

12-31-2011

The National Library of Israel and OCLC

Elhanan Adler
elhanana@savion.huji.ac.il

Marina Goldsmith
National Library of Israel, marinag@savion.huji.ac.il

Follow this and additional works at: <http://ajlpublishing.org/jl>

 Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), [Information Literacy Commons](#), [Jewish Studies Commons](#), and the [Reading and Language Commons](#)

Recommended Citation

Adler, Elhanan and Goldsmith, Marina. 2011. "The National Library of Israel and OCLC." *Judaica Librarianship* 16: 83-92.
doi:10.14263/2330-2976.1005.

ESSAYS AND RESEARCH

The National Library of Israel and OCLC

ELHANAN ADLER AND MARINA GOLDSMITH

ABSTRACT

OCLC published the following announcement in December 2008: "The National Library of Israel and OCLC have completed a pilot project that has resulted in the addition of more than 788,000 new bibliographic records and 1.1 million holdings from the national library to WorldCat." The successful completion of this project was the result of a number of policy decisions and technological developments on the part of both parties. This article describes the motivation, history, and challenges of this project.

INTRODUCTION

The National Library of Israel (NLI), formerly the Jewish National and University Library (JNUL), collects all material published in the State of Israel, as well as books, periodicals, manuscripts, documents, recordings, maps, and pictures that reflect or represent the history of the Jewish people and its culture. Beginning in 1985, all printed materials are cataloged into an integrated online catalog in one or more of four scripts: Hebrew, Latin, Arabic, or Cyrillic. A retrospective conversion of all Hebrew, Yiddish, and Arabic cards was completed in 2000–2001. In those years the JNUL also received 70,000 bibliographic records for pre-1985 Roman alphabet Judaic publications via OCLC's RetroCon program. Since then, the library has continued to convert catalog cards, but still has approximately 500,000 unconverted cards in Latin and Cyrillic alphabet scripts, mostly representing non-Judaica humanities publications. Like most large Israeli libraries, the National Library uses the Ex Libris ALEPH ILS system. While the Library was one of the first to adopt the original ALEPH, its data was converted to the fully MARC and Unicode compliant version only in 2005 (the conversion also included integration of other, previously independent catalogs such as music and man-

uscripts into the central catalog). This technological development enabled the National Library to seek membership in OCLC WorldCat.

In the Fall of 2006, after the Library was relatively settled in its new ALEPH 500 system, and following some informal discussions the Library contacted OCLC regarding becoming full members in OCLC and batch loading current and retrospective cataloging in all scripts. OCLC's response was positive and following a meeting between representatives of OCLC and the Library at the London "Online Information" conference in November 2006 a letter of agreement on a pilot project was drawn up for loading the Library's records into WorldCat and to "explore issues related to matching these records to existing records in WorldCat and adding non-matching records to WorldCat." This agreement was signed in February 2007 and announced shortly afterwards at the OCLC Members Council, in the President's Report presented by OCLC President and CEO Jay Jordan. This agreement was further announced by Mr. Jordan at the annual Israeli INFO 2007 conference in April 2007 (organized by the Teldan Company, representatives of OCLC in Israel), at which time Mr. Jordan also visited the Library.

In February 2007 the Library sent to OCLC for analysis a small sample of its MARC records in various scripts. This was followed by intensive correspondence between OCLC and Library staff regarding the unique problems of loading Israeli-standard MARC records to WorldCat. The pilot project itself took a bit longer than the projected one year, and culminated in July 2008 with the loading of some 1,164,000 records, of which 788,000 resulted in new records in WorldCat.¹ Subsequently the Library sent several major and minor update files and currently tries to send regular updates on a monthly basis.

