ADVANCED NORMALIZATION, DEDUP, AND FRBR

ANDREW WALSH | PRIMO SUBJECT MATTER EXPERT
PING GAO | PRIMO SUPPORT ANALYST

ExLibris
a ProQuest Company

Introductions

Andrew Walsh

Ping Gao

ExLibris

Objectives and Target Audience

Brief Session Description:

- A deeper look into Normalization, FRBR, and DeDup. Specific "recipes" will be shown and used to illustrated best practices.

Session Objective(s)

- By the end of this sessions you will know, understand and/or be able to:
  - Know the secrets to Primo happiness
  - Be able to solve several specific real world issues
  - Begin the journey to becoming a Primo expert

Session Target Audience:

- Experienced Customers
- Systems & Technical Staff
- Familiar with the Primo Normalization Rule Editor

3

ExLibris

Agenda

1   **The Secret**

2   **DeDup & FRBR**

3   **Recipes**

4   **Next Steps and Support Resources**

5   **Reference**

4

ExLibris

The Secret

ANCIENT TOME

A SET OF STRANGE KEYS

5

ExLibris
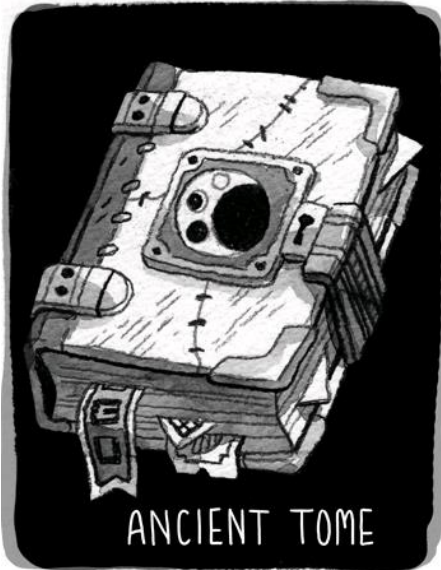
The secret keys to Primo happiness
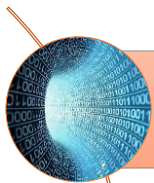
Know your data
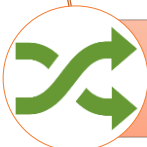
Know your NR rules

Know how they interact

6

ExLibris

A Quote

"It is you who must make the effort.

Masters only point the way."

- Buddha

h.koppdelaney

7

ExLibris
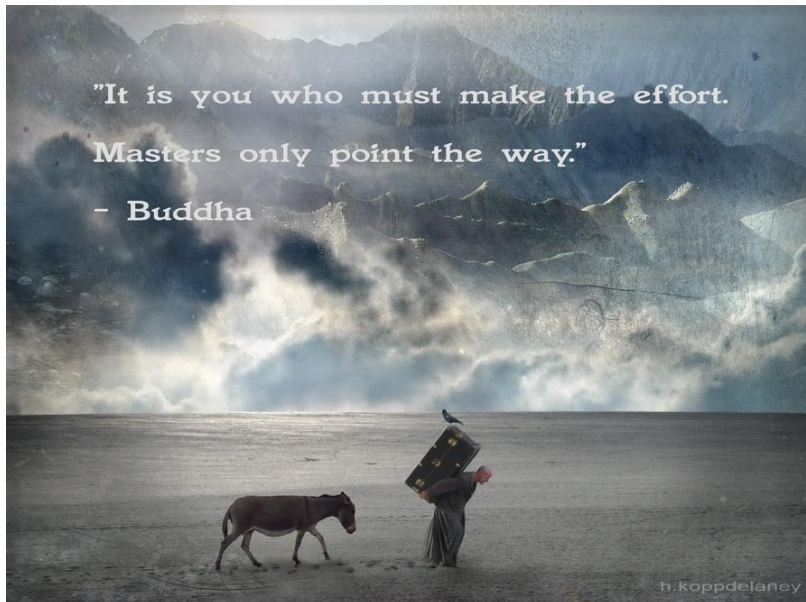
Pointing the way

Data — Always closely examine your data

- Has your cataloging always been consistent?
- If you changed how you catalog, would it display better in Primo?
- What's more important: cataloging perfection or patrons finding resources?

Rules — Never assume

- Examine the rules slowly and carefully
- Designed by programmers, not librarians
- When in doubt, open a case!

Interaction — Where the rubber meets the road

- Use the NR tester
- Test with your actual records
- Setup a micro test environment

8

ExLibris

Example One – Dates

```
<datafield tag="264" ind1=" " ind2="1">
    <subfield code="3">volume 1-8 (A-M) and 9,2 (O):</subfield>
    <subfield code="a">Lipsiae [Leipzig, Germany]:</subfield>
    <subfield code="b">In aedibus B.G. Teubneri,</subfield>
    <subfield code="c">1900-1987.</subfield>
</datafield>
<datafield tag="264" ind1="2" ind2="1">
    <subfield code="3">volume 10 (P: P-Progenies):</subfield>
    <subfield code="a">Leipzig ;</subfield>
    <subfield code="a">Stuttgart :</subfield>
    <subfield code="b">B.G. Teubner,</subfield>
    <subfield code="c">1987-1999.</subfield>
</datafield>
<datafield tag="264" ind1="2" ind2="1">
    <subfield code="3">volume 10 (P: Progenies-Pubertas):</subfield>
    <subfield code="a">München :</subfield>
    <subfield code="b">K.G. Saur,</subfield>
    <subfield code="c">2000-2006.</subfield>
</datafield>
<datafield tag="264" ind1="3" ind2="1">
    <subfield code="3">volumes 9,1 (N) and 10 (P: Pubertas- ):</subfield>
    <subfield code="a">Berlin ;</subfield>
    <subfield code="a">New York :</subfield>
    <subfield code="b">De Gruyter,</subfield>
    <subfield code="c">2007-&lt;/2016&gt;</subfield>
</datafield>
```
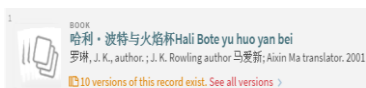
```
<controlfield tag="008">750724m19009999gw
```

```
<creationdate>1900</creationdate>
<creationdate>1987</creationdate>
<creationdate>2000</creationdate>
<creationdate>2007</creationdate>
<startdate>19000101</startdate>
<enddate>99991231</enddate>
```

