White Paper

Meeting Data Governance Requirements Starts with Data Visibility

Tackling the Enterprise Data Management Gap

By Terri McClure, ESG Senior Analyst; and Doug Cahill, ESG Senior Analyst

October 2017

This ESG White Paper was commissioned by Egnyte and is distributed under license from ESG.

Contents

Executive Summary

Data Inundation

IT organizations have been struggling with how to manage data growth since the advent of computing—from the physical world of filing cabinets and printed green-bar paper to the logical world of zeros and ones—and there is no sign of this growth slowing. Rather, growth is accelerating as the world becomes more connected. Because of this, managing data growth appears perennially among the five most commonly cited challenges across ESG research. Managing data growth is by far the most cited challenge that organizations reported dealing with as a result of archiving their data[1] and a top three most reported challenge when it comes to managing organizations' overall storage environments.[2] How fast is it growing? ESG research found that the multitude of respondents who participated in a recent survey (25%) cited overall annual storage capacity growth rates in the 11% to 20% range, with another *one-quarter reporting that their data is growing in excess of 50% per year*.[3] Data-driven business models in today's digital economy make it essential to secure data as organizational assets, an imperative challenged by these data creation rates.

It's not just about managing data as it is created; the preexisting corpus of corporate data must also be considered: Compounding the data management challenges associated with tremendous data creation velocity is the massive amount of preexisting unstructured data all companies must manage and secure. This corpus of data now includes volumes of unstructured data that previously resided on individual PCs and laptops, and shared departmental drives that users now wish to access via mobile devices. Much of this data holds tremendous business value, with potential to leverage it for data analytics, utilize visuals to improve ROI, identify new business opportunities, and help keep the business ahead of the competitive curve in the existing business lines.

Shadow IT: Driving Data Sprawl and Growth

Mobility and the cloud are also accelerating content growth and creating new opportunities and challenges when it comes to securing, managing, and leveraging that content. Mobile devices and applications are newer productivity and data creation and capture points. The ubiquity of mobile devices and easy access to applications via the cloud has resulted in the use of cloud applications without any involvement of corporate IT or a company's security team, a dynamic referred to as shadow IT. Prior to cloud and mobile device usage growth, lines of business wouldn't have dreamed of deploying applications with zero IT involvement.

Now it is commonplace. In fact, recent ESG research reveals that nearly two-thirds (65%) of IT and information security professionals believe there is a moderate to large number of shadow IT applications in existence in their organization (see Figure 1).[4]

