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

G.W. Swicord
September 27, 2019
USETDA Charleston, SC
Introduction

This presentation will cover:

- A brief history of the UF Libraries’ long-term theses and dissertations scanning project
- Things that we could have done better during the first decade
- Ambitions for our new project tracking database
UF RTDS project history

• The Retrospective Theses and Dissertations Scanning (RTDS) project began in 2008
  o Until 2012, we operated under an opt-in model
    ▪ A catalog query identified 8,163 dissertations and we received exported MARC records for them
    ▪ Titles did not enter production until we had contacted the author and secured permission to publish online
    ▪ With each opt-in, we loaded one of the 8,163 dissertation titles into the work-tracking system
      o At the end of the opt-in period, we had loaded ~6,806 titles to the tracking database
UF RTDS project history

After the 2012 shift to an opt-out model

- We had an improved catalog query that yielded 12,112 bound dissertations
  - We loaded all of those that were not already in the tracking system
- In 2015, we learned that we needed to load 1,117 Doctor of Education (EDD) records into our system
- Since adding the EDDs, better queries and catalog record corrections brought our total to 14,115
Project management and reporting would have been simpler if we had:

• Used established workflow terminology for internal scanning, e.g.
  o Tracking Box as the container for one or more related items
Standardizing field names

Project management and reporting would have been simpler if we had:

• Matched more database field names to source field names for sequential data loads
  o Catalog exports
  o Contact information lists for alumnae/i

• Synchronized related database field names
  o BibID (CMS database) vs. Bib ID (older database)

• Clearly distinguished same roles for different contexts
  o Vended scanning vs. internal scanning
    ▪ Shipment IDs vs. Tracking Box IDs
    ▪ Internal needs to match our UI, external the vendor’s process
    ▪ Keep separate vendor tables to track batches sent to them
Questions, Questions, Questions

• We want to be prepared for complex questions about work done years ago
  o Types of questions, fields for answers
    ▪ When was that added? Who added it?
    ▪ How long did we use that vendor?
    ▪ Where did we import this info from? When?
Questions, Questions, Questions

• Reporting and storytelling data needs to illustrate challenges and successes with:
  o Managing multiple physical copies to complete a digital item
  o Training new staff
  o Changing/adding workflows
  o Tracking all production aspects of vended scanning
Did I mention...

• Keep up your documentation!
  o We all know this is often very hard to do, but...
    ▪ Write a memorandum of understanding (MOU) for the project as if you knew the rough time-frame and stakeholders
      o All that will change, but you’ll have a foundation in writing
      o When major changes such as new technology or staff reduction/expansion happen, revise the MOU
    ▪ Map every field to specific workflow steps and/or queries
    ▪ Keep a change log for your data structure
    ▪ Archive at least one backup of your database every month (not strictly documentation, but creates a historical record)
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 possible 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 of them during work exceeds their value for reporting
- Line-level workers are certain to change over time, micro-workflow data like this can help identify problems with trainee workers before they affect large numbers of titles
Thank you for attending!

• Our scanned and born-digital TDs are at http://ufdc.ufl.edu/ufetd

• PowerPoint and PDF versions of this presentation will be available at https://ufdc.ufl.edu/AA00068851/
  o I hope to add a set of rich notes/mini-paper as a download on that item

• My contact info:
  G.W. Swicord
gwswicord@ufl.edu
  352-273-2763