

Multi-alphabet searching at the NLI

Elhanan Adler

elhanana@savion.huji.ac.il



US-style cataloging

- All entry points in Latin alphabet
- In recent years, option to enrich the record with parallel vernacular 880 fields
- Major advantage: works in all languages and alphabets are retrieved by a single search
- Major disadvantage: Latin alphabet heading is not always obvious – particularly for Hebrew names (various romanization schemes)
- This approach is now known as MARC Model A



Israeli style cataloging

- Works in Latin, Hebrew, Arabic and Cyrillic scripts are cataloged in the original script
- Major advantage: headings are searched in the native language of the work (no romanization)
- Major disadvantage: Multiple searches necessary to retrieve material in more than one script
- This approach is now known as MARC Model B



Example –Headings for Benjamin Netanyahu

- Netanyahu, Binyamin
- Нетаниягу, Биньямин
- נתניהו, בנימין
- نتنياهو، بنيامين •



The solution: retain separate alphabets but enable cross-alphabet searching

- In the NLI OPAC headings: currently 4 scripts
- In NLI catalog subjects: currently English only (LCSH)
- In the Index to Articles in Jewish Studies (RAMBI): currently three scripts for bibliographic data and 2 scripts for subjects
- The goal: searching any of the scripts will retrieve headings using all of them

How?

- Create single authority record for each heading with multiple 1xx fields (one for each script), e.g.
 - 1001 \$aNetanyahu, Binyamin,\$9eng
 - 1001 \$аНетаниягу, Биньямин, \$9rus
 - 1001 \$מתניהו, בנימין\$\$9heb
- Each can have its own cross references
 - 4001 \$aNetanyahu, Benjamin,\$9eng
 - 4001 \$3,נתניהו, ביביא \$9heb
- Bibliographic record contains only one form (in alphabet of cataloging) but can be retrieved by all.



Software support

- This solution is non-standard (MARC field 1xx is non-repeatable)
- It is supported by ALEPH 500 software
- It is used by other ALEPH libraries (e.g. Swiss for French/German/Italian headings)



How to locate/merge matching records

- NLI names start with VIAF clusters
- RAMBI subjects translation + locate parallel term (if exists)
- LCSH create Hebrew terms starting with existing Hebrew subject thesauri (translation only, no merge). Start with core collection subject areas (Judaica, Israelitica, Middle East)



The world's libraries.

Search:			Go
---------	--	--	----



Worldwide (English) Change

Products and Services

Research

About OCLC

Public Purpose

News and Events

Contact us

Making WorldCat better

VIAF

History

WorldCat

- Governance
- Contributors
- How to become a contributor

Search VIAF

Access recent presentations

Subscribe to OCLC updates and offers

Subscribe to OCLC distribution and discussion. lists

VIAF

Membership

VIAFTM

Virtual International Authority File

At a glance

- A collaborative effort between national libraries and organizations contributing name authority files, furthering access to information.
- All authority data for a given entity is linked together into a "super" authority record
- A convenient way for the library community and other agencies to repurpose bibliographic data produced by libraries serving different language communities

News

 Virtual International Authority File service transitions to OCLC; contributing institutions continue to shape direction through VIAF Council

More news >>

The Virtual International Authority File (VIAF) is an international service designed to provide convenient access to the world's major name authority files. Its creators envision the VIAF as a building block for the Semantic Web to enable switching of the displayed form of names for persons to the preferred language and script of the Web user. VIAF began as a joint project with the Library of Congress (LC), the Deutsche Nationalbibliothek (DNB), the Bibliothèque nationale de France (BNF) and OCLC. It has, over the past decade, become a cooperative effort involving an expanding number of other national libraries and other agencies. At the beginning of 2012, contributors include 20 agencies from 16 countries.

