Chapter 1.

Introducing IntelliChief Capture Enterprise

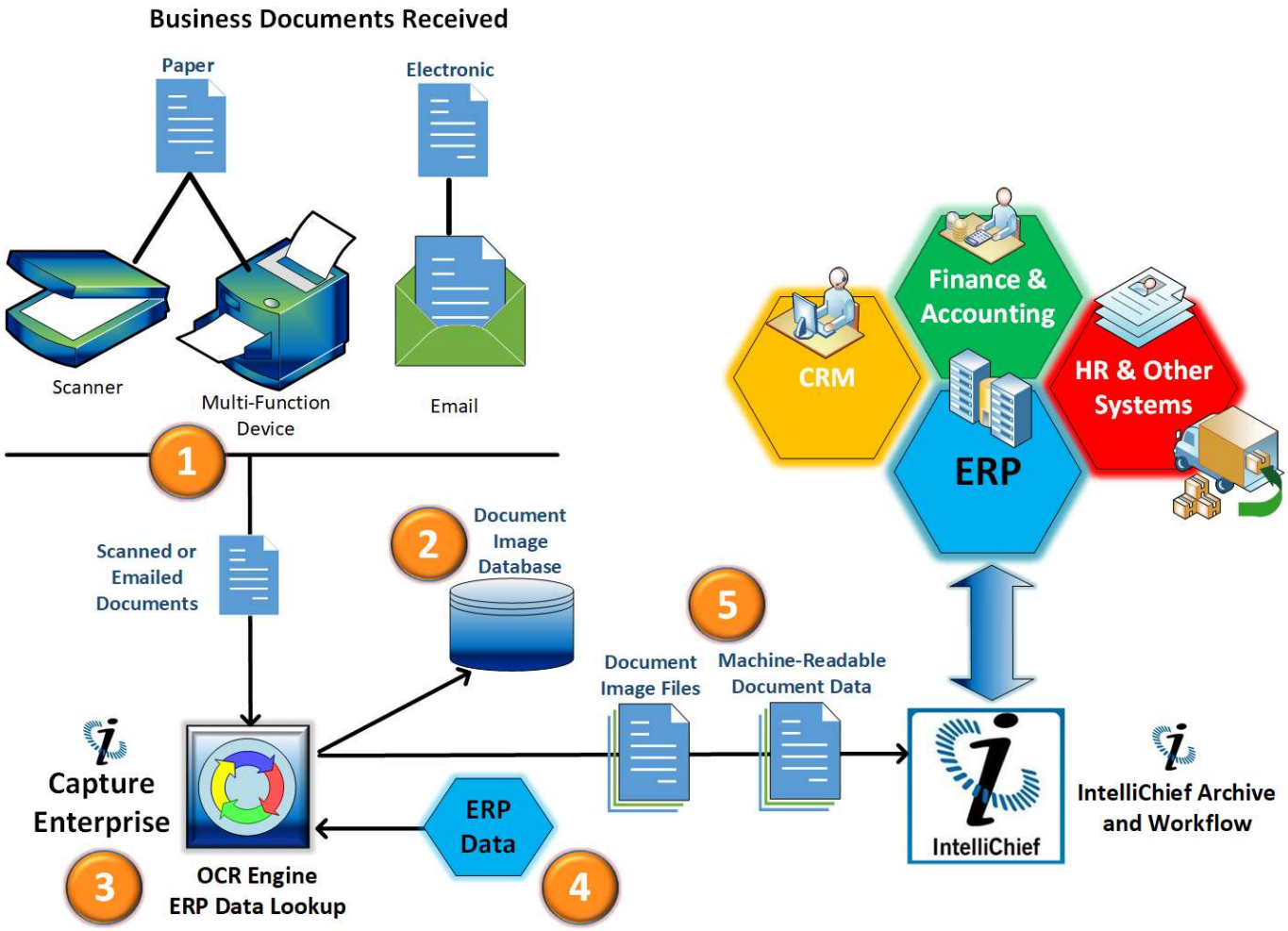
IntelliChief Capture Enterprise provides an easier and faster way to process and manage the large volume of business documents that many enterprises receive every day. This chapter introduces the technologies, components, and operational requirements of Capture Enterprise document processing.

