



## Can Your SharePoint® Backup Harm Your Business?

*Updated 5/19/08*

### Evaluation Standards and Best Practices

The native SharePoint tools for backup under SharePoint 2007 are a major step forward over the 2003 edition. But there still are major issues in the areas of granularity, performance, fidelity, and usability. By using just the native tools, organizations will still be exposed to the risk of losing or damaging critical business information. This white paper is intended for IT Professionals who are responsible for helping companies evaluate, plan and execute data backups for SharePoint.

# Table of Contents

---

- Introduction..... 3
- Evaluating Your Backup Strategy ..... 4
  - Full Fidelity of Original Data ..... 5
  - Granularity of Backup..... 7
  - Flexibility of Schedules ..... 8
  - Differentiating Storage ..... 9
- Planning Your Backup Strategy ..... 11
- Implementing Backup and Recovery with DocAve from AvePoint..... 13

# Introduction

---

With companies quickly adopting Microsoft SharePoint as their content sharing and collaboration platform, features such as *document sharing*, *team discussion* and *document-based workflow* are drawing in an ever increasing number of business users. Consequently, business-critical information traditionally stored under a variety of applications is undergoing a dynamic migration onto a new platform.

Data that was kept on intranet web sites, ERP or HR applications, e-mail servers, and other disparate sources, such as shared network folders, are now being shifted at a grass root level to SharePoint. This is great news for users and managers alike since SharePoint provides them with a complete set of tools through which they can quickly and easily access the documents and information. To storage architects and backup planners, however, things are a bit more complicated.

To better understand the issue, let's look at how data is stored under SharePoint. In order to properly support a wide range of functions such as *content search*, *business intelligence*, and *enterprise record management*, SharePoint is utilizing Microsoft SQL as the foundation of its data storage. By using this readily available and mature enterprise application, users are ensured advanced data accessing features as well as advantages in performance, reliability, scalability, and security. Then what is the issue?

**With content databases growing to tens, hundreds of Gigabytes, and even terabytes, classification of contents and implementing servicing levels are becoming ever more problematic. Without classification, a business would have to implement a service common denominator for all data types, regardless whether they are critical or not, modified frequently or static.**

How do you satisfy different Recovery Service Level Agreements to different categories of data hosted in the same database? How can you offer a higher level of recovery guarantee for a team that's working on a business critical deal? How can you both back up and recover the exact document, without being required to back up the other 99.9%? How can you ensure that the recovery of those documents does not harm your business processes? Those are the questions that this white paper explains.

# Evaluating Your Backup Strategy

---

To highlight a few major issues faced by SharePoint data architects and administrators, consider the following questions:

- Are all of the data stored in your SharePoint environment of the same value to your business?
- Are those data sets being updated/modified at the same frequency?
- Is it acceptable to you and/or your company's executive management that all associated documents' metadata (author, time of last modification, time of creation, securities, etc.) is lost forever during recovery?

If your business answers "yes" to all of the above questions, then the native backup and restore solution provided within Microsoft Office SharePoint Server 2007 or standard SQL backup methods will most likely be sufficient.

Otherwise, without proper planning, the lethal combination of growing content volume, the use of DB as primary storage, and the lack of sufficient native backup tools could very well make SharePoint a Pandora's Box for your organization.

So what should a business look for in an ideal SharePoint backup strategy? Below are some key requirements:

- Full Fidelity of Original Data
- Granularity of Backup
- Flexibility of Schedules
- Differentiating Storage

## Full Fidelity of Original Data

This *must* be your number one requirement. In a collaborative environment, data only makes sense when all its associated metadata (last time of modification, original author, etc.) is fully preserved. Therefore, the restoration must ensure that it retains both the original data and the full metadata information.

### Understanding the Importance of Metadata

With SharePoint 2007 it is easier than ever to pre-format the usage of associated metadata for specific types of documents with custom content types. These values are defined by typical business/site requirements and implemented by a SharePoint administrator to be distributed into a document library and site structure.



*For example, if you implement a **sales contract template**, created content type can define the metadata for that contract, including the account number, quote number, sales manager's name, workflows required to review and complete the contract, and policies that enforce auditing of actions related to the contract. From that point on, any document in your library to which you associate the Contract content type will include all of these features as metadata for each document and will enable authors to create new contracts based on the template.*

---

Many documents stored in SharePoint sites go through an intense document management cycle and may include many participants. Any actions taken on these documents generate and store metadata critical for business. Possible participants in any business document's life cycle include the: Creator -> Editor -> Reviewer -> User. Each associated action will record date/time and name of the person editing the document along with version control and editing and approval history.

