

New Orleans, LA • June 24, 2018

OCLC FRBR Cluster Analysis

ALCTS Cataloging and Classification Interest Group, ALA Annual 2018

Nathan B. Putnam

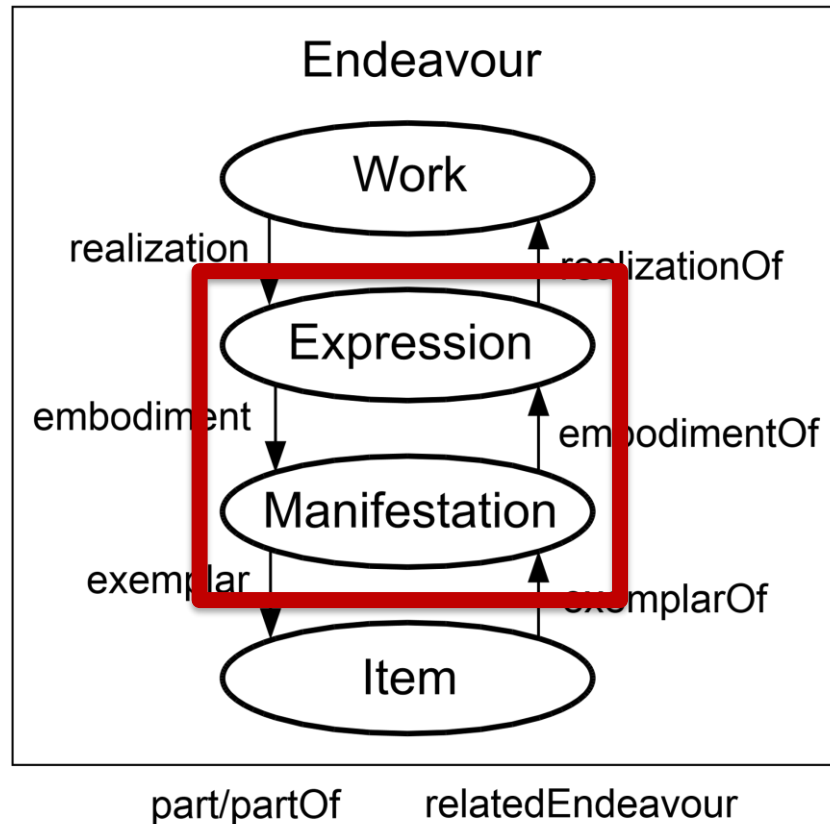
Director, Metadata Quality, OCLC

Agenda

- Overview of clustering
- Evaluating cluster groupings
- Outcomes
- Next steps

Record clusters

- Start with FRBR...
- OCLC clusters expressions and manifestations



Record clusters

- Resources within the cluster may contain
 - Different languages of the resource
 - Different translations in the same language
 - Editions with different dates and publishers
 - Different languages of cataloging (040 \$b)
- When searching, users are presented with a single, representative record
 - Includes search criteria when picking representative record
 - For example: date, publisher, editor, etc.
 - In WorldCat Discovery, includes edition the library holds

starbucked

[Advanced Search](#)

[Find a Library](#)

Search results for 'starbucked'

☒ **Format**

☒ **All Formats (24)**

☐ Article (12)

☐ Chapter (1)

☐ Book (7)

☐ Print book (7)

☐ eBook (1)

☐ Music (5)

☐ CD (3)

☐ Cassette (1)

Results **1-10** of about **24** (.11 seconds)

[Select All](#) [Clear All](#)

Save to:

☐ 1.



[Starbucked : a double tall tale of caffeine, commerce, and culture](#)

by Taylor Clark;

 Print book [View all formats and languages »](#)

Language: English

Publisher: New York : Back B...

