. This request was made in anticipation of a future visit to New Haven to listen to the actual tapes.

¹² <http://hdl.handle.net/10079/fa/mssa.ms.1880>



3. Digitizing Joel Sumner Smith's Collection of Slavic Literature

Joel Sumner Smith was a member of Yale Class of 1853, a librarian, and one of the few students of Slavic languages and literatures in the mid-nineteenth century. In 1896 he donated his impressive Slavic collection to Yale, which has the distinction of being the first of its kind not only in North America but in the entire Western Hemisphere. The nucleus of the collection consists of some 6,000 volumes focusing on literature, language, history, and geography. Although Russian publications are the backbone of the collection, it also includes many titles in other Slavic and East European languages, such as Polish, Croatian, Serbian, Czech, Slovak, Estonian, Latvian, and Kashubian. The collection is based mainly on what we would today call Slavic philology—the study of languages, folk literatures, bibliographies, grammars, and dictionaries. The project's goal has been to digitize the Joel Sumner Smith collection so that these titles from the 18th and 19th century could be available to Slavic students and scholars worldwide.

Nearly 2,000,000 pages in the collection have been scanned, processed, and ingested into a Fedora-based digital repository; the digital page images are now accessible through a web interface (project home page in appendix J).¹³ Using Optical Character Recognition (OCR) conversion, this digital collection is indexed and searchable in various languages (tri-lingual text example in appendix K). Another 390,000 pages (scanned from microfilm) remain for ingest into Fedora, and fewer than 300 physical volumes housed in the Beinecke Rare Book & Manuscript Library need to be scanned, processed, and ingested. The full collection is expected to be in the Fedora repository by the end of 2012.

A challenge in this project has been the quantity of titles containing text in multiple alphabets and languages (e.g., Arabic and Cyrillic used to describe Tatar, a Turkic language spoken in the medieval Kazan Khanate) and this pushed the current limits of OCR software products such as ABBYY FineReader, which are still far from perfect with regard to handling minor and extinct languages, as well as linguistically multilayered texts. The project has increased our understanding of OCR processing for non-Roman texts and will inform future digitization projects. Specifically, this project has established the process by which the Divinity Library Day Missions collection, digitized in Arcadia Year 2, will be ingested into a Fedora repository beginning in January 2013.

An article on the digitization of the Joel Sumner Smith collection, written by Curator Tatjana Lorković, with Graziano Krätli and Carolyn Caizzi of the Arcadia team, is due to appear in a forthcoming issue of the peer-reviewed journal *Slavic & East European Information Resources* (a pre-print of the article is included in appendix L).

¹³ <http://www.library.yale.edu/jss>

III. YEAR 4 PROJECTS

A. Continuing Projects *Digitizing Joel Sumner Smith's Collection of Slavic Literature*

Full ingest into the Digital Joel Sumner Smith repository will be completed by December 2012

B. New Projects This spring we invited project proposals from the YUL librarians for Arcadia Year 4. The numerous proposals were rated on their alignment to the Arcadia Fund's mission to "protect endangered cultures and nature" and to the four areas of foci established in Year 1 of the grant:

- Improving global knowledge of and access to works published in non-Roman alphabets and characters
- Improving global knowledge of and access to works published in languages relatively infrequently taught and used in the United States or other countries, such that YUL has been unable to catalog and manage them well
- Improving global knowledge of and access to library materials whose cataloging metadata have been invisible or nearly invisible to users both inside and outside of Yale University
- Digitization of unique materials that are immediately needed by scholars and students and/or are in suboptimal to poor physical condition such that the materials will become "extinct" if not immediately saved

From those proposals, the following five projects have been identified for funding in Year 4, should the Arcadia Fund decide to continue its support of YUL.

1. Digitization of Hán Nôm Handwritten and Woodblock Manuscripts Held in the Maurice Durand Collection

In 1967, the Yale University Library's Southeast Asia Collection acquired the private collection of Maurice M. Durand, director of École Française d'Extrême-Orient in Hanoi from 1947 to 1957, and a well-known scholar, author, and translator of Hán Nôm during his lifetime. The collection contains 197 handwritten and woodblock texts in Hán Nôm, a writing method for the Vietnamese language adapted from, and incorporating pure and/or modified Chinese characters developed in the 13th century and used up until the 20th century. The focus of this effort is to create metadata based on the finding aid for the collection, assess material for conservation, digitize the Hán

Nôm manuscripts and notebooks housed in this collection, and present them through the Library's website.

2. *Enriching, Enhancing, and Connecting Yale's Digital Collections: Yale Persian Collection*

Based on lessons learned in the Joel Sumner Smith Arcadia project, Yale Library IT will create a specialized repository for similar digital collections of international materials. This project will select Persian titles, especially those out-of-copyright in the old Yale classification, since many have minimal cataloging and thus low discoverability. The project team will create full level MARC records for the included volumes, identify and execute any necessary preservation, create digital images using external vendors, and ingest images and metadata into the digital collections repository for use by researchers worldwide through the Library's website.