### Satisfying Compliance Policies

There are also compliance issues if full fidelity of data is not maintained. Updated metadata, even unwittingly, can be construed as falsification of records or the destruction of valuable business assets.

As stated in the Sarbanes-Oxley Act, commonly referred to as "SOX", there are several areas specifically related to the information

## “SOX” At-A-Glance

The Sarbanes-Oxley Act, (a.k.a. Public Company Accounting Reform and Investor Protection Act of 2002), was signed into law on July 30th 2002, and introduced highly significant legislative changes to financial practice and corporate governance regulation. The legislation is wide ranging and establishes new or enhanced standards for all U.S. public company boards, management, and public accounting firms.

technology, data backup, management processes and disclosures, but below are the key sections:

**SECTIONS 103 & 104:** Accounting and auditing entities must retain all documents and data to audit reports of companies required to comply with SOX for a minimum of 7 years. (Auditing investigators usually demand all versions of documents to be available for their analysis.)

**SECTION 302, 404:** Responsibility of attesting to the content, accuracy, and, most importantly, the authenticity of financial reports issued by the organization lays directly on the shoulders of executive management and the board of directors. An internal control report (statement) that acknowledges that management is responsible for maintaining an “adequate internal control structure” must be produced and attached to all yearly financial reports. It must also include an assessment by management of the control structure’s effectiveness.

Compliance with the Act much depends on reports created from sensitive and accurate data. These reports must be generated from actual documents, including SharePoint documents, with strict access and security processes, and accompanied by an executive-authorized documentation to attest to the existence of and adherence to these standards.

Proper backup is critical to the support of document storage of these reports and to ensure that no disaster or human action can prevent the organization from generating critical reports on its original data as well as keeping the original data intact. Restoration of any SharePoint document with its metadata, such as author, create time etc, are critical aspects in the document preservation policies.

### Why Losing Metadata is a Major Problem

If a SharePoint native backup solution is utilized to recover corrupted or missing document, the restored document may lose its original metadata and recovered documents will be created under the name of the person restoring it. Usually, that would be a SharePoint administrator. This unintentionally causes a falsification of records that not only leads to jeopardizing integrity of the business data, but also has legal consequences for the person responsible for backup and restoration procedures. It also creates major liability for executive management, especially for public companies, as defined by Sarbanes-Oxley Act of 2002.



**Ask yourself:** *In your business, is there a need to provide high level of service? Can you afford to manipulate critical data that your organization depends on? Can you justify changing the metadata of files that do not belong to you?*

## Granularity of Backup

Database-based backup and recovery approach is quite good at maintaining full fidelity of the content. If you backup the whole database, it makes sense that everything is kept intact. However, the problem is also that “everything” must be backed up.

Your backup solution has to allow for the backup of content at various levels, i.e. item, folder, site, site collection, etc. This is essential because some parts of a site are more important than others and therefore require different service.

Many different events can cause document destructions, starting from human actions (malicious or in error), to viruses and hardware failure. Let’s assume that your current backup procedures, as in case of SharePoint native backup solution, do not include an item-level backup and restore capabilities. In order to restore a single item, say a previous version of a corrupted document, you’d have to depend on a full database-based backup.

### **There are two ways you can restore a document:**

1. Database-based backup and recovery approach is quite good at maintaining full fidelity of the content, but it would also require that “everything” was backed up and restored. This usually presents a serious productivity issue, as restoring a database with intention to retrieve only one file would ultimately overwrite all content on SharePoint that has been edited after the last backup.



*For example, if you perform backups for a team of 100 people once a week, and one user has requested a restoration of a single document 5 days later, the restoration of the document using Microsoft native backup and restore will overwrite all documents edited in the past 5 days by the other 99 members. (Note that during the restore you can choose to “Create New” content, as opposed to “Overwrite”, but it will generate a separate duplicate copy of all content).*

---



2. A common work-around is to restore the backup data to a temporary storage and perform a move of the single item back into the production environment via standard SharePoint tools; the limitation is that the **metadata will be lost**. So unless metadata can be maintained, this method of recovery still will lead to the loss of valuable data.