MOTIVATIONS: NLI'S AND OCLC'S

It is a well-known fact that WorldCat is the world's largest and most comprehensive bibliographic database. Inclusion of bibliographic records in WorldCat increases visibility and utilization of a library's resources. Searches carried out via WorldCat.org, the free Web interface of WorldCat, allow users to access the library's holdings, without having to search that particular library's OPAC. For these reasons, NLI was interested in having their holdings incorporated into WorldCat. There is, however, another aim in joining OCLC. The National Library views as one of its objectives the granting of professional guidance to libraries in Israel and in the Jewish community outside of Israel. It seeks to provide a leadership role in the area of Judaica, Hebraica, and Middle Eastern librarianship. The inclusion of bibliographic records in WorldCat enables the duplication or export of records created by NLI to all OCLC member libraries.

¹ Editor's note: Many non-Roman records contributed to WorldCat by NLI duplicate existing WorldCat records for the same editions. Because the NLI records are structured differently from the other WorldCat records (e.g., they lack parallel Roman and non-Roman fields), they display separately.

WorldCat views itself as a global catalog. Thirty year ago, the first national library to add its holdings was Koninklijke Bibliotheek, the Royal Library of the Netherlands. Since then, 34 national libraries have contributed digital images, national files, and bibliographies to WorldCat (OCLC, 2009). In *OCLC 1998–2008: Weaving Libraries into the Web*, Jay Jordan writes, “Indeed, non-U.S. libraries now have six reasons to load their holdings into WorldCat: Unicode support, Ability to FRBRize² . . . a catalog, Open WorldCat on the Web, Group catalog capability and its customized views, WorldCat Collection Analysis service, International resource sharing. . . . In April 2008, for the first time, the number of records in languages other than English exceed those for English-language materials” (Jordan 2009). OCLC was eager to include NLI’s holdings to WorldCat, as this would add many unique records in a variety of scripts—not just Hebrew, but also Arabic and Cyrillic, providing valuable experience in working with all these scripts as language of cataloging.

THE CHALLENGES

There were a number of challenges that needed to be overcome in order to incorporate NLI’s records into WorldCat:

1. Different models for cataloging multiscript publications.
2. Different systems for providing for non-filing characters.
3. Adapting NLI’s records to more rigorous MARC standards.

“MODEL A” VERSUS “MODEL B”

The Library of Congress MARC 21 Internet site describes two different models for recording data in multiple scripts in MARC records. “Model A” provides for original script and Romanized data, through use of the 880 fields for the original, non-Roman scripts. This is the standard that is employed by most libraries in the United States and also by OCLC. However, the records submitted by the National Library were created using “Model B.” In “Model B” (simple multiscript records), “all data is contained in regular fields and script varies depending on the requirements of the data.” OCLC was not equipped to accept records based on the Model B standard.

² “Functional Requirements for Bibliographic Records—or FRBR . . . is a conceptual entity-relationship model developed by the International Federation of Library Associations and Institutions (IFLA) that relates user tasks of retrieval and access in online library catalogues and bibliographic databases from a user’s perspective. It represents a more holistic approach to retrieval and access as the relationships between the entities provide links to navigate through the hierarchy of relationships. The model is significant because it is separate from specific cataloguing standards such as AACR2 or International Standard Bibliographic Description (ISBD).” (http://en.wikipedia.org/wiki/Functional_Requirements_for_Bibliographic_Records)

The different approaches stem from a difference in ideology, the needs of the patrons and librarians, and technological considerations. When publications are cataloged according to “Model A,” non-Roman access points are integrated in one catalog. This allows users to retrieve in one search all publications by a particular author. It also enables librarians to access publications, process acquisitions, circulation, and interlibrary loans in scripts with which they are not familiar.

When library catalogs were computerized, at first only Roman script could be used, so both descriptive and access fields had to be entered in Romanization only. In the 1980s OCLC and RLIN began to introduce character sets for major non-Roman scripts, enabling catalogers to transcribe bibliographic data as it appears on the piece in hand. (ALCTS, 2009.)

