Monitoring and protecting sensitive data in Office 365

With Office 365, Microsoft corporate users can access and share data from anywhere, on any device, and be more productive by using all of its collaboration features. On the other hand, it's easier to inadvertently share sensitive information with others both inside and outside of the company.

To manage security risk, Microsoft IT created a solution that uses the Office 365 Management Activity API and the data loss prevention (DLP) features of Office 365. The solution gathers data about sharing from Microsoft Exchange Online, SharePoint Online, OneDrive for Business, and Azure Active Directory. It also includes a custom governance solution to help protect data. Microsoft Power BI dashboards visualize the data to show how Microsoft corporate users share information.

The dashboards help answer four business questions that have direct business impact on risk, and the answers help leadership make decisions that reduce risk. Microsoft IT uses an agile process to answer these questions:

1. Which sites are capable of external sharing?
2. What is the classification of externally shared sites?
3. Which files are shared externally?
4. What operations are performed by external users on those externally shared files?

Microsoft IT tests hypotheses about how various policies and programs might improve users' sharing behavior and then check the dashboards to see if the behavior has changed. Besides dashboards, the solution improves sharing behavior by giving users visual cues about appropriate sharing. The solution automatically sends email to users who violate security policies by sharing too much, asking them to change their behavior. This helps manage and respond to information security risks.

Information security policies

To protect valuable intellectual property, Microsoft has corporate policies for handling and sharing data. Using business rules based on these policies, the solution detects and reports when users share documents and if the sharing is in or out of compliance with the rules. For example, Microsoft data handling policy states that sensitive business information must be encrypted both at rest and in flight. And, when shared externally, users are accountable for who they share it with.

The solution audits the following types of sharing:

Regulated information. Regulated information includes government identification numbers such as social security numbers and passport numbers, financial data such as credit card numbers and financial records, or medical information. Regulated information must always be protected by encryption.

Business information. At Microsoft, sensitive business information is called High Business Impact (HBI) data. Users can store HBI data on SharePoint Online and OneDrive for Business if they comply with Microsoft policies for HBI data storage and transmission; however, to share HBI content externally, users must get a policy exception from the Microsoft IT security and privacy team.

Low Business Impact (LBI) and Medium Business Impact (MBI) data is permitted on SharePoint Online and OneDrive for Business with no special approval. Users must review all classifications to understand how to classify, protect, and handle data that they create, and ensure that it is properly categorized for use at Microsoft. See the Data Classification Wizard to learn more about how Microsoft classifies information.

How users share too much

Inappropriate sharing occurs when users make information accessible to others in a way that violates information security policies. There's rarely malicious intent behind inappropriate data sharing. Rather, the main reasons for it are:

- The feeling of importance associated with having sought-after, inside information.
- Lack of understanding about the sensitive nature of the information or the security level of the site where it's shared.

Users often don't grasp the implications of sharing information with many people. While some users *do* understand appropriate sharing, there are people who share all information indiscriminately.

Some common inappropriate sharing scenarios are:

- When sharing a document internally, a user doesn't set appropriate security settings to limit the ability to open or edit the document to named users or groups.
- A user shares a sensitive business document or regulated information on a SharePoint Online or OneDrive for Business site, and the site has users who shouldn't have access to that document or information. For example, with OneDrive for Business, a user might inadvertently select the "share with everyone" folder for highly sensitive information.
- A user includes a credit card number, driver's license number, password, or other regulated information in email.
- A user sends a sensitive business document in email and does not set appropriate Microsoft Rights Management permissions on the document.

Detecting inappropriate sharing

Organizations subscribing to Office 365 can use DLP to detect regulated and sensitive information that users share. In addition, Office 365 provides audit data for all file-related events, such as open, upload, download, and delete. Organizations can access audit data through the Office 365 Security and Compliance Center and use search and PowerShell cmdlets to get different views. They can also use Office 365 APIs in custom solutions.

