How to Create and Work with Scanned Documents Using OCR (Optical Character Recognition)

All State of Missouri documents available to both customers and employees which contain text must be readable as text per Web Accessibility Standards. This means that all scanned documents must use OCR (Optical Character Recognition) as well as include tags.

What is OCR?

OCR (optical character recognition) is the use of technology to distinguish printed or handwritten text characters inside digital images of physical documents, such as a scanned paper document.
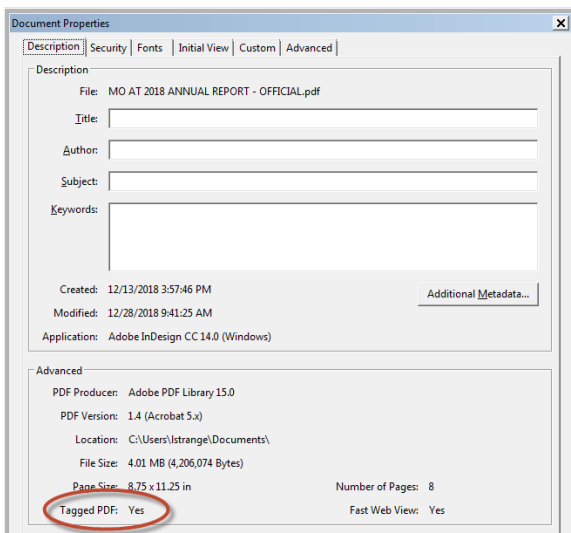
How to know if a document is an image of text and not actual text

- Open a PDF document and try to highlight any text area. If you cannot, it's a picture of text and needs to be converted to text.

How to check if a PDF document is tagged

Tagging basically gives the document structure to be read by Assistive Technology.

In Acrobat – File - Properties

Scanning tips

- Firstly, avoid scanning if at all possible. Always work from the original electronic version if you have it.
- If you have a multi-page electronic document which required the user to print off and fill out/sign one or a few of the pages:
    - Only OCR scan the hand-written pages.
    - Convert the other pages to PDF like you normally would from the electronic version.

How to OCR scan in some of the more common software products

If a paper document is your only option, do one of the following, depending upon what software you have available. Make sure that whatever scanning software you use, *it creates a tagged pdf document.*

- Scan paper documents directly from Adobe Acrobat Pro
- Settings in Omnipage that create a tagged PDF document | Scanning In Omnipage
- Settings in ABBY Finereader that create a tagged PDF document | Scanning in ABBY Finereader

If the document is readable text, but not tagged

- If you convert the document to readable text, but the document is not tagged and you don't have the paper source. (page 10)

Accessible electronic documents resources

For accessibility tips on all types of electronic documents, visit accessible electronic documents resource page.