Most large libraries maintain lists of names for people, corporations, conferences, and geographic places, as well as lists to control works and other entities. These lists, or authority files, have been developed and maintained in distinctive ways by individual library communities around the world. The differences in how to approach this work become evident as library data from many communities is combined in shared catalogs such as OCLC's WorldCat.

VIAF helps to make library authority files less expensive to maintain and more generally useful to the library domain and beyond. To achieve this, VIAF matches and links the authority files of national libraries and groups all authority records for a given entity into a merged "super" authority record that brings together the different names for that entity. By linking disparate names for the same person or organization, VIAF provides a convenient means for a wider community of libraries and other agencies to



The Virtual International Authority File (VIAF) is an international service designed to provide convenient access to the world's major name authority files. Its creators envision the VIAF as a building block for the Semantic Web to enable switching of the displayed form of names for persons to the preferred language and script of the Web user. VIAF began as a joint project with the Library of Congress (LC), the Deutsche Nationalbibliothek (DNB), the Bibliothèque nationale de France (BNF) and OCLC. It has, over the past decade, become a cooperative effort involving an expanding number of other national libraries and other agencies. At the beginning of 2012, contributors include 20 agencies from 16 countries.



VIAF uses Worldcat data including titles and 880 fields to identify various forms of names

VIAF Virtual International Authorit	y File				
Search					
Select Field:	Select Index:	Search Terms:			
All Headings	▼ All VIAF ▼	netanyahu	Search		
BENJAMIN NETANYAHU DURABLE PEACE ISRAEL AND ITS PLACE AMONG THE NATIONS	Netanyahu, Binyamin 🍱 🚝 Netanyahu, Benjamin 🕮 Netanyahou, Benjamin 1949 🗊 💶 1949 , בנימין, בנימין פוחים און איז פוחים ביימין פוחים און איז פוחים ביימין פוחים און איז פוחים אי				
	VIAF ID: 114192441 (Personal) Permalink: http://viaf.org/viaf/114	¥192441			



VIAF can supply source record IDs for all clusters Netanyahu cluster

111012494 BIBSYS|x90170933|| BNE|XX1184561|| BNF|11893234||
DNB|111070414|| LC|n 81043039|| NDL|00433976|| NLA|000035020651||
NLIara|000261992|| NLIIat|000023475 || NUKAT|n 98007803||
PTBNP|216008|| SUDOC|026742985

114192441 BIBSYS|x00022614|| BNF|12526153|| DNB|119479575|| LC|n 78049769|| NDL|00657354|| NKC|js20050815012|| NLA|000035172521|| NLIara|000174142|| NLIcyr|000154468|| NLIheb|000182292|| NLIlat|000098960|| NUKAT|n 2009147482|| SUDOC|034512438

116848601 BIBSYS|x90826611|| BNF|13483859|| DNB|118860097|| EGAXA|vtls000795786|| LC|n 83043340|| NLIara|000001709|| NLIIat|000016290|| SELIBR|32625|| SUDOC|075936267 117444950



Merging

- Clustered authority records will be merged into one
- Further clustering will be done in-house, automatically if possible, manually based on likelihood that name appears in more than one alphabet in bibliographic records (major authors, historical figures, etc.)



Example from NLI test server

Leager	LDR ^^^^nz^^a22^^^^n^^4500
Control No.	001 000098960
Date and Time	005 20090727161503.0
Fixeo Data	008 000108 a aab
Personal Name	100 1 a Netanyahu, Binyamin
Seet.Trac.Per.N	400 1 a Netanjahu, Benjamin
Seet.Trac.Per.N	400 1 a Bibi
Seet.Trac.Per.N	400 1 a Netanyahu, Benyamin
Seet.Trac.Per.N	400 1 a Netanyahu, Benjamin
Public Gen.Note	680 <u>a</u> LCN note: b. 1949
VIAF cluster	901 <u>a</u> 114192441