Microsoft IT wanted to do advanced analytics and statistical analysis on this raw data and give the results in a Microsoft Power BI dashboard. A custom solution was built to automatically detect, analyze, and report on sharing behavior. The solution uses the following types of information:

- **Sharing activities.** The solution audits how files are shared on SharePoint Online, OneDrive for Business, and Exchange Online. It also audits login activities on Azure Active Directory. To obtain audit data, it uses the Office 365 Management Activity API.
- **Regulated information.** Adhering to international information privacy regulations, Microsoft IT configured rules for DLP to monitor regulated information contained in Exchange Online email and in files on SharePoint Online and OneDrive for Business. The Microsoft IT solution uses DLP PowerShell cmdlets to create reports for further analysis and reporting. To learn more about configuring DLP rules and using the DLP cmdlets to get reports, see Data loss prevention and View DLP policy detection reports.
- **Documents containing usernames and passwords.** In addition to the DLP data about how users share regulated information, Azure Machine Learning looks for shared documents and email that contain usernames and passwords.

Technical solution components

The main components of the technical solution are:

- Office 365 Management Activity API provides endpoints for Azure Active Directory, Exchange Online, and SharePoint Online (including OneDrive for Business) from which to download audit data. The endpoints are Audit.AzureActiveDirectory, Audit.Exchange, and Audit.SharePoint. Office 365 lets organizations acquire complete audit data on their users' file actions, such as upload, download, open, close, and delete.
- DLP in Office 365 identifies regulated information shared on SharePoint and OneDrive for Business and in Exchange Online email. It informs users when their content is sensitive and, if necessary, restricts sharing.

- Get-DlpDetailReport is a PowerShell cmdlet that returns detailed information for the previous seven days about specific DLP rule matches for SharePoint Online, OneDrive for Business, and Exchange Online. The organization subscribing to Office 365 defines the DLP rules for the types of information to detect in their users' files and email messages.
- Azure Data Factory extracts, transforms, and loads DLP data.
- The Office 365 Management Activity API webhook notifies the solution's webhook endpoint when new audit data is available.
- The webhook endpoint hosts a custom API that was developed to receive notifications and acquire audit data from Office 365.
- Microsoft SQL Server 2014 running in an Azure virtual machine hosts a staging database. For security reasons and to allow data archiving, a second SQL Server virtual machine hosts the aggregated data used by the solution.
- Azure Blob Storage provides data storage.
- Azure HDInsight provides search and transformation for the raw DLP data.
- AutoSites manages SharePoint Online site classifications (LBI, MBI, or HBI) and sends users email about inappropriate sharing sensitive information. AutoSites is a governance solution that Microsoft IT developed. Design information and sample code for this solution is available on GitHub.
- Azure Machine Learning detects when files and email messages contain usernames and passwords.
- Microsoft R Server supports forensic data analysis.
- Microsoft Power BI provides reports, data visualizations, and dashboards.

How the solution works

The following diagram shows the relationship between the different components of the solution. Arrows represent data flowing through the system.