MARC 008 **MARC 264** **PNX**

Example Two – Title

BOOK
哈利・波特与火焰杯 Hali Bote yu huo yan bei
罗琳, J. K., author. ; J. K. Rowling author 马爱新; Aixin Ma translator. 2001
📚 10 versions of this record exist. See all versions >

```
<title>Hali Bote yu huo yan bei</title>
<title>Harry Potter</title>
<title>Harry Potter and the goblet of fire</title>
```

```
<datafield tag="246" ind1="3" ind2=" ">
    <subfield code="a">Harry Potter</subfield>
</datafield>
<datafield tag="246" ind1="1" ind2=" ">
    <subfield code="i">Colophon title also in English:</subfield>
    <subfield code="a">Harry Potter and the goblet of fire</subfield>
</datafield>
```

Preferred Record **search/title** **MARC 246**

Example Three – FRBR Vectors

Record 1 — Alice's adventures in Wonderland

- 100 1_ |a Carroll, Lewis, |d 1832-1898.
- 240 10 |a Alice's adventures in Wonderland
- 245 10 |a Alice in Wonderland.
- 260 __ |a New York : |b H.N. Abrams, |c 1988.
- 300 __ |a [62] p. : |b ill. (some col.) ; |c 28 cm.

Record 2 — Alice's adventures in Wonderland (in Telugu)

- 100 1_ |a Carroll, Lewis, |d 1832-1898.
- 240 10 |a Alice's adventures in wonderland. |l Telugu
- 245 10 |a Allībillīlōkaṃlō ammāyikatha |h [microform] = |b Alice in wonderland / |c mūḷaṃ, Lūyī Kerōl ; anukaraṇaṃ, Va. Durgāprasādarāvu.
- 246 31 |a Alice in wonderland
- 260 __ |a Madarāsu : |b Andhragranthamāla, |c [1951]

Record 3 — Alice's adventures in Wonderland (the play)

- 100 1_ |a Chorpenning, Charlotte B. |q (Charlotte Barrows)
- 245 10 |a Alice in Wonderland, |c dramatized by Charlotte B. Chorpenning, a play from the library of the Association of the Junior Leagues of America, Inc.
- 260 __ |a Chicago, |b The Dramatic Publishing Company |c [1946]
- 700 1_ |a Carroll, Lewis, |d 1832-1898. |t Alice's adventures in Wonderland

Record 4 — Alice in Wonderland (the video)

- 050 00 |a VXB 0196-0197 (viewing copy)
- 245 00 |a Alice in Wonderland.
- 260 __ |c 1999.
- 300 __ |a 2 videocassettes of 2 : |b sd., col. ; |c 1/2 in. viewing copy.
- 710 2_ |a Copyright Collection (Library of Congress)

Data in FRBR Vectors – Match

Records 1 & 2	Vector and keys
100 1_ |a Carroll, Lewis, |d 1832-1898. 240 10 |a Alice's adventures in Wonderland 245 10 |a Alice in Wonderland. 260 __ |a New York : |b H.N. Abrams, |c 1988. 300 __ |a [62] p. : |b ill. (some col.) ; |c 28 cm.	K1 carroll lewis 1832-1898 K3 alices adventures in wonderland K3 alice in wonderland carroll lewis 1832-1898 alices adventures in wonderland carrol lewis 1832-1898 alice in wonderland
100 1_ |a Carroll, Lewis, |d 1832-1898. 240 10 |a Alice's adventures in wonderland. |l Telugu 245 10 |a Allībillīlōkaṃlō ammāyikatha |h [microform] = |b Alice in wonderland / |c mūḷaṃ, Lūyī Kerōl ; anukaraṇaṃ, Va. Durgāprasādarāvu. 246 31 |a Alice in wonderland 260 __ |a Madarāsu : |b Andhragranthamāla, |c [1951]	K1 carroll lewis 1832-1898 K3 alices adventures in wonderland K3 allibillilokamloamayikatha carroll lewis 1832-1898 alices adventures in wonderland carroll lewis 1832-1898 allibillilokamloamayikatha

Data not in FRBR Vectors – No Match

Records 3 & 4	Vector and keys
100 1_ |a Chorpenning, Charlotte B. |q (Charlotte Barrows) 245 10 |a Alice in Wonderland, |c dramatized by Charlotte B. Chorpenning, a play from the library of the Association of the Junior Leagues of America, Inc. 260 __ |a Chicago, |b The Dramatic Publishing Company |c [1946] 700 1_ |a Carroll, Lewis, |d 1832-1898. |t Alice's adventures in Wonderland	K1 chorpenning charlotte b K3 alice in wonderland chorpenning charlotte b alice in wonderland
050 00 |a VXB 0196-0197 (viewing copy) 245 00 |a Alice in Wonderland. 260 __ |c 1999. 300 __ |a 2 videocassettes of 2 : |b sd., col. ; |c 1/2 in. viewing copy. 710 2_ |a Copyright Collection (Library of Congress)	K1 copyright collection library of congress K3 alice in wonderland copyright collection library of congress alice in wonderland

DeDup & FRBR

DeDup Identifies records representing the same item

Replaces them with a new unified record

Display comes from a preferred record **selected at creation**

Incorrect deduplication usually due to data issues

ExLibris

FRBR Gathers similar records into a group

All the group members can be seen in Primo

Displays a preferred record **selected when you search**

Incorrect FRBR usually due to what data is being normalized

ExLibris

Preventing DeDup or FRBR

Type 99

- Tells Pipe to do no processing
- **Does not mean "this record should no longer be a DeDup/FRBR member"**

New records

- Normalize with t=99
- Set *System Last Stage* of the Pipe
  - PERSISTENCE – No DeDup, No FRBR
  - DEDUP – No FRBR
  - FRBR WITHOUT DEDUP – No DeDup

Existing records

- Normalize with t=99
- Run a reload Pipe with *Force DeDup* and/or *Force FRBR* enabled
- Disable *Force DeDup* and/or *Force FRBR*