At present an attempt is being made to standardize the use of Romanized and vernacular data. In September 2009, the Program for Cooperative Cataloging (PCC) Task Force on Non-Latin Script Cataloging Documentation posted a preliminary report, *Guidelines for Creating Bibliographic Records in Multiple Character Sets*. The document recommends that “catalogers adding non-Latin forms to records are encouraged to consider the future use and international implications of their records and to include as much of the original script data as necessary to facilitate the identification and location of this often scarce material and its component parts” (LC, 2009).

While the Library of Congress is attempting to standardize use of non-Latin and Romanized data, the ALCTS Non-English Access Working Group on Romanization questioned the continued validity of “Model A” records. The Working Group was charged with the task of examining the current use of Romanized data in bibliographic and authority records and making recommendations for best practices. The Group was to take into consideration the needs of the library users for search and retrieval as well as that of the library staff. They were to address the following questions:

- ◇ Is Romanization still needed in bibliographic records, and if so, in which situations and/or for which access points? Should best or different levels of practices be adopted for Romanization?
- ◇ Can Model A & B records coexist in library systems? If so, should guidelines for usage be adopted? (ALCTS, 2009.)

The Working Group concluded that it was “premature” to make a general shift to “Model B.” They felt that further research was necessary and recommended that automatic transliteration software be employed whenever possible. It was also suggested that the Romanized fields be limited to key data fields (titles and headings). The Working Group foresaw that certain communities might decide to move to “Model B” sooner than others, due to the differences and needs of the various language/script cataloging communities.

“Model B” has always been the standard cataloging model in Israel. The Israeli tradition of maintaining separate catalogs for Hebrew, Yiddish, and Arabic publications can be traced back to the Jewish National and University Library (Lazinger 1998, p. 24). Some libraries, NLI included, also maintain a separate catalog for works in the Cyrillic script. In 1970 Bernard Hugo Rabenstein summarized the basic rationale for separate catalogs:

1. In the past, immigrants speaking and reading various languages and alphabets immigrated into Israel from many parts of the world and are still arriving.
2. The Israelis frequently read literature in non-Hebrew alphabets. Libraries are obliged to build large collections of non-Hebrew books and to establish catalogs with non-Hebrew headings as well as catalogs with Hebrew headings.
3. The Israelis can read many alphabets. They need no aid from the conversion of alphabetical characters that they can read in the original.
4. Average library workers in Israel (filers, typists, technical assistants, etc.) do not have trouble with different alphabets that they usually can read.
5. There is no ideal alphabetical form into which headings of other alphabetical forms can be converted because: 1) Hebrew is an alphabetical form into which conversion of other alphabetical forms is virtually impossible; 2) the Roman alphabet, though an important alphabet, is not native to Israelis.
6. Israeli librarians are concerned with entering author headings into catalogs under forms with which the readers are familiar. According to them a Cyrillic book ought to be given a heading in Cyrillic, rather than in Roman. (Rabenstein 1970, p. 67.)

The major disadvantage of maintaining separate catalogs is the need to perform multiple searches in order to retrieve all information pertaining to a particular person or institution. This continues to be the case, even in today’s online environment.

OCLC’s WorldCat already included much data in Hebrew, Yiddish, Arabic and Cyrillic alphabets, but prior to the incorporation of NLI’s records, this information was contained in parallel, linked 880 fields. Lacking Romanized data, and without 880 fields, non-Roman cataloging records from NLI could not be merged with existing records in OCLC. This was the case even in records where a unique identifier, such as an ISBN number, existed. According to OCLC’s Standard Merge Rules, “The field will not merge if the Language of Cataloging (040 subfield \$b) in either the retain or replace record is different from the language of cataloging in the other record(s)” (OCLC 2009a, p. 9). The solution proposed by OCLC was to convert NLI’s records from “Model B” to “Model A.” The OCLC analysis of the initial test file stated:

A large number of the records are what the Library of Congress refers to as “Model B” records [a couple examples are included below]. This is a record that contains non-Latin data in fields other than the 880. Field 880 (Alternate Graphic Representation) is not used.