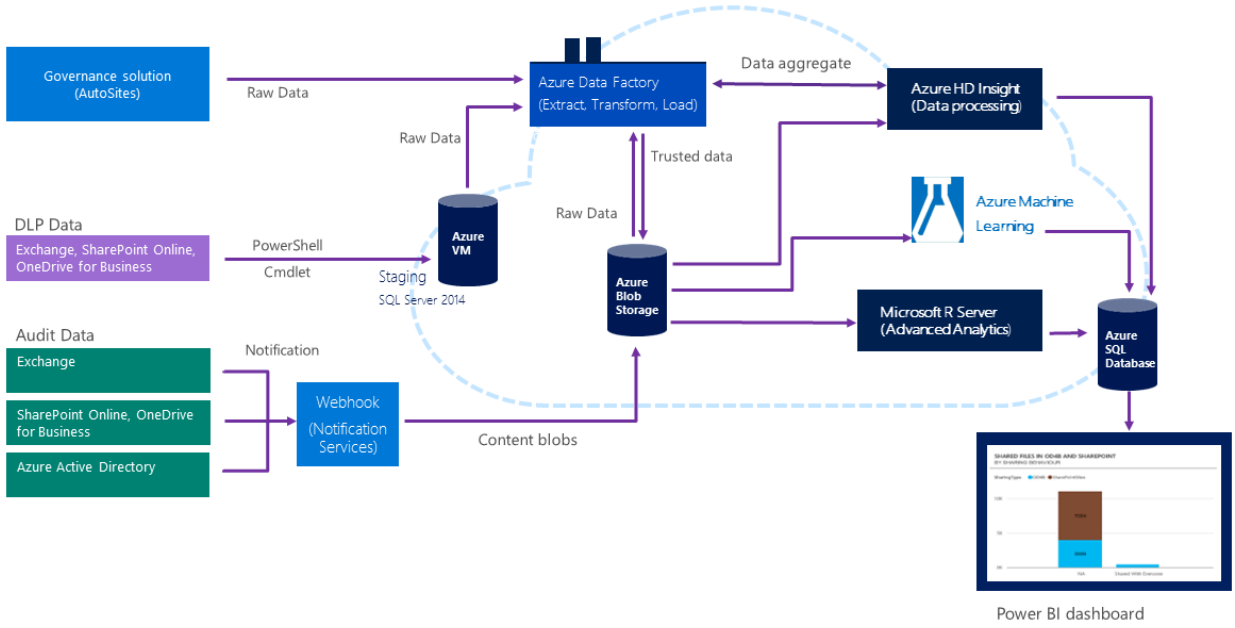

Figure 1. Microsoft IT auditing and DLP solution

To get audit data, the solution subscribes to the Office 365 Management Activity API webhook notification service. When new audit data is available, the webhook sends a notification to a webhook endpoint that hosts a custom API created by Microsoft IT. The API downloads the new audit data for Exchange Online, SharePoint Online, OneDrive for Business, and Azure Active Directory. The raw data goes to the webhook endpoint and then into Azure Blob Storage.

To acquire DLP data, the solution uses the Get-DlpDetailReport PowerShell cmdlet to move raw data to a staging database. To prepare it for further processing, the data goes to Data Factory, where it's extracted, transformed, and then loaded into HDInsight. HDInsight performs computations that aggregate the data into useful chunks, such as average number of DLP incidents. The solution then moves the data back into Data Factory, which then loads it into Blob Storage. Power BI uses the data in Blob Storage to generate reports, data graphics, and dashboards.

AutoSites reports on the number of sites that are misclassified, for example, when a site is classified as LBI or MBI, but has HBI information posted on it. AutoSites also reports on sites that have no classification at all.

The solution detects SharePoint site classification and correlates that information with DLP data and Machine Learning results to yield compliance information.

DLP in Office 365 notifies users when information they're working with is regulated. If a user attempts to share regulated information, sharing is blocked unless the user has a policy override.

Microsoft R Server allows Microsoft IT to perform advanced statistical analysis on the data to identify opportunities for further improvements in compliance.

Reporting

Power BI dashboards answer four business questions about how information is shared at Microsoft, as described earlier. They give the security and privacy team and business leaders a view of how information is shared and how many users are out of compliance with corporate information security policies. The dashboards let the security and privacy team respond to risks in a timely manner and check the effectiveness of risk reduction programs.

They are most interested in how users share HBI information. The solution detects HBI information and aggregates this data into the dashboards, as follows:

- AutoSites counts every document as an HBI document when it's stored in a site classified as HBI.
- DLP reports on the instances that conflict with its configured DLP policies.
- The Machine Learning module counts documents containing usernames and passwords when they're stored on sites that aren't classified as HBI.

