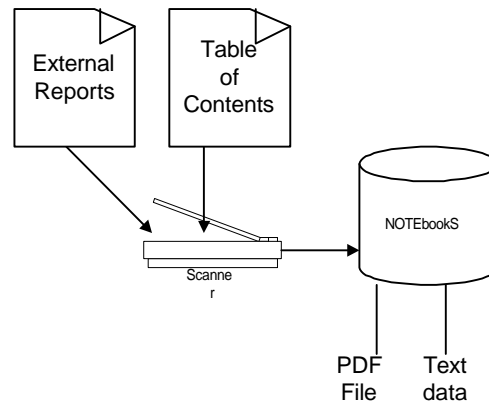


## Integrated Scanning and OCR Extensions (ISO)

ISO provides scanning and OCR capability for NOTEbookS, NORMA and can be incorporated into other Lotus Notes/Domino-based applications. It consists of a set of design elements (forms, sub-forms and a script library) that can be incorporated into any Notes application. It works with Notes version R5.02a or later.



### Features of ISO include:

- Turns scans into Adobe Acrobat PDF files and attaches the PDF files to the Notes document in a field of your choice<sup>1</sup>.
- Performs optical character recognition (OCR) on the scanned images to produce editable and searchable text. Attaches the text to the Notes document in a field of your choice.
- Will automatically extract an abstract from the scanned document, and will attach the abstract to the Notes document in a field of your choice.
- Ability to scan in Black & White, grayscale and color.
- Ability to work with any TWAIN-enabled scanner, including flat-bed, sheet-fed and those with an automatic document feeder (ADF). Can work with multiple scanners attached to the same PC.
- Automatic sensing of scanner capability, including maximum paper size, available color levels, and the presence of an ADF.
- Allows for cropping to remove headers and footers, or to limit scanning to a zone that is less than full page.

### OCR Quality:

The OCR engine used by ISO is provided by eiStream Inc. (formerly Eastman Software, a division of Eastman Kodak). It provides some of the most sophisticated OCR capabilities in the marketplace today.

### Scanning Documents

Scanning and optical character recognition (OCR) is an easy process. In most cases, it requires that you click only one or two buttons. Please remember that in order to scan documents, 1) you must have a scanner correctly attached to your PC, and 2) the scanning and OCR software must be installed.

The scanning process can result in up to three different forms of information:

---

<sup>1</sup> Creating PDFs requires a license for Adobe Acrobat, which is not included with ISO. Adobe Acrobat is not the free Adobe Reader.

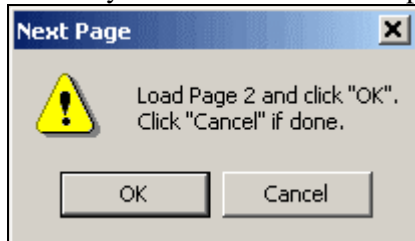
- A **PDF file** consisting of images of each page scanned. These images are usually in black and white, but they can be in color, grayscale. The PDF file precisely preserves the formatting of the original document, but contains no intelligence. The text of the PDF can not be edited, copied or searched.
- The **full text** of the scanned document extracted by OCR software. The text is complete, editable and searchable, but does not preserve the formatting of the original document.
- If the full text has been extracted, you may also obtain an **abstract** of the original document. The abstract is created by copying a specific number of lines of text from the page and location that you specify. For example, you can specify that the abstract is to consist of ten lines of text taken from the third page and starting on the fourth line on that page.

To begin scanning, create a new document or put an existing one into edit mode. You will see the buttons used to begin scanning. If you don't want to create a PDF version of the document, select "OCR Text Only." If you don't want to create OCR full text, select "PDF Only." If you are creating OCR full text, with or without a PDF, an abstract will be created, unless you uncheck the "Create an Abstract" box.

When all choices are correct, click the "Scan Document" button, and the scanning dialog will be displayed. Here is an example of what it could look like.

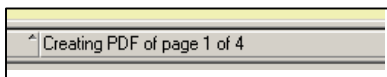
In many cases, all the choices on this screen will be correct and you can start the scan by placing the first page on the scanner's glass bed and clicking the "Start Scan" button.

As each page is completed, you will be asked if you want to scan another page:



If so, place the next page on the scanner's glass and click "OK." If there are no more pages to scan, click "Cancel." (If your scanner has an automatic document feeder, the process will be a little different. See the section below for more information on using an automatic document feeder.)

