

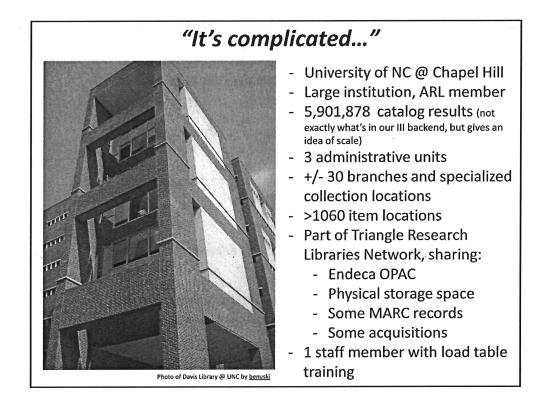
Not only is our catalog a bricolage (with data from all manner of sources), the act of building the catalog today is itself bricolage, using a variety of tools and strategies.

Today I am going to talk about some of the more useful catalog bricolage techniques that we have developed in my time at UNC-Chapel Hill.

А тар
A bit of context – my institution and my role in it
My favorite load table – gathering bib records
Extended example of "data bricolage" for cleaning/enhancing bib records
Script to verify full text access to ebooks
Script/program to summarize data exported from Millennium

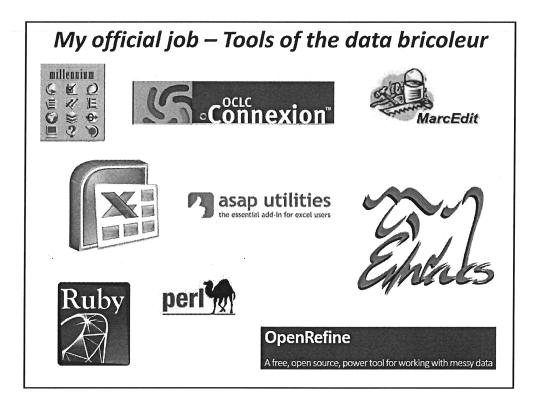
So, being more or less well-versed in all of those tools, and some more that I've either forgotten or neglected to mention, I've fallen into another role in my organization---helping with different types of problems throughout the organization.

The workflows and solutions I'll talk about today have come out of these special situations as well as my "official" work.



My official job – E-resources cataloger Managing & loading batches of MARC records for ebooks Individual cataloging of Web sites, online databases, and some ebooks Oversee maintenance of URLs in catalog records (new!) Extraction of our catalog data from Millennium for use in our Endeca OPAC

I have been at UNC Libraries since January 2010, in my current position, and I am still learning how all the local pieces work together.



The usual and expected suspects: Millennium and OCLC Connexion need no further explanation.

The free MarcEdit application, by Terry Reese, is of course indispensable for anyone working with batches of MARC records.

Microsoft Excel, while not typically thought of as a cataloging tool, is where I spend much of my time.

- Data from Millennium Create Lists is analyzed and worked on from Excel.
- Combining data from multiple sources, simple duplicate identification, title list and MARC record reconciliation, etc.
- MarcEdit outputs data to Excel and allows you to create MARC records from Excel.
- Spreadsheets serve as the input files and output formats of most of the scripts I write

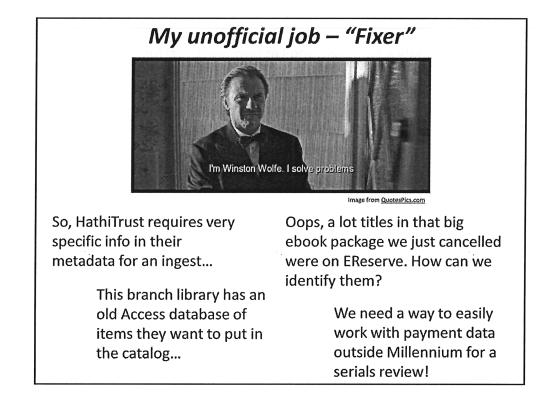
ASAP Utilities (Libraries qualify for the free edition) makes a life lived in Excel go much more easily.

Emacs is technically a text editor (in the sense that the Incredible Hulk is a strong man), but it can be used for much more than that. For cataloging, I mainly use it to search and manipulate MarcEdit Mnemonic Text files in ways MarcEdit won't let me do.