[View all editions »](#)

View clusters



[<< Return to Item Details](#)

Showing all editions for "Starbucked : a double tall tale of caffeine, commerce, and culture"

Sort by: [Date/Edition \(Newest First\)](#)

Format

☒ All Formats (16)

☐ Book (1)

☐ Print book (15)

☐ eBook (1)

Refine Your Search

Year

[2008](#) (10)







[2007](#) (6)

Language

[English](#)







Displaying Editions 1 - 10 out of 16

[Select All](#) [Clear All](#) Save to: [\[New List\]](#) [Save](#)

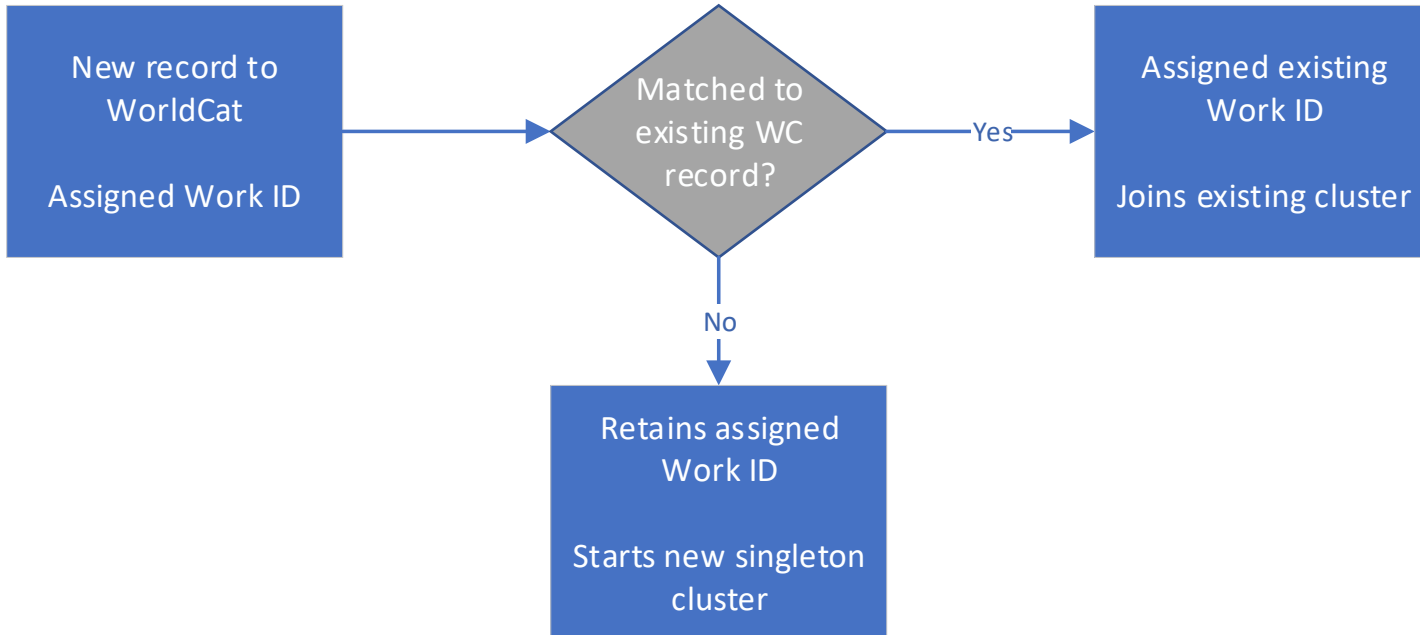
- | | Title / Author |
|-----------------------------|--|
| <input type="checkbox"/> 1. |  Starbucked : a double tall culture
by Taylor Clark; |
| <input type="checkbox"/> 2. |  Starbucked : a double tall culture
by Taylor Clark |
| <input type="checkbox"/> 3. |  Starbucked : a double tall culture
by Taylor Clark |
| <input type="checkbox"/> 4. |  Starbucked : a double tall
by Taylor Clark |
| <input type="checkbox"/> 5. |  Starbucked : a double tall
by Taylor Clark |
| <input type="checkbox"/> 6. |  Starbucked : a grande tall
by Taylor Clark |

Displaying Editions 11 - 16 out of 16

[Select All](#) [Clear All](#) Save to: [\[New List\]](#) [Save](#)

