Acting on Lessons Learned: Tracking Databases for Long-term Scanning Projects

Background

Resources are scarce for most research libraries and the enthusiasm among grant providers for digitization projects is not what it used to be. For any substantial collection of theses and dissertations, this means digitization could require many years to complete. More importantly, for planners and managers of these projects, long-term efforts like these must all agree that we can prepare for changes like newly discovered titles to add to the system, major shifts in production processes, and migration to new ILS platforms.

At UF’s Smathers Libraries, we are preparing to conclude dissertation work and begin mass digitization of our master’s theses. Over the 10+ years of our project, we have had to address many changes to workflows and resources. We are applying that experience to the design of a new tracking database to use in upcoming years. This poster provides an overview of the scope and scale of the project as well as how and why the new Access database differs from the original one.

A likely change factor you will need to manage: steadily finding more dissertations to process

Dissertations in Tracking System

When the project began in 2008, the first search for dissertation records in the catalog yielded 8,163 titles. Using the opt-in model, the team added catalog records to the tracking system as they reached out to authors, hitting a total of 12,112 in 2011. In 2012, we shifted to an opt-out model and by then the best catalog search we had yielded 13,264 records. We loaded all of those that weren’t already in the system. In 2015, we learned that all 1,117 Doctor of Education dissertations had escaped earlier searches. Through the next four years, further refinement of catalog queries and discovery of cataloging errors brought our total number of dissertations to 14,115. We expect to find a few more.

The numbers story

The original master table has 176 fields, only 79 are still active

- The Tracking role encompasses 51% of the active fields

In the new structure, many tracking fields are in related tables

- This added a new field role, Relationships

The current version of the new structure includes fields (and an entire table) that we might eliminate after we have processed a few hundred titles.

Field Roles in New Master Table

<table>
<thead>
<tr>
<th>Role</th>
<th>Original</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracking</td>
<td>51%</td>
<td>37%</td>
</tr>
<tr>
<td>Author Contact</td>
<td>19%</td>
<td>19%</td>
</tr>
<tr>
<td>Production</td>
<td>17%</td>
<td>12%</td>
</tr>
<tr>
<td>Relationships</td>
<td>10%</td>
<td>17%</td>
</tr>
<tr>
<td>Filtering/Sorting</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>Mapping/Production</td>
<td>4%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Our core planning advice for project tracking databases:

Apply relational database basics:

- Reduces duplicate data
- Can help improve query performance
- Avoids scrolling through long field lists when building new objects

Do some ‘future proofing’:

- Assume that your library will move to a new ILS during the project
- If you start with vended scanning only, plan for eventual in-house work; vice versa if you start with only in-house work

Add many timestamps and user identifiers:

- You can always sunset them later if the effort of using them during work exceeds their value for reporting
- Line-level workers are certain to change over time, data like this can help identify problems with trainee workers before they affect large numbers of titles

Poster available at [http://ufdc.ufl.edu/AA00068847](http://ufdc.ufl.edu/AA00068847) and related presentation at [https://ufdc.ufl.edu/AA00068851](https://ufdc.ufl.edu/AA00068851)