

Providing full and equal access to resources is a primary responsibility of libraries in multi-cultural societies, but when these cultures find expression in multiple languages and scripts, libraries can find themselves stymied by the limits of technology and international standards.

Israel is one such multi-cultural, multi-lingual hotbed, with two official languages (Hebrew and Arabic), one commonly accepted language of academia (English), and, as a nation with a high percentage of immigrants, several smaller language groups. The primary scripts in use are Hebrew, Arabic, Latin, and Cyrillic. No one language is the native tongue of a majority of Israelis – only 49% of Israelis over the age of 20 are native Hebrew speakers, and a further 18% have Arabic as their native tongue and 15% Russian (Central Bureau of Statistics, 2013). A solution such as practiced abroad – romanization – is unworkable for access by the general public, both because of the system’s emphasis on grammatical purity over relevance to spoken language and because it uses an alphabet which is not native to the vast majority of Israelis.

Cutter’s (1904, p. 12) objectives for a dictionary catalogue hold that the function of the catalogue is to show what the library has:

- (a) by a given author
- (b) on a given subject
- (c) in a given kind of literature.

With resources available in nearly every language and four languages (Arabic, English, Hebrew, and Russian) used in cataloguing, how could Israeli libraries provide access to the breadth and width of their collections? The traditional answer, which was used in the era of card catalogues, was having separate catalogues for each script. To find all resources on a topic or by a creator the patron had to run four separate searches. This is inefficient, unwieldy, and utterly unsuited to the digital age.

Libraries have tried various technological solutions to overcome this difficulty. The most common work-around is to give an added entry for names in additional scripts. While this allows for discovery in multiple scripts it distorts the bibliographic record by recording the same entity twice (the “phantom tracing” warned against by Malinconico in 1974 [1985]) and does not solve the problem of simultaneous searching across scripts. Added entries cannot be updated automatically from unilingual authority records, so cataloguing staff must do more manual bibliographic file management. This system also places an added cognitive burden on patrons, who need to learn that added entries sometimes refer to separate agents and sometimes to the same, requiring them to distinguish between the two cases and know when they must follow up on the added entry to see all of a creator’s work.

In any case, added entries for agents already named in the record leads to the number of resources in the catalogue being inconsistent, since the same resource is counted in each language, leading to duplicate mentions of the same manifestation.

רשימת חיפוש לפי: Authors+Titles		רשימת חיפוש לפי: מחברים	
ערך	מספר רשומות	ערך	מספר רשומות
Kedar, Rochelle	2	קדר רחל	10
Kedar, Shelley	1	קדר רחל (תדפיסים)	75
Kedar, Yig'al	1	קדר רמי	2
Keddie, Kenneth M. G.	1	קדר רנה - רמיזות	1
Keddie, Nell	1	קדר רקפת	1
Keddie, Nikki R	17	קדר שאול	2
Keddie, Nikki R Roots of revolution	1	קדר שדמה	3
Kedem, Benjamin	1	קדר של - רמיזות	1
Kedem-Cuomo, Dina	2	קדר שמאל - רמיזות	
Kedem, Michal	1	ראה: קדר שמאל אליקים	
		קדר שמאל אליקים - רמיזות	30

Fig. 1: The goal of full access is achieved, but not of equal access. Two of the 10 resources attributed to the Hebrew form of Kedar, Rochelle are English-language expressions. Since the Hebrew form was added to those cataloguing records, they show up in both the Hebrew-language form and the English. Users would be justified in thinking the author created 12 manifestations: 2 in English and 10 in Hebrew.

The National Library Law (2007) made the National Library of Israel (NLI) independent of its academic parent body. Part of the new mandate was “to make the Library’s collections reasonably accessible to the general public, in Israel and abroad, also through advanced technologies.” The lack of full and equal access to NLI’s multilingual collection went from being an annoyance to being a potential legal problem; the search for a solution was on.

Though many European libraries have multilingual collections, they are, primarily, in a single script. Therefore they were of limited help in deriving a solution. The libraries of the Arabic-speaking world experience problems similar to Israel’s, but they, too, had no answers which could be adopted.

Though the Bibliotheca Alexandrina uses “the Library of Congress MARC 21 Model B, where the transcribed text in the bibliographic record is entered only in the script in which it appears” (El-Sherbini, 2013), their authority records are uniscript. The same agent may have multiple name authority records and the problem of simultaneous access is not solved.