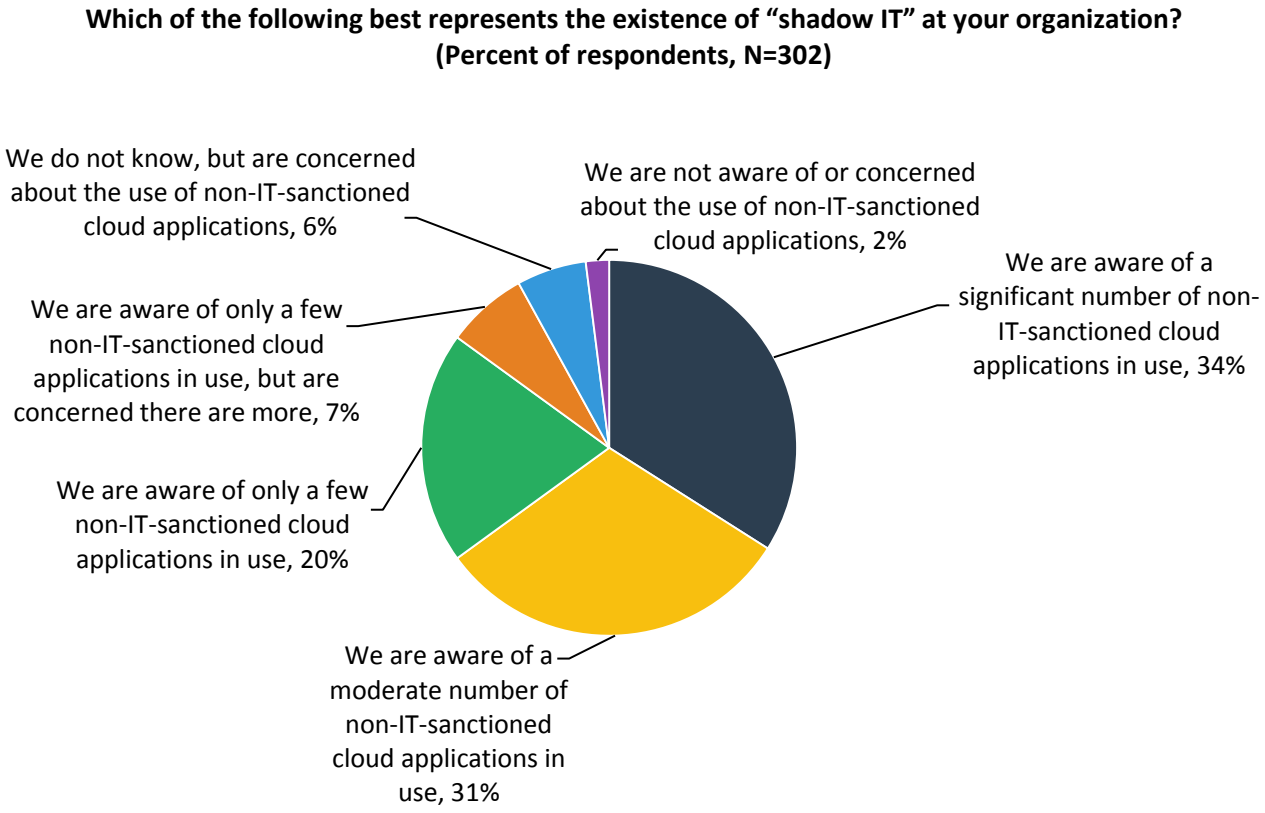
65%

Shadow IT is now commonplace.

Recent ESG research reveals that nearly two-thirds (65%) of IT and information security professionals believe there is a moderate to large number of shadow IT applications in existence in their organization.

[1] Source: ESG Brief, *Long-term Data Retention Drivers and Trends*, April 2017.
[2] Source: ESG Brief, *2017 Storage Trends: Challenges and Spending*, August 2017.
[3] Source: ESG Research Report, *2015 Data Storage Market Trends*, October 2015.
[4] Source: ESG Brief, *Shining a Light on Shadow IT*, September 2016.

Figure 1. The Prevalence of Shadow IT



Source: Enterprise Strategy Group, 2017

The challenge with shadow IT? Lack of visibility into the data created by and shared via the use of unauthorized cloud applications introduces greater risk to the business, as IT can't secure what it can't see. The prevalent use of unsanctioned cloud applications creates multiple application data silos. The use of numerous cloud applications exacerbates an existing problem IT is already struggling with—managing a plethora of siloed application data repositories—making it much more difficult to manage data growth, and ensure that, regardless of the application used, the data is protected, available, secure, and retained in relation to regulatory requirements. And all of these tasks need to be completed in a cost-efficient manner. Perhaps the greatest of these data management challenges are securing data and ensuring proper data governance, challenges greatly compounded in recent years with the rise in the number of applications managed by IT, the consumption of cloud services, and shadow IT, all of which widen the visibility and control gap. Increasing cybersecurity was selected by 39% of ESG research respondents as one of the business initiatives they believe will drive the most technology spending in their organization for the year, making it the most-cited response, and strengthening cybersecurity tools and processes was selected by 32% of respondents as the *most important* IT initiative for their organizations for 2017.[5]

