



AnyDoc Software is a global provider of innovative document and data capture solutions which have been the industry standard for more than a decade.

Thousands of companies worldwide rely on its software to eliminate millions of man-hours of manual data entry while boosting data accuracy and productivity levels. AnyDoc provides a complete front-end solution for total information capture with its flagship OCR for AnyDoc® software. Any paper form, invoice and document can be automatically processed with full data extraction without the need for manual keying. AnyDoc Software's solutions are the preferred choice of more than 2,000 companies worldwide.

# Easy, Effective Data Processing

## The seven keys to an unstructured and semi-Structured document solution

*Charles W. Jackson*

Here are seven characteristics to consider when selecting a solution for your unstructured and semi-structured documents.

based on those parameters. Taken further, the software should be able to cross-check against your existing databases to ensure the information is accurate, valid and complete.

### 1. Intelligent Data Collection

Look for a solution that uses logic, keywords and business rules to locate the correct data, no matter where it may be found within the document's page(s). Look for helpful tools like a user-friendly wizard to guide you through the setup process.

### 5. Remote Scanning and Correction Capabilities

Many organizations have offices scattered across the country or around the world. An Internet-based software solution that can remotely scan documents and/or validate data for distribution to a central office can reduce labor costs, eliminate shipping costs and greatly improve processing time.

The efficiency of standard OCR-based software has increased dramatically in recent years. With the help of template design, an organization can readily automate the extraction of data from static locations on structured documents.

Standard OCR solutions use a template-based process, where users define the type of data to be extracted based on their location on the form. Other elements on the form, such as company logos, field descriptions (e.g., 'Street Address') and lines are excluded from processing. This is a solid method of collecting the data you need from consistent forms.

Often, however, an organization's crucial data is received on unstructured or semi-structured documents – such as invoices and Explanation of Benefits (EOB) statements. The data you need often appears in a different location on these documents, from one vendor to the next. Technology now exists that streamlines this process for these unstructured and semi-structured documents. To save time, money and frustration, an organization should select a solution that offers better automation flexibility for all document types.

### 2. Single or Multiple Page Processing

Many unstructured documents, such as invoices and EOBs, contain more than one page, so a solution that can handle multiple page documents is essential. A complete solution treats this as one logical document throughout the entire process – from scan to data capture to storage.

### 6. All-in-One Solution

Structured and unstructured forms are two sides of the same coin. No business receives one form type and not the other, so look for a single software solution that can automate both.

### 3. Detail Line Processing

You also need a solution that can extract line-item detail data (e.g., unit price, quantity ordered/shipped, terms, part number, etc). Don't limit yourself to software that captures only general document information.

### 7. Growth and Longevity

Finally, look for adaptability. The solution should be able to grow along with the volume and application needs of your company. And it should keep up with emerging technology for years to come.

### 4. Business Rule Compliance

Let the software work for you! You should be able to define precisely which data to capture, based on your specific business rules. The purchase order number, for example, might be a 12-character, alpha-numeric field that always begins with the numbers '216.' Your solution should be able to locate data

A solution that meets all of these criteria will simplify the automation of data capture from your document types – structured, semi-structured or unstructured. In turn, you'll enjoy lower costs and more accurate data, and you'll be positioned to accommodate future business needs.

*Charles W. Jackson is president of AnyDoc Software, Inc. He can be reached at 813-222-0414 or via e-mail at info@AnyDocSoftware.com.*