OCLC cannot process “Model B” records as such. However, OCLC can modify these records so that any field containing non-Latin data would be moved to 880 fields. In addition to the 880s being created an appropriate 066 [Character Sets Present] field could be supplied. Fields containing Latin data would be unaltered. The software needed to accomplish this is currently not in production and would have to be developed as we move forward with the project.

In practice, the OCLC batch record-ingest software was adapted to create a minimal 245 title field consisting of angle brackets only: <>—and changing all non-Roman fields to unlinked 880 fields, for example:

100 Hebrew author
245 Hebrew title
260 Hebrew imprint

These became:

No field 100
245 <>
No field 260
880 (subfield 6 100) Hebrew author
880 (subfield 6 245) Hebrew title
880 (subfield 6 260) Hebrew imprint

Users of OCLC Connexion, however, see the 880 fields as regular MARC tags, but this is a display convention and not the real coding:

040		J9U #b heb #c J9U #d OCLCQ
066		#c (2
016	7	001915481 #2 IsJNL
092		#b
049		J9UA
100	1	גרוסמן, דוד, 1954 -d
245	1 0	מישהו לרוץ אתו / #c דויד גרוסמן.
245	0 0	<>
260		[בני ברק] : #b הקיבוץ המאוחד, 2000 #c.
300		341 ע' ; 21 #c ס"מ.
440	1	הספרייה החדשה למנויים, 2000 ; #w 1

NON-FILING CHARACTERS

Traditionally, initial articles are recorded but ignored in filing in headings such as titles (“The history of science” files as “History of science,” ignoring the article “The”). In traditional card catalogs, the person filing the cards with such headings did this “on the fly,” but with the advent of computerized catalogs it became necessary to deliberately code articles to be ignored in filing. The MARC solution for such cases was to assign a numerical “indicator” to the relevant fields (primarily titles and series) providing the number of non-filing characters to be ignored, for example:

245 04\$aThe Year book of medicine.

Here, the second indicator “4” indicates that the word “The” and following space are to be ignored, and filing should begin at the word “Year.”

This solution, while perhaps adequate for most European languages, is insufficient for Hebrew: The non-filing indicator solution can be applied only where there was a “spare” indicator available for this purpose. The MARC 246 “Variant form of title” field does not have an available indicator and therefore initial articles are simply dropped from the heading, which is not acceptable in Hebrew.

It is necessary to suppress initial articles not only in titles and series, but in other headings as well, for example, corporate entries such as “ha-Universiṭah ha-petuḥah.” LC practice with initial articles at the beginning of corporate headings is to drop them entirely (“Universiṭah ha-petuḥah”), creating a form that is grammatically incorrect. (In Hebrew the article is attached to both the noun and the adjective and must appear with both.)

Israeli practice is also to display but suppress initial articles at the beginning of secondary subfields such as: “Israel. ha-Lishkah ha-merkazit li-staṭiṣṭikah.” (Here, also, LC uses a grammatically incorrect heading: “Israel. Lishkah ha-merkazit li-staṭiṣṭikah.”)

With the advent of automated systems, the initial pre-MARC Israeli ALEPH system developed a solution to handle non-filing characters by using double angle brackets to delimit the text to be suppressed in filing, for example:

<<The>> Year book of medicine

<<ha->>Universitah ha-petuḥah

Israel. <<ha->>Lishkah ha-merkazit li-staṭiṣṭikah

As the ALEPH system developed into the fully MARC-compliant “ALEPH-500” it provided for both methods of handling non-filing elements—either non-filing indicators or delimiting with <<>>.