Four separate authority records before merge

8	Leager	<u>LDK</u>	~~~~~nz~~azz~~~~n~~4500
q	Control No.	001	000154468
ı	Fixeo Data	008	020116 a aab
	Personal Name	100 1 a	Нетаниягу, Биньямин
ı	Seet.Trac.Per.N	400 1 a	Нетаньягу, Биньямин
å	Seet.Trac.Per.N	400 1 a	Нетаниягу, Беньямин
ī	Seet.Trac.Per.N	400 1 a	Нетаньяху, Биньямин
4	Public Gen.Note	680 a	LCN: Netanyahu, Binyamin; note: b. 1949
ı	Public Gen.Note	680 a	БЭС,1997: Нетаньяху, Биньямин
1	VIAF ciuster	9 <mark>01 a</mark>	114192441
Ē	Leager	LDR ^^	^^^nz^^a22^^^^n^^4500

000182292

008

100

Control No.

Fixeo Data

Personal Name

נתניהו, בנימין, -1949

באנגלית: NETANYAHU , BENJAMIN

```
Nonpub.Gen.Note
                      667
                      901
                                a 114192441
VIAF cluster
                                 ^^^^nz^^a22^^^^n^^4500
                     LDR
Leager
                     001
Control No.
                                 000174142
                     005
                                 20060502122150.0
Date and Time
                     008
                                050523|||a||||aab||||||||||||||||||
Fixeo Data
                     \overline{100} \overline{1}
Personal Name
                     400
Seet.Trac.Per.N
                              a 114192441
VIAF cluster
```

000108|||a||||aab||||||||||||||||||

نتنیاهو، بنیامین (دردددرر حددهرر) نیتانیاهو، بنیامین



After merge

	7 (10) 1110190	
Leager	LDR ^^^^nz^^a22^^^^nn^^4500	
Control No.	001 000098960	
Date ano Time	005 20120521143735.0	
Fixeo Data	008 000108 a aab	
Personal Name	100 1 a Netanyahu, Binyamin	
	<u>9</u> eng	
Personal Name	100 <u>1 a</u>	נתניהו, בנימין,
	$\frac{\mathbf{d}}{\mathbf{d}}$	1949-
	100 t ⁹	heb
Personal Name	100 1 <u>а</u> Нетаниягу, Биньямин	
Personal Name	9 rus	(2)(2)27 (2)272)
Personal Name	$\frac{100}{1} \frac{1}{9}$	نتنیاهو، بنیامین (נתניהו, בנימין) ara
Seet.Trac.Per.A	400 1 a Netanjahu, Benjamin	ai a
Jeen des et sv	9 eng	
Seet.Trac.Per.N	400 0 a Bibi	
Seet.Trac.Per.N	400 1 a Netanyahu, Benyamin	
Seet.Trac.Per.N	400 <u>1</u> <u>a</u> Netanyahu, Benjamin	
	$\frac{9}{2}$ eng	
Seet.Trac.Per.N	<u>400</u> <u>1</u> <u>а</u> Нетаньягу, Биньямин	
Cool Tree Cor A	9 rus	
Seet.Trac.Per.A	400 <u>1 а</u> Нетаниягу, Беньямин 9 rus	
Seet.Trac.Per.A	400 1 а Нетаньяху, Биньямин	
Seemingen en av	9 rus	
Seet.Trac.Per.N	400 1 <u>a</u>	نيتانياهو، بنيامين
	$\frac{1}{2}$	ara
Nonpub.Gen.Note	667 Viaf cluster heading merged from record 000182292	
Nonpub.Gen.Note	667 <u>a</u>	באנגלית: NETANYAHU, BENJAMIN
Nonpub.Gen.Note	667 Viaf cluster heading merged from record 000154468	
Nonpub.Gen.Note	Viaf cluster heading merged from record 000174142	
Public Gen.Note	680 <u>a</u> LCN note: b. 1949	
Public Gen.Acte	680 a LCN: Netanyahu, Binyamin; note: b. 1949	
Public Gen.Note	680 <u>а</u> БЭС,1997: Нетаньяху, Биньямин	
VIAF ciuster	<u>901</u> <u>a</u> 114192441	