Ruby is my go-to scripting language, and I'm now using Perl for our catalog data extraction scripts.

OpenRefine (was Google Refine) is something I've used less often, to clean up and transform really messy data. The situations in which it's useful are more rare for me, but it has been a lifesaver when I needed it.

(All of the images are links, if you want to find out more about any of these tools.)



So, being more or less well-versed in all of those tools, and some more that I've either forgotten or neglected to mention, I've fallen into another role in my organization---helping with different types of problems throughout the organization.

The workflows and solutions I'll talk about today have come out of these special situations as well as my "official" work.

gathering, cleaning, & enhancing records MY FAVORITE LOAD TABLE

Movembry of 1996 beach to durant of conduction of the term of characterial and a search effect extrinsion and the new avgramment of the terms of graned conductor for the term of the term of the case of the grant for the Contexts with the terms

aar areesta joo Armaan aree www.co. damenia in soon a awaan aareeta mooro a areeta areeta areeta areeta areeta waan waar ista di beskaan joo ahaan ya officia areen officia eeska siifa eeskaan oo waan 1000 ameen wifed. Beshtuu di diibeed waari oo eeskaa

> a akti barta, Mercíka (h. 1936) ereka konstanto ereka erekti bilaka (h. 1996) erek elektrolaka (h. 1997) 1969 - Anarak Jahor Karisha anton ar eserek eremendi erekteti erekteti anton (h. 1917).

ison to site. Case no to success on the treather body rest for burnet (Case aprecident et aus *reak** base** server on,

BACKGROUND: a pre-ex	xisting workflow
Spreadsheet from Internet Archive Scribe IDENTIFIER TITLE CALL_NUMBER VOLUME hespokenowlhey He spoke, now they spead b2058719 agriculturaladdr0(Agricultural address of Col b2409093 extractsfromorigCExtracts from original lette b2564550 soulcitynorthcar0 Soul City North Carolina b2709614	Ų
Spreadsheet >> MarcEdit Delimited Text =LDR 00000nam 22000001a 45e0 =530 \\\$aAlso available via the World Wide Web. =856 41\$yFull text available via the UNC-Chapel Hill LibrariesSuhti	
=907 \\\$a.b2058719a =949 \1\$g1\$lebnb\$h0\$m\$t11\$u-\$jOCA electronic book	
=LDR 00000nam 2200000a 45e0 =530 \\\$aAlso available via the World Wide Web. =856 41\$yFull text available via the UNC-Chapel Hill Libraries\$uhtt =907 \\\$a.b2409093a =949 \1\$g1\$lebnb\$h0\$m\$t11\$u-\$jOCA electronic book	p://archive.org/details/agriculturaladdr00cann

Г

Years ago, UNC became home to multiple Internet Archive Scribe scanning machines and began contributing digitized works to Internet Archive as part of the Open Content Alliance.

Hundreds, and sometimes over a thousand items scanned each month. Our public services staff believed the print bib record for each title scanned should contain a link to the digitized version online.

Thankfully, the Scribe manager was able to provide technical services with a spreadsheet associating a bnum with each digitized item and its URL.

MarcEdit's Delimited Text Translator function was used to turn that spreadsheet into "stub" MARC records.

BACKGROUND: A pre-existing workflow

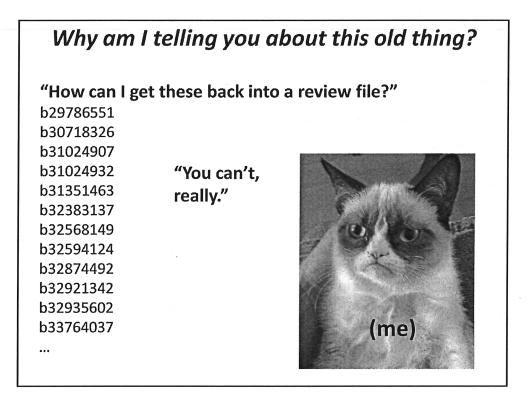
=LDR 00000nam 2200000la 45e0
=530 \\\$aAlso available via the World Wide Web.
=856 41\$yFull text available via the UNC-Chapel Hill Libraries\$uhttp://archive.org/details/hespokenowtheysp00yate
=907 \\\$a.b2058719a
=949 \\\$g1\$lebnb\$h0\$m\$t11\$u-\$jOCA electronic book
=LDR 00000nam 2200000la 45e0
=530 \\\$aAlso available via the World Wide Web.
=856 41\$yFull text available via the World Wide Web.
=856 41\$yFull text available via the UNC-Chapel Hill Libraries\$uhttp://archive.org/details/agriculturaladdr00cann
=907 \\\$a.b2409093a
=949 \\\$g1\$lebnb\$h0\$m\$t11\$u-\$jOCA electronic book