Microsoft IT works closely with attorneys and privacy experts to make sure that the solution is ethical and that a balance is maintained between individual privacy and the organzation's needs for information security. Only authorized users can view the dashboards. Management and security team members get different views according to the type of information they need. Authorized dashboard users are:

- **Chief Information Security Officer and leadership team.** They use the dashboards to make strategic decisions about appropriate security risks.
- **Online service manager.** The service manager decides how to deploy service features in a way that reduces risk. Service managers are responsible for the security of information that's shared on their services. They use the dashboard to measure outcomes of different approaches to improving security.
- **Security Operations team members.** Security Operations team members monitor dashboards and drill down into more detail when exceptional events occur, such as cases of extreme sharing or access by users from blacklisted IP addresses. They then provide details to the appropriate manager for further action.

Evaluating dashboard data

Leadership looks at aggregated numbers and trends in the dashboards to see how well policies are working and the impact of policy changes. To learn when and where sensitive information is shared inappropriately, dashboard data is evaluated, such as:

- Number of SharePoint Online sites
- Number of SharePoint Online sites that are set up for external sharing
- The number of sites that are actually shared
- The number of external users of these sites
- Operations performed on shared files

They get some of this data from this summary dashboard:

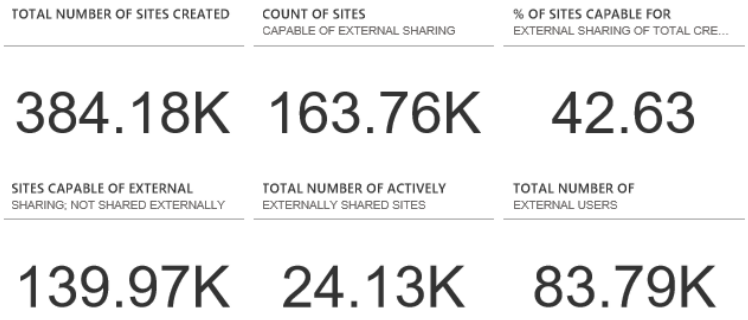

Figure 2. Summary dashboard data

This data shows that most sharing is appropriate. Less than 10 percent of SharePoint sites have externally shared content, even though many more are set up for it.

Another dashboard shows file operations.

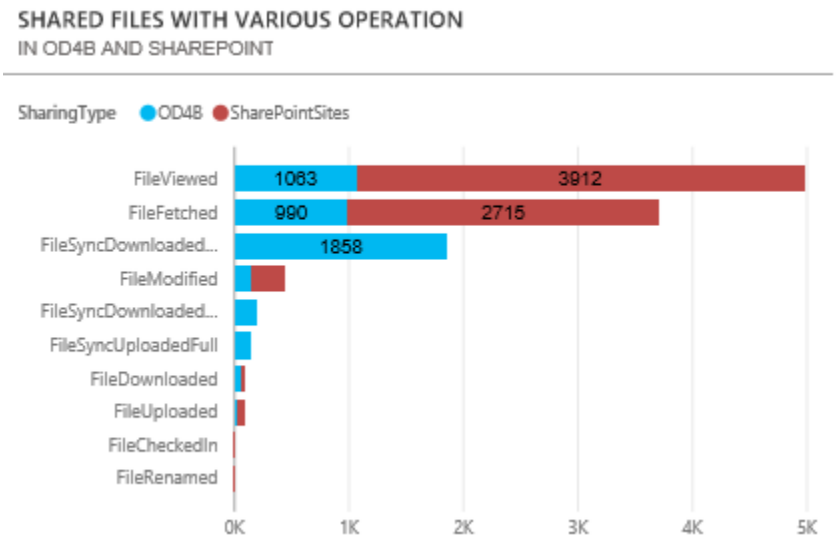

Figure 3. Operations on files shared on OneDrive for Business and SharePoint

The security team is most interested in HBI sharing and if the sharing is appropriate. Authorized users can drill down into the dashboards to get more detailed information, such as the groups sharing the most HBI information.

The following dashboard shows that few external users have access to HBI as compared to LBI and MBI.