Browse index – before merge (10+11+2+11 records)

Browse List: LC Subjects

No. of Records

Entry

Next Page >

10

Netanyahu, Benyamin - [References] See: Netanyahu, Binyamin

Netanyahu, Binyamin - [References]

Browse List: LC Subjects

No. of Records

1

11

Entry

Entry

< Previous Page

< Previous Page

Next Page >

נתניה. קרית צנז

נתניהו, בנימין, References] - -1949

נתניהו, יונתן, References] - 1946-1976

Browse List: LC Subjects

No. of Records

Hетаниягу, Беньямин - [References] See: Нетаниягу, Биньямин

2

Нетаниягу, Биньямин - [References]

< Previous Page

Next Page >

Browse List: LC Subjects

No. of Records Entry

< Previous Page Next Page >

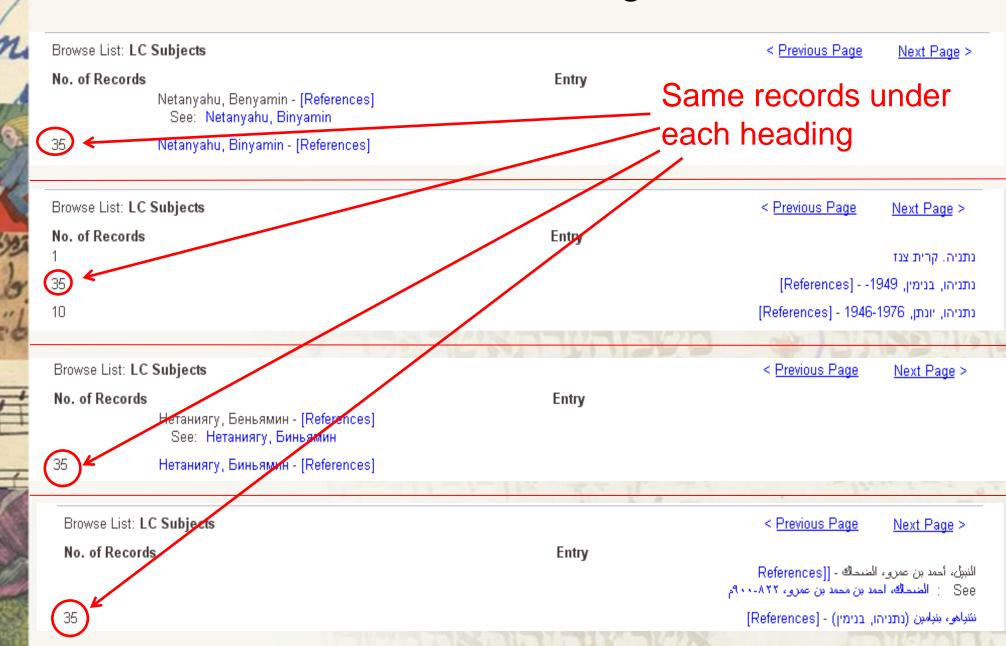
النبيل، أحمد بن عمرو، الضحاك - [[References : مرو، ۸۲۲ - ۹۰ م

نتتياهر، بنيامين (נתניהו, בנימין) - [References]

11)