Compiled to .mrc and loaded with locally-created load table that:

- Matches on bnum (907) for overlay
- Protects ALL fields in existing record (LDR, Cat Date, etc... everything)
- Inserts any fields from the new stub record (will create dupe fields)
- Creates new item

The stub MARC records were loaded using a locally-created load table with the above qualities.

This resulted in an efficient workflow for adding URLs to many hundreds of bibliographic records each month.

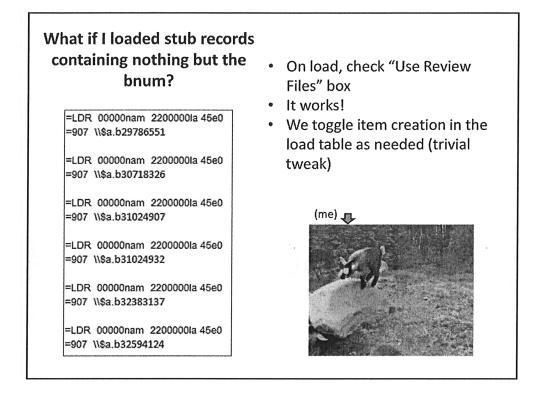


I didn't work in my position very long before I went to a colleague with that common question: how do I create a review file from a list of bnums?

The answer was, of course, "That's not a supported function."

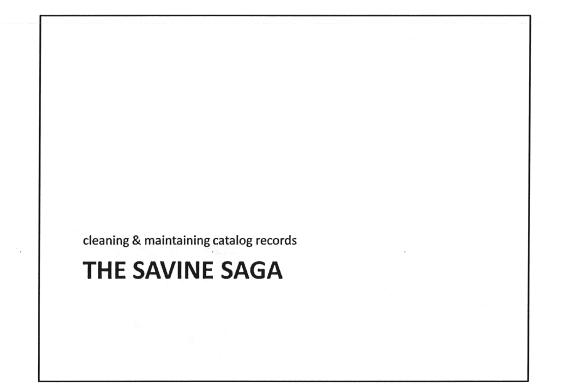
•••

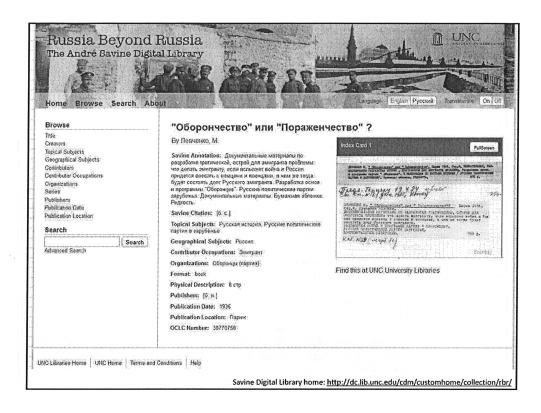
I've since learned of some other workarounds for this using Expect scripts and so on, but here is what dawned on me the next time I processed a batch of the Internet Archive/OCA records...



After I began thinking outside the established workflow for this particular load table, we began using it for all kinds of things. The next section of my talk illustrates another use.

I understand that not all III libraries have the ability or desire to create and edit their own load tables. But I have wondered if there is some generalized version of this load table that III could make available in order to allow much greater flexibility in gathering, correcting, and enhancing records in our catalogs.



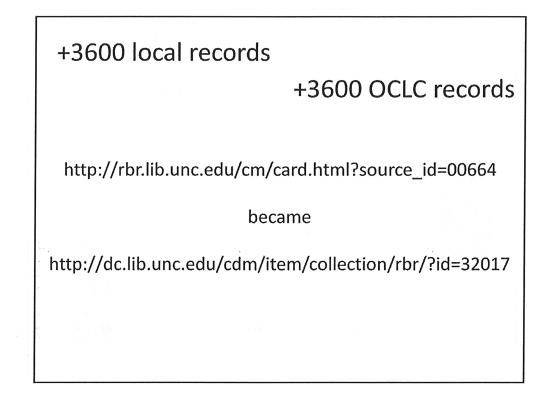


André Savine was the owner of an antiquarian bookstore in Paris specializing in materials relating to the Russian Diaspora. In 2002, UNC acquired the André Savine collection, which includes the stock of the bookstore and Savine's extensive private collection dealing with Russian emigration.