	Title / Author	Type	Language	Date / Edition	Publication
<input type="checkbox"/> 11.	 Starbucked : a double tall tale of caffeine, commerce, and culture by Taylor Clark	eBook : Document	English	2007 First edition	New York : Little, Brown
<input type="checkbox"/> 12.	 Starbucked : a double tall tale of caffeine, commerce, and culture by Taylor Clark	Print book	English	2007	New York : Little, Brown
<input type="checkbox"/> 13.	 Starbucked a double tall tale of caffeine, commerce, and culture by Taylor Clark	Print book	English	2007 First edition	New York : Little, Brown
<input type="checkbox"/> 14.	 Starbucked : a double tall tale of caffeine, commerce, and culture by Taylor Clark	Print book	English	2007 1st ed	New York : Little, Brown
<input type="checkbox"/> 15.	 Starbucked : a double tall tale of coffeine, commerce, and culture by Taylor Clark	Print book	English	2007 1. ed	New York : Little, Brown
<input type="checkbox"/> 16.	 Starbucked : a double tall tale of caffeine, commerce & culture by Taylor Clark	Print book	English	2007	London : Hodder & Stoughton

Assigning Work/cluster IDs



Record clusters

- Not all clusters are created equal
 - Varying levels of data fullness
 - Different data elements (number of authors, title variations)
 - Missing data = records in wrong clusters; clusters not coming together
- PCC/ARL community desires identifiers for the clusters
 - An identifier for the cluster is essentially an identifier for the Work
 - Several pilot projects from the community: Cornell, U of MN
 - Potential use for moving to BIBFRAME or graph-data

Cluster review goals

- Goal – populate (select) Work IDs in MARC records
 - Field 758, Resource Identifier
 - Use “accurate” clusters
- Determining cluster accuracy
 - Review of cluster categories
 - Assign confidence ratings for the cluster categories
 - Analysis covered 90% sampling of clusters

Cluster review definitions

- X names → in a record
 - 1 name → all records in cluster have exactly 1 name
 - 0 names → some records had no names fields
 - Multi names → some records had more than 1 name field
- Y titles → in a cluster
 - 1 title → exactly 1 unique title (no variations) across all title fields in all records

Cluster review sheets

- Cluster categories examples
 - 1 personal name, 1 title
 - Any names, 2 titles
 - 1 name, 10 titles in cluster
- Clusters reviewed in groupings of 500
 - Spreadsheet with cluster link and stats
 - Clusters chosen randomly
 - 14 reviewers assigned cluster ranges
 - 7 cluster groupings reviewed