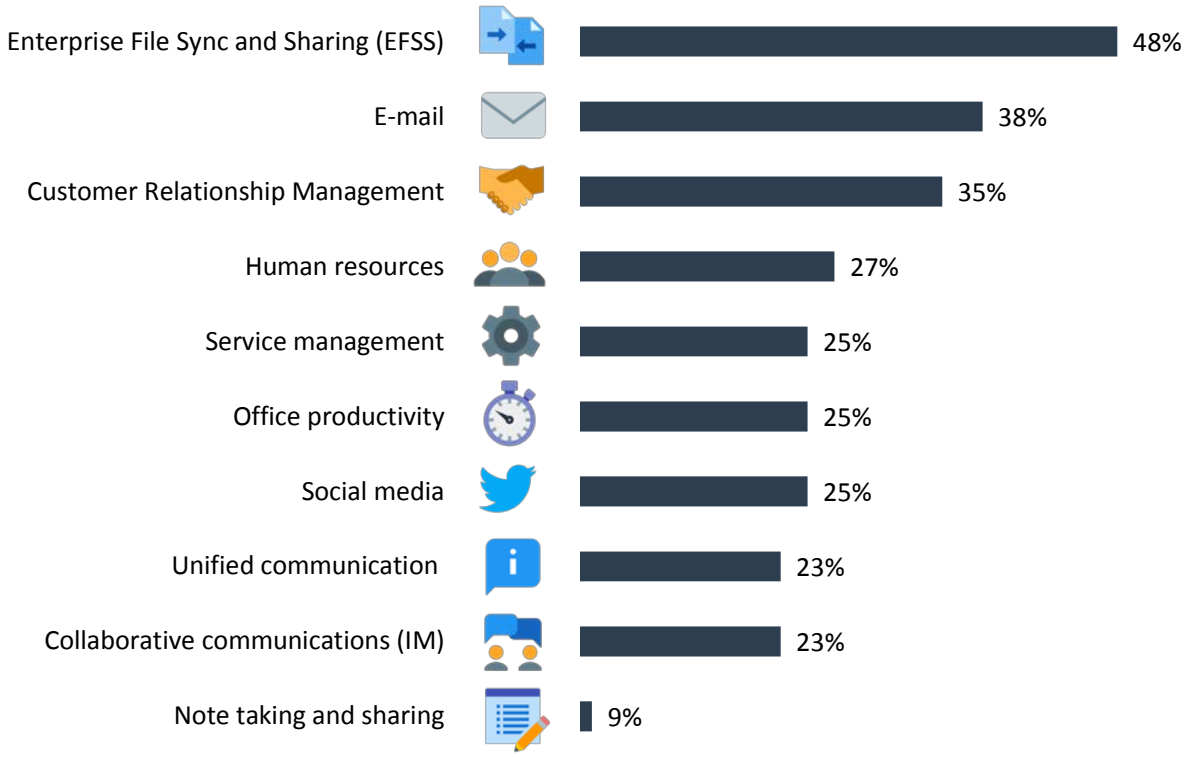
But regarding data security, which applications are causing the most concern for organizations? According to ESG research, nearly half (48%) of enterprise organizations believe that enterprise file sync and share (EFSS) is among the applications most in need of security controls and monitoring oversight (see Figure 2).[6] But IT is also concerned about email, customer relationship management (CRM), human resources, and office productivity applications such as Microsoft Office 365

[5] Source: ESG Research Report, *2017 IT Spending Intentions Survey*, March 2017.
[6] Source: ESG Research Report, *The Visibility and Control Requirements of Cloud App Security*, May 2016.

(including SharePoint). The key point about these applications? Many reside in the cloud, creating yet more disconnected silos of application data.

Figure 2. Applications Requiring the Most Security Attention

In your opinion, which of the following types of applications require the most security controls and monitoring oversight? (Percent of respondents, N=302, three responses accepted)



Source: Enterprise Strategy Group, 2017

The productivity impact needs to be considered. With data sprawled across many applications and content repositories, such as SharePoint, email, CRM, endpoint devices, multiple shared file servers, and loads of siloed cloud apps with their own cloud repositories, IT and knowledge workers need to deal with data sprawl and duplicate data that may reside across these silos of disconnected data. There is no easy way to arbitrate which systems contain the latest versions of files and individual productivity takes a hit. End-users spend excessive time searching for information and comparing document versions and revisions rather than being productive. And of course with questionable revisions and fragmented data, the company itself has no way to leverage information for competitive advantage—you can't leverage and analyze data if you don't know what you have or what data is current.

The Insider Threat

The threat of insider action must be considered and guarded against. Identifying bad actors can be challenging. The insider threat is privilege misuse perpetrated by different entities, inside and outside of the organization. Insider action could be intentional, whether malicious or for financial gain, or unintentional and the result of an unwitting insider being manipulated by a malicious external party. True insiders can be the most difficult to detect, for multiple reasons: they are

familiar with the environment and take a purposeful approach; they use legitimate credentials, often with escalated privileges; they are invisible to perimeter controls since they are already inside; and they have access to cloud-based file storage services as an exfiltration target. Because of the difficulty in detecting malevolent insider actions, breaches most often go undetected for months or even years.