In 2008, the "Russia Beyond Russia Core Module" database of Savine's extensive bibliographic research--thousands of index cards and notebook pages--was completed.

MARC Leader		1467 1a 4500		Local
LOCATIONS b	,wc			Millennium
Bib Utility No. Misc. Misc. LCCN Misc. Misc. Misc. Misc. Misc. Misc. Misc. Author This Publication Info. Description Note	003 OC 005 20% 008 80% 010 3 035 00% 040 DL 043 e-u 050 4 050 4 050 4 050 4 050 10 245 1 200 89 300 89 300 9	17 ICS 93351 b,M5 N 180-01 aMTTakk 180-03 aMTTakk 180-03 aBerlin; [1] p. [c20 cm. [p. [4] of cover; 1	w 000 Obrusr doCLIdOCLCQIdOCLCG vin, V. A.(qVenediki Alaksandrovich).(d1867-1937. shkin i dekabristy/JcVA. MTTakolin. IoVataga i JaPraga : joPlamTTa (c1923)a(Berlin :)oTip. Kummer i Ko) zdanie "Vataga" i "PlamTTa", Berlin, Praga.	Record
Note LC Subject LC Subject Added Author Misc. Misc. Misc. Misc. DRA No. Misc.	600 1 0 668 651 0 Ru 710 2 And 856 4 2 3R http://www.com/second/s	100 1 100 1 245 1 0 245 1 0 260 1 260 1	ournale - History Philosophy Theology (S15) Мяхотин, В. А <u>М'а'Nokin, V. A</u> +q <u>(Venedikt Aleksandrovich)</u> , +d <u>1867-1937</u> А.С. Пушин и декабристы / +с В.А. Мяхотин. А. S. Pushkin i dekabristy. Берлин : +b Barara ; +a Прага : +b Пламя, +с 1923 +е (Berlin : +b Тип. Куммер и Ко) Berlin +b [Продгайі'а" Киттег i ко] +с 1923. 89. (11), +с 19 //2 cm.	
79(20).	<i>aut</i> (0)	500 500 500 500 500 1 0	in Cyrillic characlers. At head of title: V. A. Mi'a`kotin. On p. [4] of cover: Izdanie "Valaga" ("Plami'a"", Berlin, Praga. Пушкин, Александр Сергеевич, +d 1799-1837.	
	OCLC Master Record	651 0	Eushkin, Aleksandr Sergeevich, ±d <u>1799-1837.</u> Russia ±x History ±y <u>December Uprising, 1825.</u> ±3 Related electronic resource: Russia Beyond Russia Core Module (RBR-CM) based on An http://thr.lib.unc.edu:80/cm/card.html?source_id=05378	dré Savine's research 1

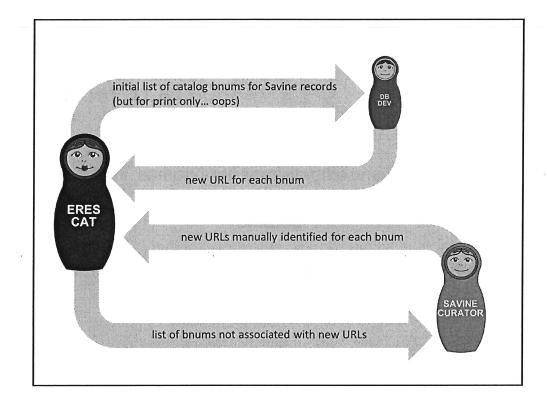
Where the Core Module database entries described materials held by UNC, a link to the database entry was included in the catalog record. These links were also added to the corresponding master records in OCLC WorldCat.



These links were added to over 3000 records.

And then in 2012, the Core Module database was redesigned and became the Savine Digital Library.

All the database entry URLs changed.



