

11-6-2015

FRBR & Dedup in CUNY OneSearch

Joan Kolarik

City University of New York, New York, NY-USA - 1003

Follow this and additional works at: http://academicworks.cuny.edu/lacuny_events



Part of the [Cataloging and Metadata Commons](#)

Recommended Citation

Kolarik, Joan, "FRBR & Dedup in CUNY OneSearch" (2015). *CUNY Academic Works*.
http://academicworks.cuny.edu/lacuny_events/4

This Presentation is brought to you for free and open access by the Library Association of the City University of New York at CUNY Academic Works. It has been accepted for inclusion in Events by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

FRBR & Dedup in CUNY OneSearch

Joan Kolarik

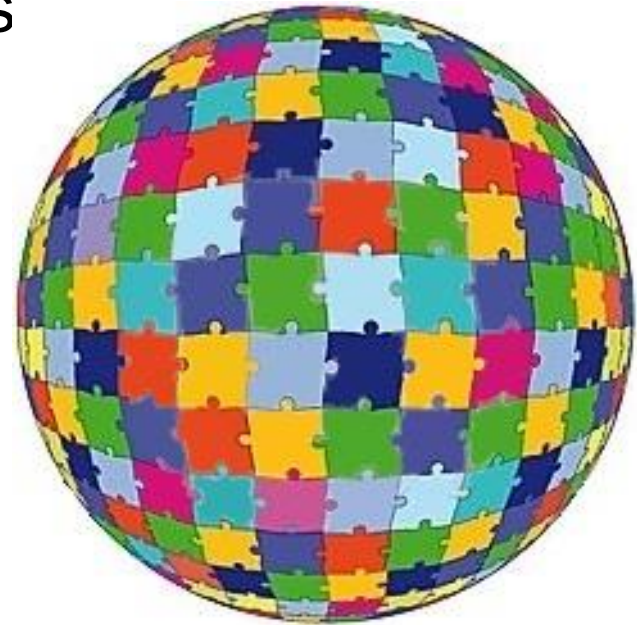
LACUNY Cataloging Roundtable

Baruch College

6 November 2015

It's Complicated!

- Lots of record sources coming together
- Many, many duplicate records



(let's face it, everything is complicated at CUNY!)

What are Dedup & FRBR?

- De-duplication
 - "intelligent compression" or "single-instance storage"
 - removes redundant data, eliminates "noise," saves space
- FRBR
 - structured ideas about what bibliographic records must contain to meet user needs
 - NOT data format (MARC 21), rule set (AACR2), mark-up language (XML)
 - access to 1 instance of title in results while obvious that other versions available



Dedup Overview

- Deduplication in OneSearch is
 - option, not a requirement
 - configured out-of-the-box
- Essential to know about dedup:
 - Merged record created and indexed
 - Individual records not shown
 - Access to duplicates available through merged record
 - Can be skipped
 - Test utility available
 - Rules can be tweaked

FRBR Overview

- FRBRization in OneSearch is
 - option, not a requirement (follows dedup)
 - configured out-of-the-box
- Essential to know about FRBR
 - FRBR records show a “version” = summary of member records
 - Individual FRBR members can be discovered and viewed
 - Can be skipped
 - Test utility available
 - Rules can be tweaked and specific fields boosted

Is it Dedup or FRBR?

- Is it Dedup?
 - Multiple links of same type?
 - Locations tab >1?
 - Add *&showPnx* to URL
 - *doc=dedupmrg**
 - [PNX Record](#) shows multiple records
- Is it FRBR? It's simple:

Example

Example



Multiple Versions

Optical spectroscopy of xenon-related defects in diamond

Dziashko, Yury; Anshel Gorokhovskiy; City University of New York. Physics. degree granting institution.

Multiple versions found. Click on title to see them all.

What Does Dedup Merge?

Merged record overview

- Full PNx of **preferred** record
- Search, Facet and Delivery sections of **all** members
- Merged availability information
- Delivery category – **online** preferred over physical item
- Fun fact: a record with more than 150 dedup candidates, won't be deduped

Merged PNX Record

Section	Fields
Control	Most fields included
Display	Duplicate fields removed All source and avallibrary included availinstitution and availpnx recalculated
Links	<div data-bbox="1265 678 1883 978" style="border: 1px solid green; padding: 5px;"> <p>Note: At CUNY, we limit display of local, subscription links and also notes -- these fields only display for the OWNing institution</p> </div>
Search	
Facets	
FRBR	
Additional Data	
Sort	Only preferred fields included
Delivery	All fields included
Ranking	Highest value among all records
Enrichment	Not relevant