Data Assets at Risk

But the biggest issue is corporate risk. With data sprawled across the enterprise and the cloud (including both sanctioned IT applications and shadow IT), it becomes harder and harder to know where business-critical or regulated information is stored, who can access it, and how to protect it. It is highly likely that this information exists in multiple places, increasing the risk of unprotected data or a security breach. How can data encryption policies, critical when it comes to cloud adoption, be applied if the organization has no visibility into where data that should be encrypted resides? And when it comes to compliance, data sprawl has a number of ripple effects. Limited visibility into data assets makes it difficult to set proper data retention policies; cleanse data after project completion; and ensure e-discovery, legal hold, or other regulatory obligations are met.

This backlog of existing data combined with the accelerated rate of creation makes understanding the data with respect to its contextual business value incredibly difficult. Companies with remote employees and offices add levels of complexity with distributed applications—more silos of disconnected data.

And amidst these challenges, of course, is the fact that not all data is created equal. Data security and compliance approaches to date have struggled to embrace data of differentiated value due to the challenge of classifying massive corpuses of unstructured data. Until data can be accurately classified, it is nearly impossible to properly undertake a data management strategy that encompasses cloud. This inability to classify data introduces risks regarding data quality and compliance. These challenges need to be met. Fortunately, tools emerging on the market can help. Employing such tools to secure sensitive data assets starts with visibility.

Visibility: The First Step to Ensuring Proper Data Governance

Data governance typically refers to the overall management of the availability, usability, integrity, and security of the data employed in an enterprise. It is often directly related to an organization's need to meet and maintain compliance with an industry regulation. This includes ensuring that the integrity of the data can be trusted and that people can be made accountable for any adverse event that happens because of low data quality. It's all about implementing tools and processes that ensure data management meets these ends, including activities like verifying that access controls are properly set up and implemented, and ensuring that there are audit trails to record who accessed what data assets and when.

The combination of the rate at which today's businesses create and thus have to manage and secure data, the fact that all data needs to be better understood with respect to its business value, and the fact that highly motivated adversaries are intent on stealing those assets represents a need, if not an imperative, for a more proactive approach to data security and governance.

Prior approaches to taming ever growing data sets have included hierarchical storage management (HSM) and tiered storage with limited success due to the fundamental (heretofore unmet) requirement to be able to manage data based on its business value and time sensitivity, and thus the need to align location and policy accordingly. Such an infrastructure-centric view, however, does not consider the security-related requirements for these data assets. IT needs to manage at the application layer. To address this issue, IT needs a solution that can both provide visibility into the backlog of legacy data and tag the new data as it's created by the application, which allows for the identification and organization of

sensitive data. That need can be met by employing technologies that provide visibility into the data assets stored across siloed applications.

Cloud Storage and File Sharing Services Complicate Data Security

The shadow IT phenomenon—in other words, the ease with which cloud-delivered SaaS applications can be utilized by business units without the involvement of IT—is a relatively new reality for IT. The end result? More corporate data being stored in cloud storage services, and that data residing outside of normal corporate data governance oversight. The absence of corporate governance due to the use of shadow IT applications creates both a security blind spot and compliance exposure.

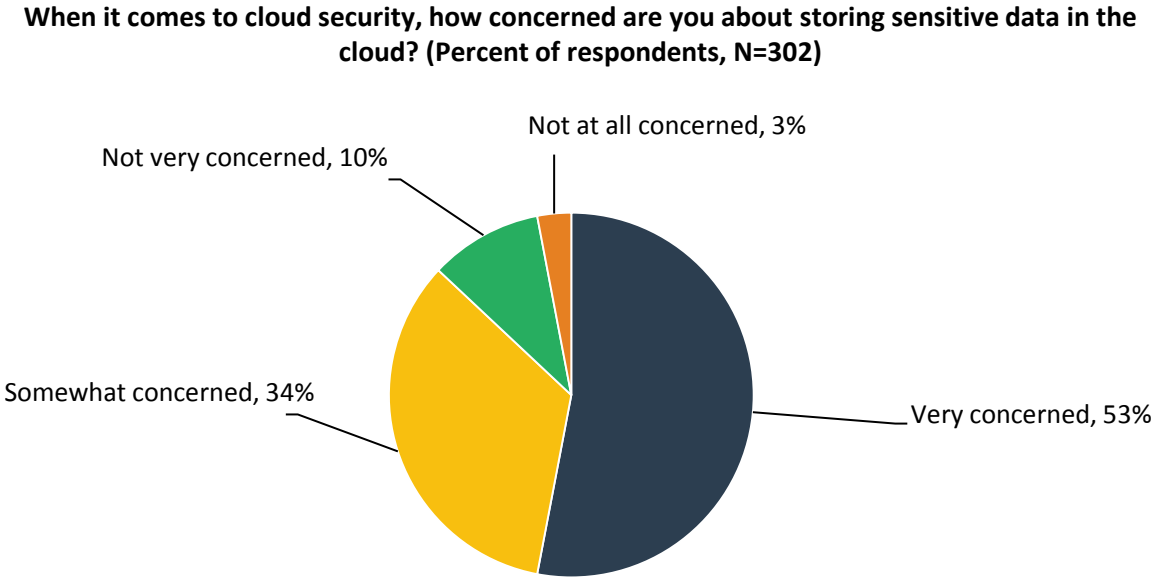
But more organizations are looking to leverage the cloud for its cost savings and efficiency. The challenge is that yet more silos of data are getting created in the cloud, in both shadow deployments and IT-led deployments. But since not all data is created equal, and some companies are hesitant to store sensitive data in the cloud, it is imperative for IT to have the visibility and control to determine what data is suitable to be stored in the cloud and restrict data that shouldn't.

