E-Resource Cataloging Challenges and Solutions

- Challenge 1: **SLIDE 4**
    - How to make e-resources discoverable?
        - To begin, we need to consider how to make collections of e-books, streaming video, and streaming audio discoverable to our library users.
- Solution 1:
    - MARC records
        - One of the most common ways of proving access is by loading MARC bibliographic records into the library catalog.
    - Discovery Services
        - Alternatively, it may be possible to make e-resources discoverable without loading MARC records by using a discovery service.
        - For example, EBSCO can make your EBSCO e-books automatically discoverable in EBSCO EDS.
        - If you use WorldCat Discovery, e-resources can be discoverable by selecting the collection in the OCLC knowledge base or setting up an automatic feed from the vendor.
    - At UNT Libraries, we provide access to our e-resources by loading MARC records, so that will be the focus of our presentation.

- Challenge 2: **SLIDE 5**
    - From where to obtain MARC records?
- Solution 2:
    - When the Collection Development Department at UNT acquires a new electronic resource, they notify me as the Head of Cataloging and Metadata Services that MARC records need to be loaded. We usually obtain our MARC records directly from the vendor or from OCLC's WorldShare Collection Manager.
    - E-Resource Vendor:
        - Almost all vendors provide free MARC records.
        - However, the quality of the records can vary significantly from vendor to vendor. Some vendors provide excellent MARC records that conform to standard cataloging practices and have proper Library of Congress Subject Headings. On the other hand, some vendors provide machine-generated MARC records that are, shall we say, less desirable and may only provide uncontrolled keywords instead of proper subject headings.
        - When compared to WorldShare Collection Manager, there are a few drawbacks. For instance, each vendor has a different location where you have to go and download records. Each site functions differently and may require a unique login. The site may change without notice or may not work as expected.
        - There are also a few advantages to vendor records over Collection Manager. There is usually not a lot of initial setup and records can be downloaded immediately. Vendor records also tend to be more uniform, which can make batch editing easier.
    - WorldShare Collection Manager: **SLIDE 6**
        - WorldShare Collection Manager is included as part of your OCLC cataloging subscription. When a collection is selected in Collection Manager, you can

enable the delivery of WorldCat MARC records for titles in the collection with an OCLC control number.

- Unlike most vendor records, all records obtained from Collection Manager are obtained from a single location. There are a number of customizations that can be applied across all collections, such as adding or deleting certain MARC fields. For each collection, you can determine the delivery frequency of records from daily, weekly, monthly, or quarterly.

- For knowledge base collections, you will receive new, updated, and deleted records. You may customize the settings so you only receive updated records for changes to fields in the WorldCat master record that are important to your library.

- While there are many positives to using WorldShare Collection Manager, there are a number of drawbacks. First, when selecting a new collection, there is some initial setup and it can take a couple days for the records to be delivered. Further, records will only be delivered for titles with an OCLC control number in Collection Manager. If an incorrect OCLC number was added, the wrong record will be delivered.

- Challenge 3:
  - How to organize e-resource MARC files and supporting documentation? **SLIDE 7**
- Solution 3:
  - So you've downloaded the file of MARC records, but where should it be saved?
  - Currently, at UNT Libraries, we have a main E-Resources folder in the department's shared drive that contains subfolders for each package or collection. The folder for the collection always includes a procedure and a MARC Records folder.
  - For new collections, I create the initial folders and procedure before delegating to a cataloger. The initial procedure includes instructions for obtaining the MARC records, adding the local collection note for tracking the records in the catalog, and adding the number of simultaneous users. The cataloger completes the procedure by documenting the edits made to the records before loading.
  - To reduce the number of files in the main folder and make it easier to edit and load new files, records that have been loaded are kept in the MARC Records folder.
  - This file setup makes it easy to shift record loading responsibilities to accommodate changing workloads or extended absences.

- Challenge 4:
  - How to track e-resources cataloging? **SLIDE 8**
- Solution 4:
  - With numerous collections requiring MARC records to be loaded into the library catalog, it can be extremely difficult to keep track of them all.
  - Some institutions have used project management software like Trello, Jira, Basecamp, or Freedcamp to help manage their electronic resources cataloging. These tools have features that make it easy to assign tasks and set deadlines. It is also possible to schedule tasks, such as checking for the latest update file. However, it can be very time-consuming to setup the software and it can be difficult to get people to use it and keep it updated.