Once the scanning is complete, the creation of the PDF and/or full text extract will begin. This process takes a little time, which depends on how many pages have been scanned. Watch the information area at the bottom of your Lotus Notes screen to see messages about the progress of the PDF and OCR process. Once the OCR process begins, there will also be a screen that shows you the progression



of the OCR.

When the entire process is complete, the “Document Scan” screen will close and you will be returned to the document from which you initiated the scan.

## Modifying your selections – Document Information:

This section allows you to change the settings related to what information is sent back to your Notes document.

**PDF Attachment Name:** The PDF file that gets created will be attached to your document with the file name shown here. You can change the name, but bear in mind that each file attached to a document must have a unique name. If the name is not unique, an alternative will be suggested when you click the “Start Scan” button.

PDF Attachment Name: attachment\_.pdf (don't include the ".pdf")  
(Required before scanning)

**Automatic Abstract:** If you have chosen to create an abstract from the extracted text, fill in the information in this section. There are three pieces of information that are required – how many lines to abstract, what page to start on, and how many lines to skip on that page. A typical example could be this:

Automatic Abstract:  Create an Abstract  Don't Create an Abstract

Lines To Abstract: 10

Pages to Skip: None

Lines On Page to Skip: None

You will be scanning a document that has a title page and 2 pages of table of contents – three pages to be skipped. On the first page with text for the abstract, there are two lines (title and by...) that have to be skipped. Then, you want to create an abstract from the next 8 lines. You would change “Lines to Abstract from “10” to “8”. Change “Pages to Skip” from “None” to “First 3 Pages,” and change “Lines On Page to Skip” from “None” to “2”

A note on counting lines: Count lines of text only. Do not count blank lines.

## Modifying your selections – Scanner Preferences

This section allows you to modify information that controls the scanner’s settings.

**Scanner:** You should find your scanner already listed here. If you have several scanning devices attached to your PC, you may find the wrong one listed, and it will be necessary to make a change. Click the button labeled “Change scanner,” and select the correct scanner from the devices listed.

Scanner: HP PrecisionScan Pro 3.0

Select/Change Scanner

**B&W/Color/Grayscale:** Here you can choose the kind of color that will be used in creating the PDF attachment. The list of choices may be different from the ones shown in this example, because the list is customized for the capabilities present in your scanner. If, for example, your scanner doesn’t support grayscale, then that choice will not be present. The choices that you find may include:

B&W/Color/Grayscale:  Blk & Wht  256-Color Palletized  
 Grayscale - 256-level  24-Bit True Color

- **Blk & Wht** – Black and white produces a very high contrast monochrome image in which each pixel (dot) is either totally black or totally white. This is usually the best choice for text and line drawings. It produces the most compact (fewest number of bytes) PDF files.
- **Grayscale – 256-Level** – This is a monochrome image in which each pixel can range from pure white to pure black or any of 256 shades of gray in between. This tends to produce images that require a large number of bytes to store.
- **256-Palletized** – This is a full-color image with limited color depth. The scanning process attempts to choose a pallet of colors that best match the colors present in the original material. The resulting image tends to look broken up if the original is shaded, such as a photograph. This works well with

line drawings having solid colors and with colored text. The image is less compact than black and white, but much smaller than true color or grayscale images.

- **24-Bit True Color** – This is a full-color image with smooth color transition. It can reproduce more than 16-million colors. While the image is quite pleasing, 24-Bit True Color produces very large file sizes.

## Possible Problems

You do now see the Scan button on a form even through it should be there:

If you do not see the scanning button when you are in edit mode on a document, it may be because the computer you are working on is not set up to scan. To modify this go into the configuration document. Into the Scanning, OCR and PDF Generation section. Click on Set Current Workstation (to Scan). The scanning buttons should now appear on the form.

Computer Freezes Up:

The software that is used to integrate the scanning into Lotus Notes is a 3<sup>rd</sup> party product (eStream). It is the only software available at the publication of this document that offers extensions that work with Notes and provide this functionality.

We have encountered instances where computers run out of memory while performing all the tasks necessary to scan, generate PDF files and OCR, especially when scanning in color mode.

If you encounter problems where the process “sticks” or hangs up, it is probably memory. To solve this problem,

- reboot your computer
- scan in a single page.
- Save the page
- Repeat for all the pages to be scanned in
- Go into Adobe Acrobat and consolidate the pages.

While this is not a very clean “fix” it is a work around until eStream resolves their memory issues.