

**White Paper**

# **Understanding the Difference Between Structured and Unstructured Documents**

**By: Randy Van Ittersum and Erin Spalding CDIA+  
[www.disusa.com](http://www.disusa.com)**

## **Differences Between the Two Document Types**

---

What is the difference between structured and unstructured documents? With a structured document, certain information always appears in the same location on the page. For example, in an employment application the applicant's name always appear in the same box in the same place on the document. In contrast, an unstructured document has the opposite characteristics – information can appear in unexpected places on the document. An example would be in a hand written note or a whitepaper.

Some documents share the characteristics of both types of documents, such as invoices. For example, suppliers' invoices feel like a structured document because they have a consistent appearance from one billing period to the next. However, when viewed in aggregate by an accounts payable department that receives thousands of invoices daily in a myriad of different formats; they seem more like structured documents.

---

## **What About Template-Based OCR Systems**

---

Some document imaging systems advocate template-based OCR (optical character recognition) to capture the information needed to identify the document for later retrieval. They call this pixy dust, where you don't need to do anything with the documents other than to load the automatic document feeder. Unfortunately this solution only works well with structured documents, and it is not 100% accurate even under the best conditions. (For more information on the accuracy of OCR, read our whitepaper on that subject).

Needless to say, you will need to have a different method to capture the key information needed to retrieve documents that are unstructured. In many organizations unstructured documents represent the majority of the documents that will be imaged with a document imaging system.

---

## **Characteristics of Structured and Unstructured Documents**

---

Type of Document	Structured	Unstructured
Characteristics:	<ul style="list-style-type: none"><li>• Familiar data appears in the same place every time.</li></ul>	<ul style="list-style-type: none"><li>• Data appears in unexpected places in the document.</li></ul>
Examples:	<ul style="list-style-type: none"><li>• Insurance claim form</li><li>• Employment application</li></ul>	<ul style="list-style-type: none"><li>• A letter</li><li>• A hand-written note</li></ul>
Used by Organizations:	<ul style="list-style-type: none"><li>• Low volume operations</li><li>• Internally created invoices</li></ul>	<ul style="list-style-type: none"><li>• High volume operations</li><li>• Invoices received from outside the organization</li></ul>

---

## **Conclusion**

---

Every organization will have both structured and unstructured document with which to contend. It is generally a good idea to purchase a document imaging system that offers the maximum capabilities to deal with both types of documents, rather than purchasing a system that caters only to a single document type.