In 2011 NLI ramped up efforts to solve the problem. The strongest contender for a solution was using the Virtual International Authority File (VIAF) to create a linked data solution, but at that point VIAF was not yet considered a stable product. In any case, using VIAF as the mechanism to link various forms of an agent’s name would mean losing all notes and justifications in the local database. Using VIAF would have also meant giving up local control, being unsure what sort of results patrons would get, and would have involved a lot of manual bibliographic file maintenance as VIAF clusters changed. Most importantly, existing library systems could not make sophisticated enough use of VIAF data to give patrons a unified and uniform search experience.

The problem remained: How could NLI, and in its wake other Israeli libraries, allow all resources to be discovered in a single, multilingual vernacular search? The answer: NNL10, a multilingual, multiscript authority database that combines MARC encoding with local subfields to create robust equivalent terms.

NNL10 started its life as NLI's authority file. It contains two major divisions:

- name authorities (people, corporate bodies, families, titles, series)
- subjects (Library of Congress Subject Headings [LCSH], translations of LCSH, localized subjects, geographical entities)

In keeping with NLI's initial research, VIAF clustering was used to bring together the original unilingual authority records. The Swiss National Library's use of multiple 1XX fields with a local subfield (\$9) to separate authorized access points in different languages (all in one script) was adopted to bring together the four legal scripts of authorized and variant access points in a single authority file.

010	—	a	n 50051544	
040	—	b	mul	
		r	rda	
046	—	f	1939-05-04	
		z	edtf	
100	1	a	Oz, Amos,	
		d	1939-	
		g	lat	
100	1	a	Os, Amoc,	
		d	1939-	
		g	cyr	
100	1	a		עוז, עמוס,
		d		1939-
		g		heb
100	1	a		عوز, عاموس,
		d		1939-
		g		ara
370	—	a	Jerusalem (Israel)	
374	—	a	Author	
375	—	a	male	
400	1	a	Ouzi,	
		d	1939-	
		g	lat	
400	1	a	Ao, Zi,	
		d	1939-	
		g	lat	
400	1	a	Klausner, Amos	
		g	lat	
400	1	a	Ozs, Amoss	
		g	lat	
400	1	a	Os, Amos	
		g	lat	
400	1	a	Oz, Amosz	
		g	lat	
400	1	a	Klozner, Amos	
		g	lat	

Fig. 2: Part of the NNL10 record for Israeli author Amos Oz, showing four instances of the 100 field and use of the local subfield \$9 to distinguish between scripts.

Along with information from VIAF, Israeli databases were mined for access points and attributes. Among these sources: the Bibliography of the Hebrew Book, Imagine (database of Israeli artists), and ACUM (database of Israeli musicians).

Latin character authorized access points match Library of Congress records and Arabic and Cyrillic access points are designed to match the Latin character access point as closely as possible. Hebrew access points, on the other hand, follow the agent's preferred spelling or the best of local scholarship. Therefore there may be an imperfect match between the four authorized access points representing a single agent, but each will be the access point most recognizable to the user.

The other side of NNL10 is multilingual access LCSH. At present available in Hebrew and English, NNL10 offers the same full and equal access to subjects as it does to name authorities. In addition, NNL10 provides a translation of every subject and subdivision, precoordinates millions of subject headings, and has tables governing which combinations are legal and warnings against those which are not. Indexers need not study the complex rules of the Subject Heading Manual to accurately construct compliant subject headings.

NNL10 provides an answer to a question which “*has been recognized as an issue to be solved by the Association for Library Collections and Technical Services (ALCTS) Task Force on non-English Access as far back as 2007*” (El-Sherbini, 2016).

The problem with using LCSH for a multilingual database is well-known: LCSH reflects the American culture, vocabulary, and law, while a multilingual society is likely to be multi-cultural. In the interest of standardization it was decided to retain American decisions about general subject headings, but as in the name authority records, it was decided that Israeli headings would be subject to adaptation. Jerusalem, Israel’s capitol, is not left stateless in NNL10 as in the Library of Congress database. Israeli place names were determined by an Israeli government committee; the decision on whether to qualify a geographical location as Israel, West Bank, or Gaza Strip was based on official Israeli, and not United States, governmental policy. The pre-State geographic area which is now Israel is given the subdivision “Eretz Israel,” the Jewish term for the area, rather than the Greek term “Palestine”. Local conflicts and wars are given their Israeli names; no matter what their native language, no Israeli would think of the Six Day War as “Israel-Arab War, 1967”.