ExLibris

Customizing the DeDup algorithm

It can be done

- Requires server access
  - Local
  - Direct-Dedicated
- Very complex

To update existing records

- Run a re-load Pipe for each Data Source
- Enable *Force DeDup*

Important considerations

- Approved but not Supported
- NR or cataloging changes are usually easier

ExLibris

Transitivity

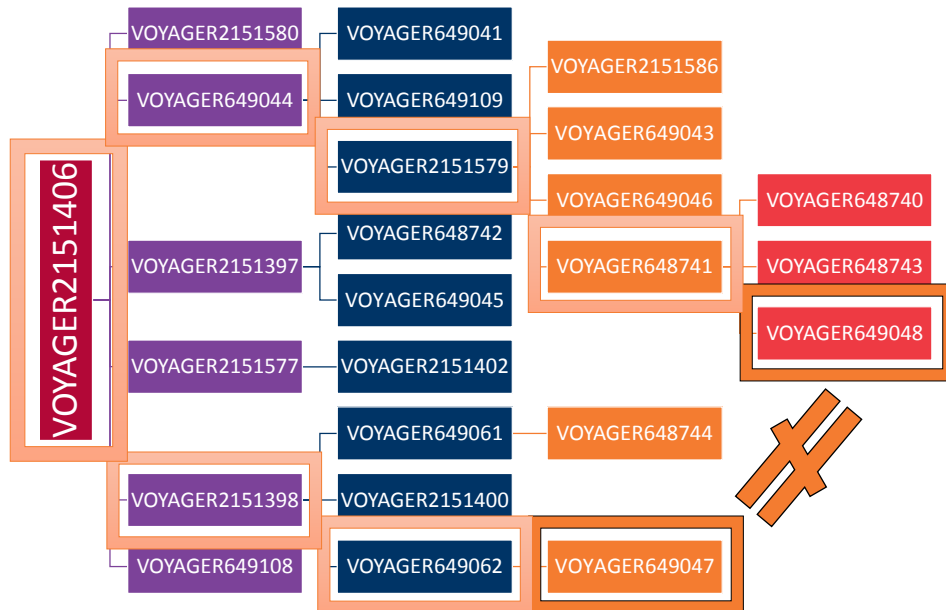| Dedup & FRBR	**not** fully transitive
	Depends on order records are processed
Transitive	$$A = B, B = C, A = C$$
Non-Transitive	$$A = B, B = C, A \neq C$$

Dedup Transitivity Example

- VOYAGER2151406
  - VOYAGER649044
    - VOYAGER2151579
      - VOYAGER2151586
      - VOYAGER648741
        - VOYAGER648740
        - VOYAGER648743
        - VOYAGER649048
      - VOYAGER649043
      - VOYAGER649046
    - VOYAGER649041
    - VOYAGER649109
  - VOYAGER2151580

- VOYAGER2151398
  - VOYAGER2151400
  - VOYAGER649062
    - VOYAGER649047
  - VOYAGER649061
    - VOYAGER648744
- VOYAGER2151577
  - VOYAGER2151402
- VOYAGER2151397
  - VOYAGER648742
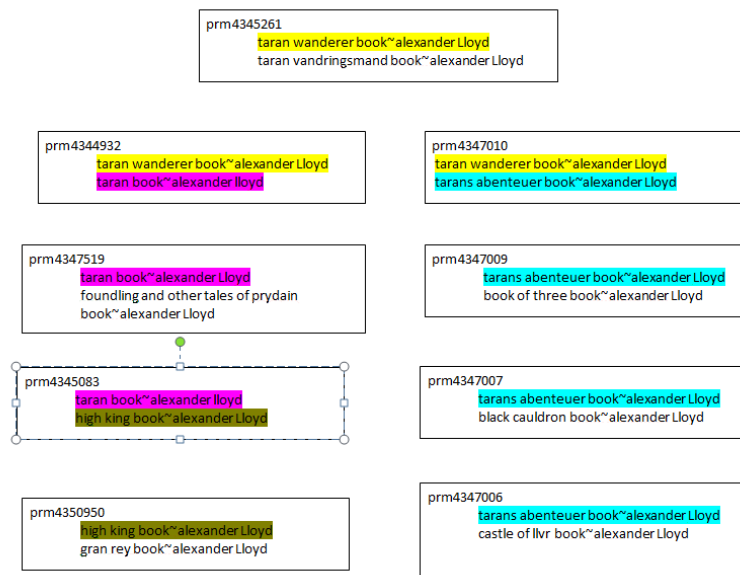 - VOYAGER649045
- VOYAGER649108

Dedup Transitivity Visualization

FRBR Transitivity Example

prm4345261
taran wanderer book~alexander Lloyd
taran vandringsmand book~alexander Lloyd

prm4344932
taran wanderer book~alexander Lloyd
taran book~alexander lloyd

prm4347010
taran wanderer book~alexander Lloyd
tarans abenteuer book~alexander Lloyd

prm4347519
taran book~alexander Lloyd
foundling and other tales of prydain
book~alexander Lloyd

prm4347009
tarans abenteuer book~alexander Lloyd
book of three book~alexander Lloyd

prm4345083
taran book~alexander lloyd
high king book~alexander Lloyd

prm4347007
tarans abenteuer book~alexander Lloyd
black cauldron book~alexander Lloyd

prm4350950
high king book~alexander Lloyd
gran rey book~alexander Lloyd

prm4347006
tarans abenteuer book~alexander Lloyd
castle of llvr book~alexander Lloyd

Recipes

GOLDEN SHOVEL
OF GREATER DIGGING

A QUIVER OF ARROWS

ExLibris

Recipes

1   Subject Capitalization

2   Normalizing Search ISBNs

3   No Universal ID for DeDup

4   DeDup Vector Over 4000 bytes – Single Source Element

5   DeDup Vector Over 4000 bytes – Multiple Source Elements

ExLibris

Subject Capitalization – Problem

```
<datafield tag="650" ind1=" " ind2="0">
  <subfield code="a">Americanisms</subfield>
  <subfield code="z">north Carolina</subfield>
  <subfield code="z">Cape Fear River Valley</subfield>
  <subfield code="x">Dictionaries.</subfield>
  <subfield code="0">(uri) http://id.loc.gov/authorities/subjects/sh85004429</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="0">
  <subfield code="a">popular culture</subfield>
  <subfield code="z">North Carolina</subfield>
  <subfield code="z">cape Fear River Valley.</subfield>
  <subfield code="0">(uri) http://id.loc.gov/authorities/subjects/sh85104904</subfield>
</datafield>
```

