

Capture By Scan Image Services Ireland

Input/Capture

Document capture is the conversion of a paper document into an electronic image of that document while data capture extracts information from a business form.

What Type of Document Can You Scan?

The Document type to be scanned will determine what kind of scanners are needed - photos or glossy items, color documents, hand written correspondence, etc. will need extra handling to ingest.

Documents such as invoices, applications, cheques allow the scanning speed to be optimised and are generally higher volume. You will also need to determine your scanning requirements such as the size, color or not, average number of pages, quality, and volumes.

The Conversion Process

Capture

Documents can be captured by:

- Scanning
- Importing electronic documents (word files, video, spreadsheets etc.) for sharing or archiving
- Converting existing electronic documents into unalterable images

Document Preparation

Paper documents have to be manually prepared for scanning: torn pages are repaired, sort documents, remove sticky notes (or tack down so they can be scanned too) paper clips and staples etc. This task is time-consuming and too often underestimated. However poor document prep will slow down document throughput if not properly executed.

Scanning

The transformation of paper documents into an electronic image. Can also mean the digitising of microfilm. Electronic images can also be captured by:

- **Fax** - Software can read from the fax server. Be Aware that image quality will be lower, which may negatively affect recognition accuracy
- **Multi-Function Device (or Peripheral) - Network** - Connected MFDs can suffice for low-volume imaging needs.
- **Camera Phone** - Higher-resolution cameras in mobile phones, and software designed to work on a mobile phone, allow the capture and conversion of documents on the go - whether a conference whiteboard or a restaurant menu.

Type of Scanner

Document throughput in the real world will be slower than the scanner's rated speed. Plan accordingly. Scanner categories:

- **Workgroup:** 10 - 25 ppm
- **Departmental:** 26 - 40 ppm
- **Mid-volume production:** 40 - 120 ppm
- **Production:** 120 plus
- **Large format:** for over-sized documents and engineering drawings
- **Cheque scanners:** Will read the account number on a cheque, speeding processing.
- **Microfilm:** for digitizing film-based documents

Document Imaging

Document images can be saved as one of a number of file formats, including:

- **TIFF (Tagged Image File Format)** - Generally used for monochrome office documents. Can be compressed. ISO standard. Multiple types of TIFF.
- **JPEG** - Often used for color documents. Also a compression method. ISO standard. Multiple types of JPEG.
- **PDF - De Facto standard.** Replica of documents. With OCR text can be full-text indexed and searchable.
- **PDF/E** - Recent standard for large format and multi-layer engineering drawings.
- **GIF (Graphics Interchange Format)** - Exchange and display for high-quality/resolution graphics.

Forms Processing

When forms are scanned, either the data or the entire form, can be captured - depending on your business requirements. Data captured from a form can be entered seamlessly into the appropriate database and be linked to other enterprise applications such as ERP.

Image Clean-up

Many products include image-enhancement features to increase the quality of the scanned documents - de-skew, de-speckle, crop, rotate, and/or blank page and double feed detection, etc. This step can be performed by software or an image processing board in the scanner.

Recognition

Recognition is valuable for indexing each image.

- **OCR (Optical Character Recognition):** recognizes machine-printed characters
- **Zonal:** used where only specific fields on a form are required
- **Full-Text:** free form document conversion allowing search on all words in the document
- **ICR (Intelligent Character Recognition):** for hand-printed characters
- **OMR (Optical Mark Recognition):** recognizes check boxes, filled-in bubbles, etc.
- **Bar Codes:** read and extract information from a pre-printed bar code

Indexing and / or Data Extraction

Indexing is NOT optional. There is no other way to find and manage documents.

1. **Key from index fields (document type, date, customer name, etc.)** - a data entry person manually indexes documents
2. **Barcodes allow auto-indexing.** By storing form information on a bar code before scanning a batch of documents, certain index values can be automatically populated.
3. **Zone OCR** - also automatic
4. **Ingest from other applications** - email, word processing, etc. (Metadata for the document, i.e., subject line, sender, etc., become the index fields.) The index can be either key fields or full text. A combination is generally best.

Quality Control/Assurance

Electronic images must be double-checked. In key-from-image, data can be validated by a second operator or via automated processes like database look-ups. Bad images are flagged and re-scanned.

Output

After going through the capture process, electronic content is released to:

- Storage
- Document Filing/management system
- Records management/Retention plan
- Print
- Email
- Workflow/Business Process: a customer claim could launch an insurance business process, for example
- Enterprise Content Management: access and collabora-