File Home Insert Page Layout Formulas Data Review View Help Search																		
A1																		
	B	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	Link	Title	workid	sibs	holds0-1	holds2-8	holds9more	rda	tiny	dlc	good	thesis	fiction	eholds	date1	mtype	mType1	
2	http://worldcat.org/oclc/246277981/editions	aO komičeskom / cJurij Borev. Akademija Nauk SSSR	newwork150297975	4	1	3	0	2 (5,1)	0	0	0	0	0	0	1957	1	(u'book', 4)	
3	http://worldcat.org/oclc/67934541/editions	aCollectieve besluiten, belangen en wetgeving : bd	newwork150317143	2	1	0	1	2 (24,1)	0	1	2	0	0	0	1999	1	(u'book', 2)	
4	http://worldcat.org/oclc/62408648/editions	aContemporary college algebra : bdata, functions, n	newwork150341029	3	0	3	0	3 (11,1)	0	2	0	0	0	0 (2004/2008)	1	(u'book', 3)		
5	http://worldcat.org/oclc/781530690/editions	aHigh-order finite-difference methods for Poisson's	newwork150342091	8	5	2	1	4 (6,1)	0	1	6	0	0	0	1971	1	(u'book', 8)	
6	http://worldcat.org/oclc/15034264/editions	aAn annotated bibliography on pornography : bcurr	newwork15034264	4	2	1	1	4 (13,1)	1	1	0	0	1	1987	1	(u'book', 4)		
7	http://worldcat.org/oclc/150348383/editions	aTheist delusions : bthe Christian revolution and it	newwork150348383	15	7	4	4	11 (9,1)	1	4	0	0	821 (2007/2010)	1	(u'book', 15)			
8	http://worldcat.org/oclc/220146641/editions	aDokimes / cGiōrgos Sepherēs.	newwork150350434	2	0	1	1	2 (11,1)	0	1	0	0	0	1974	1	(u'book', 2)		
9	http://worldcat.org/oclc/921216/editions	aPeter Gayleigh flies high.	newwork150360686	10	6	3	1	9 (8,1)	0	1	0	10	0	0 (1951/1980)	1	(u'book', 10)		
10	http://worldcat.org/oclc/181557968/editions	aNeurotheologie zwischen Religionskritik und -affir	newwork150370982	4	4	0	0	3 (11,1)	0	0	4	0	0	0 (2005/2006)	1	(u'book', 4)		
11	http://worldcat.org/oclc/5416743/editions	ala difesa francese / cGiorgio Porreca.	newwork150387191	4	3	1	0	3 (11,1)	1	1	0	0	0	0 (1977/2005)	1	(u'book', 4)		
12	http://worldcat.org/oclc/150388857/editions	aKu dan mu / cKarsono H. Saputra.	newwork150388857	2	2	0	0	1 (8,1)	0	0	0	2	0	2005	1	(u'book', 2)		
13	http://worldcat.org/oclc/66439629/editions	aLower respiratory tract infections in general practi	newwork150391475	2	1	1	0	1 (17,1)	0	0	1	0	0	2005	1	(u'book', 2)		
14	http://worldcat.org/oclc/34000609/editions	aThree octave scales and arpeggios / cadapted for v	newwork150395985	3	2	1	0	1 (10,1)	0	1	0	0	0	0 (1988/1995)	1	(u'scor', 3)		
15	http://worldcat.org/oclc/150398702/editions	aRinga ett telefonsamtal eller logga in på chatten? :	newwork150398702	5	3	2	0	3 (10,1)	0	0	0	0	0	2006	1	(u'book', 5)		
16	http://worldcat.org/oclc/150409411/editions	aICARUS-4 h[elekttronische middelen] : bsector stud	newwork150409411	2	2	0	0	0 (9,1)	0	0	0	0	0	2001	1	(u'book', 2)		
17	http://worldcat.org/oclc/67393795/editions	aScary warnings and rational precautions : ba study	newwork150410527	2	1	0	1	2 (20,1)	0	1	2	0	0	2000	1	(u'book', 2)		
18	http://worldcat.org/oclc/676181959/editions	6245-01/\$1 aウェブを進化させる人たち : bITの?	newwork150411913	4	4	0	0	1 (15,1)	0	0	0	0	0	2007	1	(u'book', 4)		
19	http://worldcat.org/oclc/674443536/editions	6245-01/\$1 aフォトジャーナリスト吉田ルイ子	newwork150412804	2	1	1	0	1 (20,1)	0	1	0	0	0	1998	1	(u'book', 2)		
20	http://worldcat.org/oclc/2173150/editions	aPopulation densities, habitat selection, and foliage	newwork15041392	3	2	1	0	3 (15,1)	0	3	3	0	0	0 (1976/1979)	1	(u'book', 3)		
21	http://worldcat.org/oclc/67108833/editions	aTo solve or not to solve? : beffects of problem solv	newwork150414262	2	1	0	1	2 (23,1)	0	1	2	0	0	2002	1	(u'book', 2)		
22	http://worldcat.org/oclc/150416819/editions	aPlangebied Meester van de Bergstraat : bgemeent	newwork150416819	2	2	0	0	0 (11,1)	0	0	0	0	0	0 (2006/2007)	1	(u'book', 2)		
23	http://worldcat.org/oclc/3756124/editions	aThe English image of the Ottoman Empire, 1580-17	newwork150427231	13	11	2	0	11 (8,3)	0	1	10	0	0	0 (1976/1982)	1	(u'book', 13)		
24	http://worldcat.org/oclc/8049587/editions	aSugar and slavery in Puerto Rico : bthe municipalit	newwork150427682	10	3	7	0	10 (9,1)	0	3	7	0	0	0 (1978/1982)	1	(u'scor', 10)		
25	http://worldcat.org/oclc/150432901/editions	aUltrasound used as an environmental sensor by bli	newwork150432901	2	2	0	0	2 (10,1)	0	0	2	0	0	1982	1	(u'book', 2)		
26	http://worldcat.org/oclc/8525556/editions	aThe medieval background of Lazarillo de Tormes a	newwork150435250	7	4	3	0	6 (8,1)	0	2	7	0	0	0 (1982/1986)	1	(u'book', 7)		
27	http://worldcat.org/oclc/187456450/editions	aDesign for the controllability and operability of ch	newwork150436097	2	1	1	0	2 (13,1)	0	0	2	0	0	0 (1991/1993)	1	(u'book', 2)		
28	http://worldcat.org/oclc/610787/editions	aDas Leiden unsers Herrn Jesu Christi nach dem Eva	newwork150436141	2	1	1	0	2 (18,1)	0	1	0	0	0	1938	1	(u'scor', 2)		
29	http://worldcat.org/oclc/150436592/editions	aHistoire de Belgique / cConstant Leclere.	newwork150436592	4	4	0	0	1 (6,2)	0	0	0	0	0	0 (1927/1947)	1	(u'book', 4)		
30	http://worldcat.org/oclc/15043949/editions	aA revision of the genus Scincella (Reptilia, Sauria,	newwork15043949	4	0	3	1	4 (15,1)	0	1	0	0	0	1986	1	(u'book', 4)		
31	http://worldcat.org/oclc/15045451/editions	aTranslating data flow graphs to architectures / cby	newwork15045451	2	1	1	0	2 (15,1)	0	0	2	0	0	1985	1	(u'book', 2)		
32	http://worldcat.org/oclc/150459362/editions	aBasic themes in the prose of Malcolm Lowry : bwit	newwork150459362	2	2	0	0	2 (14,1)	0	0	2	0	0	1973	1	(u'book', 2)		
33	http://worldcat.org/oclc/4537937/editions	aResponses on the Davidson Adjective Check List as	newwork150474789	6	3	1	2	5 (10,1)	0	2	5	0	0	0 (1967/1968)	1	(u'book', 6)		