- o **SLIDE 9** At UNT, we use an Excel spreadsheet. While it doesn't include the notifications of the fancy project management software, it is extremely easy for people to use and update on a regular basis and we've customized it suit our needs.
- o Calendar **SLIDE 10**
  - The Calendar tab of the spreadsheet tracks all collections that have had MARC records loaded since 2016, including collections that have updates, that are complete (without updates), and that we cancelled. Including complete and cancelled collections is useful when delegating new collections. It is also possible that collections may have updates, even if we were not expecting them.
  - For each collection, we include:
    - how often we check for updates,
    - the date we last checked for updates,
    - the date to check for the next update,
    - what the expected next update is (for example, the month of the update),
    - the cataloger responsible for editing and loading records for that collection, and
    - where to obtain the MARC records.
    - Not shown on the slide is a notes column, which may include login information.
  - Having the login information included in the spreadsheet makes it efficient for the catalogers to check for updates and update the spreadsheet.
  - When checking for updates, the cataloger can easily filter the spreadsheet to their name and sort by date to identify collections that need checking.
- o New Records **SLIDE 11**
  - We also track records that have been loaded by recording the file, number of records in the file and loaded, the date they were loaded, and by whom. For new records, we also track the starting and ending bibliographic record numbers.
  - This information is helpful for a number of reasons. First, our Calendar tracks how frequently we check for updates, but it doesn't indicate how often updates were loaded or how many records were included, which can be useful if record loading for a collection needs to be reassigned. Second, if a problem is discovered after a record is loaded, we can use the information in the spreadsheet to help investigate. For example, we can go back to the original file and see if the problem was part of the original file or was introduced during the editing process.
- o Similar to New Records, we also track records that were overlaid and records that were withdrawn.
- o A blank spreadsheet is available as a handout as part of this presentation.

- Challenge 5:
  - o How to delegate to catalogers for editing and loading? **SLIDE 12**
- Solution 5:
  - o As the Head of Cataloging and Metadata Services, all new e-resources come to me first. I decide where to obtain the MARC records, create the necessary folders in the shared

drive, begin the initial procedure, add the collection to the Calendar tab of our spreadsheet, and then delegate to a cataloger in the department to edit and load the records.

- There are a number of factors that I consider when delegating a collection, but the two biggest are probably the format of the resource and the source of the MARC records. For instance, streaming video or streaming audio collections may be assigned to catalogers with experience cataloging those formats. Vendor records may be delegated to a cataloger with experience editing records from that vendor. I also consider if the cataloger has experience editing similar types of records. If a cataloger was assigned a collection from one vendor, they may be assigned a similar collection from a different vendor because there may be similarities between the records. I try to be aware of existing and upcoming responsibilities of the catalogers so the work can be spread across the department. If a collection has updates, I want to ensure that the cataloger is able to handle the number and frequency of the updates. If it is a very large file, it can be more difficult and time-consuming to edit and load. Lastly, catalogers in the department naturally have different strengths, so the complexity of the edits may influence to whom I assign the collection.
- These factors are considered holistically. Depending on the collection and department priorities, greater emphasis may be placed on certain factors over others.
- When delegating a collection to a cataloger, I send them an email. In addition to letting them know the name of the collection, I let them know if I noticed anything in the records that may be problematic, if there is a procedure for another collection on which they can base the procedure for the new collection, as well as any special instructions for loading.

Now Stacey will discuss some of the decisions the catalogers consider when determining the level of edits to make to the records.

1) **SLIDE 13 Are these Discovery Records?**
   a) The records are just there so users can discover resources we could have access to. Most of our discovery records also get updated when triggered for purchase as part of the invoicing processes.
   a) So we decided not to spend much time on the discovery records and wait until they are purchased before doing more thorough clean up.
   b) So for these packages, we
       i) Create as many tasks as possible to make them quicker to edit.
       ii) To prevent the task from causing or introducing problems, keep the edits simple: like add/delete fields, or simple find/replace actions.
       iii) Occasionally, there just isn't a good way to edit these records in a batch process and you have to decide if you'll accept the problems or spend the time to correct each record individually.
       iv) In these situations, think of the patron and how the record might affect their access to the resource.

2) **SLIDE 14 How does the vendor provide updated or new records? Does the vendor separate updates, additions, and deletes into separate files? How will you load these into your ILS?**
   a) Depending on your load tables in your ILS, separated updates and additions can allow you to make different edits to the different files.
   b) If all of the records come in one big file, and you don't know what's new, updated, or deleted, then it may make more sense to overlay all records in a collection every time.
      i) We treat these collections similarly to discovery records, simplifying and automating the editing process as much as possible.