Capture Enterprise – Automated Document Management

The business processes that many companies use to handle the large volume of business documents they receive are often mostly manual and therefore very labor-intensive. Capture Enterprise is the IntelliChief multi-channel solution that automates the processes involved in incoming document data capture and management. Capture Enterprise greatly reduces—and in many cases eliminates—the need for manual data entry of the data from documents received such as invoices, customer purchase orders, contracts, and more.

Capture Enterprise Document Processing Overview

The following figure shows an overview of Capture Enterprise incoming document processing operations:



Capture Enterprise Document Processing Operations

As shown in the preceding figure, Capture Enterprise automated document processing is comprised of the following component operations:

1. **Document Input:** Capture Enterprise can process any document that arrives through electronic communication channels. This includes paper documents that are scanned to a digital image file or digital documents sent as email attachments.
2. **Document Image Database:** All documents sent to Capture Enterprise are saved to a digital image database before processing.
3. **Data Capture/Conversion:** After archiving, the Capture Enterprise Optical Character Recognition (OCR) engine captures and converts document data into machine-readable form.
4. **Data Validation:** After data conversion, Capture Enterprise uses its integration to the Enterprise Resource Planning (ERP) system to look up and validate the documents with current ERP data.
    - **Data Verification:** Documents with any data that cannot be automatically validated with high confidence against ERP data are placed into a Verify task queue for verification, validation and processing by company-assigned Capture Enterprise users.
5. **Document Routing:** Once automatically validated or verified and validated by a user, Capture Enterprise sends the machine-readable document data and document image files to IntelliChief, where Workflow actions and rules-based criteria can make the machine-readable document data and image accessible to the ERP and other business systems.

Purpose of This Document

This document provides the information required by Capture Enterprise customers and IntelliChief personnel to understand the details of the component operations described above, so they can efficiently and effectively support Capture Enterprise document processing. In addition, this document provides the task instructions and best practices for Capture Enterprise users to follow when user action is required. The ultimate purpose of this document is to help companies get the most out of Capture Enterprise functions and capabilities, to greatly reduce the time and effort required to process incoming business documents and greatly increase the efficiency of their document management operations.

Document Topics

The following table summarizes the major topics covered in the remainder of this document.

Chapter	Topics	See
Document Input	Best practices for preparing Capture Enterprise document input, use cases for various document types, input methods available for getting a document into Capture Enterprise, and file formats supported.	Chapter 2
System and Technology Architecture	OCR, Client applications, Server functions, client-server architecture, system configuration, databases, application folder/file structure, location and use, including key files.	Chapter 3
Operations and Processes	Detailed description of CE document processing operations and processes from document enter to exit.	Chapter 4
Processing Document Batches – User Tasks	User task instructions and best practices for processing document batches when user action is required, including fingerprinting and document verification considerations and procedures.	Chapter 5
Administration	User management, onboarding new document sources, fingerprint maintenance, Capture Enterprise Settings (definition, setup and use), start/stop servers, controlling input scan & batch size, key file management, logging, DB management, folder management, batch/file purge and cleanup.	Chapter 6
Troubleshooting	Verifying the Health of Capture Enterprise, document tracking, fingerprint troubleshooting	Chapter 7

Chapter 2.

Document Input

This chapter describes the impact of document input quality on Capture Enterprise performance, the best practices for the available methods used to get documents into Capture Enterprise and the document input types and business applications supported by Capture Enterprise.

Key Terms

The following table defines the key Capture Enterprise terms used in this chapter.

Term	Definition
OCR engine	Optical Character Recognition (OCR) engine used by Capture Enterprise to recognize text characters on a document image and convert those character images to machine-readable text.
OCR grid	Horizontal and vertical lines used by OCR engine to align and identify image words and lines.
OCR word	Set of contiguous text characters.
OCR line	Set of words that align along a single OCR horizontal grid line of the document.
Skewed image	Document image that is neither parallel nor at right angles to a specified or implied line (“crooked” or “tilted”).
Contiguous text	Text that has no additional blank space between characters. Contiguous text in a document image is converted as a single word by the OCR engine.
Greyscale (in document)	Shaded grey background areas in documents. Text on a greyscale background can lead to issues with the OCR scan.
Greyscale (scanner mode)	Preferred scanner mode for desktop and Multi-Function Peripheral (MFP) scanners providing document input to Capture Enterprise.