Spreadsheet stats

- Direct link to the records
- Title
- Work ID
- Siblings in cluster
- Holdings count on records
- RDA
- DLC
- Dates
- Material type(s)
- Language of resource
- Language of cataloging

- Review the stats for the items (columns in the spreadsheet; the key is in the second tab of the spread sheet)
- Review the clusters (the links in column B will take you directly to them)
- Rate the cluster with one of the following values
 - Yes – this is a good cluster; users would expect to find these grouped together
 - Maybe – this may or may not be a good cluster, it's unclear if the users would expect this together
 - No – this is a bad cluster, the items are unrelated
- For Maybe and No ratings, add a comment as to why you have given that rating
- Caveats to the ratings
 - **Translations** – we would expect these to be in the same cluster
 - **Formats** – the algorithm is set-up to cluster formats, so we would NOT see a print book and an audio book in the same cluster. We would expect to see different SMDs together, so DVD and VHS in the same cluster, book and thesis in the same cluster. Related, e-books and print books will be in the same cluster.
 - **Editions** – the algorithm will group different editions together and is considered a good match

starbucked

[Advanced Search](#)

[Find a Library](#)

Search results for 'starbucked'

☒ **Format**

☒ **All Formats (24)**

☐ Article (12)

☐ Chapter (1)

☐ Book (7)

☐ Print book (7)

☐ eBook (1)

☐ Music (5)

☐ CD (3)

☐ Cassette (1)

Results 1-10 of about 24 (.11 seconds)

[Select All](#) [Clear All](#)

Save to:

☐ 1.



[Starbucked : a double tall tale of caffeine, commerce, and culture](#)

by Taylor Clark;

[View all formats and languages »](#)

Language: English









Publisher: New York : Back B...

[View all editions »](#)

View clusters

<http://www.worldcat.org/title/starbucked-a-double-tall-tale-of-caffeine-commerce-and-culture/oclc/227018423/editions>









Select All Clear All Save to: [New List] ▼

	Title / Author	Type	Language	Year	Location
<input type="checkbox"/>	1.  <u>Starbucked : a double tall tale of caffeine, commerce, and culture</u> by Taylor Clark	 eBook Document	English	2007 First edition	New York : Little, Brown
<input type="checkbox"/>	2.  <u>Starbucked : a double tall tale of caffeine, commerce, and culture</u> by Taylor Clark	 Print book	English	2007	New York : Little, Brown
<input type="checkbox"/>	3.  <u>Starbucked a double tall tale of caffeine, commerce, and culture</u> by Taylor Clark	 Print book	English	2007 First edition	New York Little, Brown
<input type="checkbox"/>	4.  <u>Starbucked : a double tall tale of caffeine, commerce, and culture</u> by Taylor Clark	 Print book	English	2007 1st ed	New York : Little,

Select All Clear All Save to: [New List] Save

Title / Author

040 \$b contains information on the **language of cataloging**, the predominate language of the cataloging agency.

