

Managing SharePoint - Know Your Data

By Steven Pogrebivsky, CEO at MetaVis Technologies

The only way you can truly manage your content is to understand it. Depending on who you are, that may sound really simple. The reality? It's something organizations struggle with every day.

SharePoint libraries, lists, file shares, Dropbox folders, Google Drive, Amazon EC2 storage, SkyDrive -- need I go on? Your employees can, and likely already do, store documents in one or more locations. So how do you ensure that your data is secure? How do you ensure that compliance is enforced? That your governance processes are being followed? The answer is simple - you monitor all the locations where your content can be and you manage it appropriately.

Have You Tagged Your Data Lately?

One of the most common issues with SharePoint today -- lack of taxonomy. SharePoint environments have been thrown into place and handed out to employees like candy. Without even realizing it, organizations have thrust their own data into chaos by not putting proper taxonomy structures in place for employees to use. The result is a ton of data dropped into SharePoint whenever and wherever it made sense to the employee who put it there.

Here's a common scenario: SharePoint is set up in an IT consulting firm with the intent to help manage projects. That firm is ISO 9001 compliant, which means it has strict rules for the documents it must create for a project and store after the project is over. The required documents and associated metadata is standard across project size and type.

Unfortunately for this organization, no time was put into creating a taxonomy based on those ISO requirements and each project that sets up its own project team site has done things differently: different folder structures, different file names, etc.. In fact, some of the people working on the project aren't even using the project team site for their work in progress documents. Instead they are storing them in their own personal MySite, or in a file share, or in SkyDrive so they can easily work from home. In this common scenario, the company can barely find its data, let alone properly governing it.

Tagging and classification are important capabilities that SharePoint offers, especially in SharePoint 2010 and SharePoint 2013. This is not the first time I've spoken about [proper information architecture](#), and to be honest, it's unlikely to be the last. I also won't go into the [challenges related to migration](#), but that's another situation in which you really need to know your data.

Not Only a SharePoint Problem

Proper tagging and classification of information are only one part of the problems organizations are facing today. It's also important to know what information you are storing in your documents, especially when you are dealing with personally identifiable information such as social security numbers, credit cards or health information. Financial institutions and health organizations are two great examples of organizations that need to maintain a lot of personal information. This means they are heavily regulated by industry standards such as [HIPPA](#), [FISMA](#) and [GBLA](#), among others. Each of these standards have stringent rules for how personal information is stored, who can use it and how it can be used.

For example, consider the credit card company that takes all that personal information when an credit application is completed. Now consider all the data stored as that card is used and reports are made to help decide who should get credit increases, or warning notices.

Although SharePoint gets special attention when it comes to governance of data, (especially considering all the reports about how it's used -- or misused), it isn't the only place we see governance and compliance concerns. With so many organizations (and individual employees) using cloud storage technologies to store documents, the whole of idea of "know your data" and "control your data" becomes even more important.

Don't get me wrong, I'm all for cloud-based technologies, we offer some of them ourselves, but I understand how easy it can be to lose track of your data when you don't have the proper tools in place to manage it. Which means you may be losing track of someone else's personal information and as a result putting your company at risk of lawsuits or worse.

Data Management is Critical

There are any number of products on the market that can scan the content in your SharePoint repository to look for certain keywords or phrases you specify in the titles and metadata of your content. That's standard functionality for SharePoint governance solutions. That's the first step.

But you need to go further if you really want to understand your content. And that means being able to scan the content itself. This means a full-text review - something search engines do all the time, but many governance solutions aren't built for. What you want to be able to do is define policies or rules and then scan the full text of your data, plus the metadata to be certain you are monitoring everything completely.

For example, you might want to scan for credit card numbers stored inside documents, social security numbers, health information or a person's name. Often, the most important information you need to track and control is not going to be found in document titles or metadata, but in the actual document text itself.

In the Cloud - Even More So

What if your data is spread across multiple locations (so not just SharePoint, but say, Dropbox or SkyDrive)? Companies need to go out and scan documents in those locations and have one report that identifies all issues or concerns related to its governance and compliance policies. That kind of capability



is hard to come by for cloud-based services like those mentioned above, and for even bigger services like Office 365.

The problem is, many governance tools require access to the database for the service you want to monitor. Or they need to be installed within the same environment. You aren't going to get this kind of permission in a service located in the cloud -- so you need to be looking for a tool that can do in depth data reviews, as well as one that will work remotely.

Act, Don't Just Monitor Your Data

Now let's take this a step further. You've monitored and you know where your policies aren't being fully met. You now have a manual process to actually do something about it. That's typical and in some cases is likely fine. But if you have a lot of content to monitor and deal with, it would be ideal to put an automated processes in place to do that -- whether that's moving the content, changing permissions, tagging and classifying differently or something else. Now, you are not only monitoring your data, you can have processes in place that can automatically enforce compliance and mitigate risk.

I want to note here that I don't think it's necessarily easy to just drop these automated processes in place and walk away. You'll want to analyze and understand everything coming out of your monitoring reports before you put automated processes in place. Automation can happen for consistent issues you know happen and you know exactly what to do when they do happen. There will no doubt be instances where you can't automate the actions that need to be done. What's important to understand is that you should automate as much as you can, so that you have time to look after the bigger issues.

The Right Tool Goes a Long Way to Help

At MetaVis we spend a lot of time learning about the challenges our customers face when dealing with their data within SharePoint. In many cases, there are very serious concerns that data is not being handled properly (and often not on purpose).

It's our goal to help our customers resolve their governance and compliance issues with technology that not only helps track their data, but also acts to ensure that data is in the right place with the right access.

And that's what our latest tool does. [MetaVis Informant for SharePoint](#) allows you to analyze all your content, right down to full text reviews. It will produce reports for you to view and then take action against. Maybe you need to move the content, tag or classify it differently, encrypt or back it up, or modify permissions on it. You can choose to do this manually or you can create what we call rule packs - policy groups that are packaged into a license. When a report is run, the results are run against the rule packs and automated actions can be performed.

What you are able to do is focus your attention on the bigger risks identified, allowing Informant to take care of smaller (yet still important) issues. For many this type of tool is a great time saver, helping to quickly prevent data loss/leakage and ensuring governance policies are clearly followed. I won't deny that there are other products on the market that can do similar things, but you aren't going to find one



that not only monitors your data down to the last detail and also allows you to create automated rules to act on it -- that is what makes our product the most useful one available for SharePoint today.

Final Thoughts

People share a lot information with organizations today and they need to be able to trust that information is being properly secured and used. In most cases, employees do not intentionally put private information in places where it can be misused. It's more a matter of employees using tools that are easy to work with. Cloud storage is one of those ways that helps people work easily from anywhere.

What that means is that organizations need to have the tools in place to ensure that information is being managed appropriately and is done in compliance with what their industry standards demand. In the end, it all comes down to one rule: know your data!

About the Author

Steve Pogrebivsky is an expert in information management and content management systems with over 20 years of experience. As co-founder and CEO of MetaVis Technologies, he is responsible for overall management and product strategy for the company. Steve has worked at a number of technology companies developing software for content management systems and enterprise integration. Steve was a co-founder of Stelex Corporation, a software and services vendor to FDA regulated industries. Steve developed and managed many of the company's product and service offerings from its launch in 1991 to its acquisition in 2002.

Steve holds Computer Engineering and MBA in Information Systems degrees from Drexel University. You can follow MetaVis on Twitter@MetaVisTech or read the MetaVis blog at <http://blog.metavistech.com/>