Document Input Quality Considerations

The goal of Capture Enterprise is to process incoming documents automatically into IntelliChief and thereby make them available to your company ERP system, with no user action or manual data entry required. The quality of the documents received from your document sources (suppliers/vendors, customers, or other sources) is the most critical factor for achieving this goal.

Low-Quality Document Input and OCR

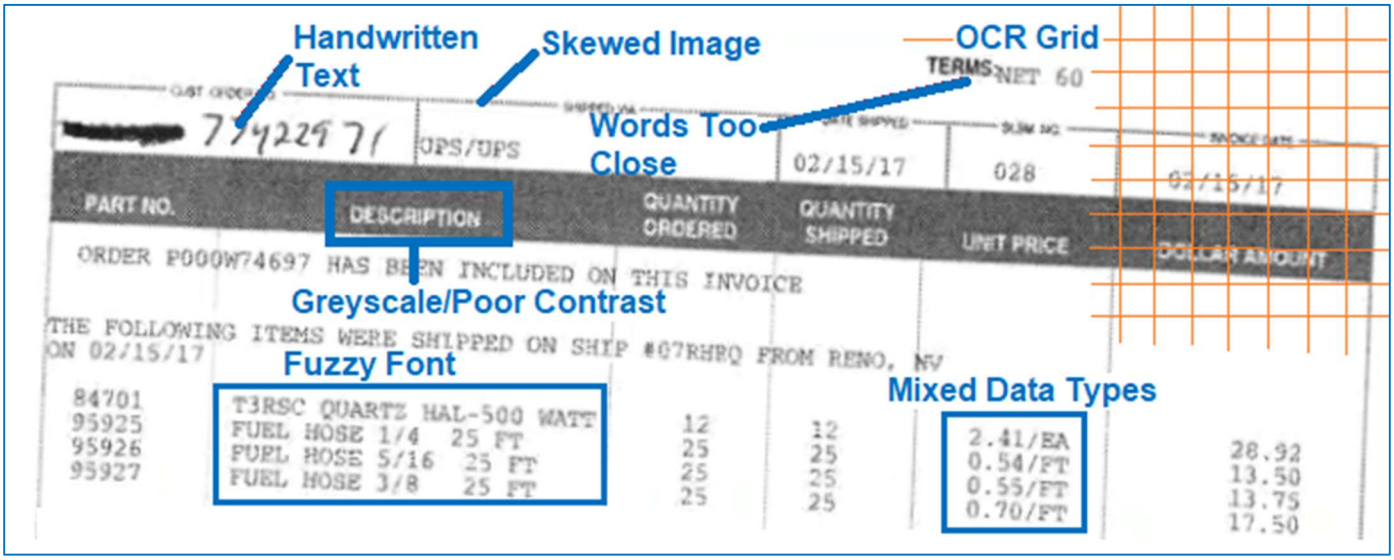
Low-quality document input makes it difficult for the Capture Enterprise Optical Character Recognition (OCR) engine to accurately recognize the text on the document image. Low-quality document input issues are usually associated with documents brought in through manual scanning of the paper originals, and not found in documents that originated electronically. Capture Enterprise has functions that can enhance low-quality document images and can often still capture the document data, but low-quality document input will often require some user action to correct or at least to confirm the accuracy of the document data.