Note that the MOSS 2007 Recycle Bin would be able to recover deleted documents or libraries, but it would not be possible to restore a deleted subsite or top-level site collection. Also, as it only captures "deletion" events, it is impossible to restore a document that has been corrupted or lost due to the hardware failure. Thus similar to the analogy of corporations backing up individual workstations despite the fact that every Windows workstation has recycle bin installed as part of the OS, the introduction of MOSS 2007 Recycle Bin, while a significant improvement over previous versions of SharePoint, does not relieve SharePoint administrators from scheduling reoccurring SharePoint content backups.

**Ask yourself:** *Does your scheduled backup take an all-or-nothing approach? Can you make granular restores with all of the associated metadata? Can you do it quickly?*

## Flexibility of Schedules

Different parts of your SharePoint environment can vary in business importance. In addition, their frequency of modification will be vastly different.



*For example, a project site that's being used by a large merger and acquisition team for an ongoing deal will have significantly more modifications than a quiet blog page. It would make sense for your backup solution to offer the ability to backup that portion daily or even hourly whereas the blog page can be included in the weekly backup schedule.*

The limitation of the MOSS 2007 native backup is that not only does it not allow for finer granularity than an entire content database, but also it **does not contain any scheduling options**. The administrator would need to manually kickoff the backup processes each time backup needs to run. Unfortunately, this is not an efficient way to



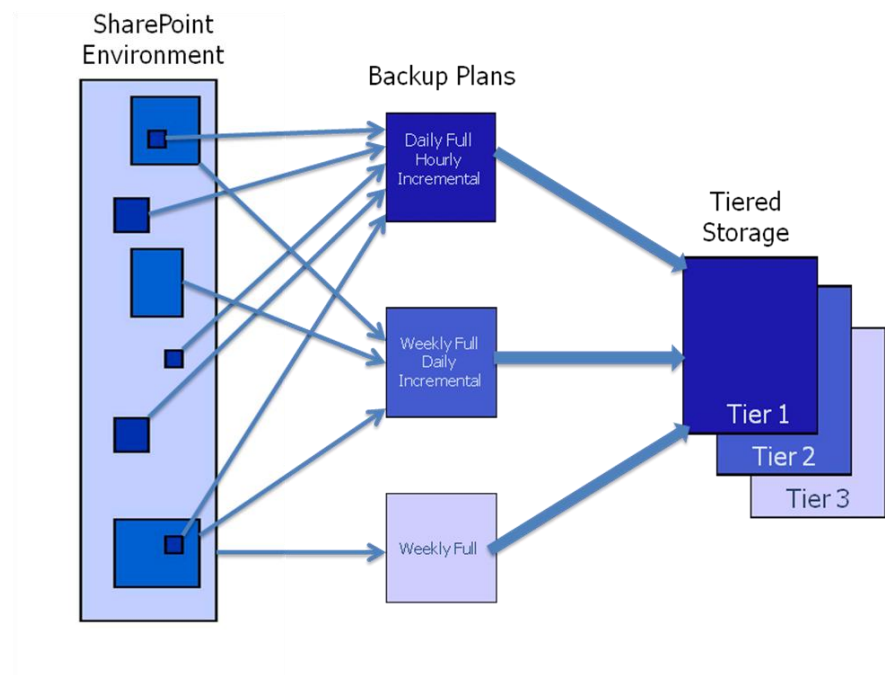
backup the content, as both the backup and recovery should be efficient in time and resources. Also, if you only want to backup a few important documents, it is probably not very logical for you to back up the entire content database.

**Ask yourself:** *In your business, is there a need to provide different levels of service for data of varying business importance? Is there a need to schedule different backup times for different sites? For example, a busy team site or a dormant personal page? Should the site which hosts the quarterly sales document that the VP of Sales has just spent an entire day on be on the same backup schedule as the announcement for the return of casual Fridays?*

## Differentiating Storage

Your SharePoint backup solution should be able to take advantage of existing tiered storage capabilities in an organization. For business critical sections of your SharePoint environment, the backups should also be reflected in the storage tier, to make sure the backup data is reliable.

Below we describe backup standards that can be applied by your business to evaluate the functional strategy and scheduling.





In the above illustration, the left side is your SharePoint environment, or the backup source. The right side is the Shared Storage Pool or the backup destination. In this particular case, the storage pool is tiered into three levels based on the business criticality of data. The thin arrows represent scheduling based on the granularity of the backup schedules, while the thick arrows represent storage based on the various tiered storage support. The middle section is the flexible scheduling option which glues the whole strategy together.

