



Converting to MS Word and Text Files

SensusAccess will convert image-based documents into MS Word, RTF, and text files. You may also find it useful with some image-based documents to convert initially to Tagged PDF and then copy and paste the text from the Tagged PDF into MS Word. This may result in a better reading experience and may remove non-essential content.

With the MS Word version of the document, you can more accurately "clean" the content for conversion into MP3 audio or for use with assistive technologies. Most conversions will take just a few seconds within MS Word and involve the use of the Find and Replace tools.

Please note - in the Find and Replace examples below, replace the <space> value with one space bar and do not include the quotes.

Image-File to Tagged PDF to MS Word Document

1. Submit the image-based document to SensusAccess and select Tagged PDF as the output option.
2. Open the Tagged PDF and select all the text. Copy and paste this into a MS Word document (Open Office may also be used).
3. Using Find and Replace:
 - a. Search for ".<space>^p" and replace with ".^p^p" .
 - b. Search for "<space>^p" and replace with "<space>" .
 - c. Search for "<space>•<space>" and replace with "^p•<space>" .
 - d. Search for "-<space>" and replace with no value.
 - e. Save the document in your preferred text format.

Image-File to MS Word Document

To clean-up a MS Word file for use with assistive technology or for creating MP3 files, perform a "search and replace" to remove optional hyphens and section breaks. Identify the special character you wish to find in the "Find:" box and leave the "Replace with:" box empty.

1. Submit the image-based document to SensusAccess and select Microsoft Word as the output option.
2. Open the converted Microsoft Word document (Open Office may also be used).
3. Using Find and Replace:
 - a. Search for "Optional Hyphen" under Special Formatting and replace with no value.
 - b. Search for "Section Breaks" under Special Formatting and replace with "^p^p".
 - c. Search for "Manual Page Breaks" and replace with "^p^p".
 - d. Save the document in your preferred text format.