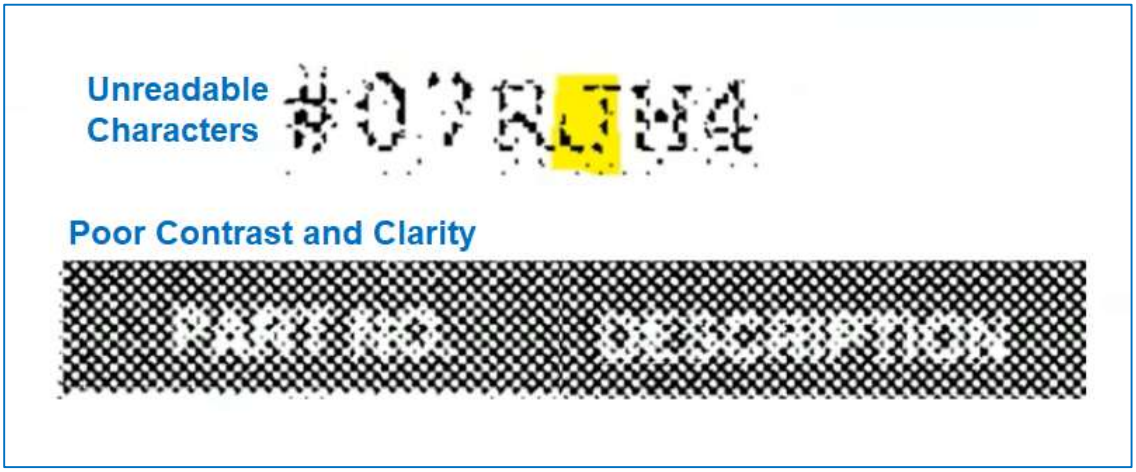
Examples

OCR engines use character lines and spacing to identify character images and convert them into machine-readable text. Examples of low-quality document input issues that can impact OCR text conversion include:

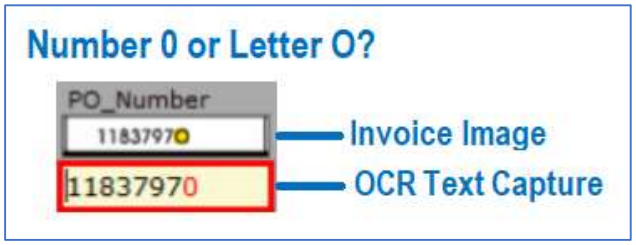
- **Lack of contrast:** Text characters that are not clearly distinguishable from the document background are difficult for the OCR engine to recognize. This includes “greyscale” shading or highlighting in areas of documents that do not have sharp black and white contrast between the characters and their background.
- **Lack of clarity:** Text characters that are “fuzzy” and do not have sharp edges make it difficult for the OCR engine to distinguish the lines of the characters from each other.
- **Inconsistent spacing:** The OCR engine uses the amount of blank space between characters to recognize the end of one word and the start of the next. Words that are too close together make it difficult for the OCR engine to identify where one word stops and the next begins. Characters that are spaced out too far within the same word can make the OCR engine incorrectly identify a character as the start of a new word that should just be the next character in the same word.
- **Handwritten characters:** The Capture Enterprise OCR engine does not recognize handwritten characters.
- **Poor image alignment:** The Capture Enterprise OCR engine uses a vertical and horizontal grid to align the words and lines in a document image. Document images that are skewed and/or have words and lines that do not align along the normal vertical and horizontal alignment grid may be incorrectly recognized.
- **Mixed/unclear data types:** Some documents may have two or more different data types as part of the same word. For example, a price may show on an invoice document as “9.97/EA”. The OCR engine would recognize and convert this contiguous text as one word. Capture Enterprise would then not identify this pricing data as a price (because the trailing “/EA” characters are not monetary data values).
- **Text font characteristics:** Some text fonts have characteristics that make it difficult for the OCR engine to accurately convert character images to machine-readable text. For example, some fonts may have a zero that looks like the letter “O” or a lower-case letter “L” that looks like the number 1, or vice versa. Other fonts have characters that are very close together so that it is difficult for the OCR engine to distinguish the separate characters.

The following figures show examples of these low-quality document input issues.