Transformations

Transformation	Parameter
Replace Characters ▾	:@@^
Remove characters from the end ▾	.
Upper case every first letter (lower case others) ▾	

```
<subject>Green, Paul, 1894-1981 -- Homes and haun
Americanismsmmmmnonth Carolinammmmcape Fear Riv
Valley -- Dictionaries; Popular Culturemmmmnorth
Carolina -- cape Fear River Valley; Cape Fear Riv
</subject>
```

MARC 650   **NR Rule**   **Problem**

Subject Capitalization - Solution

Replace Characters ▾	:@@^
Remove characters from the end ▾	.
Define subfield delimiter ▾	^^
Upper case every first letter (lower case others) ▾	

```
<subject>Green, Paul, 1894-1981 -- Homes and haunts
Americanisms North Carolina Cape Fear River Valley
Dictionaries; Popular Culture North Carolina Cape F
Fear River Valley; Cape Fear River Valley (N.C.) --
```

Add new transform   **Good Display**

Normalizing Search ISBNs

```xml
<search>
    <isbn>9792224567896</isbn>
    <isbn>1113456787</isbn>
    <isbn>9781113456789</isbn>
    <isbn>0913239666</isbn>
    <isbn>9780913239667</isbn>
    <isbn>9781473980495</isbn>
    <isbn>1473980496</isbn>
</search>
```

```xml
<datafield tag="020" ind1=" " ind2=" ">
   <subfield code="a">9792224567896</subfield>
</datafield>
<datafield tag="020" ind1=" " ind2=" ">
   <subfield code="a">666</subfield>
</datafield>
<datafield tag="020" ind1=" " ind2=" ">
   <subfield code="a">9781113456789</subfield>
</datafield>
<datafield tag="020" ind1=" " ind2=" ">
   <subfield code="a">978-0-913239-66-7 (set)</subfield
</datafield>
<datafield tag="020" ind1=" " ind2=" ">
   <subfield code="a">1473980496 (Cloth)</subfield>
</datafield>
<datafield tag="020" ind1=" " ind2=" ">
   <subfield code="0">(uri) http://www.isbnsearch.org/i
</datafield>
```

Goal

- Enable search on both ISBN10 & ISBN13

Challenges

- Quality of data
- ConvertISBN13toISBN10

Normalizing Search ISBNs – Solution

Rule group	search_isbn						
	Type		**Field**	**Ind1**	**Ind2**	**Subfield**	
Source	MARC ▾		020			Include ▾	a ☑ Enabled
	Conditions logic						
Conditions ⊕	True ▾						

Condition 1 - Logic	True ▾

Condition 1 - Source	**Type**		**Field**	**Ind1**	**Ind2**	**Subfield**		**Success If**
	MARC ▾		020			Include ▾	a	Match Current ▾

Condition1 - Routines ⊕ General Parameter

Routine	**Parameter**
Take first words ▾	1
Delete characters ▾	!"#$%&'()*+,-./:;<=>?@[\]^_`{|}~@@
Validate Length ▾	10

Transformations ⊕

Transformation	**Parameter**	
Take first words ▾	1	+ −
Delete characters ▾	!"#$%&'()*+,-./:;<=>?@[\]^_`{|}~@@	+ −
ConvertToISBN13 ▾		+ −

	Behavior
Action	ADD ▾

Normalizing Search ISBNs – Solution

Normalizing Search ISBNs – Solution

No Universal ID – Problem

Fields	Values1	Values2	Points
f1			0
f2			0
f3			0
f4			0

Fields	Values1	Va
f10	iowa state university college of engineering	iowa

Fields	Values1	Values2	Points
f11	cowles harold a	cowles harold a	125

Points	Upper	Lower	Decision
890	875	null	MATCH

No Universal ID is the **single most common reason** records incorrectly dedup

DeDup requires 875 points to match

- No values adds 0 points
- Mismatch adds -325 points
- 890 – 325 = 565 => No Match

ExLibris

No Universal ID – Solution

Rule group	dedup_f1		
	Type	Value	
Source	PNX	control/recordid	
	Conditions logic	Conditions relation	
Conditions	True	And	
Condition 1 - Logic	True		
Condition 1 - Source	Type		Value
	PNX		dedup/t
Condition1 - Routines	General Parameter		
	Routine		Parameter
	Check string equals string		1
Condition 2 - Logic	False		
Condition 2 - Source	Type		Value
	PNX		dedup/f2
Condition2 - Routines	General Parameter		
	Routine		Parameter
	Input exists		

Use control/recordid!

Condition 3 - Logic	False		
Condition 3 - Source	Type		Value
	PNX		dedup/f3
Condition3 - Routines	General Parameter		
	Routine		
	Input exists		
Condition 4 - Logic	False		
Condition 4 - Source	Type		Value
	PNX		dedup/f4
Condition4 - Routines	General Parameter		
	Routine		
	Input exists		

The "false/Input Exists" logic is the key

ExLibris

16

No Universal ID – Solution (cont.)

Transformations +

Transformation

Copy As Is ▼

Behavior

Action OR ▼

To use with serials

• Copy the rule
• Change the first condition from 1 to 2
• Add a fifth condition for dedup/f5

ExLibris

DeDup Vector 4000 byte limit – Single Source

group dedup_f7

	Type	Field	Ind1	Ind2	Subfield	
Source	MARC ▼	245			Include ▼ a,b,n,p	☑ Enabled

Conditions + No conditions Specified

Transformations +

Transformation		Parameter
Drop String (use reg. exp)	▼	<<.*>>
Drop non-filing text	▼	@@ind2@@
Drop String (use reg. exp)	▼	ˆ.*‰
Delete characters	▼	'
Replace Characters	▼	!@#$%^&*()_+-=[]|:";<>?,./~`@@^
Character Conversion	▼	
Lower Case	▼	
GetHeadTail	▼	3000@@1000

Behavi

ion OR

Edit the rule

Lower Case

GetHeadTail

Behavior

Action OR ▼

Add a GetHeadTail line

Set the value to 3000@1000

3000@@1000

ExLibris

DeDup Vector 4000 byte limit – Multiple Sources

Combine source data
- Use a custom PNX element
- Recommend addata, then display

DeDup Vector
- Disable existing rules
- Add a new rule

New Rule
- Source is the combined data field
- Apply a transformation of GetHeadTail