Browse index – after merge





#	Author		Title	Year	Lib/Items
1 □			éveil du faucon : portrait ique de Benjamin hou /	2011	All copies
2		כתב גלוי	ל תלך בדרך האיוולת של אוסלו :מ לבנימין נתניה		All copies
3		קורות	שיאי המדינה וראשי: הממשלה [ערכה] :ערכת תצלומים,		All copies
4			שינוי עמדות מנהיגים בסכסוך קיומי מורכב :בנימין נתני		All copies
5			הפוליטיקאי החדש בזירה המערבית בישראל :קלינטון, בליי		All copies
6		- אל	סגנון פוליטי של ראשי ממשלה בישר מנחם בגין, יצחק רב		All copies
7			с премьер-министрами : Рабин, Шимон Перес,	2001	All copies
8		ממשלה	נית-ספר לראשי-ממשלה :10 ראשי- שראליים, הסיפור האי		All copies
9		כה]ערכת	שיאי המדינה וראשי הממשלה [ער: דצלומים ועלונים		All copies
10		Netanya	hu :the road to power /	1999	All copies
11 🗆			cult road to peace :Netanyahu, d the Middle East peace /	1999	All copies
12 \Box		Netanya	hu's legacy /	1999	All copies
13		Verhältn	elisch-palästinensische is seit dem Regierungsantritt n Netanyahus [elec	1999	All copies
14		Нетания выпуск	гу Ликуд :информационный	1999	All copies
15			/ האיש שניצח את עצמו	1999	All copies
16		99' וירות	קסם על שפתי מלך :ישראל ערב בח י	1999	All copies
17 \square		Т	שיאי המדינה וראשי הממשלה :אוג: דצלומים ועלונים.		All copies
18			لسلام الضبائح :نتنباهو بحرئ واشنطن /	1998	All copies
19			us of Jewish settlements in stinian territories during the	1997	All copies
Library Info	Preferences	Feedback	Expert Search Or	ıline order	s help



Implementation stages

- RAMBI subjects: done
- NLI catalog names: planned Summer 2012
- NLI catalog subjects: planned Winter 2012
- N.b. "Implementation" means technical changes made and linking of headings underway
- Different procedures for each implementation



RAMBI subjects: background

- Hebrew subjects for Hebrew articles, English subjects for English articles
- Subjects assume Jewish content (e.g. "Paris, France" means "Jews in Paris, France)
- Not all subject headings existed in both languages
- Not same level of detail in both languages



RAMBI subjects: linking - 1

- Change format to LCSH-style subfields (a/x/v/y/z)
- Create authority records for <u>all</u> subjects (including compound with subdivisions)
- Identify most frequent headings and subdivisions and create translation table:
 - Holocaust = שואה
 - Bibliography = ביבליוגרפיה
- Using the table, search for matching headings, e.g.
 \$\$aHolocaust\$\$vBibliography and \$\$aאואה \$\$vביבליוגרפיה \$\$
- Merge identified parallel headings into one authority record



RAMBI subjects: linking - 2

- For remaining headings:
 - Produce lists by frequency of term and check manually
 - Many terms appear only in 1 language
 - For many terms translation is probably superfluous, e.g.
 - Coverdale Bible
 - Journal of Jewish Studies (periodical)
 - Buffy, the Vampire Slayer (TV-series)

and a mechanism exists to flag them as not needing translation

If you were wondering...

"That Was Nifty": Willow Rosenberg Saves the World in *Buffy the Vampire Slayer*

Matthew Pateman University of Hull

The extent to which ethnicity permeates an understanding of identity in Buffy the Vampire Slayer is apparent from the ways in which species difference stands in for racial or ethnic difference. However, among its many points of contact with ethnicity, one that is especially curious is the case of Willow Rosenberg's disappearing Jewishness. As a character, she shifts from nerd, to poster-child for geek-chic, she suffers from major addiction, is a lesbian and ends up as something approaching a goddess. She is also, intermittently, Jewish. The paper encourages a reading of Willow that sees her as cool and occasionally as Jewish but not necessarily as Jewish cool. To that extent, Buffy the Vampire Slayer can be read as an index of shifting sensibilities in relation to representations of Jews in popular culture whereby old stereotypes and perceptions are largely ignored, but there are not yet the necessary store of images and discourses available for young Jewish womanhood.