Luckily, this was not completely news to me. A couple of months earlier, the new database developer had requested a list of bnums of Savine collection items. At the time, providing him with two bnums per title (records for print and digitized) sounded like it would be problematic, so I only provided print bnums.

The developer provided me with a spreadsheet of all the bnums I'd provided and the new URL for each. It was then that I realized the folly in not sending him e-bnums.

Luckily, Excel came to the rescue, as I'll show in the next slide.

For the items that I could not match up with a new URL, the Curator of the Savine Collection provided the new URL, and I merged that into my spreadsheet.

- Create review file of all bib records with 856 matching old db URL

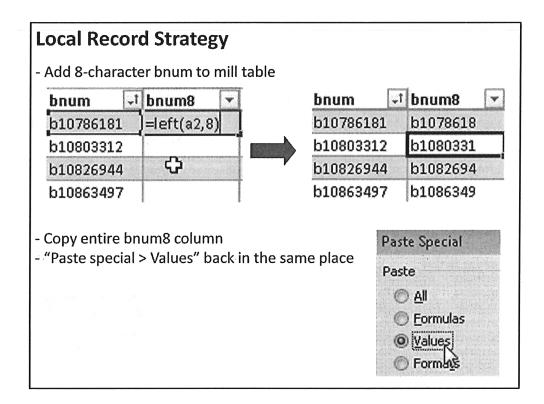
- Export data from Millennium/open in Excel... (table name = mill)

£	Successing		generating		Summer		general g		
bnum	-1	oclcnum	*	blvl	*	MAT T	¥	856	u

		2	Andread Andread Andread Andread	3 · · · · 3 · · · · · · · · · · · · · ·
b10786181	295324	m	print	http://rbr.lib.unc.edu/cm/card.html?source_id=03642
b10803312	302595	m	print	http://rbr.lib.unc.edu/cm/card.html?source_id=00854
b10826944	312034	m	print	http://rbr.lib.unc.edu/cm/card.html?source_id=00982
b10863497	326011	m	print	http://rbr.lib.unc.edu/cm/card.html?source_id=04588
b10983430	367020	m	print	http://rbr.lib.unc.edu/cm/card.html?source_id=03384
b10999103	372388	m	print	http://rbr.lib.unc.edu/cm/card.html?source_id=00606
b11005890	22261636	m	print	http://rbr.lib.unc.edu/cm/card.html?source_id=05363

- New worksheet w/new DB info (table name = contdm)

catalog_id	y url	
b1080331	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=31495	
b1082694	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=19923	
b1086349	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=18694	Hmm these
b1098343	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=21187	bnums won't
b1099910	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=21873	
b1100589	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=17388	match
b1102608	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=28027	
b1127652	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=23409	



mili tai	ole (some	colui	nns hi	idden)
А			В		G
bnum	. ↓	bnum	8 💽	NEW	URL
b1078618	31	b1078	618		#N/A
b10803312 b1080331		http:/	//dc.lib.unc.edu/cdm/item/collection/rbr/?id=31495		
b1082694		b10820			//dc.lib.unc.edu/cdm/item/collection/rbr/?id=19923
				http:/	//dc.lib.unc.edu/cdm/item/collection/rbr/?id=19923
			694	http:/	//dc.lib.unc.edu/cdm/item/collection/rbr/?id=19923 B
	14	b10826	694	A og_id 🖓	//dc.lib.unc.edu/cdm/item/collection/rbr/?id=19923 B
b1082694	14	b10826	694 catal	A og_id	//dc.lib.unc.edu/cdm/item/collection/rbr/?id=19923 B url

- Identify pattern in missing new URLs

MATT	2856 u	NEW URL
print	http://rbr.lib.unc.edu/cm/card.htm ?source_id=00019	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=34559
ebook	http://rbr.lib.unc.edu/cm/card.htm ?source id=00019	#N/A
print	http://rbr.lib.unc.edu/cm/card.html?source_id=00020	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=34638
print	http://rbr.lib.unc.edu/cm/card.html?source_id=00022	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=31502
print	http://rbr.lib.unc.edu/cm/card.htm ?source_id=00035	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=34607
ebook	http://rbr.lib.unc.edu/cm/card.htm ?source_id=00035	#N/A
print	http://rbr.lib.unc.edu/cm/card.html?source_id=00036	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=19394

- Create new table (name = urlmatch)