DeDup PNX

002722845 & 002943430

003490203

<t>1</t>		<t>1</t>
<c1>44047006</c1>	Match	<c1>44047006</c1>
<c3>newwizardofoz</c3>	Match	<c3>newwizardofoz</c3>
<c4>1944</c4>		<c4>1944</c4>
<f1>44047006</f1>		<f1>44047006</f1>
<f5>newwizardofoz</f5>		<f5>newwizardofoz</f5>
<f6>1944</f6>	Match	<f6>1944</f6>
<f7>new wizard of oz</f7>		<f7>new wizard of oz</f7>
<f8>xx</f8>		<f8>nyu</f8>
<f9>4 p. l., 209 p.</f9>		<f9>209 p.:</f9>
<f10>bobbs merrill company</f10>		<f10>grosset dunlap</f10>
<f11>baum l frank lyman frank 1856 1919</f11>		<f11>baum l frank lyman frank 1856 1919</f11>

Computation

Type	Name	Info																				
HANDLER	SINGLE_MATCH	<table border="1"> <thead> <tr> <th>Fields</th> <th>Values1</th> <th>Values2</th> <th>Points</th> </tr> </thead> <tbody> <tr> <td>f20</td> <td></td> <td></td> <td>0</td> </tr> </tbody> </table>	Fields	Values1	Values2	Points	f20			0												
Fields	Values1	Values2	Points																			
f20			0																			
THRESHOLD	tr3	<table border="1"> <thead> <tr> <th>Points</th> <th>Upper</th> <th>Lower</th> <th>Decision</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>800</td> <td>null</td> <td>CONTINUE</td> </tr> </tbody> </table>	Points	Upper	Lower	Decision	0	800	null	CONTINUE												
Points	Upper	Lower	Decision																			
0	800	null	CONTINUE																			
HANDLER	CDLID	<table border="1"> <thead> <tr> <th>Fields</th> <th>Values1</th> <th>Values2</th> <th>Points</th> </tr> </thead> <tbody> <tr> <td>f1</td> <td>44047006</td> <td>44047006</td> <td>200</td> </tr> <tr> <td>f2</td> <td></td> <td></td> <td>200</td> </tr> <tr> <td>f3</td> <td></td> <td></td> <td>200</td> </tr> <tr> <td>f4</td> <td></td> <td></td> <td>200</td> </tr> </tbody> </table>	Fields	Values1	Values2	Points	f1	44047006	44047006	200	f2			200	f3			200	f4			200
Fields	Values1	Values2	Points																			
f1	44047006	44047006	200																			
f2			200																			
f3			200																			
f4			200																			
HANDLER	CDLShortTitle	<table border="1"> <thead> <tr> <th>Fields</th> <th>Values1</th> <th>Values2</th> <th>Points</th> </tr> </thead> <tbody> <tr> <td>f5</td> <td>newwizarfofoz</td> <td>newwizarfofoz</td> <td>450</td> </tr> </tbody> </table>	Fields	Values1	Values2	Points	f5	newwizarfofoz	newwizarfofoz	450												
Fields	Values1	Values2	Points																			
f5	newwizarfofoz	newwizarfofoz	450																			
HANDLER	CDLDate	<table border="1"> <thead> <tr> <th>Fields</th> <th>Values1</th> <th>Values2</th> <th>Points</th> </tr> </thead> <tbody> <tr> <td>f6</td> <td>1944</td> <td>1944</td> <td>200</td> </tr> </tbody> </table>	Fields	Values1	Values2	Points	f6	1944	1944	200												
Fields	Values1	Values2	Points																			
f6	1944	1944	200																			
THRESHOLD	tr1	<table border="1"> <thead> <tr> <th>Points</th> <th>Upper</th> <th>Lower</th> <th>Decision</th> </tr> </thead> <tbody> <tr> <td>850</td> <td>850</td> <td>0</td> <td>MATCH</td> </tr> </tbody> </table>	Points	Upper	Lower	Decision	850	850	0	MATCH												
Points	Upper	Lower	Decision																			
850	850	0	MATCH																			

What Does FRBR Merge?

- Nothing! All records indexed and retrieved
- Preferred Record calculated dynamically based on active search scope and most highly ranked record

Example CUNY

Example Kingsborough

- Standard function = matches and groups based on title and author fields
- In OneSearch, FRBR = “Multiple versions...”



Multiple
Versions

The Wizard of Oz

Baum, L. Frank, (Lyman Frank), 1856-1919.; Denslow, W. W, (William Wallace), 1856-1915.; Hearn, Michael Patrick

Multiple versions found. Click on title to see them all.