In 1997 the MARBI (Machine-Readable Bibliographic Information) advisory committee on the MARC standard noted:

The omission of initial articles to deal with not being able to handle them otherwise is not totally acceptable to some USMARC users. European and Middle Eastern libraries have been particularly vocal in their call for a generalizable technique, like the UNIMARC control character technique, for indicating non-filing characters. Their chief argument has been that the simple omission of articles corrupts the cataloging data grammatically and yields title strings that the public finds unacceptable. (MARBI, 1997.)

Several years of MARBI discussion on this issue led to the addition to the MARC format of the option to use delimiting characters for article suppression (LC, 2004) however, while this technique is now permitted, it is not implemented by LC or OCLC, and the Israeli records containing <<>> for article suppression were not acceptable to OCLC. The solution proposed by NLI, approved by OCLC and implemented by the NLI as part of its record export procedure has three parts:

1. For MARC fields which have non-filing indicators, the NLI drops the non-filing delimiters <<>> and sets the standard MARC non-filing indicator accordingly.
2. MARC fields which lack non-filing indicators:
 - a. For Roman and Cyrillic alphabet headings beginning with suppressed articles, the article is totally dropped. This does create grammatically incorrect headings (as noted above) but since this is what LC is doing anyway and such headings already appear in WorldCat it was felt that this is the lesser of the two evils.
 - b. For vernacular Hebrew and Arabic character fields it was decided to maintain the grammatically correct forms even if this caused the headings to file under the article.

OTHER ADAPTIONS OF NLI'S RECORDS TO MEET OCLC STANDARDS

The MARC format for cataloging was partially implemented in Israel in the year 2000. Prior to this, Israeli libraries, including NLI, used ALEPH's initial system of mostly two-letter mnemonic field codes (e.g. TL for the MARC 245 title field) (Lazinger 1998, p. 185). This system allowed and even encouraged much creativity and flexibility on the part of the catalogers. There were no set fields (although most libraries chose TL for titles and AU for personal authors) and each library was free to define the cataloging codes in accordance with their needs. At the beginning of the twenty-first century, NLI had nine separate catalogs, representing special collections, in addition to its main catalog. Each cata-

log had its own set of codes. So for example, the Music Department catalog had special codes for female composers, male composers, directors, performers, and authors. Much, but not all, of this excess “creativity” was removed in the NLI’s conversion to MARC format and in merging the special collection catalogs into one central file. The Library also did not adhere to all the coding standards of fixed length fields, such as the Leader and 008 fields. Following data testing by both the NLI and OCLC many of these aberrations were corrected in the NLI’s own catalog. Others, primarily Israeli-standard practices (such as additional general material designations [gmd]) are removed or changed as part of the export-to-OCLC procedure.

FUTURE COOPERATION BETWEEN ISRAELI LIBRARIES AND OCLC

Building on the experience of the NLI and OCLC in loading Israeli-standard bibliographic records to WorldCat, an agreement was reached in 2009 between OCLC and the Israeli academic MALMAD consortium to gradually load to WorldCat the holdings of some forty libraries which participate in the Union List of Israel (ULI). These records will be centrally exported from ULI to WorldCat, beginning in 2010. Initially Roman-character records (their language of cataloging is English) will be uploaded, followed at a later stage by the other alphabets. OCLC does not currently have the capability to match records in non-Roman scripts, so the export of Hebrew, Arabic, or Cyrillic records that do not have unique identifiers (such as ISBN or OCLC system numbers) would create duplicate records in WorldCat. OCLC and MALMAD are working to resolve this problem—either at the WorldCat ingest stage, or possibly within the Israeli union list before export of the records.

CONCLUSION

Almost twenty years ago, one of the authors of this article wrote:

As networking improves and increases, it is more and more likely that Israeli libraries will wish to mesh into international bibliographic networks as well—giving as well as receiving. Israel should be the source of cataloging data relating to Israeli publications in all languages. This will require adherence to MARC format and standards, at the very least at the export stage, and ultimately internally as well. The problems are not trivial: they include not only character and format conversion, but also creation of data fields not commonly used in Israel. (Adler 1991, p. 11.)

