**Transcription discussion 10/15/15**

**Mike, William, Trish, Joel, Bianca, Carolyn, Joe**

Text - Includes OCR, transcriptions

How does markup relate?

Should we add functionality to adm dashboard?– text editing of a single page or upload an OCR page.

Uploading page Steps: Specify format (from a drop down list), upload, review and approve

Editing a page – searching title, items, select a page, edit text of page. Save and cancel

Questions:

* Should we define a text format for others to use?

We could start with the 4 sources we have so far: games, SI transcription, DigiVol, FromThePage. The different formats each require changes to the importer to be able to recognize. They then get all converted as .txt files and overwrite the BHL version of the OCR file. Lets go with formats we know for now and address custom as they come in

* Do we Need a history log of who uploaded and changed what in adm dashboard?

Should adm be able to view the history? We upload at an item level but each page will have information about its history. Joel says we need the full history. With SIL webpage Drupal keeps track of every change and when and can revert backwards. Do we save full text or each version is up for grabs.

What features do we want an adm dashboard and how much info do we want to save? Last version only, last 10 version?

Mike says we really only need Record of when a change occurred and who by. We agreed to keep whole history of the page, who did upload, when it happened, source of data.

Not a matter of quality we are concerned about but completeness.

* IF pages are missing in the upload should we allow it? Yes because some books have illustrations and don’t need to be uploaded (only the text pages). This might argue for a defined format so we can require structuring or sequencing of pages.

Should we allow users to re-align pages after upload? No. fix it in the file before uploading

Should we allow users to revert to original OCR file? WE should table this until we get the versioning worked out.

* Other validations? Integrate into paginator?

Strange Characters? We need to know the encoding of the file before upload. Show back to user to verify if it looks correct. We can recommend they use certain encodings e.g. UTF8 but can’t enforce it because the machine won’t know. We can validate file types though .

Integrate into paginator or stand-alone?

How sophisticated does text editing need to be? Big text box. This will not be a transcription tool but just for small simple edits.

Read comments from Nicole’s email and we can meet again to go over their process