

Best Practices for Scanning Tax Documents

This document details the best practices for scanning tax documents for use with GruntWorx. You'll learn which scanner settings provide optimal results, what detracts from quality images, and how to avoid scanning imperfections that can hinder GruntWorx productivity.

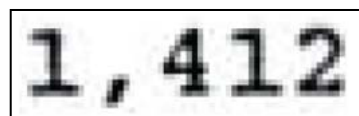
Know Your Scan Settings

The image quality of scanned documents starts at the scanner level. If your scan settings are not optimal for scanning tax documents, the image quality of the resulting scanned document will follow suit. Set your scan software to the following basic image specifications to ensure you always scan your documents right the first time.

- ✓ **Scan to a PDF.** To create a PDF from your scanner, choose File > Create PDF > From Scanner in Adobe® Acrobat®, or any industry-standard PDF writer. You can also choose whether you'd like to create a new document or append the scanned pages to a current document.

Insider Tip - Scanning Simplex vs. Duplex: A simplex scanner scans one side of a page, while a duplex scanner scans both sides at once. If you have a duplex scanner, we recommend automatically scanning both sides of each document at once, as GruntWorx will create bookmarks for both back pages and blank pages.

- ✓ **Scan at 600dpi.** While GruntWorx recognizes thousands of variations of tax forms with industry-leading accuracy, the performance of GruntWorx products rely on the quality of the documents you submit. For best results scan at 600dpi (300dpi at minimum). This is not the same as scanning at a lower resolution and then saving at 600dpi. GruntWorx may not be able to identify and extract the text on low resolution images. If GruntWorx cannot read the text on a document, data will not be extracted and the document will be classified as: "Poor Quality/Handwritten Notes".
- ✓ **Scan in Black and White.** Though color and grayscale may look more readable to the human eye, scanning your documents in black and white will provide clearer images for GruntWorx processing – and black and white files will be up to ten times smaller than color or grayscale. Smaller files will take up less space on your hard drive or DMS, and take less time to upload to and process with GruntWorx.



The image to the left shows what GruntWorx "sees" when processing a document that was scanned in black and white. The image to the right shows what GruntWorx "sees" when processing a document that was scanned in grayscale. Notice the noise in the grayscale image, which makes the numbers harder for the software to read and interpret.

- ✓ **No Read/Write/Password Protection.** Documents saved with read, write, or password protection cannot be processed by GruntWorx *Organize* or GruntWorx *Populate*. If your PDF is saved with these settings, be sure to remove them before uploading to either program.

Creating the Highest Quality Images from Client Source Documents

Now that you know what settings to use before you start scanning, it's time to understand what you can do to achieve maximum image quality from the actual source documents and ensure document classification by GruntWorx.

Scan the Original Document. Every time a document is copied, the image quality is degraded. The further removed you get from the original document, the “noisier” the document becomes.

For best results, scan the original source document, as received by the client. Avoid using copies of client documentation. Faxed copies are particularly low quality and should be avoided.

This is a "noisy" document. Noise makes documents harder to read, and interferes with GruntWorx's ability to classify documents and extract data.

Insider Tip – The “Readability Test”: After you’ve scanned your documents, give them the “Readability Test”. If your eyes are straining to read a document, GruntWorx will likely do the same. Noisy, hard to read documents will most significantly affect GruntWorx *Populate* users. If GruntWorx *Populate* is unable to read field labels and data contained therein, the data will not be extracted and populated into the tax return. Any document that cannot be read will be bookmarked as “Poor Quality/Handwritten Notes”.

Scan the Document at its Original Size. For best results, the size of the scanned document should correspond roughly to the size of the original document. GruntWorx recognizes scanned documents as large as legal size documents, with a maximum scanned image size of 8.5”x14”.

If the size of a document is greatly reduced or expanded when scanned, it increases the chances that the scanned image will be distorted. A document scanned at the original size settings will preserve readability of the document and minimize distortion.

Distortion may also be caused by scanning documents that are folded or crumpled, and by documents that were caught in a paper jam while being scanned.

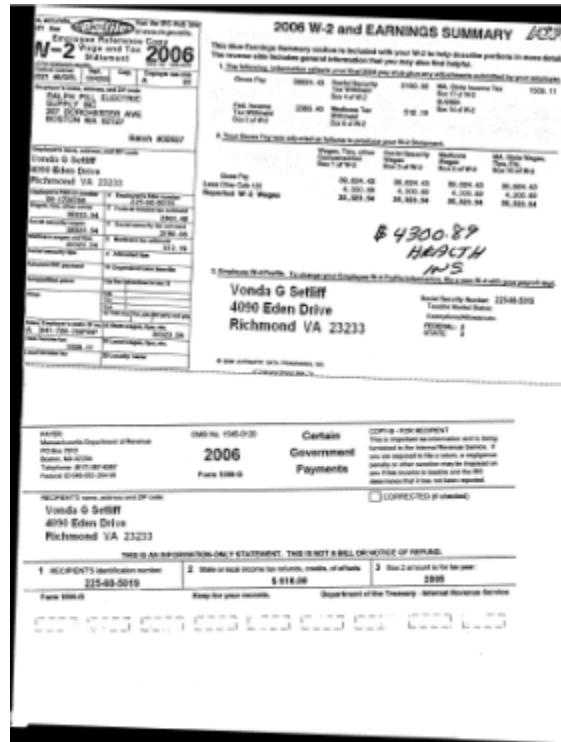
This scanned image is distorted. Distortion can happen when the size of the scanned document does not correspond to the size of the original document.

