Book Data Cleaning, Spot Checking, and Filtering

= Have not enacted

Alicia’s Initial Findings

1. 6011 Records with Non-English Language
   1. Many have English metadata - Could do some manual work here to get the most we can
2. 1338 are blank in the description field and will need to be removed
3. 4624 do not have Genre information
   1. Mix of fiction and non-fiction and different languages in these
      1. Will make it very difficult to filter what's for us unless we do it manually
   2. Does include titles for your enjoyment such as: The Evil Hobo, As the Matzah Ball Turns, What they aren't telling you about Vaccines, and many Christian liturgical
4. 314 books with a first publish date of 2023-2029.
   1. This could be a misprint or they haven't been released yet.
   2. These still have descriptions, but should we suggest books that haven't been released yet?
5. Will need to filter on Bookformat
   1. Album, Newsprint, Newsletter Serial, and Wattpad will need removed
   2. Wattpad had a little smut and some poor writing
   3. Board Books - removed all
      1. So some of these are mislabeled as board books and are good
      2. Could filter on if not Young Adult and Is Board Book then remove
         1. Complicated and not worth saving the three books
6. What are your thoughts on:
   1. Poetry - it's not fiction but it's not nonfiction and much harder to tag with a life event
   2. Multiple books listed in a single entry
      1. Some of these are Boxed sets or a series which we can just use everything it gives us or remove
      2. Some books are in there under the single and under the set
   3. Swear word title filter
7. Genre for BDSM and Erotica will most likely need removed, but I’ll review some
   1. 2028 Erotica Books
      1. Some also labeled as New Adult and Young Adult and as Erotica
   2. Must remove Kama Sutra