FRBR Matched PNX

dedupmrg223787760		006547171
<t>1</t>		<t>1</t>
<k1>\$\$Kbaum l frank lyman frank 1856 1919\$\$AA</k1>	Match	<k1>\$\$Kbaum l frank lyman frank 1856 1919\$\$AA</k1>
<k1>\$\$Kbolger ray\$\$AA</k1>	Match	<k1>\$\$Kbolger ray\$\$AA</k1>
<k1>\$\$Kmorgan frank 1890 1949\$\$AA</k1>	Match	<k1>\$\$Kmorgan frank 1890 1949\$\$AA</k1>
<k1>\$\$Kgarland judy\$\$AA</k1>	Match	<k1>\$\$Kgarland judy\$\$AA</k1>
<k1>\$\$Kkarlen harold 1905 1986\$\$AA</k1>		<k1>\$\$Kwalshe pat 1900 1991\$\$AA</k1>
etc		<k1>\$\$Kblandick clara 1881 1962\$\$AA</k1>
		etc
<k2>\$\$Kwizard of oz motion picture 1939\$\$ATO</k2>	Match	<k2>\$\$Kwizard of oz motion picture 1939\$\$ATO</k2>
<k3>\$\$Kwizard of oz\$\$AT</k3>	Match	<k3>\$\$Kwizard of oz\$\$AT</k3>

FRBR UNmatched PNX

007817695

006547171

<t>1</t>

<t>1</t>

<k1>\$\$\$Kborder rosemary\$\$\$AA</k1>

**NO
Match**

<k1>\$\$\$Kbaum l frank lyman frank 1856 1919\$\$\$AA</k1>

<k1>\$\$\$Kbolger ray\$\$\$AA</k1>

<k1>\$\$\$Kmorgan frank 1890 1949\$\$\$AA</k1>

<k1>\$\$\$Kgarland judy\$\$\$AA</k1>

<k1>\$\$\$Kwalshe pat 1900 1991\$\$\$AA</k1>

<k1>\$\$\$Kblandick clara 1881 1962\$\$\$AA</k1>

etc

**NO
Match**

<k2>\$\$\$Kwizard of oz motion picture 1939\$\$\$ATO</k2>

<k3>\$\$\$Kwizard of oz\$\$\$AT</k3>

Match

<k3>\$\$\$Kwizard of oz\$\$\$AT</k3>

FRBR Group = Title + Author

Field ID	Source (value of \$\$K for K fields)	Key part type (value of \$\$A for K fields)	Note
T	Always 1		
K1	100 or 110 or 111 or 700, 710, 711	A	Single occurrence of 100, 110, 111; multiple occurrences of 700, 710, 711 Take subfields a, b, c, d, q and n for 111
K2	130	TO	Subfields a, d, m, n, p, r, s
K3	240	T	Subfields a, d, m, n, p, r, s
	and (unless format is SE) 245 or 246 or 247		Subfields a, b, f, g, n, p
	or 740		Subfields a, n, p
	or 245		Subfield k

- 130 - Main Entry-Uniform Title (NR)
- 240 - Uniform Title (NR)
- 246 - Varying Form of Title (R)
- 247 - Former Title (R)
- 740 - Added Entry - Uncontrolled Related/Analytical Title

Clarity?

- 130 stands alone (TO = "title-only") = powerful
- Author + Title always paired
- Title
 - 240 used, if exists
 - **AND** 245, 246 or 247 used, if exist
 - If 245, 246, or 247 exist, then 740 and 245\$k are not used
 - If no 240 / 245/6/7, then 740 used
 - If no 240 / 245/6/7 / 740, then 245\$k used
- Author
 - **Only one** (1) 1xx used
 - If 1xx exists, then 7xx's **not** used
 - If there no 1xx, then **ALL** 7xxs used

130 - Main Entry-Uniform Title (NR)
240 - Uniform Title (NR)
246 - Varying Form of Title (R)
247 - Former Title (R)
740 - Added Entry - Uncontrolled Related/Analytical Title

How to Impact?

Both Dedup and FRBR

- Prevent for all metadata (OLS/pipe configuration)
- Prevent for groups of records = Type 99 (OLS/normalization rules)
- Modify PNX content
 - Change metadata (Local control!)
 - Normalization rules (OLS)
 - add universal id like record number
 - repurpose another id universal id, like ISBN

Problem?

Report it!

- ☞ CUNY Service Desk:
646-664-2311
service.desk@cuny.edu
- Provide as much information as possible!
- Include permalinks and EZproxy credentials (where applicable)



Thank You

Joan Kolarik

joan.kolarik@cuny.edu

Digital Initiatives Manager

CUNY