

Unstructured Data Doubles Every 3 Months

What Is Unstructured Data and Why Is It Important?

Most people believe having information in digital format means they can find it quickly and easily. The reality is that finding information in the business world today is more difficult than ever. While the digital revolution has made creating information easier, our ability to find this information when we need it decreases exponentially over time. According to Gartner more than 80% of the information generated consists of unstructured data¹ and the amount of unstructured data doubles every three months².

People use unstructured data in the normal course of business every day. While they may not be aware, they use it for creating, storing and retrieving reports, e-mails, spreadsheets and other types of documents. Unstructured data consists of any data stored with no identifiable structure where there is no conceptual definition and no data type definition - in textual documents, a word is simply a word³.

Unstructured data can be textual or non-textual. Textual unstructured data is generated in forms such as email messages, PowerPoint presentations, Word documents, blog posts, and social media posts. Non-textual unstructured data is generated in media like JPEG images, MP3 audio files and Flash video files.

If left unmanaged, the sheer volume of unstructured data that is generated each year within an enterprise can be costly in terms of storage. Unmanaged data can also pose a liability if information cannot be located in the event of a compliance audit or lawsuit³. The information contained in unstructured data is not always easy to locate. Several studies show that 25%-35% of a knowledge worker's time is spent looking for information and about half the time they can't find. This means that they either have to recreate it (if this is even possible) or deal with the consequences of critical information being missing.

The key is how to leverage this data to control access while sharing it amongst the individuals and teams who need it to enable collaboration.

The growth of unstructured data is expected to far outpace the growth of structured data. While transactional data is projected to grow at a compound annual growth rate (CAGR) of 21.8%, it's far outpaced by a 61.7% CAGR prediction for unstructured data⁵.

This is not a problem that will go away. To the contrary, this problem will just keep getting bigger and bigger. The longer an organization waits to deal with this problem, the more unmanageable and unwieldy it will become.

How Is Unstructured Data Managed Today

In most organizations unstructured data is simply either stored on a local or network drive or, in some cases, in an enterprise content management system. With the amount of unstructured data doubling



every three months, just remembering what you called something or where you put it quickly becomes difficult. How often do you search for a document that you worked on just a few months ago that you and either can't find it or waste a lot of time before you finally do? The document is still there taking up valuable storage space, you just can't find it.

So maybe you emailed it to five of your coworkers, so you check with them to see if they saved a copy or still have the email. Many of them probably did save a copy just "in case" they needed it, but they are just as likely not to be able to find it as well. As you create or save more and more documents, the "haystack" continues to grow the likelihood that you find "THE needle" decreases substantially.

From the IT manager's perspective, this scenario creates a nightmare as the volume of information they have to maintain quickly becomes untenable. They have no budget for expanding the live storage at the rate new information is being generated. They don't dare delete any files for fear they'll delete the one file someone was actually looking for, so instead they archive information which hasn't been accessed in a period of time off to external media, the most economical still being tape.

The reality is they could have just deleted it because the "haystack" through which they'd have to search to find the "needle" is now so enormous the time and cost to find it is astronomical, so these archives sit there mostly untouched until...

A lawsuit has been filed against the company. Legally, the company is bound to protect and produce anything that may be relevant to the lawsuit. The only problem, no one really knows what is really on those old archive tapes, let alone on the company shared drive, so in come the e-discovery processors.

You'll soon find out the time and expense required to search through all this unstructured data and will wish you'd done something to manage it long ago. Case in point: DuPont spent \$12 million going through old computer tapes just verify there was nothing relevant on them they had to produce because they didn't know what was on the tapes⁶. Furthermore, they determined more than half the documents could have been destroyed because they were beyond their required retention period.

Who Should Be Concerned About Managing Unstructured Data?

Everyone should be concerned about managing unstructured data because we all create information, but the needs are greatest where there are regulatory requirements that mandate record keeping and areas that are at the greatest risk of potential litigation and those are Human Resources and Finance.

The area of a company at the greatest risk from both regulatory audits and potential lawsuits is Human Resources. Federal law requires that employers maintain I-9 forms for every current and former employee and the Federal government has recently stepped up audits. In May 2012 the Wall Street Journal reported that the Obama administration recently enhanced efforts to seek employment records from hundreds of companies. Since January 2012, the Obama administration has audited over 7,533 employers and doled out over \$100 million in fines⁷.