3) **How often are the records updated or how often do we get new records in the package?**
   a) For packages that are updated or added to monthly or frequently, we might do less editing.
      i) It's a waste of time if the records are constantly updated since you lose your edits with every update
      ii) You want to limit the number of edits as well because you will have to do them every time you load records
      iii) Tasks are another great tool in this instance and allow you to do more edits in less time.
   b) However, if there are no updates or new records coming, it may mean we can spend a little more time on the edits without the fear of losing the work we put into it later on down the line.

4) **SLIDE 15 Are the problems with the records consistent?**
   a. Are there patterns to the data that make it easy to make changes using Regular Expressions or other features of MarcEdit?
   b. Is it going to require touching every record to make edits one by one?
   c. If so, how important is that field to discovery for the patron?
   d. Basically, is it something that can be changed without introducing more problems to the record?

In the end, really, it depends on your local standards, knowledge, and willingness to crack open a file of records and manipulate the data.

Kevin's Section

1) **SLIDE 16 Why we need to edit**
   a. Help users and enhances discovery
   b. Helps create a clean and uniform catalog
      i. Users know what they are looking at

1. If every record looks similar in style and layout, users can more quickly and clearly understand what information is being presented to them
2. Example, If punctuation is not correct, a name could display in reverse order and confuse the user

ii. Ensure cataloging rules are being followed
   1. Such as, cataloging using Provider-Neutral E-Resource MARC record guide: P-N/RDA version (2019) from the Program for Cooperative Cataloging
      a. This document is linked to in our resources slide and handout

c. Delete obsolete data
   i. 245 $h
      1. 245 = Title Statement
      2. $h = general material designation or "medium"
         a. Example "[electronic resource]"
   ii. Correct errors
      1. 33X fields
         a. The Content, Media, and Carrier types
      2. If we know a record should have XYZ, we can more easily identify when it does not
      3. If we know a particular edit should edit a field in a certain way, we can troubleshoot the field (or the edit) if it does not perform like it should
      4. Remove non-standard subject terms
         a. Either because they are incorrect or because we have chosen to not use them
   iii. Supply missing information
      1. Call numbers for organizational purposes and collection evaluation purposes
      2. Easily inserted standard fields
         a. 006 and 007
            i. 006 = Additional Material Characteristics
            ii. 007 = Physical Description Fixed Field

   iv. Insert local fields
      1. Edit URLs to conform to local needs
         a. Proxy

b. Display text
2. 9XX fields to include local information
3. 9XX fields that allow for the creation of records or setting of standard data
   a. For us
      i. 949 creates and set fields in an item record
      ii. 999 sets standard fixed fields in the bibliographic record

2. **SLIDE 17** Common edits/significant fields
   a. If there are edits that can be easily made, either because they are easy or because they are important, they should be included in most workflows
      i. Easy edits
         1. **SLIDE 18** 006 and 007
            a. 006 = Additional Material Characteristics
            b. **SLIDE 19** 007 = Physical Description Fixed Field
            c. Depending on the type of material you are cataloging, these can be relatively standard
            d. If you are getting records from vendors, maybe they can be automatically included. It does not hurt to ask
            e. If you are adding them yourself, determine if there is a standard format you can use if records do not already have them
         2. **SLIDE 20** Spacing in all fields
            a. Removing extraneous spaces or ensuring things are in the correct can be achieved as a batch process
            b. There are tools and methods out there, more of these will be discussed by my colleague in the next section
         3. **SLIDE 21** Regardless, before deciding to make even these easy edits, talk to people
            a. See if the 006 and 007 can be used by your OPAC
            b. How does your system handle extra spaces in fields?
               i. Are they included?
               ii. Are they ignored?
               iii. Do they affect the public display?
   b. **SLIDE 22** Significant fields
      i. These fields are determined either by common library practice (Library of Congress, RDA) or local practice
      ii. In UNT's case, we have a mixture of both
         1. We base decisions off of RDA, LC, etc, but we also have made local decisions based around our needs and/or past practices
         2. Regardless, significant fields typically revolve around user access to information and because they affect the user so much, they usually require more editing time and/or attention to detail
         3. **SLIDE 23** Examples of UNT's significant fields (not exhaustive)
            a. **SLIDE 24** 100 – Main Entry – Personal Name
               i. Should be the authorized form

ii. Correctly punctuated

iii.   Controllable in OCLC if applicable

b. **SLIDE 25** 245 – Title Statement

    i. Spacing and punctuation are key

    ii. If a subfield is miscoded, display is affected

c. **SLIDE 26** 264 - Production, Publication, Distribution, Manufacture, and Copyright Notice

    i. If a subfield is miscoded, display is affected (check)

    ii. Important for those needing specific versions of an item

    iii.   Important to the user

d. **SLIDE 27** 300 – Physical Description

    i. This is important because it can help the user know if this is a resource they need

    ii. Does it include content that will be helpful?