DeDup Vector 4000 byte limit – Multiple Sources Example

9783110975253;3110975254;3110975254;9783110975253;9783110960617;3110960613;3110960613;9783110960617;9783110955507;3110955504;3110955504;9783110955507;9783110953305;3110953307;3110953307;9783110953305;9783110953299;3110953293;3110953293;9783110953299;9783110953282;3110953285;3110953285;9783110953282;9783110953275;3110953277;3110953277;9783110953275;9783110953268;3110953269;3110953269;9783110953268;9783110953251;3110953250;3110953250;9783110953251;9783110957273;3110957272;3110957272;9783110957273;9783110954852;3110954850;3110954850;9783110954852;9783110964509;3110964503;3110964503;9783110964509;9783110964493;311096449X;311096449X;9783110964493;9783110964486;3110964481;3110964481;9783110964486;9783110964479;3110964473;3110964473;9783110964479;9783110964325;3110964325;3110964325;9783110964325;9783110964318;3110964317;3110964317;9783110964318;9783110964301;3110964309;3110964309;9783110964301;9783110964295;3110964295;3110964295;9783110964295;9783110964288;3110964287;3110964287;9783110964288;9783110964271;3110964279;3110964279;9783110964271;9783110964264;3110964260;3110964260;9783110964264;9783110964257;3110964252;3110964252;9783110964257;9783110960594;3110960591;3110960591;9783110960594;9783110960587;3110960583;3110960583;9783110960587;9783110960570;3110960575;3110960575;9783110960570;9783110965179;3110965178;3110965178;9783110965179;9783110964202;3110964201;3110964201;9783110964202;9783110979626;3110979624;3110979624;9783110979626;9783110932102;3110932105;3110932105;9783110932102;9783110932096;3110932091;3110932091;9783110932096;9783110955866;3110955865;3110955865;9783110955866;9783110955859;3110955857;3110955857;9783110955859;9783110953435;3110953439;3110953439;9783110953435;9783110953428;3110953420;3110953420;9783110953428;9783110938517;3110938510;3110938510;9783110938517;9783110938500;3110938502;3110938502;9783110938500;9783110953411;3110953412;3110953412;9783110953411;9783110938494;3110938499;3110938499;9783110938494;9783110953404;3110953404;3110953404;9783110953404;9783110937534;3110937530;3110937530;9783110937534;9783110953398;3110953390;3110953390;9783110953398;9783110953381;3110953382;3110953382;9783110953381;9783110953374;3110953374;3110953374;9783110953374;9783110956344;3110956349;3110956349;9783110956344;9783110956337;3110956330;3110956330;9783110956337;9783110954968;3110954966;3110954966;9783110954968;9783110954951;3110954958;3110954958;9783110954951;9783110954944;311095494X;311095494X;9783110954944;9783110954937;3110954931;3110954931;9783110954937;9783110954920;3110954923;3110954923;9783110954920;9783110958805;3110958805;3110958805;9783110958805;9783110958652;3110958651;3110958651;9783110958652;9783110957662;3110957663;3110957663;9783110957662;9783110957655;3110957655;3110957655;9783110957655;9783110957648;3110957647;3110957647;9783110957648;9783110967425;3110967421;3110967421;9783110967425;9783110967418;3110967413;3110967413;9783110967418;9783110967401;3110967405;3110967405;9783110967401;9783110934595;3110934590;3110934590;9783110934595;;9783110959246;3110959240;3110959240;9783110959246;9783110959222;3110959224;3110959224;9783110959222;9783110948097;3110948095;3110948095;9783110948097;9783110952483;3110952483;3110952483;9783110952483;9783110967012;3110967014;3110967014;9783110967012;9783110952285;3110952289;3110952289;9783110952285;9783110960044;3110960044;3110960044;9783110960044;9783110932843;3110932849;3110932849;9783110932843;9783110973945;3110973944;3110973944;9783110973945;9783110973518;3110973510;3110973510;9783110973518;9783110254969;3110254964;3110254964;9783110254969;9783110308501;3110308509;3110308509;9783110308501;9783110345803;3110345803;3110345803;9783110345803;9783110479102;3110479109;3110479109;9783110479102;9783110977158;311097715X;311097715X;9783110977158;9783110979688;3110979683;3110979683;9783110979688;9783110979671;3110979675;3110979675;9783110979671;9783110979664;3110979667;3110979667;9783110979664;9783110979657;3110979659;3110979659;9783110979657;9783110979640;3110979640;3110979640;9783110979640

Next Steps and Support Resources

37

ExLibris
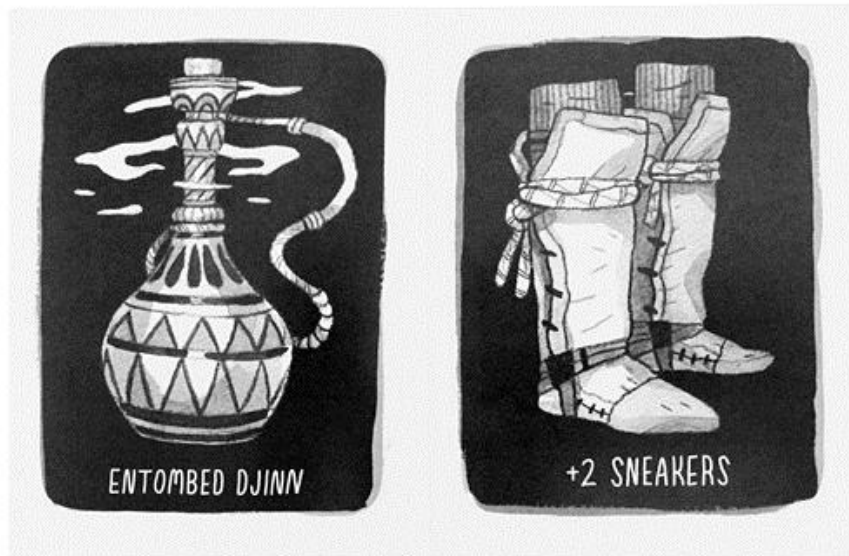
Next Steps and Support Resources

- Customer Knowledge Center
    - Working with Normalization Rules
    - Duplicate Detection Process
    - FRBRization
    - Harvesting a record fails with an UncategorizedSQLException error on the P_DEDUP_VECTOR table

- Additional support resources within the ExLibris Ecosystem:
    - Idea Exchange
    - System Status Pages: Single Tenant ENV / Multi-Tenant ENV
    - Developer Network

- Technical Seminar Presentations
  (located in the Cross-Product section of the CKC)

38

ExLibris

Q & A

Any Questions?