- | | | |
|--|---|---|
| <input type="checkbox"/> 1.  | <p>Starbucked : a double tall tale of caffeine, commerce, and culture
by Taylor Clark</p> <p>040 \$b eng</p> | <p> eBook English 2007 New York :
: First Little,
Document edition Brown</p> |
| <input type="checkbox"/> 2.  | <p>Starbucked : a double tall tale of caffeine, commerce, and culture
by Taylor Clark</p> <p>040 \$b swe</p> | <p> Print English 2007 New York :
book Little,
Brown</p> |
| <input type="checkbox"/> 3.  | <p>Starbucked a double tall tale of caffeine, commerce, and culture
by Taylor Clark</p> <p>040 \$b eng</p> | <p> Print English 2007 New York :
book First Little,
edition Brown</p> |
| <input type="checkbox"/> 4.  | <p>Starbucked : a double tall tale of caffeine, commerce, and culture
by Taylor Clark</p> <p>040 \$b fre</p> | <p> Print English 2007 New York :
book 1st ed Little,
OCLC</p> |

[Select All](#) [Clear All](#) Save to:

	Title / Author	Type	Language	Date / Edition	Publication
<input type="checkbox"/> 1.	 <u>Starbucked : a double tall tale of caffeine, commerce, and culture</u> by Taylor Clark	eBook Document	English	2007 First edition	New York : Little, Brown
<input type="checkbox"/> 2.	 <u>Starbucked : a double tall tale of caffeine, commerce, and culture</u> by Taylor Clark	Print book	English	2007	New York : Little, Brown
<input type="checkbox"/> 3.	 <u>Starbucked a double tall tale of caffeine, commerce, and culture</u> by Taylor Clark	Print book	English	2007 First edition	New York Little, Brown
<input type="checkbox"/> 4.	 <u>Starbucked : a double tall tale of caffeine, commerce, and culture</u> by Taylor Clark	Print book	English	2007 1st ed	New York : Little,

040 \$b eng









040 \$b swe

040 \$b eng









040 \$b fre

[Select All](#)[Clear All](#)

Save to:

	Title / Author	Type	Language	Date / Edition	Publication
<input type="checkbox"/> 11.	 Starbucked : a double tall tale of caffeine, commerce & culture by Taylor Clark 040 \$b ger	 Print book	English	2008	London : Sceptre
<input type="checkbox"/> 12.	 Starbucked : a grande tale of caffeine, commerce & culture by Taylor Clark 040 \$b eng	 Print book	English	2008 1st ed	London : Hodder & Stoughton
<input type="checkbox"/> 13.	 Starbucked : a double tall tale of caffeine, commerce & culture by Taylor Clark 040 \$b eng	 Print book	English	2008	London : Sceptre
<input type="checkbox"/> 14.	 Starbucked : a grande tale of caffeine, commerce & culture by Taylor Clark 040 \$b eng	 Print book	English	2008	London : Sceptre

[Select All](#) [Clear All](#) Save to:

	Title / Author	Type	Language	Date / Edition	Location
<input type="checkbox"/> 11.	 <u>Starbucked : a double tall tale of caffeine, commerce & culture</u> by Taylor Clark 040 \$b ger	 Print book	English	2008	London : Sceptre
<input type="checkbox"/> 12.	 <u>Starbucked : a grande tale of caffeine, commerce & culture</u> by Taylor Clark 040 \$b eng	 Print book	English	2008 1st ed	London : Hodder & Stoughton
<input type="checkbox"/> 13.	 <u>Starbucked : a double tall tale of caffeine, commerce & culture</u> by Taylor Clark 040 \$b eng	 Print book	English	2008	London : Sceptre
<input type="checkbox"/> 14.	 <u>Starbucked : a grande tale of caffeine, commerce & culture</u> by Taylor Clark 040 \$b eng	 Print book	English	2008	London : Sceptre

Results (generally speaking)