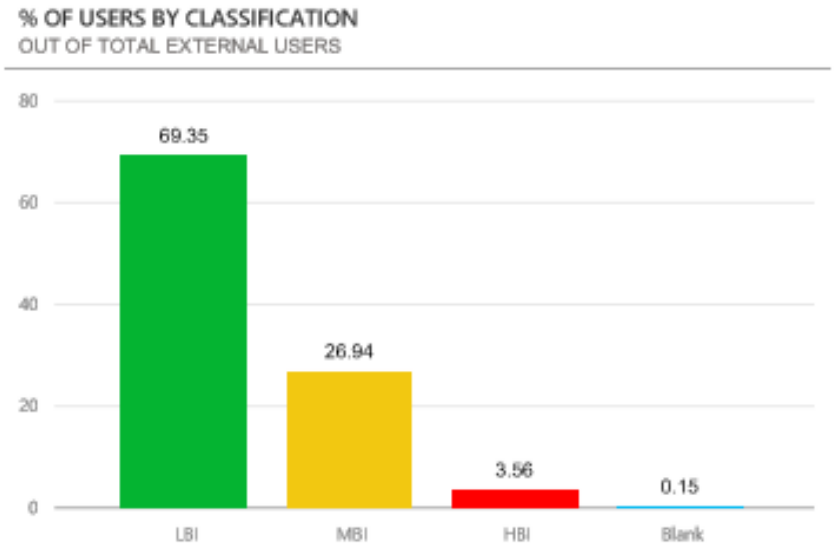

Figure 4. Percent of external users with access to HBI

While there are about 80,000 external users, most of the information shared with them is LBI. This means that employees are collaborating outside the company, which is desirable, but mostly with information that isn't highly sensitive.

The security team is more interested in sharing on SharePoint Online versus OneDrive for Business. Because the scope of sharing is broader on SharePoint sites, which often host group projects with multiple users, it's easier to inadvertently share too much. The security team prefers sharing on OneDrive for Business because users explicitly share a single document. The following figure shows that most sharing is, in fact, on OneDrive for Business.

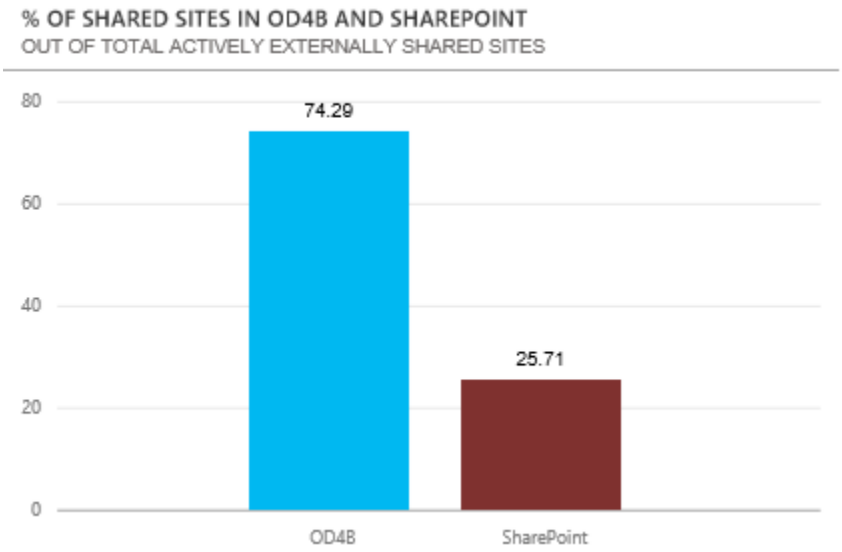

Figure 5. Percentages of shared sites on SharePoint Online and OneDrive for Business, out of active, externally shared sites

The team also wants to know who does the most sharing. The next dashboard shows the distribution of sharing.

