

## **Dege Kangyur Proofreading Workflow**

As of 1 June 2013. Note that we are evaluating this and modifications might be made, particularly in the number of manual inputs being used.

Esukhia: 18 proofreaders (May 2013) and 12 proofreaders (June 2013) and 2 supervisors

Each proofreader proofs an entire volume

### **Approved workflow steps**

1. THL converts Word docs of input 2 into XML
2. THL exports the input 1 and input 2 XML files into a simplified XML format: a single XML file for each volume. The only markup will be square brackets for page and line numbers, and text begin and text end (and if the inputters marked any text as illegible or unclear).
3. Esukhia strips markup from THL input (either first or second input; volumes 1-61 already done w/ first input as of 26 June 2013) and in Word compares against Unicode OCR (the Russian OCR; this OCR has been "proofed" by a Python script that makes systemic changes) of the Dpe sdur ma. For each difference, look at LOC scan and decide which reading is correct, and edit Word doc. Output is .docx files (one file for each volume)
4. Esukhia compares in Word the proofed text (this is output of input 1 compared to OCR) with second THL input, checks each difference against LOC Dege Kangyur scan, and corrects the base edition (the proofed version)
5. Copy and paste compared text from Word file into InDesign
  - a. Manually add carriage return for line breaks in InDesign
  - b. Word-by-word proofreading against LOC scans in InDesign
6. Copy and paste from InDesign into a single column of Excel file
  - a. Add a second column to the left of column with text; add page and line markup.

- b. Combine the two columns; check to make sure page and line numbers are consistent with scans
  - c. Export into .txt file
- 7. Communication feedback loop - volume report from Esukhia. Paldor/Steven spot review 5 pages per volume; each volume is published in its own directory in Dropbox with a review file that contains the reviewer and comments.
- 8. Esukhia incorporates feedback.
- 9. Esukhia delivers final text file with tags on a volume by volume basis.
- 10. THL converts final text with tags volume into TEI/XML with bibliographic control from TBRC.
- 11. THL break each volume file into individual text files.
- 12. Final source control (volume vs. text, XML vs TXT) form still under discussion.

For now, we will source control everything (TXT, XML, PDF) NT's team outputs.
- 13. Return the merit of this activity to the benefit of all sentient creatures.

### **Needed**

- From Jeff: All LOC scans in a light format from TBRC
- Bad quality scans of a Derge edition held by UVA FROM THL
- 1st UVA input by volumes without milestones - THAN