After creating authority files for local subject headings, existing partial translations of the remaining LCSH terms were imported. These translations were made over the past two decades by two Israeli universities and followed two different orthographic principles – defective and plene spelling. Batch conversion routines changed nearly all defective spelling to plene, and crowdsourced quality control is quickly finding the outliers to mark for manual correction.

Though this effort is expected from a national library, it was felt that the authority database would only truly fulfill its potential if it became a cooperative effort; not only would libraries use the information as the basis of their own local authority files, but they would install NNL10 in place of their local files. Thus, patrons searching the national union list of bibliographic records would enjoy the same full and equal access to names and subjects as they do when searching NLI’s catalogue.

As a result of outreach efforts, in 2014 two major universities, Haifa University and the Hebrew University of Jerusalem, joined the NNL10 cooperative. In 2015 four colleges (Beit Berl, David Yellin, Kinneret, and Oranim) joined, and in 2016 three more universities, Ben Gurion University of the Negev, Bar Ilan University, and Tel Aviv University, began the switch from local authority records to NNL10.

Libraries which join have a local copy of NNL10 uploaded to their servers. An automated routine written by NLI runs on the local catalogue and adds the appropriate \$9 subfield to all authority controlled fields in the bibliographic records and to 1XX, 4XX, and 5XX fields in the authority records. Another routine identifies overlaps in 1XX or 4XX between the local catalogue and NNL10. These overlaps are then converted, in the bibliographic record, to the NNL10 authorized access point in the script of the bibliographic record.

The percentage of out-of-the-box matches between local catalogues and NNL10 depends on the quality of each library’s authority work and its adherence to national and international cataloguing standards. To create more matches additional routines are run; one highly successful routine compares 1XX which are similar but not exactly the same (a name with and without dates, for example) and creates a match if the two headings share bibliographic works.

Partners receive individualized training and then are authorized to create new 1XX, 4XX, and 5XX fields for name authority records. They may not change or merge existing 1XX fields but must request such changes from NLI cataloguing department managers. Only NLI is allowed to make changes to subject heading authority files.

Working with the various institutions, NLI has managed to overcome compatibility issues between different versions of Ex Libris's Aleph program and to solve the problem of labelling authorized access points in Aleph authority records to automatically update bibliographic records in Ex Libris's Alma program via the Community Zone, thus allowing NNL10 partners who have switched to Alma to completely dispense with their own local authority files.

With over two thirds of Israeli universities now partners in NNL10, NLI decided to make the database publicly available through the national union list (http://uli.nli.org.il/F/7B9BTLNCJD6RVVU4HSK79SFV8GPM9KTC1GQE8N2B6JPSN536MF-35582?func=find-b-0&local_base=nnl10&con_lng=eng). The cooperative is in the midst of expanding the database (over a quarter million name authorities are expected to be added over 2017) and NLI is currently working on plans to release both name and subject information as linked data.

References:

- Central Bureau of Statistics. (2013). Selected Data from the 2011 Social Survey on Mastery of the Hebrew Language and Usage of Languages. Retrieved November 11, 2016, from http://www.cbs.gov.il/reader/newhodaot/hodaa_template.html?hodaa=201319017
- Cutter, C. A. (1904). *Rules for a dictionary catalog* (4th ed.). Washington: Government Printing Office. Retrieved from <https://ia800205.us.archive.org/13/items/rulesforadictio06cuttgoog/rulesforadictio06cuttgoog.pdf>
- El-Sherbini, M. (2013). Bibliotheca Alexandrina's model for Arabic name authority control. *Library Resources & Technical Services*, 57(1): 4–17.
- El-Sherbini, M. (2016). Improving resource discoverability for non-Roman language collections. In *IFLA WLIC 2016, 11-12 August 2016, Columbus, Ohio in Session Subject Access: Unlimited Opportunities*.
- Malinconico, S. M. (1985). The role of a machine based authority file in an automated bibliographic system. In M. Carpenter & E. Svenonius (Eds.), *Foundations of cataloging: A sourcebook* (pp. 211–233). Littleton, Colorado: Libraries Unlimited.
- The National Library Law (2007). Retrieved from http://web.nli.org.il/sites/NLI/English/library/aboutus/renewal/Documents/National_Library_Law.pdf