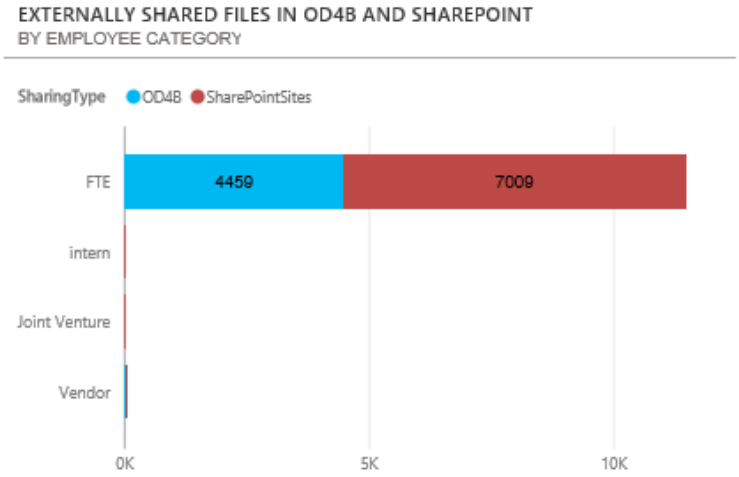

Figure 6. Sharing by category of user

A DLP dashboard gives summary data and details about instances where regulated data is shared.

Figure 7. Summary data on the DLP dashboard

This dashboard reports the number of documents found daily that contain regulated data. Other DLP dashboards give the number of OneDrive for Business and SharePoint instances by user category—employee, intern, or vendor—and also file type.

The dashboards reveal that most users at Microsoft share HBI appropriately, in keeping with company policies. Even so, the less HBI shared, the lower the risk of sharing too much. The following dashboard shows sharing trends since 2014, when the solution was implemented.

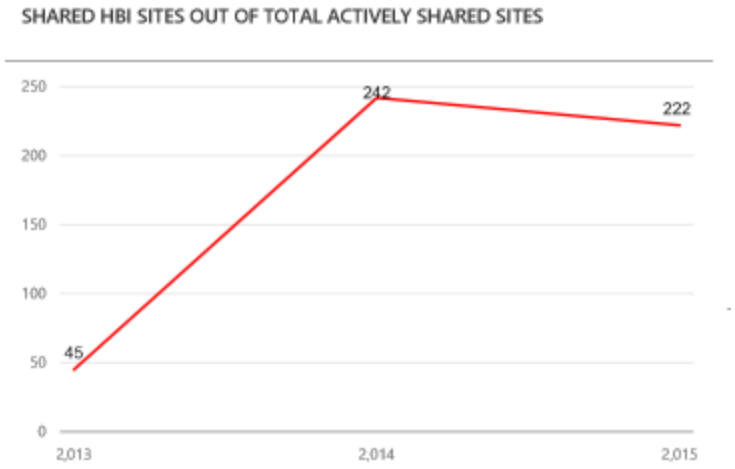

Figure 8. Percent of all shared sites that are classified HBI

Healthy collaboration—with controls

At Microsoft, we expect employees to use good judgment and common sense—and we want them to collaborate. Instead of shutting off their ability to share information, we believe it's more effective to teach them to avoid sharing too much. As an extra security step, if necessary, DLP may also prevent sharing of regulated and/or sensitive business information.

The Microsoft IT solution influences and modifies users' sharing behavior in these ways:

Site classification and labeling

AutoSites requires site owners to classify SharePoint sites according to the type of information that may be posted on it: LBI, MBI, or HBI. When creating a new site, the site owner picks the type. This applies the appropriate security settings to the site and labels it according to its classification. The levels of information are clearly defined in the user interface, as shown here.