39
ExLibris

Session Survey Evaluation

Please use the following link
https://www.surveymonkey.com/r/techsem2017
to provide feedback on your sessions.

40
ExLibris

THANK YOU
Andrew.Walsh@exlibrisgroup.com
Ping.Gao@exlibrisgroup.com

ExLibris
a ProQuest Company

Image Credits

Cards

Portal 2

Math

Alchemy

Primo Central

Icon made by Freepik from www.flaticon.com is licensed by CC 3.0 BY

ExLibris

The Alchemy of DeDup

ExLibris
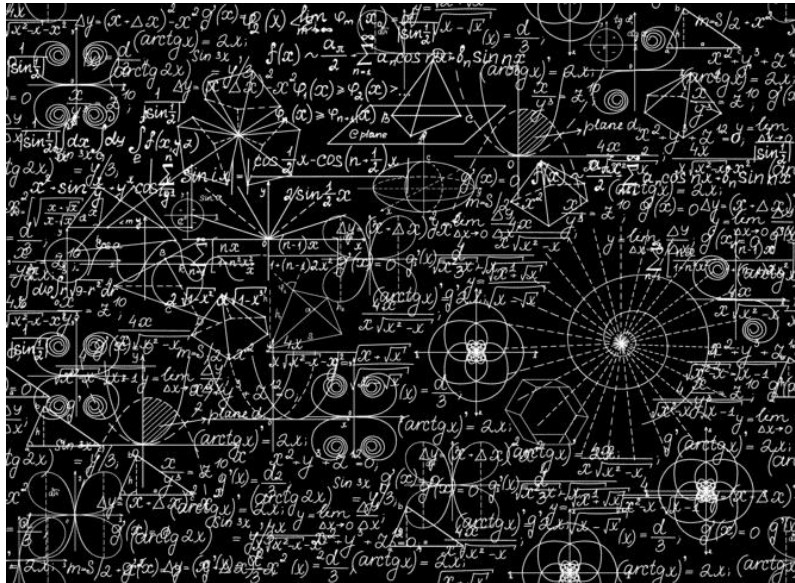
DeDup 101

Has the `<dedup>` section changed?

Does `<dedup/t>` have a value of 99?

Does any record have a matching `<dedup/c5>`?

Does a *Quick Match* give **850** points?

Does a *Full Match* give **875** points?

ExLibris

DeDup 102 – Quick Match

Fields for Comparison	Result	Points
UnivID <dedup/f1> UnivID_invalid <dedup/f2>	Match on valid UnivID	200
	Match on invalid UnivID	50
	Match between valid and invalid	100
	Field present in both records but no match	-320
	Either record or both records missing	0
ISBN <dedup/f3> ISBN_invalid <dedup/f4>	Match between valid ISBN	85
	Match between invalid ISBN	10
	Match between valid and invalid	30
	Field present in both records but no match	-225
	Either record or both records missing	0
Short-Title <dedup/f5>	Exact match on first 25 characters	450
	Non match	0
Date <dedup/f6>	Exact match	200
	+/- 2 years	-25
	No match	-250
	Value missing	0

ExLibris

DeDup 102 – Quick Match Example

HANDLER   CDLID

Fields	Values1	Values2	Points
f1	95022038	95022038	200
f2			200
f3	0415135168;9780203206614	9780203206614;0415135168	200
f4	9780415135160		200

HANDLER   CDLShortTitle

Fields	Values1	Values2	Points
f5	financialsystemunderldeconomy1	financialsystemunderldeconomy1	450

HANDLER   CDLDate

Fields	Values1	Values2	Points
f6	1996	1996	200

THRESHOLD   tr1

Points	Upper	Lower	Decision
850	850	0	MATCH

ExLibris

DeDup 103 – Full Match, Part 1

Fields for Comparison	Result	Points
UnivID <dedup/f1>	Match on valid UnivID	200
UnivID_invalid <dedup/f2>	Match on invalid UnivID	50
	Match between valid and invalid	100
	Field present in both records but no match	-320
	Either record or both records missing	0
ISBN <dedup/f3>	Match between valid ISBN	85
ISBN_invalid <dedup/f4>	Match between invalid ISBN	10
	Match between valid and invalid	30
	Field present in both records but no match	-225
	Either record or both records missing	0
Date <dedup/f6>	Exact match	200
	+/- 2 years	-25
	No match	-250
	Value missing	0

ExLibris

DeDup 103 – Full Match, Part 2

Fields for Comparison	Result	Points
Full-Title <dedup/f7>	Exact match	600
	Either title contained within the other title	350
	Either title shorter than nine characters	0
	Matching keywords	450 x (% of matching words)
	Matching keywords in order	450 x (% + 50)
	Non-match	-600
Country of Publication <dedup/f8>	Exact match	40
	Either one missing	0
	Non-match	-205
Pagination <dedup/f9>	Match exactly and value > 10	100
	Match exactly and < 10	50
	Match within 10 and both are value > 10	50
	Match within 10 and either < 10	20
	Non-match (by more then 10)	-225

ExLibris

24

DeDup 103 – Full Match, Part 3

Fields for Comparison	Result	Points
Publisher <dedup/f10>	Exact match	100
	Either missing	0
Main Entry <dedup/f11>	Exact match	125
	Both main entries missing	75
	Half (or more) of the main entry keywords are common and <u>in the same order</u>	$80 \times (\% \text{ common keywords}) + 10$
	Half (or more) of the main entry keywords are common, but are <u>*not*</u> in the same order	$80 \times (\% \text{ common keywords})$
	Present in one record but missing in the other	-25
	Non-match	-200

ExLibris

DeDup 103 – Full Match Example

HANDLER	CDLID	Fields	Values1	Values2	Points
		f1		90010855	100
		f2	90010855		100
		f3		0816618623;9780816618620;0816618631;9780816618637	100
		f4	9780816618637;0816618623;0816618631		100

