

What is the difference between raster and vector content in PDFs?

Answer

Vector is the preferred format for PDF drawings and specifications, as it will lead to the most accurate results when populating information from Procore's Optical Character Recognition (OCR) technology. We recommend requesting vector PDFs from the design team (instead of raster PDFs or PDFs with a mix of vector and raster content).

Note: See the explanations below for more information on the differences between raster and vector content:

Raster Content

Raster files consist of a series of pixels that are a static grid of colored squares, rather than actual lines or letters. Procore's OCR and text parsing technology attempts to identify lines and letters in the PDF from the shape of the pixels. If you have received a raster-based PDF, contact the design team, who can save the original content as a vector-based PDF.

- You can identify files with raster content by the following:
 - Attempt to highlight text in the PDF with your mouse. If you cannot highlight the text, the content is raster.
 - Zoom in on the PDF. If the image or text gets blurry or pixilated, it is a raster file.
 - If the file was scanned to your computer, it is a raster file.

Vector Content

Vector files are created from a mathematical model which creates links between two points, or a series of points, and then displays the line segments between them on your computer. Vector content can be zoomed into almost indefinitely, and the lines and text will still remain sharp. Procore's OCR and text parsing technology was built with this logic in mind, and can most easily identify text and shapes in a vector-based PDF.

- You can identify files with vector content by the following:
 - Attempt to highlight text in the PDF with your mouse. If you can highlight the text, the content is vector.
 - Zoom in on the PDF. If the image and text remain sharp, it is a vector file.

See Also

- [How can I improve the accuracy of OCR on my drawings?](#)
- [How can I improve the accuracy of Specification Section Identification?](#)