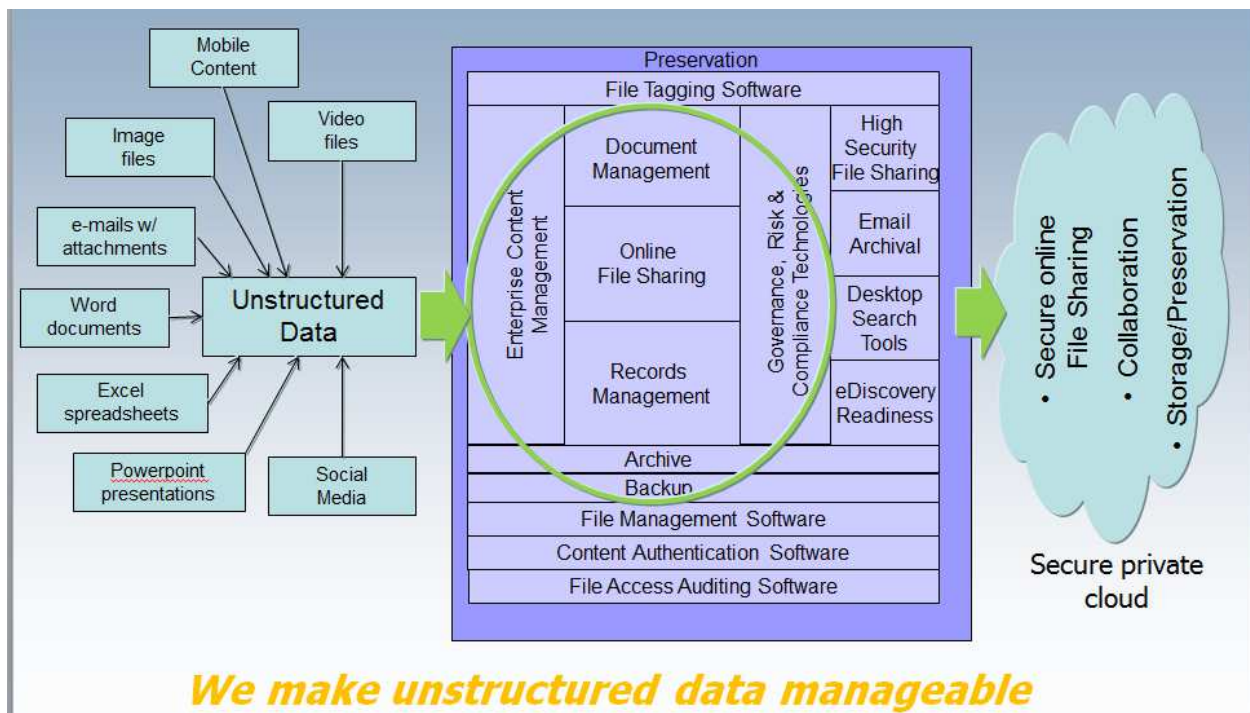
While the I-9 forms are in the forefront because of the frequent audits, limited time to produce the forms, and possible hefty fines, pretty much every piece of information that HR touches is a potential for either a government audit or potential lawsuit. This information can include information such as other government forms like W-4's, regulatory training compliance records, harassment and discrimination

issues and training, performance reviews, disciplinary actions, or termination records. You quickly get an idea why HR should care.

While the audits and lawsuits related to finance may not be as frequent as they are for HR, the number of records that must be kept required for regulatory compliance is far greater. Since most are in the form of unstructured data and there are a lot more records that must be kept, finding them when required or needed in the sea of information becomes more difficult.

Since Sarbanes-Oxley went into effect the number of new regulatory compliance mandates continues to grow which creates another issue. While records must be kept, they are only required to be kept for a certain amount of time. The issue is once you can finally purge records, being able to find them in order to get rid of them becomes problematic. At best, it makes search for and locating the records you do need more difficult and, at worst, you now have records that are discoverable that you could have purged.

Again, while HR and Finance probably have the greatest needs, managing unstructured data is an issue across the entire enterprise.



Searchable. Simple. Secure – File Sharing, Collaboration, and Storage

These are three important words to live by when dealing with unstructured data. The InfoPreserve Vault is the first comprehensive private Cloud-based platform that solves all your key problems in dealing with unstructured data. It allows you to securely manage your information yet makes it accessible anywhere you have a secure Internet connection. By allowing you to designate any folder you can access on your computer as a “watched folder,” information is automatically uploaded to the



Vault every time you save something to a watch folder. Or you can use traditional drag and drop or right-click to upload any individual file into the Vault.

To search for information you don't need to remember what you named a file or where you put it. You search for it based on something that is contained in the document or a tag that you assigned it. When you upload a file, you automatically designate whom else can see the file, so when someone else needs to find information you created, as long as they have permission, they can search and retrieve it as well as export or preserve it.

Because there is no software to license, hardware to install, or hierarchy to create, you can begin using the Vault at your own pace. For example, Human Resources can decide they want to tackle I-9 forms first and then move on to other areas within HR. Finance can start using the system at the same time as HR or they can begin moving their critical information in at their own pace. The same holds true with other departments within the company. This is the beauty of using a Cloud-based solution.

If you are ready to start managing your unstructured data using the InfoPreserve Vault, send an email to salesupport@infopreserve.com or call (585) 542-4183.



¹ "IBM: 80 percent of our global data is unstructured (so what do we do?)", Computer Weekly, Adrian Bridgwater, October 26, 2010.

² "Introducing the High-Performance Workplace: Improving Competitive Advantage and Employee Impact", Gartner, May 16, 2005.

³ "Two Worlds of Data – Unstructured and Structures", Information Management, Geoffrey Weglarz, September 1, 2004.

⁴ "What is unstructured data? – Definition from WhatIs.com", TechTarget, April 2010.

⁵ "IDC Enterprise Disk Storage Consumption Model", IDC, Fall 2009.

⁶ "Proactive eDiscovery: The Key to Reducing Litigation Risks and Costs", Information Security Systems, Ursula Talley, 2008.

⁷ "Fresh Raids Target Illegal Hiring", Wall Street Journal, Miriam Jordan, May 2, 2012.