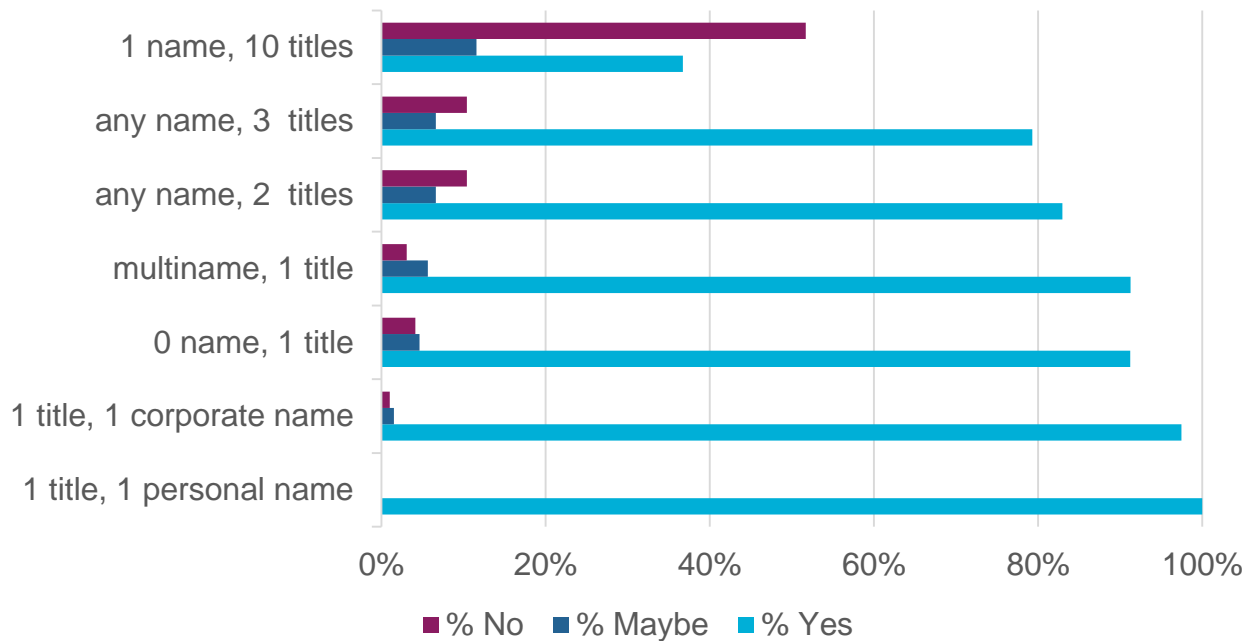
- Cluster groupings

- | | |
|---|------|
| • 1 title, 1 personal name | 100% |
| • 1 title, 1 corporate name | 97% |
| • 1 title, 0 names | 91% |
| • 1 title, multiple names | 91% |
| • 2 distinct titles, any names | 83% |
| • 3 distinct titles, any names | 79% |
| • 1 name, at least 10 titles in the cluster | 37% |