OLD URL	. †	NEW URL
http://rbr.lib.unc.edu/cm/card.html?source_id=00001		http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=22832
http://rbr.lib.unc.edu/cm/card.html?source_id=00003		http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=23896
http://rbr.lib.unc.edu/cm/card.html?source_id=00004		http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=20896
http://rbr.lib.unc.edu/cm/card.html?source_id=00007		http://dc.lib.unc.edu/cdm/item/cpllection/rbr/?id=19960

- In mill table, clear out NEW URL column

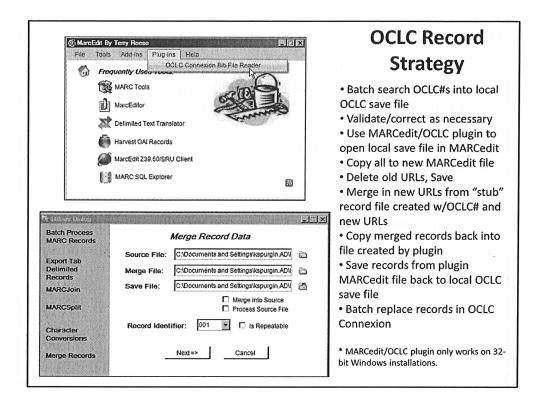
MATT	856 u	NEW URL
print	http://rbr.lib.unc.edu/cm/card.htm ?source_id=00019	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=3455
ebook	http://rbr.lib.unc.edu/cm/card.htm ?source_id=00019	#N/A
print	http://rbr.lib.unc.edu/cm/card.html?source_id=00020	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=34638
print	http://rbr.lib.unc.edu/cm/card.html?source_id=00022	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=3150
print	http://rbr.lib.unc.edu/cm/card.htm ?source_id=00035	http://dc.lib.unc.edu/cdm/item/collection/rbr/?id=3460
ebook	http://rbr.lib.unc.edu/cm/card.htm ?source_id=00035	#N/A
print	has 1/4 when a data the deal has 12 and a data the	the state of the second state of the state o
	http://rbr.lib.unc.edu/cm/card.html?source_id=00036	
MATT	Ţ	at NEW URL
MATT -	Ţ	J NEW URL
orint	856 u	JI NEW URL
orint ebook	856 u http://rbr.lib.unc.edu/cm/card.html?source_id=0001	9 9
orint ebook orint	856 u http://rbr.lib.unc.edu/cm/card.html?source_id=0001 http://rbr.lib.unc.edu/cm/card.html?source_id=0001	-1 NEW URL 9 9 0
orint 2book orint orint	856 u http://rbr.lib.unc.edu/cm/card.html?source_id=0001 http://rbr.lib.unc.edu/cm/card.html?source_id=0001 http://rbr.lib.unc.edu/cm/card.html?source_id=0002	-1 NEW URL 9 9 0 2
	856 u http://rbr.lib.unc.edu/cm/card.html?source_id=0001 http://rbr.lib.unc.edu/cm/card.html?source_id=0001 http://rbr.lib.unc.edu/cm/card.html?source_id=0002 http://rbr.lib.unc.edu/cm/card.html?source_id=0002	-1 NEW URL 9 9 0 2 5

	1999 1999 1999 1999 1999 1999 1999 199	*****	**************************************
C	D		
MAT T	▼ 856 u	-1	NEW URL
print	http://rbr.lib.unc.edu/cm/card.html?source_id=0	0019	=VLOOKUP(d2, urlmatch, 2, FALSE)
ebook	http://rbr.lib.unc.edu/cm/card.html?source_id=0	0019	
print	http://rbr.lib.unc.edu/cm/card.html?source_id=0		
print	http://rbr.lib.unc.edu/cm/card.html?source_id=0		
print	http://rbr.lib.unc.edu/cm/card.html?source_id=0		
ebook	http://rbr.lib.unc.edu/cm/card.html?source_id=0		
print	http://rbr.lib.unc.edu/cm/card.html?source_id=0	inforced block burged of	
			·. ·
с	D		E
MAT T 👻			
print	http://rbr.lib.unc.edu/cm/card.html?source_id=00019 http://de		
ebook	http://rbr.lib.unc.edu/cm/card.html?source_id=00019 http://do	crossronone-orden	
print	http://rbr.lib.unc.edu/cm/card.html?source_id=00020 http://do	*****	andrean and a first subscription of the second s
annananan	http://rbr.lib.unc.edu/cm/card.html?source_id=00022 http://do		
print	http://rbr.lib.unc.edu/cm/card.html?source_id=00035 http://do		
			na adu ladus likaus ladi adu adu di da 1916
ebook print	http://rbr.lib.unc.edu/cm/card.html?source_id=00035 http://do http://rbr.lib.unc.edu/cm/card.html?source_id=00036 http://do	******	***************************************