Scan Multi-Page Documents Together. Multi-page documents, like brokerage statements and K-1's, should be submitted in logical order, where possible. GruntWorx will not reassemble a multi-page document that has been scattered throughout an input PDF.

Scan Each Document to Its Own Page. If more than one document is scanned to the same page, only the dominant form, as determined by the GruntWorx software, will be bookmarked. For GruntWorx *Populate* users, data will not be extracted from either form. Visibility of the less dominant form(s) will be lost, as only one bookmark will be generated per page.

Avoid Submitting Documents with Faint or Faded Text. Not all original tax documents are suitable for processing by GruntWorx *Organize* and GruntWorx *Populate*. Going back to the "Readability Test", faint, faded text is hard to read and therefore may not be properly classified. If a document cannot be classified, the data cannot be extracted by GruntWorx *Populate*.

Insider Tip – Documents that Contain Handwriting: Though GruntWorx *Organize* and GruntWorx *Populate* recognize handwriting and classify handwritten notes as "Poor Quality/Handwritten Notes", neither product is able to *read* the handwriting. Forms that contain important identifying information that is handwritten, like the form to the right, may not be properly classified and will not extract tax data.

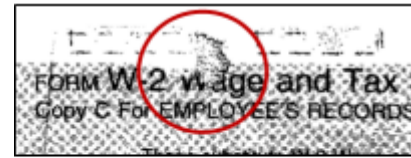


The top form shown is a W-2; the bottom form is a 1099-G. GruntWorx will classify only the dominant form on this page, as determined by the GruntWorx software.



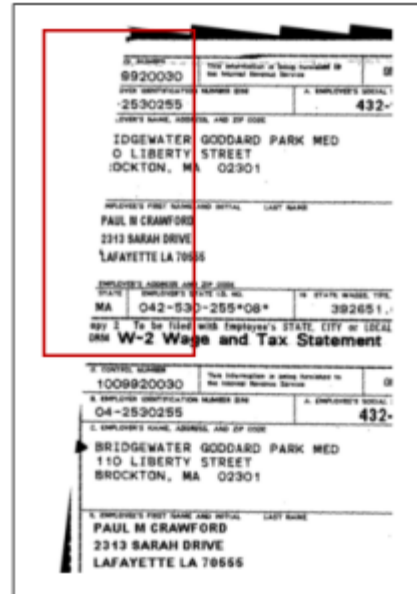
This document has faint, faded text, making it difficult to read and therefore difficult to classify with high confidence.

Avoid Black Backgrounds, Ink Bleeding, and Smudging. Black backgrounds can be created by leaving the tray cover open during single page flatbed scanning. Black backgrounds, ink bleeding, and smudging are considered “noise” and can slow GruntWorx processing.



Ink bleeding and smudging is considered "noise" and can slow GruntWorx processing.

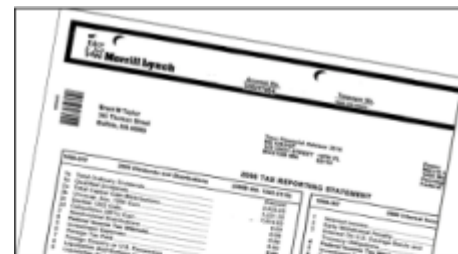
Avoid Submitting Clipped or Cut Forms. Clipped or cut forms may be missing important data that GruntWorx *Organize* and GruntWorx *Populate* need to identify and classify the form, and to extract the tax data.



Clipped or cut forms may be missing important information that GruntWorx needs to identify, classify, and extract the tax data.

Other Tips and Tricks to Help You Easily Navigate Through GruntWorx *Organize* and GruntWorx *Populate*

- ✓ **Naming Your PDF Input File.** Do not use special characters when naming your PDF files. PDF filenames containing ampersands (&) and apostrophes (') can create problems when being processed by GruntWorx.
- ✓ **File Size.** If your PDF file size is greater than 30MB, we recommend you split the document into two or more PDFs before submitting for GruntWorx processing. You can upload up to four PDF documents to GruntWorx per client.
- ✓ **Upside Down and Sideways Pages.** GruntWorx will automatically rotate all upside down (180°) and sideways (90° and 270°) pages to the proper page orientation during processing. However, if a scanned document is skewed more than 10° from any of these positions, page orientation will not be corrected and data will not be extracted.



This document is skewed. Though GruntWorx will rotate documents scanned upside down or at a right angle, documents that are skewed 10° or more from absolute N, S, E, or W, will not be corrected and data will not be extracted.

Best Practices for Scanning Tax Documents

Quick Reference Guide

Scan Settings

- ✓ Scan to PDF
- ✓ Scan at 600dpi
- ✓ Scan in Black & White
- ✓ Do not upload files with Read/Write/Password protection

Scanning Source Documents

- ✓ Scan the original tax document – avoid uploading copies and faxes
- ✓ Scan the document at its original size
- ✓ Do not upload documents larger than 8.5”x14”
- ✓ Scan multi-page documents together, in logical order
- ✓ Scan each document to its own page
- ✓ Do not leave the tray cover open when scanning single pages on a flatbed scanner

Source Documents to Avoid

- ✓ Documents with faint or faded text
- ✓ Documents that contain important identifying information that is handwritten
- ✓ Documents with ink bleeding or smudging
- ✓ Clipped or cut forms that exclude important identifying information

For additional resources, visit our [Frequently Asked Questions](#) and [Tutorials & Guides](#) pages. To contact Customer Support, email: support@copanion.com or call 800-291-2136 ext 3