The ability to maintain full fidelity throughout this process is a prerequisite for this illustration. Meanwhile efficiency is evidence since the backup source can be of any level within the SharePoint environment. In the illustration, small sets of business critical data are backed up more frequently and this backup data is managed separately in regard to storage.

**Ask yourself:** *Do you differentiate backup storage tiers from one SharePoint site to another? What if they are on the same DB?*

For more information on the limitations of the native tools, please refer to the AvePoint white paper "MOSS 2007 Backup Strategies" at [www.avepoint.com](http://www.avepoint.com).

# Planning Your Backup Strategy

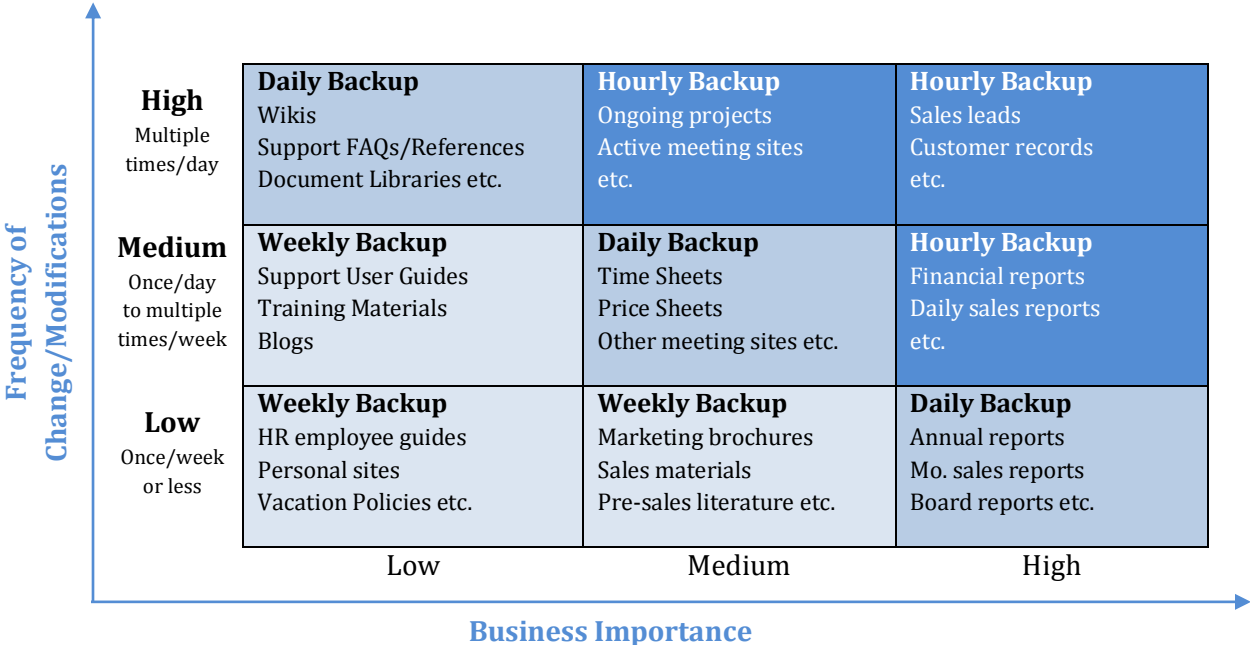
## Not all Data is Created Equal

Understanding of this key concept will be critical in planning your SharePoint backup strategy. The most challenging business decision to be made by a SharePoint backup administrator during the design of a competent backup architecture is how to determine the scheduling of your backups for various data types. This decision is critical to meet proper service levels. To achieve this, two important items need to be evaluated:

- The business importance of the data
- How often is the data being modified

The following SharePoint Backup Planning Criticality Matrix can help *businesses* in determining the right mixture of backup scheduling strategy for their SharePoint environment. **Note that this information is provided as a sample guideline only – you should adjust the schedules to fit your specific business needs.**

## SharePoint Backup Planning Criticality Matrix



**Ask yourself:** *Is some data in my SharePoint environment more critical to my business than other data? Do I need to back up the entire database to provide adequate backup scheduling for the critical data? Do my backup schedules match the frequency of change of my data? Am I backing up critical, frequently changing data often enough?*

## **Features Worth Considering**

In addition to the requirements and questions discussed above, these are some nice-to-have features you should consider while evaluating an enterprise backup solution:

### **Ease of Use**

The solution should provide administrators with a single pane of glass GUI and allow for the configuration and saving of various scheduled backup plans.