- Use I	MarcEdit	d Strategy Delimited Text eate "stub records"	MarcEitt By Teny Recze File Tools Add-ins Plug-ins Help Frequently Used Tools: MARC Tools MarcEditor Delimited Text Transistor Marcest 0A Records	ALS .	
	A		E		
	bnum	NEW URL		*	8
	b18307097	http://dc.lib.unc.edu/cdm/ite	m/collection/rbr/?id=34559		
	b60224307	http://dc.lib.unc.edu/cdm/ite	m/collection/rbr/?id=34559		
	b57385762	http://dc.lib.unc.edu/cdm/ite	m/collection/rbr/?id=34638		
	b40325805	http://dc.lib.unc.edu/cdm/ite	m/collection/rbr/?id=31502		
	b57474412	http://dc.lib.unc.edu/cdm/ite	m/collection/rbr/?id=34607		
	b60285126	http://dc.lib.unc.edu/cdm/ite	m/collection/rbr/?id=34607		
856 42\$3Mor 907 \\\$a.b612	11503x m 22000001a 45e0 e information about 1	-	é Savine Digital Library\$uhttp://dc.lib.unc.edu/cd é Savine Digital Library\$uhttp://dc.lib.unc.edu/cd		

- Global update on review file of Savine records

- Delete all old 865s containing |uhttp://rbr.lib.unc.edu
- Load stub records with my favorite load table
 - New URLs added



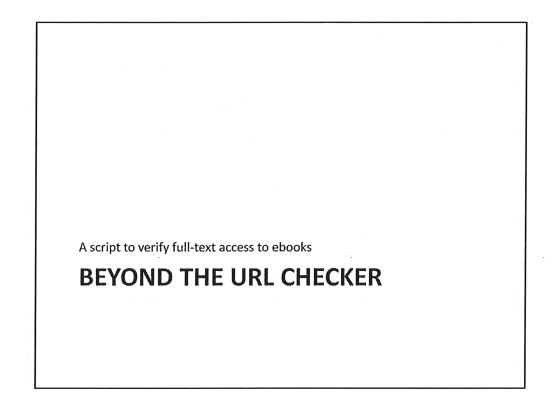
Other bricolage projects using my favorite load table

- SpringerLink ebook records

- 950s (subject module) were deleted from many records
- In SpringerLink title list: DOI url, Subject module
- In Millennium: bnum, DOI url
- Stub records with bnum (907) and new 950

- Alexander Street Press (ASP) records released without OCLC nums

- From ASP: ASP record ID, OCLC num
- From Mill: bnum, ASP record ID
- Stub records with bnum (907) and new 035



The previous examples show some ways we've cleaned and enhanced our catalog via data bricolage, moving data out of and back into Millennium.

I'm shifting gears here a bit to talk about tools I've developed to let us work with data exported from Millennium, with a focus on verifying and summarizing data from the ILS---not necessarily having to get the results back into the system.

Access checker:

The problem addressed

- Ideally, vendors would provide us with:
 - MARC records for ALL items to which we have full access
 - NO MARC record for items to which we have restricted access
- Reality is not ideal.
- Example: SpringerLink e-books
 - 250-560 new MARC records a month

http://dx.doi.org/10.1007/97	<u>'8-3-540-87631-1</u>
	doi> The DOI*System
	Error - DOI Not Found
	The DOI you requested
	10.1007/978-3-540-87631-1
	cannot be found in the Handle System.
	Possible reasons for the error are:
	 the DOI has not been created the DOI is cited incorrectly in your source the DOI does not resolve due to a system problem.
	If you believe you have requested a DOI that should be found, you may report this error by filling o the form below:
	Missing DOI: 10.1007/978-3-549-87631-1
	Referring Page:
	E-mail address: Please enter your email address
	Comments:
	Submit
	Dot Web Site