

# Consolidating Labelled PDF Files

a.k.a. preparing a collection of  
identified tax slips for my accountant

April 4, 2024

John Nash  
for  
Virtual Meeting of Linux-Ottawa

# Present motivation

- Accountant would like tax slips consolidated into single pdf, with preferred order
- But when there are queries, it is useful to have a reference to original slip (assuming downloads or scans to single files)
- Useful to have a naming convention for slip files  
e.g., JNT3RBC-HarvestHouseETF.pdf

# Possible tools -- interactive

- **Xournal** allows addition of text to PDF files
- **pdfarranger** allows combining, ordering and rotating of pages of PDF files.

- **BUT**

have to do everything manually. Can we script?

# CLI Labelling of PDF files

- Labels from (text) file names
  - **pandoc**
  - **enscript**
- Use the above to create temporary pdf “overlay” files
- Merge name file onto slip file with **pdftk**
- BUT – many arcane details

# Annoying details

- Name might overlay important information – positioning!
- (Semi-)transparent colour for name text could be helpful, but not trivial to arrange
  - Set up a LibreOffice template.odt with dummy name that is suitably coloured and named
  - Rename template.odt to template.zip; Unzip
  - Use automated Find and Replace (use **sed**?) to change to desired filename
  - Re-zip; rename to template.odt.
  - Use **pdftk** to merge name into file
  - Still b/w if we print on most Laser printers

# Annoying details (2)

- Spaces and dashes in filenames cause trouble
  - Be sure to quote filename variables and names
  - Check and verify, then do it again
- Some agencies put a password on pdf files (e.g., to stop modification)
  - Can print to pdf and copy back (e.g., **cupspdf** )
  - Seems to work. Does NOT need the original password

# Merging into composite file

- Could use **pdfarranger**
- But **pdftk** does the job quite nicely.
- Even an intro page

# Check files present

```
#!/bin/bash
#
# chkfilelist.sh -- see if (commented) file list is OK text file in order
# File list file is $1
# J C Nash 2024-3-30
#
set -e
set -o pipefail
# Uncomment list file
sed '/^[[:blank:]]*#/d;s/#.*//' "$1" > "tflist.txt"
while IFS= read -r fn; do
    echo "Text read from file: $fn"
    cp "$fn" ../tcpy/
done < "tflist.txt"
echo "Finished"
```



# Script tax2label.sh

```
#!/bin/bash
#
# tax2label.sh -- add filename labels to files in a text file in order
# This file is $1
# J C Nash 2024-3-30
#
set -e
set -o pipefail
# Uncomment list file
sed '/^[[:blank:]]*#/d;s/#.*//' $1 > "tflist.txt"
while IFS= read -r fn; do
    echo "Text read from file: $fn"
    echo "$fn" > name.txt
    pandoc -V geometry:paperwidth=8.5in -V geometry:paperheight=11in -V geometry:margin=0.1in
    "name.txt" -o name.pdf
    pdftk "$fn" stamp name.pdf output stamped.pdf
    mv stamped.pdf "./labelled/$fn"
done < "tflist.txt"
echo "Finished"
```

# Script taxcombineslips.sh

```
#!/bin/bash
#
# taxcombineslips.sh -- add filename labels to files in a text file in order
# This file is $1
# J C Nash 2024-3-30
#
sed '/^[[:blank:]]*#/d;s/#.*//' "$1" > "tflist.txt"
fn="John and Mary Nash 2024 tax slips"
echo $fn | enscript -B -f Courier-Bold24 -o- | ps2pdf - > cmbfile.pdf

while IFS= read -r fn; do
    echo "Text read from file: $fn"
    pdftk cmbfile.pdf ".$fn" output temp.pdf
    mv temp.pdf cmbfile.pdf
done < "tflist.txt"
echo "Finished"
```