

# Thoughts about PDF annotation architecture

## **File areas:**

- original area - specified by the plugin that wants annotation doing (in the first instance, this would be the file area used by the 'Files' submission type, but left flexible enough to work with other plugin types in the future).
- **annotatepdf** - stores all the PDFs from the original area, converted to PDF v1.4 (if needed) + (optional, future development) converted from doc/odt/etc. Needs to prioritise the 'coversheet', if set, so it always appears first. Will need, associated with it (by pathnamehash?) a list of page counts (to save having to open each file individually to get this information).
- **annotatepdf\_image** - for the generated page images (needs clearing from time to time)
- **annotatepdf\_combined** ? - one combined PDF of all the submitted files (to allow downloading of a student's submission as a single file - only generated when requested, but then kept for future downloading; will need regenerating if creation time is < creation time of any file in 'pdf' area)
- **annotatepdf\_response** - store individual annotated PDFs (? only generated on demand, then kept? - assume that mostly students will view online?)
- **annotatepdf\_response\_combined** ? - stores combined response file as a single PDF (only generated on demand, then kept?)

All of these areas should use the same *context*, *component* + *itemid* as the 'original area', but with different filearea.

As for the *filename* + *filepath*:

- **annotatepdf** + **annotatepdf\_response** - these should use the same *filepath* + *filename* as the original area (except in the future when 'convert docs to pdf' is added, in which case the filename should match, apart from the extension).
- **annotatepdf\_images** - *filepath* = '/<pathnamehash of the file in annotatepdf>', *filename* = 'pageNN.png'
- **annotatepdf\_combined** - *filename* localised?, only ever one file in there, *filepath* '/'
- **annotatepdf\_response\_combined** - *filename* localised?, only ever one file in there, *filepath* '/'

## **Database tables:**

mdl\_annotatepdf - store the pagecounts for all of the PDFs (1 pagecount per record, or comma-separated list in one record per submission?)

Somewhere to store the current list of pdfs in order (so that on any given session this stays consistent as the teacher presses 'next' through the documents) - probably store by pathnamehash (and include page counts?). Store in mdl\_annotatepdf as well?

Switch annotations/comments to have pathnamehash + pageno as identifiers (rather than submissionid + pageno) & identify page images by pathnamehash + pageno as well.

## **Frontend <=> backend**

Should frontend just ask by page (with the backend doing the work of converting this into a file+page), or should the frontend ask by pathnamehash+page?

e.g. pdfa - 10 pages, pdfb - 12 pages, pdfc - 10 pages

1. frontend asks for page 24 - backend returns page 2 of pdfc
2. frontend asks for 'pdfc, page 2' - backend returns this

option 1 simplifies the front end slightly + the API to the backend, but complicates the backend slightly and reduces flexibility?

option 2 makes the frontend + API slightly more complex, but gives more flexibility (e.g. frontend can jump to a particular file directly - would allow a 'contents list' of files to be shown)