**Prime OCR Processing:**

PrimeOCR provides the best OCR (optical character recognition) by polling 6 OCR engines. For the text in documents and theses that are primarily text, PrimeOCR also provides efficient JBIG2 compression (only applies to black and white images/pages). Converting high res images that are all text to Bitonal (Black/White, 1bit) provides a significant size savings and a better print version while retaining text readability on screen. For theses and documents that are primarily 8 bit color or grayscale images, Acrobat or ABBYY may be more effective. However, PrimeOCR provides better text recognition for text in color images especially on color backgrounds.

Recommended SOP for PDF creation:
1. For documents that is primarily text
   a. Convert text to Bitonal (Black/White, 1bit) and images to 300dpi grayscale/color TIFFs:
      i. See RTD-ImagingPrep document for details
      ii. Source images should be at least 400dpi TIFFs for best quality
   b. Use the RTD_template (see Appendix 1: Create PrimeOCR Template)
   c. Use the job creation script for batches of documents to be processed (see below)
2. For document that all 8 bit gray scale or color images:
   a. Convert images to 300dpi JPEGs
      i. See RTD-ImagingPrep document
   b. Use the Catalog_template type (see Appendix 1: Create PrimeOCR Template)
      i. for images with columns of text (e.g. journal article, catalog)
   c. Use the Yearbook_template type (see Appendix 1: Create PrimeOCR Template)
      i. for images with clusters of text (e.g. year book, engineering/architectural drawings)

**Prime OCR Processing:**

1. **Open the program:**

   ![Icon](icon located at bottom toolbar)

2. **Create jobs for Batch processing from Script:**
   a. RTD Job files can be created by a script from a series of FI#’s in a CSV (comma separated values text file)
   b. Copy the FI#’s from the RTD Project Data spreadsheet and paste them as values to FNumber.csv
   c. Run RTDjob.cmd by double clicking on the file in C:\Prdev\job\to-do
   d. Move the created job files to C:\Prdev\job
      i. Remove other job files from C:\Prdev\job

RTDjob.cmd: (job creation script)

```bash
@Echo off
Set d=S:\RTD_current
for /F "tokens=1 delims=," %%l in (FNumber.csv) do {
   Echo Prime Recognition Job File >RTDjob.txt
   Echo Version=3.90 >>RTDjob.txt
   Echo 1 >>RTDjob.txt
   Echo %d%\%%\%TIFF2OCR\*tif^| APPEND_DIRNAME ^| OUTPUTPATH %d%\%%\%l >>RTDjob.txt
   Echo C:\Prdev\Templates\RTD_template.ptm >>RTDjob.txt
   copy /y RTDjob.txt "%%l.jjob"
}
```

FNumber.csv: (FI # input to the Job creation script)
3. **Manual Job Creation:**
   a. Go to Wizard > Only Create Job > Select Directory (i.e. Source Folder)
      i. This is where you’re choosing the specific FI folder where the TIFFs are located in.

   ![Image of Template & Job Wizard](image)

   b. Click "Select Template"
      i. Navigate to C:\Prdev\Template then select the appropriate template (see page 1) and click “Open”
         1. The RTD project uses primarily the **RTD_template.ptm** (a copy of this template can also be found in RTD_Current/RTD Documentation/PrimeOCR folder)
      ii. Click Next

   ![Image of Job Wizard](image)

   c. Click "Select Directory" to select the directory where the OCR’d PDF file will saved:
      i. into (i.e. Output Folder) –
      ii. make sure the 3rd option "Combine all output from a directory in one output file, file name is directory name" is checked on.
d. Click on "Save Job". Use the FI# as the name, to keep it simple.

e. Click "Exit" to close the job wizard

4. Running the Jobs:
   a. After the job has been created (through use of a script or manually), return to main server window, and click "Play"
   b. Prime OCR will run the job(s) you just created.
   c. Be sure JBIG2 compression is enabled (see Appendix: PrimeOCR Settings)
   d. Make sure only the jobs to be run are in the C:\Prdev\Job directory
      i. Please note that PrimeOCR will run all the job files in the C:\Prdev\Job directory
      ii. Move completed jobs to C:\Prdev\Job\done
      iii. Move pending jobs to C:\Prdev\Job\to-do
Appendix 1: Create PrimeOCR Template
PrimeOCR uses templates to set some of the parameters for creating PDFs with OCR text recognition.

1. Create PrimeOCR Template
   a. Start up PrimeOCR and click on the “Wizard...” button:
   b. Click on “Template”
   c. Images are processed in Photoshop so none of the following settings are required:
   d. Select the type of Zoning:
      i. Full page zoning is suitable for most thesis documents which are single-column and text only
ii. For multi-column or image intensive documents, choose Auto-zoning.
iii. Click Next.

Create PrimeOCR Template
a. If “Full Page Zone” was selected, skip to step 2.

iv. Used for RTD_template.ptm and suitable for most theses.

b. If “Auto Zoning” was selected, select the type of zoning:

v. Select “Reading Order” for multi-column documents such as catalogs (used for Catalog_template.ptm).

vi. Select “Clump Text” for graphic intensive documents such as yearbooks (used for Yearbook_template.ptm).

vii. Do not select “Top to Bottom” or “Reading Order” as output will not be PRO, RTF, or PDF Normal.

c. Click Next

2. Set Language and level of accuracy:

a. The defaults below are suitable for most documents. Click Next

b. 6 OCR engines are combined for better text recognition, enabling PrimeOCR to recognize text in images with a colored background better than Acrobat Pro.

3. Select the OCR output Format

a. “PDF Image+HiddenText” creates larger PDF files but retains the original look and feel of the document.
Appendix 2: PrimeOCR Settings

The PrimeOCR server can be configured for a number of output settings beyond the setting available in the templates. Some settings such as PDF/A output and JBIG2 compression must be configured in the PrimeOCR server settings.

1. Configure PrimeOCR server settings:
   a. Under Configuration select Setup (Job Server)
   
   ![Configuration screen](image)

   b. Select the Output tab, check PDF and click on Details:
c. The PDF Output settings are used to configure PDF files created by Prime OCR
1. Configure PrimeOCR server settings:

d. Configure PrimeOCR server settings JBIG2 Compression or PDF/A output:

i. Settings for JBIG2 compression:

ii. Settings for JBIG2 compression and PDF/A-1b: