Disco Open Forum – VE Dedup issues [update] [October 8 2021]

The problem

Several CSU-wide reports of e-inventory erroneously 'merging' with other bib records. Titles are similar in that they share common words in the 245 $a, but the works are <u>clearly unrelated</u>.

What does this look like?

Well, a classic (aka utterly horrendous) example may be found [here](#)

What initially seems to be a singular work (a streaming video entitled **Becoming American**) actually also displays access portfolios for **two** other unrelated eBooks (**Becoming American***: Why Immigration Is Good for Our Nation's Future* and **Becoming American***?: The Art and Identity Crisis of Yasuo Kuniyoshi*)

Notably, the *unrelated* bib records remain intact, un-merged and un-nested ([see example here](#))

So VE is broken then, yes?

Hmm. Seemingly yes and no. Maybe.

Here's the relevant [documentation](#) entitled 'Understanding the Dedup and FRBR Processes (Primo VE)'

Regarding Dedup: "In the search results, records that are marked as duplicates are displayed as a **single record**. The metadata is displayed from the first record in the result set and **the delivery related information is created from all the records in the group**."

That sounds both familiar and very disappointing, doesn't it?

But…

The documentation goes on to explain that the Dedup algorithm creates keys from – among other data elements including various unique identifiers – title "245 a,b,n,p"

Brandon posed the excellent Q – does that mean 245 a **and** b subfields etc., OR 245 a **or** b subfields etc.?

We're not sure, and have queried Exl Support on this in an ongoing case (00996668), but I suspect each and every one of us would say it <u>should</u> be subfields a **and** b etc.

Indeed, I've not seen a broken example so far that would display such issues if 245 a **and** b were both taken into account

So, what can we do?

First of all, we need ExL to reply to the latest query in the SalesForce case.

Also, it would be useful to obtain a better understanding of the scale of the problem. Title-matching (especially limited to the $a) is by definition very imprecise, so it is likely the issue described above may be affecting a large number of records. Though bad, this could also assist us in terms of leverage to change VE behavior if indeed it is 'expected'.

So, please feel free to share examples with me:

sbarber@fullerton.edu

samuel_fullerton (CSU Slack)