Willow Rosenberg (Alison Hannigan), in the very last episode of Buffy the Vampire Slayer (WB/UPN, 1997–2003), saves the world. She does so by summoning her extraordinary magical powers and performing a spell of such magnitude that it transforms all of the girls and young women who might have become the Slayer on the death of the previous one into a Slayer now. In other words, it alters millennia of (super)natural law and lore and, in the process, allows Buffy (the current vampire slayer, Sarah Michelle Gellar) to have an army strong enough to defeat the First Evil. As a by-product, Willow glows in translucent white light, seemingly becoming a goddess, and she comments in happy, exhausted satisfaction at the conclusion of all of this that it was "nifty."

^{1&}quot;Chosen" 7.23.



RAMBI subjects: linking - 3

- As of June 1, 2012:
 - Paired subjects: 10,327
 - Single subjects not needing translation: 1,991
 - Still unresolved subjects: 8,211 (632 topical, 1,435 title, 6144 geographic)
 - Total records in RAMBI: 321,274
 - Total subject headings in RAMBI: 718,671
 - Of those, unresolved: 20,654 (2.9%)



RAMBI – future plans

- Multi-alphabet author searching (linking names to NLI authority file)
- Convert RAMBI subjects to bilingual LCSH for single search in NLI/RAMBI
- Both dependent on next 2 projects



NLI catalog - Names

- Priority names which are in multilingual use
- Start with clustering done by VIAF (previous examples)
- Identify high frequency names and merge or add alternate a/b form
- Identify names that exist in multiple scripts (e.g. Hebrew literature in translation)
- Do not create headings not justified by publications
- Project will begin this summer



NLI Subject Catalog

- In 2010 the NLI switched from a classified catalog (unique classification based on Dewey) to LCSH.
- Currently (June 2012) NLI catalog contains
 1.98m records (all types of materials)
- 1.47m of these have one or more subject headings (LCSH or LCSH-style)
- [some collections still have unique subjects, these are being gradually converted to LCSH]



But these are in ENGLISH!







Slide from presentation at AJL 2010

Why not Hebrew?

- Creation of a Hebrew subject thesaurus for a large research library (even based on existing thesauri) is a major, expensive professional task and would take years
- It is unlikely other Israeli academic libraries would adopt it (another unique system!)
- It would not serve non-Hebrew speaking scholars (the Library has a universal mission)
- Maybe just add Hebrew cross references to LCSH headings?



NLI LCSH translation project

- Use multi-alphabet authority functionality to create (selectively!) parallel Hebrew terms to LCSH subjects.
- Indexing will remain in English only
- Primary goal, to cover the NLI's 3 core areas: Judaica, Israel, Near East
- Project to begin before the end of 2012



Sources of Hebrew Terminology for LCSH

- Bar-Ilan thesaurus (Bar-Ilan uses LCSH for Latin a/b materials & Hebrew terms for Hebrew materials with see-alsos)
- Other Israeli Hebrew thesauri (Haifa Index to Periodicals (IHP), Szold Institute database, etc.
- Commercial translation services
- National advisory committee
- Translate unique headings and subheadings.
- Use RAMBI routines to translate compound headings



Non-LCSH subjects at the NLI

- Maps and Scholem collections converted to LCSH
- Archives collection
- Music/National Sound Archives collections (extreme detail of Judaica and musical traditions) – conversion has begun
- Manuscripts (NLI + Institute of Microfilmed Hebrew Manuscripts) - unique subjects
- RAMBI very different from LCSH. Jewish context assumed. Geographic facet primary.
- UN and EU publications (deposit collections) use collection thesauri



Some other technical projects underway or in planning

- Assign LCC numbers to some public collections for shelving (Edelstein collection completed, General Reading Room and Bibliography collection being considered)
- Integrate (link or merge) the Bibliography of the Hebrew Book (BHB) into the NLI public catalog (Authority information, bibliographic records). Major challenge – very different orthography.
- Linking either the NLI catalog or the BHB to hebrewbooks.org
- The NLI catalog as the national bibliography.
- New user interface.