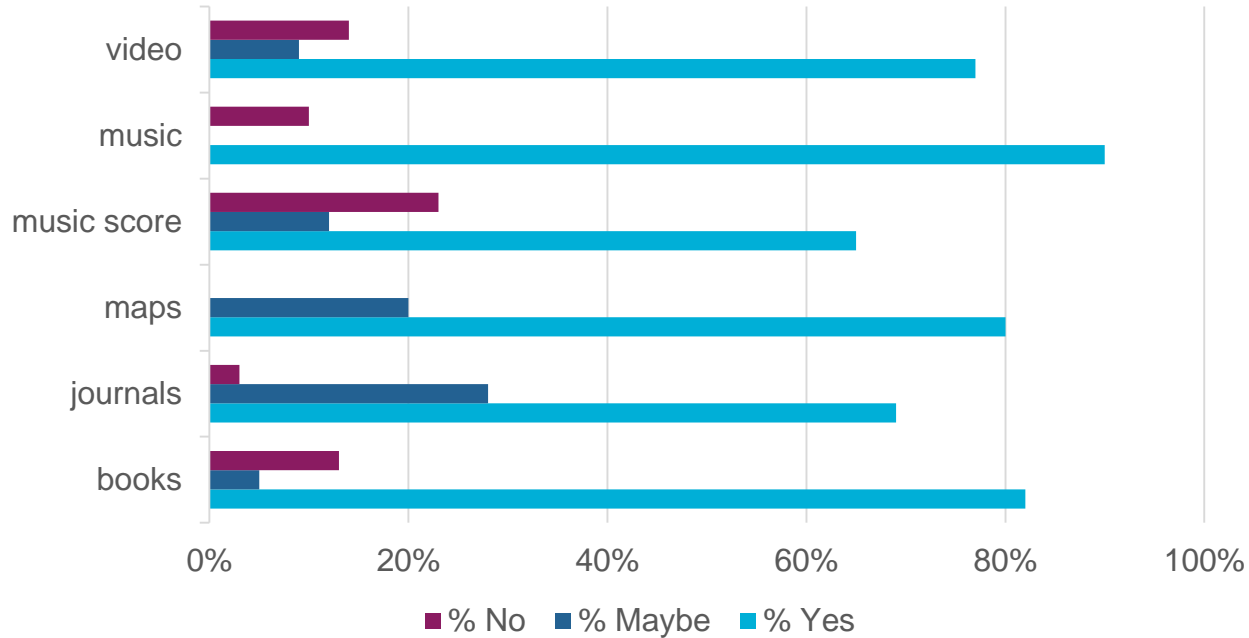


Accuracy of cluster

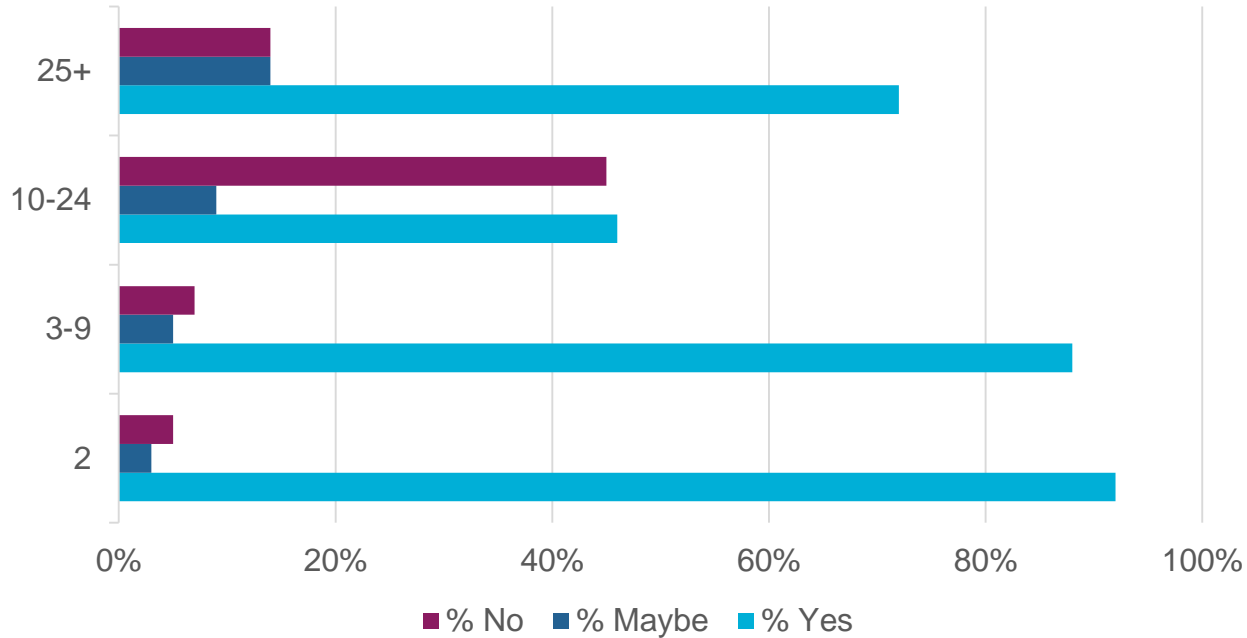
Cluster analysis – by groupings



Cluster analysis – by format



Cluster analysis – by sibling count



Next steps

- Finalize which Work IDs to add to bib records in 758 field
- Develop implementation options
 - Populating data
 - Keeping data up-to-date
- Work with PCC on best practices
- Future review – look at under-clustering

OCLC Research Activities and IFLA's *Functional Requirements for Bibliographic Records*

<https://www.oclc.org/research/activities/frbr.html>

Starbucked cluster example

<http://www.worldcat.org/title/starbucked-a-double-tall-tale-of-caffeine-commerce-and-culture/oclc/227018423/editions>

Works in Progress Webinar: Using OCLC Work IDs for Discovery

<https://www.oclc.org/research/events/2017/12-12.html>