Handwritten Text
Skewed Image
OCR Grid
Words Too Close
Greyscale/Poor Contrast
Fuzzy Font
Mixed Data Types
TERMS NET 60
UPS/UPS
02/15/17
028
PART NO.
DESCRIPTION
QUANTITY ORDERED
QUANTITY SHIPPED
UNIT PRICE
DOLLAR AMOUNT
ORDER P000W74697 HAS BEEN INCLUDED ON THIS INVOICE
THE FOLLOWING ITEMS WERE SHIPPED ON SHIP #07RHBQ FROM RENO, NV
ON 02/15/17
84701
95925
95926
95927
T3RSC QUARTZ HAL-500 WATT
FUEL HOSE 1/4 25 FT
FUEL HOSE 5/16 25 FT
FUEL HOSE 3/8 25 FT
12
25
25
25
12
25
25
25
2.41/EA
0.54/FT
0.55/FT
0.70/FT
28.92
13.50
13.75
17.50


Unreadable Characters
Poor Contrast and Clarity


Too Much Space
120.00
1st OCR "Word"
120.
2nd OCR "Word"
00


Number 0 or Letter O?
PO_Number
11837970
11837970
Invoice Image
OCR Text Capture

Document Input Best Practices

This section describes the best practices for getting documents into Capture Enterprise that will help to prevent the low-quality document input issues described in the preceding section.

IntelliChief Email Import Service

The **IntelliChief Email Import Service** is the best practice for document input to Capture Enterprise and is the easiest way to prevent low-quality document input issues. Your incoming document sources (vendors/suppliers, customers or other sources) can send their digital original documents as email attachments directly to an inbox monitored by the IntelliChief Email Import Service. The service will automatically extract and convert the attached file before sending it to Capture Enterprise for automatic processing. The operational details on this process are presented in the Operations and Processes chapter.

Digital Originals

Documents sent to the IntelliChief Email Import Service inbox should be “digital originals”. This means that the documents must come from the application that created the invoice, whether directly as in a Microsoft Word document file (*.doc or *.docx) or in a digital publishing format such as Portable Document Format (PDF) files. Printing a document and then scanning the printed document into a digital format is considered a paper original and is not the best practice (see Paper Document Scans). To ensure the fastest processing by Capture Enterprise with the least amount of user intervention required, request that all your document sources provide digital originals whenever possible.

Best Practices for Email Attachments

Documents attached to emails sent to the email inbox monitored by the IntelliChief Email Import Service can be in any of the following formats:

- Portable Document Format (*.pdf) **Best Practice**
- Microsoft Word (*.doc or *.docx)
- Microsoft Excel (*.xls or *.xlsx)
- Tagged Image File Format (*.tif)

Emails processed by the IntelliChief Email Import Service can have multiple attachments, but each attachment should only contain one document. Multiple documents within a single attachment can be processed, but this processing is not as reliable and may end up requiring user action to complete the document processing.

Paper Document Scans

Paper documents received through the mail or other physical delivery method can be scanned using any kind of scanner or multifunction peripheral (MFP) device that can convert paper documents to the Tagged Interface File Format (TIFF) digital document format. Once scanned, documents can be sent to Capture Enterprise either through email to the inbox managed by the IntelliChief Email Import Service or by saving the document to the designated Capture Enterprise automatic processing folder. The operational details on these processes are presented in the Operations and Processes chapter.

While digital originals provide the most reliable high-quality input for Capture Enterprise, observing the best practices described in the following sections when scanning paper documents for input to Capture Enterprise can greatly improve the chances that manually scanned documents can be processed without further user action.

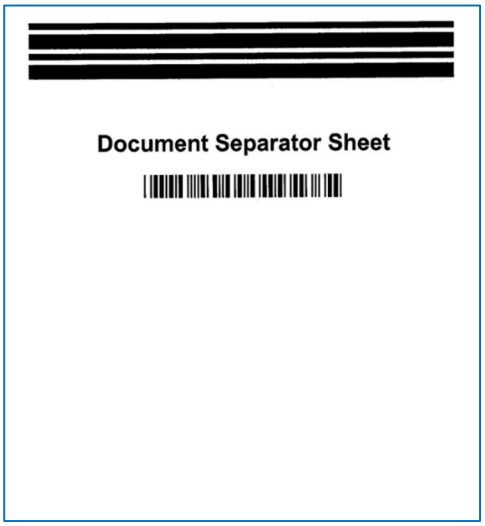
Paper Document Batch Preparation

Observe the following best practices when preparing to scan multiple paper documents for Capture Enterprise input:

- **Batch size:** Batches should not contain more than 50 separate documents.
- **Multi-page documents:** If a batch contains multi-page documents, use the Document Separator Sheet (DSS) as described in the following section.
- **Multi-page and single-page documents:** Do not mix multi-page and single-page documents within the same batch. If a single batch does contain both single-page and multi-page documents, the documents should be arranged and loaded so that all the single-page documents are scanned first followed by the multi-page documents, with a DSS placed before the first page of each multi-page document.

Using the Document Separator Sheet (DSS) for Multi-Page Documents

Capture Enterprise has processing logic that can identify the end of one document and the start of another in a batch that contains more than one multi-page document. However, this logic is not infallible and relying on it may result in having to later manually verify the documents. Therefore, the best practice for batches that contain multi-page documents is to use the Document Separator Sheet (DSS). The DSS is not required for single-page documents.



The DSS is provided as a TIFF file with the Capture Enterprise installation in the following location:

<Capture Enterprise Install>\Images\SeparatorPage.tif

Print out a copy of this page for each multi-page document in the paper batch you plan to scan. A DSS should be placed before the first page of each multi-page document in the batch, including the first one.

Paper Document Scanning

Observe the following best practices when scanning paper documents for Capture Enterprise input:

- **Digital Output File Format:** Tagged Image File Format (*.tif) is required
- **Scan Resolution:** 300 dpi (dots per inch) yields the best quality scan for OCR
- **Scan Mode:** Color/Grayscale mode yields the best quality scan for OCR
- **Document Alignment:** Monitor the document alignment when you place the documents into the scanner or MFP device and be sure they do not skew (become "crooked") in the scanner. Capture Enterprise contains image enhancement functions that will de-skew images, but better input from the beginning will lead to better processing results.

Scanned Digital File Management

Observe the following best practices for managing the digital output file (*.tif) of the paper batch scan that will be used for Capture Enterprise input:

- **Unique Input File Naming Convention:** Establish a file naming convention for paper batch files that provides a unique identifying name for each batch file. This is important as it will enable IntelliChief Document Capture Tracking if you later need to find out the status of a batch that may not have completely processed in Capture Enterprise. Without unique batch file names, document capture tracking is not possible.
- **Sending Input File to Capture Enterprise:** How you send your input file to Capture Enterprise will vary based on whether you are using a dedicated desktop scanner or an MFP device.
    - **MFP Devices:** MFP devices have capabilities that most desktop scanners do not. For example, some MFP devices allow you to add a button to their scan screen that will enable users to simply press that button and automatically place the scanned output file in a designated network folder such as the Capture Enterprise automated input processing folder. Many MFP devices also allow you to email scanned files, so you could email your MFP paper batch scan directly to the IntelliChief Email Import Service inbox where it can be processed.
    - **Desktop Scanners:** If you are using a desktop scanner, you will most likely have to manually browse and save the input file for each batch to the Capture Enterprise automated input processing folder.

How Capture Enterprise automatically processes document input is explained in the Operations and Processes chapter.

Addressing Document Input Quality Issues

Even if you follow the best practices described in this section, you may still encounter document input quality issues described in the Low-Quality Document Input and OCR section. Using the low-quality examples from that section, Capture Enterprise offers the following capabilities to address document input quality issues:

- **Document image enhancement:** Capture Enterprise has a built-in capability to automatically recognize and fix document issues with image contrast, clarity, and alignment/skewing.
- **Capture Enterprise Settings:** These settings allow users to address issues with document input including inconsistent character spacing, mixed data types, and font-related issues. These settings must be created and configured using the Capture Enterprise Settings tool in the IntelliChief Management Console. The function and use of this tool are explained in the Administration chapter.
- **Document capture zone adjustments:** The function and use of document capture zones and document "fingerprinting" are explained in the Processing Document Batches – User Tasks chapter.