**Processing Masters Theses RTDs for ingest into the DR**

**1.** After receiving notice via email or Trello notification that a new set of files is ready, enter the batch or shipment information into the “masters-RTDs-workflow.xlsx” tracking spreadsheet.

**2.** Locate files on the “DR sharing” network drive. Copy the file path.

**3.** Run the Python script “pdf\_pages.py” to get a page count for each pdf, using the file path from step 2.

**4.** The file names in this output data should correspond to the Aleph Bib number of the physical Masters theses it was created from. Use these file names to create a list of Bib numbers to search to pull MARC records for these documents. Save this list as a plain text file (.txt).

**5.** Use MarcEdit to extract these records from the larger file of all theses records from Eric P. (from Aleph) called “kt\_thesis\_list.mrc”. DO NOT delete these files from the original record set; save them as a separate record set. **Use the MarcEdit syntax *FILE#:[filepath]* to search for the bib numbers in the .txt file from step 4.**

**6.** Convert these MARC records to a delimited text file using MarcEdit. Open this delimited text file in Excel. To ensure you didn’t get false hits on the Bib number search, insert a new column into the far left of the spreadsheet and paste in the bib numbers from the old plain text list. Delete any rows in the spreadsheet that are of MARC records not intended to be in the set.

**7.** Use this MARC data to create metadata suitable for upload to the digital repository. Use the “rtd.xls” template. You will need to do some data transformations on titles, names, and dates.

Use information from the 245 field for title, author names, and rights holder fields.  
Use the 260 field for publication and copyright dates.  
Use the pages data from the python script output for the file\_size.  
For years 1858-1959, institution should be Iowa State College ; For years 1959 to present, institution should be Iowa State University.  
Use aleph\_control\_number, not the field for aleph\_bib\_number.  
Use the 502 field to obtain abbreviations of degree names. Use “degguide.txt” as a reference to find/replace these abbreviations with the full degree names.   
Use the values of any 6XX fields concatenated together with commas for the keywords field.  
Document type should be “thesis”, language should be “en”, publication date format should be “YYYY”, publisher should be “Digital Repository @ Iowa State University, http://lib.dr.iastate.edu/”.

**8.** Upload the files to the lamp server at <http://lamp.lib.iastate.edu/DR/> . Place them in a folder with the batch or shipment number. Update the URLs in the metadata spreadsheet you have prepared and check one to ensure they’re working.

**9.** Once you have the metadata and file access prepared, you can assign the quality control and subject analysis steps to someone in the Metadata and Cataloging department. There is a separate set of procedures written especially for this step called “new masters ETD procedures.docx”.

**9.** Once you have received this metadata back, check any lists of theses with Interdisciplinary Majors that you have to confirm that these documents have any interdisciplinary majors listed in the “major” column and to account for them on these lists.

**10.** Do a final quality control check by glancing over the metadata to ensure completeness and consistency. Upload the metadata to the “rtd” series in the digital repository.

**11.** Update the workflow tracking spreadsheet along the way to track progress through the workflow. Once the RTDs have been uploaded into the DR, spot check a few from the list to ensure they are appearing correctly in the RTD series.

**12.** Move the corresponding batch or shipment card in the Trello “Masters Theses Workflow” board to the “Uploaded to DR and Done!” column. Celebrate that you have assisted in providing open access to research.