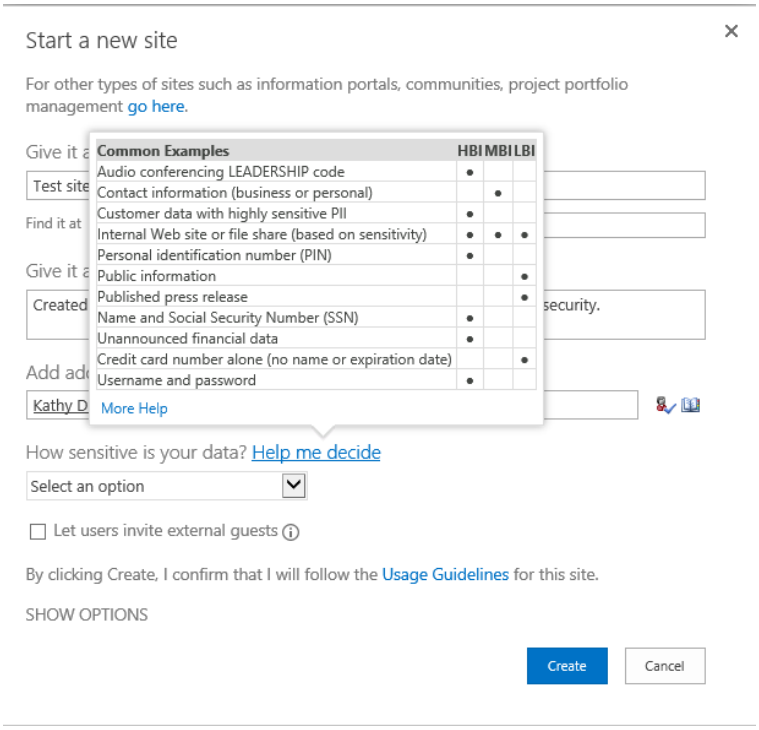

Figure 9. Information classification in SharePoint Online

When a site is created, it's labeled based on what the information type that the site owner specified: LBI, MBI, or HBI. This tells SharePoint Online users what type of information they should post. Users are expected to honor the classification and post only the type specified. If HBI information that is posted on a site labeled LBI or MBI or on a site that hasn't been labeled, AutoSites detects the classification and includes this information in a dashboard report.

Figure 10. SharePoint site labels

Signaling

A user who shares files inappropriately automatically receives a signal that helps teach them the desired behavior. A signal can be a Policy Tip or an email message. And, if necessary, the sensitive content is blocked.

Policy Tips

DLP includes policies for sharing regulated information that administrators can use out of the box and customize for their specific company needs and region. Information covered under these policies includes credit card and social security numbers and their international equivalents. DLP displays Policy Tips in the user interface that inform users about potential policy violations. At Microsoft, Policy Tips display when the content of an Exchange Online email or a file that's been uploaded to a SharePoint site or OneDrive for Business doesn't comply with Microsoft sharing policies.