### **Speed**

The strategy should be based on an architecture where the speed of both backup and recovery will be acceptable to the business while ensuring minimum impact to the business' day-to-day operations.

### **Scalability**

The solution should be able to scale as the company's SharePoint environment grows.

### **Enhancement Features**

- Compression – This should be supported by the solution to save backup storage requirements.
- Encryption – This should be supported to provide backup data security.
- Data de-duplication – This should be an optional advanced feature to save storage capacity.

# Implementing Backup and Recovery with DocAve® from AvePoint

---

The DocAve Software Platform is a one-of-a-kind solution for complete SharePoint infrastructure management. DocAve's item-level backup and restore functionality in **DocAve Backup and Recovery Module** lets Administrators backup their SharePoint environment at any level, from an entire server farm down to a specific list item or document version. Upon restore, utilizing an intuitive keyword/criteria-based search engine, Administrators can navigate to a specific folder/list to locate an exact item – whether an attachment, a single document, or even a single version of a document.

Key business-requirements addressed by *DocAve Backup and Recovery* include:

### **Full Fidelity of Original Data**

DocAve ensures that restored data retains **both the original data and all metadata**.

### **Granularity of Backup and Restore**

Granular backup lets Administrators take advantage of **existing tiered storage capabilities** in an organization. Business criticality of data should determine the appropriate usage across storage tiers. Granular recovery allows restoration of individual items, directly into its original location without the need to restore the complete database.

### **Flexibility of Schedules**

DocAve lets Administrators perform backups on different components (i.e. site, sub-site, folder, document library, etc.) of the environment at **customizable frequencies/schedules**, based on business importance and frequency of modifications and/or access.

When it comes to recovering from a disaster, there are simply no native backup solutions for SharePoint platform recovery. If a component fails, Administrators must reconfigure it manually. **Most components have numerous configurable parameters, requiring a very time-consuming manual reconfiguration** during disaster recovery. (That is...if the configurations have been saved at all.)

With *DocAve Backup and Recovery*, Administrators utilize a simple “point-and-click” approach to recover all relevant SharePoint data and configurations.

### **All Servers, All File Systems, All Configurations**

**SQL Server backups only cover databases**, and will not protect items such as Shared Service Provider (SSP) applications, the Search Index, or File System Resources such as Web Parts, Solutions, Features, and other configurations.

### **SharePoint-centric Logic Ensures Consistent Backup**

SQL backups **do not have any SharePoint-centric logic** to avoid inconsistencies during backup. Each individual database backup can be consistent atomically, but if configuration changes are introduced between backups, the farm backup will not be consistent, thereby invalidating the entire backup image. *DocAve Backup & Recovery*, on the other hand, maintains database configuration consistency.

### **In-Place and Out-of-Place Restore for Flexible Platform Recovery**

*DocAve Backup & Recovery* allows migration of an entire Web Application to a different farm. This is a unique feature that is not supported by any other tool, native or otherwise.

Without a proper data protection strategy, organizations expose themselves to unnecessary downtime and lowered productivity. AvePoint's *DocAve Backup & Recovery* Module are designed to help enterprises address these crucial issues.

To learn more about DocAve software for SharePoint and to purchase, visit [www.avepoint.com](http://www.avepoint.com).

**Copyright**

2001-2008 AvePoint, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by *any* means, electronic, mechanical, photocopying, recording or otherwise, without the prior written consent of AvePoint, 3 Second Street, Suite 202, Jersey City, NJ 07311, USA

**Trademarks**

AvePoint DocAve®, AvePoint logo, and AvePoint, Inc. are trademarks of AvePoint, Inc.

Microsoft, MS-DOS, Internet Explorer, Microsoft Office SharePoint Servers 2007, SharePoint Portal Server 2003, Windows SharePoint Services, Windows SQL server, and Windows are either registered trademarks or trademarks of Microsoft Corporation.

Adobe Acrobat and Acrobat Reader are trademarks of Adobe Systems, Inc.

All other trademarks are property of their respective owners.

**Changes**

The material in this document is for information only and is subject to change without notice. While reasonable efforts have been made in the preparation of this document to assure its accuracy, AvePoint assumes no liability resulting from errors or omissions in this document or from the use of the information contained herein. AvePoint reserves the right to make changes in the product design without reservation and without notification to its users.

AvePoint  
3 Second Street, Suite 202  
Jersey City, NJ 07311  
USA

**For AvePoint Products and Services, please visit our website at [www.avepoint.com](http://www.avepoint.com).**