        1. Illustrations

        2. Portraits

        3. Maps

    iii.   Is it online?

e. **SLIDE 28** 856 - Electronic Location and Access

    i. This treads line between and easy edit and a significant field

        1. Easy, because URLs have a lot of standardization in formatting

        2. Significant, because they provide access

        3. If a URL does not link to the correct place or is "broken" due to incorrect formatting, it is useless to the user

        4. Also, as is the case at UNT, we can format our URLs to pop up in a new tab

            a. We do this so that Users are not taken away from the catalog and can keep browsing from where they were instead of having to navigate away from the e-resource or start their search over

MarcEdit- this will not be a guide to how to use MarcEdit, it's about how *we* use MarcEdit.

- **MarcEdit main page slide (30)**
- Key to batch editing—could never catalog that many new resources (hundreds/thousands) in a timely manner otherwise.
    - **ME Lots of records w/highlight. (31)**

- Especially important given how bad many vendor records are. It's not just the quantity of records, it's also the quantity of fields in any given record that need to be edited.
  - **SpringerPages3 (24 pages) (32)**
- Very granular.
  - Can edit:
    - Fields
      - **Field Edit (33)**
    - Subfields
      - **Subfield Edit (34)**
    - Indicators
      - **Indicators Edit (35)**
  - Editing options:
    - Add
    - Delete
    - Swap
    - Add/Delete/Edit only if field/subfield contains particular pieces of data.
  - Call numbers and Cutters
    - Call numbers might not seem to matter for eBooks, but they're important for collection assessment.
      - **ME Call Number and Cutter (36)**
    - MarcEdit can retrieve call numbers from OCLC. Note options for which fields, subfields, and indicators to use.
      - **ME Call Number Assign (37)**
    - MarcEdit can create Cutters. Note option for which field to use.
      - **ME Cutter Assign (38)**
  - Validate Headings
    - You can make sure that names, corporate names, and subject headings are correct.
  - Programmable Tasks:
    - Automate things—example of 6xx Deletes
      - 6xxs are the subject fields. UNT has local policies that mean we delete a bunch, since there are terms from some controlled vocabularies that we don't accept.
        - **6xx Deletes2 (39)**
          - This is about 2/3 of the deletes the task does. It takes maybe 2 or 3 seconds.
    - But you can also automate the entire editing process if, as Stacey said earlier, you have packages that you update regularly and lose all the edits anyway.
      - This is especially important if you have a small staff. Lose the ability to check each field individually, but you can save a ton of time.
- RegEx
  - You can do some pretty specific things with MarcEdit as is, but you can manipulate it even more powerfully using Regular Expressions.

- RegExs are "a sequence of characters that define a search pattern" per Wikipedia.
- Here, no comma between the name and $e
- **RegEx Slide (40)**
  - Especially handy for fiddly punctuation and changing things at particular points like 008, which has 40 different variable positions.
    - If ME's already a scalpel, this makes it laser surgery.
      - **RegExCheatsheet (41)**

Excel

- Speaking of 008, that one's tough to eyeball.
  - You can line up 006 and 007 and pretty easily see if they're uniform, but 008 varies in length for acceptable reasons like a single publication date.
    - **006 and 008 lineups (42)**
  - Excel lets you create separate columns for a position or group of positions. Then you can sort and filter.
    - You can either copy or download the data.
      - **Copy and download slide (43)**
    - **Data from text slide (44)**
    - **Text To Columns (45)**
  - Example of postions 23 and 32, which we always check for in eBook records—23 should be o for online, and 32 should be undefined.
- Also good for finding duplicate data—e.g. records with multiple URLs, check for abnormalities, see what's changed.
  - **Multiple Fields (46)**

Resources

- MarcEdit listserv- ask Qs of the user community, gets very high level.
- PCC Provider Neutral Guide- for creating PCC-grade records.
- MarcEdit Development- keep up on new versions and tweaks.
- Worklog- detailed accounts fo what's new and planned with ME.
- YouTube channel- ME videos.
  - **Resources (47)**

Back to Stacey to wrap up.

- **Credits (48)**