We are pleased to see that this vision is now being realized.

SOURCES

- Adler, Elhanan (1991). "Judaica and Hebraica Cataloging: The Hebrew Bibliographic and Israeli Traditions," *Judaica Librarianship* 6 (1991–1992): 8–12.
- [ALCTS, 2009] Association for Library Collections and Technical Services. Task Force on Non-English Access. *Draft Report*. Nov. 24, 2009. <http://connect.ala.org/node/88651> (accessed January 1, 2010).
- Jordan, Jay (2009). "OCLC 1998–2008: Weaving Libraries into the Web," *Journal of Library Administration* 49 (2009): 727–762.
- Lazinger, Susan S.; Adler, Elhanan (1998). *Cataloging Hebrew Materials in the Online Environment: A Comparative Study of American and Israeli Approaches*. Englewood, Colorado: Libraries Unlimited, 1998.
- [LC, 2004] Library of Congress, Network Development and MARC Standards Office. *Guidelines for the Non-Sorting Control Character Technique*, Feb. 2004. <http://www.loc.gov/marc/nonsorting.html> (accessed February 25, 2010).
- [LC, 2009] Library of Congress. *PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets and Preliminary Report: Task Force on Non-Latin Script Cataloging Documentation*. September 1, 2009. <http://www.loc.gov/catdir/pcc/scs/Non-Latin-TF-Prelim-Report-Guidelines.pdf> (accessed December 31, 2009).
- Lovins, Daniel (1991). "The Changing Landscape of Hebraica Cataloging," *Judaica Librarianship* 14 (2008): 1–13.
- MARBI (1997). *Discussion Paper no. 102: Non-Filing Characters*. 1997. <http://www.loc.gov/marc/marbi/dp/dp102.html> (accessed February 25, 2010).
- National Library of Israel. *About the Library: Objectives and Functions*. <http://jnul.huji.ac.il/eng/about-en.html> (accessed December 31, 2009).
- OCLC (2008). *News Releases: National Library of Israel Adds 788,000 Records to OCLC WorldCat*. <http://www.oclc.org/news/releases/200848.htm> (accessed January 1, 2010).
- OCLC (2009a) *OCLC Batch Services Guide: Appendix C, Merging Specifications*. http://www.oclc.org/support/documentation/batchprocessing/using/guide/C_mergingspecifications.pdf (accessed January 2, 2010).
- OCLC (2009b). *National Library Participation in the OCLC Global Cooperative*. <http://www.oclc.org/WorldCat/catalog/national/default.htm> (accessed January 1, 2010).
- Rabenstein, Bernard Hugo (1970). *A Survey on the Use of Alphabetical Forms in Author and/or Title Headings in the Catalogs of Israeli Libraries*. Master's Thesis, Washington, DC: Catholic University of America, 1970.
- Weinberg, Bella Hass (1991). "Judaica and Hebraica Cataloging: Anglo-American Traditions," *Judaica Librarianship* 6 (1991–1992): 13–23.

Elhanan Adler recently retired from the National Library of Israel, where he was Deputy Director for Information Technology. He is currently head of the M.Ed. program in Information Science and Librarianship at the David Yellin College of Education in Jerusalem, part-time coordinator of the Israeli Academic Libraries Network, lecturer at the Bar-Ilan University Department of Information Studies, and a consultant on library and information systems and processes. He is the author (with Aviva Shichor and Rochelle Kedar) of the standard Israeli text on cataloging (three editions), and (with Susan Lazinger) of the book *Cataloging Hebrew Materials in the Online Environment*. He has also written articles and lectured extensively on organization of information, computers in libraries, and the connection between them. elhanana@savion.huji.ac.il

Marina Goldsmith is Head of the Foreign Language Cataloging Department at the National Library of Israel. She is also chair of the Israeli Inter-University Cataloging Committee. marinag@savion.huji.ac.il