3. *Arabic and Persian Medicine: Middle Eastern Texts in a Western Collection*

While Yale's Medical Historical Library is primarily known for its Western emphasis, it also has an important collection of early Arabic and Persian medical manuscripts and books that are underutilized and unknown to scholars who might not think to look for this collection at Yale. This collection reflects the Arabic and Persian intellectual efforts that translated, augmented, and transmitted Greek and Roman medical knowledge to Western societies during the Renaissance. It includes iconic works by Arabic authors such as Avicenna and al-Razi. The team will extract existing MARC records from ORBIS representing approximately 25,000 pages, convert these records to the Dublin Core format, digitize and ingest the page images and metadata into the AMEEL collection, an existing digital repository of academic resources focusing on the Middle East.¹⁴

4. *Cataloging Indigenous African Language Pamphlets*

Yale Library has identified approximately 1,170 uncataloged pamphlets representing about 90 African languages, many of which are considered endangered. The identified pamphlets were published from 1945 to the 1980s, with some of the pamphlets in non-Roman, non-Arabic script. This project will add cataloging records to ORBIS and WorldCat, identify languages, and classify items without copy cataloging. As the materials are being cataloged, a preservation assessment will be made for each item. Once cataloged, the materials will be discoverable by scholars worldwide through ORBIS and WorldCat.

¹⁴ AMEEL (Arabic and Middle Eastern Electronic Library) began with funding from the U.S. Department of Education. It holds approximately 250,000 pages of full text, indexed and searchable in the language of publication including Arabic and Western scripts. Please see: <http://www.library.yale.edu/ameel>.

5. Processing the North African Jewish Collection

The Jewish community of North Africa dates back to pre-Roman times. Though historically there were Jews living in Tunisia, Algeria, and Libya, the Moroccan community was by far the largest and most important. The language spoken by most Jews in Morocco was Judeo-Arabic, composed of Arabic and Hebrew. North African Jews also spoke Haketiya (Judeo-Spanish), a merging of Hebrew and Spanish. As a result of a mass exodus of Jews from North Africa in the mid-20th century, their culture and languages are in danger of dying out. The Yale University Library Judaica Collection has collected 15,000 pages in 2,000 titles of manuscripts from Jewish North Africa covering legal, rabbinic, business, personal, and professional matters. Moshe Bar-Asher, one of the world's leading scholars in the field and professor emeritus at the Hebrew University in Jerusalem, will advise as the team reviews, deciphers, and translates the manuscripts so that short summaries of their contents can be provided. The collection will also undergo a preservation review and will be rehoused as necessary. An online finding aid will be created so that the collection can be discovered via ORBIS and WorldCat.

IV. BUDGET SUMMARY

The following budget summary details the use of the Year 3 Arcadia funds by expenditure type and includes the total expenditures of Year 1 and Year 2. The Arcadia Endowment, into which \$250,000 of each year's grant has been deposited, will begin generating income during Arcadia Year 4. To ensure that the achievements of the Arcadia projects are truly enduring, each year the full array of Arcadia projects will be reviewed to determine which would most benefit from the annual income generated by the endowment. Future reports will include a statement about how the income from the Arcadia Endowment has been used.

Yale University is deeply appreciative of the opportunities made possible by the Arcadia funds. Our vision of transforming Yale's "treasure house" into a global resource is taking shape as we improve global knowledge of and access to YUL's remarkable collections. We are pleased to be moving swiftly and efficiently toward achieving our goals of identifying and executing projects that reflect Yale University's shared commitment with Arcadia to preserve and make accessible the world's cultural treasures.

APPENDIX A: ROMANIZATION TABLE FOR
MOROCCAN TAMAZIGHT

APPENDIX B: ROMANIZATION TABLE FOR VAI

APPENDIX C: ROMANIZATION TABLE FOR CHEROKEE

APPENDIX D: MICROSOFT LOCAL LANGUAGE
PROGRAM EBRIMA FONT ANNOUNCEMENT

APPENDIX E: CAMBODIAN NEWSPAPER PROJECT
LIBGUIDE HOMEPAGE

**APPENDIX F: SAMPLE CAMBODIAN NEWSPAPER
PROFILES & HOLDINGS SUMMARIES**

APPENDIX G: ROMANIZATION TABLE FOR KHMER
LANGUAGE

APPENDIX H: SAMPLE PAGES FROM THE GUIDE TO
THE “AL CANTÍO DE UN GALLO” RADIO PROGRAM
AUDIO RECORDINGS

APPENDIX I: SAMPLE PAGES FROM THE GUIDE
TO THE ISLAMIC FUNDAMENTALIST AUDIO
RECORDINGS COLLECTION

APPENDIX J: DIGITAL JOEL SUMNER SMITH
COLLECTION HOMEPAGE

APPENDIX K: SAMPLE TRI-LINGUAL RECORD FROM
JOEL SUMNER SMITH COLLECTION

APPENDIX L: ARTICLE FROM SLAVIC & EAST
EUROPEAN INFORMATION RESOURCES ON
JOEL SUMNER SMITH DIGITIZATION PROJECT

APPENDIX M: ARTICLE FROM SLAVIC & EAST
EUROPEAN INFORMATION RESOURCES ON
JOEL SUMNER SMITH DIGITIZATION PROJECT