HANDLER	CDLDate	Fields	Values1	Values2	Points
		f6	1990	1990	200

HANDLER	CDLLongTitle	Fields	Values1	Values2	Points
		f7	nationalism colonialism and literature	nationalism colonialism and literature	600

HANDLER	CDLCountryOfPub	Fields	Values1	Values2	Points
		f8	mnu	mnu	40

HANDLER	CDLPagination	Fields	Values1	Values2	Points
		f9	102 p.	102 pages ;	100

HANDLER	CDLPublisher	Fields	Values1	Values2	Points
		f10	university of minnesota press	university of minnesota press	100

HANDLER	CDLMainEntry	Fields	Values1	Values2	Points
		f11	eagleton terry 1943	eagleton terry 1943	125

THRESHOLD	tr2	Points	Upper	Lower	Decision
		1265	875	null	MATCH

ExLibris

The dedupmrg record

- **A new record**
  - Created during the Pipe
  - Replaces the member records
  - Contains (almost) everything from all members
- **Preferred record**
  - Doesn't change
  - Selection criteria (in order)
    - delivery/delcategory
    - # of fields in display

ExLibris

---

dedupmrg Example – Display

Member Record

Fields for Record ID VOYAGER1793211

Section	Field	Value
display	type	book
display	title	Homeless children : problems and needs
display	contributor	Panos Vostanis; Stuart Cumella
display	publisher	London ; Philadelphia : Jessica Kingsley Publishers
display	creationdate	1999
display	format	202 pages ; 24 cm

Member Record

Fields for Record ID VOYAGER3512864

Section	Field	Value
display	type	book
display	title	Homeless children problems and needs
display	contributor	Panos Vostanis; Stuart Cumella, ebrary, Inc.
display	publisher	London ; Philadelphia : Jessica Kingsley Publisher
display	creationdate	1999
display	format	202 p. : ill.

Dedupmrg Record

Fields for Record ID dedupmrg460417855

Section	Field	Value
display	type	book
display	title	Homeless children problems and needs
display	contributor	Panos Vostanis; Stuart Cumella, ebrary, Inc.
display	publisher	London ; Philadelphia : Jessica Kingsley Publishers
display	creationdate	1999
display	format	202 p. : ill.

ExLibris

dedupmrg Example – Facets

Member Record			Member Record		

Fields for Record ID VOYAGER1793211

Section	Field	Value
facets	topic	Homeless children
facets	topic	Homeless children–Services for
facets	topic	Great Britain

Fields for Record ID VOYAGER3512864

Section	Field	Value
facets	topic	Homeless children
facets	topic	Homeless children–Great Britain
facets	topic	Homeless children–Services for–Great Britain

Dedupmrg Record		

Fields for Record ID dedupmrg460417855

Section	Field	Value
facets	topic	Homeless children–Great Britain
facets	topic	Homeless children–Services for–Great Britain
facets	topic	Homeless children
facets	topic	Homeless children–Services for
facets	topic	Great Britain

DEDUP test utility

Primo Home > Primo Utilities > System Tests & Monitor

Home	Deploy & Utilities	General	Local Data

> **System Tests & Monitor**

» **DEDUP test**
DEDUP test utility

» **FRBR test**
FRBR test utility

DEDUP test utility - Candidates

Candidates: **Candidates**

Result: **Match**

Dedup type: **nonSerialMatcher**

Please notice that a record with more than 150 dedup candidates, won't be deduped

Field Name	Value 1	Value 2	Is Match
C1	EMPTY	00020033	No Match
C2	0195136047	9780195136043	No Match
C2	0195136047	1628701625	No Match
C2	0195136047	9781628701623	No Match
C2	0195136047	0195136047	Match
C2	9781628701623	9780195136043	No Match
C2	1433701081	9780195136043	No Match
C2	1628701625	9780195136043	No Match
C2	9781433701085	9780195136043	No Match
C2	9780195136043	9780195136043	Match
C3	introductiontoelectrngineering	introductiontoelectrngineering	Match
C4	2001	2001	Match

DEDUP test utility – Alma ID

Type	Name	Info			

		Fields	Values1	Values2	Points
HANDLER	SINGLE_MATCH	f20			0

		Points	Upper	Lower	Decision
THRESHOLD	tr3	0	800	null	CONTINUE

DEDUP test utility – Quick Match

Fields	Values1	Values2	Points
f1		00020033	85
f2			85
f3	1433701081;9781433701085;9781628701623;1628701625;0195136047;9780195136043	9781628701623;1628701625;9780195136043;0195136047	85
f4			85

HANDLER   CDLID

HANDLER   CDLShortTitle

Fields	Values1	Values2	Points
f5	introductiontoelectrngineering	introductiontoelectrngineering	450

HANDLER   CDLDate

Fields	Values1	Values2	Points
f6	2001	2001	200

THRESHOLD tr1

Points	Upper	Lower	Decision
735	850	0	CONTINUE

85+450+200 = 735

DEDUP test utility – Full Match

HANDLER   CDLSubShortTitle

Fields	Values1	Values2	Points
f5	introductiontoelectrngineering	introductiontoelectrngineering	-450

HANDLER   CDLLongTitle

Fields	Values1	Values2	Points
f7	introduction to electrical engineering	introduction to electrical engineering	600

HANDLER   CDLCountryOfPub

Fields	Values1	Values2	Points
f8	nyu	nyu	40

HANDLER   CDLPagination

Fields	Values1	Values2	Points
f9	1 online resource (xxiv, 869 pages) :	xxiv, 869 pages :	100

HANDLER   CDLPublisher

Fields	Values1	Values2	Points
f10	oxford university press	oxford university press	100

HANDLER   CDLMainEntry

Fields	Values1	Values2	Points
f11	sarma mulukutla s 1938	sarma mulukutla s 1938	125

THRESHOLD tr2

Points	Upper	Lower	Decision
1250	875	null	MATCH

735
− 450
+ 600
+ 40
+ 100
+ 100
+ 125
= 1250

The Simple Magic of FRBR

silver1 silver2 silver3 silver4 silver5 silver6 silver7 silver8 silver9
silver10 silver11 silverspirit skull soapstone1 soapstone2 soapstone3 sol soot
spirit spiritofmercury spiritoftin spring steel still stone sublimation sugar
sulphur1 sulphur2 sulphur3 sulphur4 sulphur5 sulphur6 sulphur7 sulphur8 sulphur9
sulphur10 summer take tartar1 tartar2 terra tin1 tin2 tin3
tin4 tin5 torrefactionofsilver uranus urine vapourbath verdigris vinegar vitriol1
vitriol2 vitriol3 water1 water2 wax1 wax2 week wick winespirit
winter wood year yields zinc1 zinc2 zinc3 zinc4 zinc5

FRBR 101

Create FRBR keys from the vectors

Every combination of $$AA and $$AT

Every $$ATO

Does any record have a matching FRBR key?

