Procedures for processing ETDs received from ProQuest

1. Notification email that files are ready from Proquest. IT moves them to the DR drive and sends second notification email.

2. Make a copy of the folder of files and place it in the DR_Sharing drive. Unzip the zipped folders in the batch using the first step of the ETD-CON tool (or a command line utility).

3. Add the batch to the "new-ETD-workflow" tracking spreadsheet.

4. Make a copy of the XSLT "ETD-ProQuestXML2bepressXML" and add to the file name a hyphen followed by the batch ID (something like "r20150618"). Edit the XSLT with the correct file path for the full-text URL.

5. In oXygen XML editor, configure a transformation scenario for all files in the XML folder, using the XSLT from step 4 and the Saxon-PE processor. Specify the output location to be a new folder in the batch's root folder called "transformed XML" and for the output file name to be ${cfn}-out.xml [current file name with 'out' appended]. Apply the transformation.

6. You can take a look at this data to make sure it looks normal. You should check a few ETDs with embargos specified to make sure that the embargo date transformation worked correctly.

7. Split the received files into smaller batches for QC by the catalogers. Upload the corresponding PDF files to the LAMP server, making sure you put them in a folder that will agree with the full-text URL you specified in the XSLT in step 4, or you will need to edit this in the data.

8. Compile the XML files for the batch into one XML document using the "merge.xsl" transformation. You only need to configure the transformation scenario on the first file in the folder, since the XSLT will iterate through the files at the folder level. Name the output file "[batch name]-merged.xml".

9. Add the <documents> tag from a merged XML file from a previous batch to add the correct namespaces, etc. (I know this should be in the XSLT, but I can't figure out how to make that work, so this is just as efficient for me right now! Pull requests welcome. ☺) Example:

```
<documents xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="http://www.bepress.com/document-import.xsd">
```

Validate each batched file to ensure the vocabularies were transformed satisfactorily. If errors are present, edit the "ETD-ProQuestXML2bepressXML" XSLT as needed.

10. Run the XSLT "ETD-bepressXML2Txt.xsl" on the new merged XML file. Use Excel to open this file as data. Be sure to select UTF-8 as the character encoding in the first screen of the data import wizard.

11. Format the spreadsheet for optimal viewing.

12. Assign the batch to a staff member for quality control "copy-cataloging". Place the spreadsheet for them to work on in their folder on the DS_sharing network drive in the "ready for catalogers" folder. Note this information on the workflow tracking spreadsheet. There is a separate protocol for the work the support staff will do at this stage entitled "Procedure for Quality Control of ETD Metadata.docx" which details the quality control checks to be performed.

13. After receiving notice from the support staff that they have completed the QC check, make a copy of their spreadsheet and place it in the batch's folder on the DR_sharing drive. Manually input the necessary changes marked in yellow on this spreadsheet in the 'merged' xml document from step 8. Validate the XML and save the file.

14. Upload as XML to the repository per usual protocols.

15. After receiving email notice that the series in the DR has updated, spot check a few titles to ensure everything seems ok. Check in particular an ETD with an interdisciplinary major to see if it is showing up on the appropriate page, as well as the home department page. Mark as done on workflow spreadsheet and celebrate providing Open Access research to the world.

16. At some point, files received from ProQuest should be archived for digital preservation, but that aspect of this workflow is not currently developed.