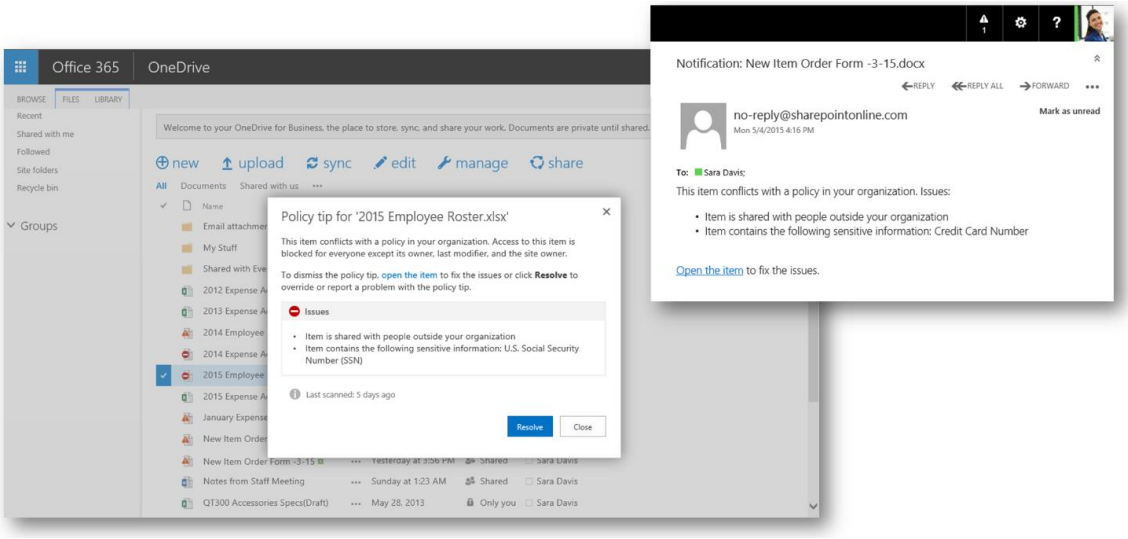

Figure 11. Policy Tips in SharePoint and OneDrive

In addition, when a user posts a file on a SharePoint site or OneDrive for Business that contains regulated information, DLP displays an icon in line with the file that indicates the file contains regulated information. DLP also blocks other users from viewing or accessing the file unless the administrator has configured a policy override for the site. Microsoft IT has a business process for users to request this override.

Email messages

Both DLP and AutoSites send email messages to users who share too much, as follows:

- **DLP for Office 365.** If a user shares regulated information on SharePoint or OneDrive for Business or in an Exchange Online email message, the DLP system locks down the document or rejects the message. It then sends an email message letting the user know about it. The email message contains the same information as the Policy Tip. If there's a valid business reason to share the information, the user can request a policy override.
- **AutoSites.** When a user shares other types of sensitive information, such as usernames and passwords, AutoSites sends an email message asking the user to correct the issue. AutoSites also sends an email message to a user who shares HBI information on an LBI or MBI SharePoint site, or one that has no label. The SharePoint site owner receives the same message.

  Rather than pointing out that users are doing something wrong, the AutoSites messages are positive. If a user doesn't change the sharing behavior on SharePoint or OneDrive for Business, AutoSites automatically delivers another message. If the user and the site owner still haven't corrected the issue after receiving three email messages, the site is locked down. If the site remains out of compliance, it is removed.

Training programs

Training programs educate users about information security policies and how to handle sensitive business information. All employees receive formal security training and have access to reference information about information security policies. Microsoft IT also works with particular groups that share a lot of HBI information to make sure they're trained on how to handle it properly. One important thing the security team has learned is that all users are not alike when it comes to sharing information. There are the two extremes—inappropriate sharing or not sharing at all. Then, there are the people who fall in the middle. This means that not all approaches and training programs will fit all users. Therefore, the information that the auditing and DLP solution provides about groups of users who share inappropriately makes it possible to tailor future training programs to just these people.

How the solution reduces risk

This section describes how the auditing and DLP solution is reducing information security risks at Microsoft.

Product design is improved

Information security managers, service managers, and Microsoft IT take dashboard data to product teams. The data influences changes to services and features that improve information security. For example, a SharePoint Online service process was changed to require users to obtain permission before sharing HBI information externally. This process tells users about their responsibility to handle HBI information appropriately. It acts as an extra reminder that they're accountable for their actions.

Fewer HBI documents are being shared

Dashboard data is shared with service owners and works with them to reduce inappropriate sharing. So far, the number of HBI documents being shared has been reduced by about a third. In most cases, the issue was corrected when users got the required policy exception to post HBI information.

Executives can make informed decisions about security policy

Decisions about securing information are no longer made based on guesses and gut feelings, but are informed by concrete data in reports.

Data security strategies are tuned based on actual data

When a strategy is implemented, its relative success or failure can be seen in the audit data. The strategy can be honed and audit results checked to see if the change was effective. This creates an ongoing process of improvement because the impacts of decisions are measurable.

Best practices

Account for diverse attitudes on information sharing

Some users share too much—others don't. Some are in the middle. Microsoft IT needed to account for the whole gamut of attitudes and behavior. They plan to target training to the users who need it the most.

Verify regional privacy regulations

DLP and auditing activities can have privacy implications around the world. Notice and consent are fundamental privacy principles that apply here, but before implementing these security controls be sure to check with your legal advisor and works councils in the European Union.

For more information

Microsoft IT

microsoft.com/itshowcase

Additional resources

Auditing in Office 365

Data loss prevention

Microsoft Rights Management sharing application for Windows

Office 365 Dev Center

Office 365 Patterns and Practices

Office 365 Trust Center

View DLP policy detection reports