YES: **Create a FRBR group**

NO: **STOP**

FRBR Vectors & Keys – Example

FRBR Vectors

Title Only ($$ATO)
hawaiian sugar manual 1978

Title ($$AT)
hawaiian sugar manual

Author ($$AA)
• hawaiian sugar planters' association
• h s p a
• Hspa

FRBR Keys

hawaiian sugar manual 1978

hawaiian sugar manual~hawaiian sugar planters' association

hawaiian sugar manual~h s p a

hawaiian sugar manual~hspa

ExLibris

FRBR Group – Preferred Record

☆ **Hawaiian sugar manual.**
Hawaiian Sugar Planters' Association.; ©1978-©1995; 1978-1995.
● Available at   (HD9107.H3 H76) and other locations
View 2 versions of this title
Journal
Get This Item   Locations   Details   Virtual Browse

Refined by: List of Versions: x
☆ **Hawaiian sugar manual.**
Hawaiian Sugar Planters' Association.; ©1978-©1995; 1978-1995.
● Available at   (HD9107.H3 H76) and other locations
Journal   Get This Item   Locations   Details   Virtual Browse

☆ **Hawaiian sugar manual.**
Hawaiian Sugar Planters' Association.; 1974-1976]; 1974-1976.
● Available at   (HD9104 .H39) and other locations
Journal   Get This Item   Locations   Details   Virtual Browse

Brief Results

• Detailed information
• Tab links
• Resource icon

View All Versions

ExLibris

FRBR Group – Generic Record

Brief Results

- Limited information
- No tab links
- FRBR icon

View All Versions

ExLibris

OTB FRBR Vector

Field ID	Source ($$K)	Type ($$A)	Note
T	Always 1		
K1	100 or 110 or 111 or	A	Single occurrence of 100, 110, 111; Take subfields a, b, c, d, q and n for 111
	700, 710, 711		multiple occurrences of 700, 710, 711
K2	130	TO	Subfields a, d, m, n, p, r, s
K3	240	T	Subfields a, d, m, n, p, r, s
	(if not SE) and 245 or 246 or 247		Subfields a, b, f, g, n, p
	or 740		Subfields a, n, p
	or 245		Subfield k

ExLibris

FRBR test utility

Primo Home > Primo Utilities > System Tests & Monitor

Home   **Deploy & Utilities**   **General**   **Local Data**

> **System Tests & Monitor**

» **DEDUP test**
DEDUP test utility

» **FRBR test**
FRBR test utility

FRBR test utility – Single

Single: ● Multiple: ○

Record ID1: * VOYAGER1857744

Matched record Id's
VOYAGER1857744
VOYAGER2103523
VOYAGER1857744
VOYAGER2103523
VOYAGER1857744
VOYAGER2103523
VOYAGER1857744

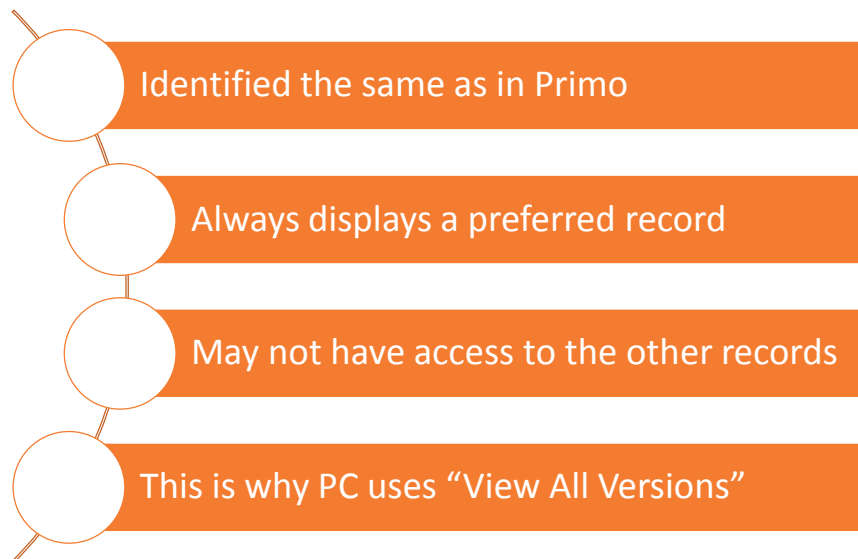| Record 1 keys	Key Hash
hawaiian sugar manual 1978	-4608125168587464567
hawaiian sugar manual~hawaiian sugar planters' association	2269858978978342289
hawaiian sugar manual~h s p a	-8898579990928581650
hawaiian sugar manual~hspa	-4608125167092699814

FRBR test utility – Multiple

Single: ○ Multiple: ◉

Record ID1: * `VOYAGER1857744`   Record ID2: * `VOYAGER2103523`

Result: **Match**

Record 1 keys	Key Hash
hawaiian sugar manual 1978	-4608125168587464567
hawaiian sugar manual~hawaiian sugar planters' association	2269858978978342289
hawaiian sugar manual~h s p a	-8898579990928581650
hawaiian sugar manual~hspa	-4608125167092699814
Record 2 keys	**Key Hash**
hawaiian sugar manual~hawaiian sugar planters' association	2269858978978342289
hawaiian sugar manual~h s p a	-8898579990928581650
hawaiian sugar manual~hspa	-4608125167092699814

ExLibris

DeDup & FRBR in Primo Central

ExLibris

DeDup in Primo Central

NONE

No dedupmrg record

Must maintain each publisher's information

ExLibris

FRBR in Primo Central

Identified the same as in Primo

Always displays a preferred record

May not have access to the other records

This is why PC uses "View All Versions"

ExLibris

FRBR in Primo Central – Example

PC FRBR Group

PC FRBR Group Versions