A Hybrid Approach Is Strongly Desired

Therefore, IT requires a solution that unites governance across multiple repositories.

A hybrid approach where the most important data assets stay on-premises and under the visibility and control of corporate IT is a pragmatic one desired by many companies, but is also one that *requires* organizations to have the ability to identify and classify data across such a hybrid deployment to be able to determine what data is suitable to be stored in the cloud, determine encryption policies, and prioritize migration to the cloud. Storing sensitive data in the cloud is a concern for 87% of those IT managers surveyed about cloud usage, with 53% very concerned (see Figure 3).[7] IT requires a solution that will unite governance across multiple repositories to mitigate this concern. As is the case with all facets of data security, customers first need to understand their data to be able to define policies including location, user access, and the use of encryption.

Figure 3. Storing Sensitive Data Is a Cloud Security Concern for the Majority of Enterprise Organizations



Source: Enterprise Strategy Group, 2017

[7] ibid.

Classification: The Second Step to Proper Data Governance

Security and governance are best applied where the data resides with a rich set of controls that work in concert and context to ensure data is classified appropriately. Yet with disparate applications and multiple content repositories to manage both on-premises and in the cloud, it has become an ever more complex task. When securing data across the application environment, IT must consider the following tenets:

- Not all data is equal with respect to its intrinsic value.
- Not all data is equal with regard to performance and retention requirements.
- IT cannot secure, manage, and retain what it cannot see.
- Visibility is paramount and a prerequisite for control.
- Controls need to be policy-based and such policies should be data classification specific.

The Immutable Aspects of Securing Data Assets

Technologies that allow companies to apply data value-specific security and governance policies need to provide the following levels of classification levers:

- Classifying data based on its relative value requires visibility beyond file system attributes of name, time, size, and access date. The ability to see inside unstructured data and inspect content is needed to determine whether files contain sensitive data and thus require specific security treatment or are subject to retention requirements. Examples of such content include data in a credit card format, personally identifiable information (PII) such as personnel records, confidential product pricing information, competitive intelligence, financial spreadsheets, and anything with information that represents intellectual property. Both security *and* retention need to be considered.
- To make such classification for security work at scale, custom metadata tags are essential. Tags enable not only organizing and searching data corpuses, but also applying data security controls specific to a type of data such as encryption and data loss prevention (DLP) policies.
- Information about who accesses sensitive data assets requires multidimensional analysis. In addition to seeing what is inside our data, we need a 360 degree view of who is accessing our most important data assets, their job functions, whether they are authorized for access to a project, what content types they're looking at, when they access data, and even from where they are doing so. A normalized demographic view of data access can then be employed to determine whether such access is authorized or potentially that of an insider threat or a bad actor who gained access to data assets via stolen credentials or malware. This normalized demographic data could be leveraged for user behavior analytics (UBA) to identify anomalous access and usage that could be indicative of a data breach.
- Knowledge about where data resides is also a relevant attribute of classifying data. Source is often times an indicator of data type. In order to factor location into classification, governance solutions need to support breadth of coverage by integrating seamlessly in hybrid clouds that span on-premises and cloud resources.

Control and Implementation: The Third Step to Proper Data Governance

Once these levels of visibility and classification are achieved, IT has a basis upon which they can set about performing unified governance across the application environment. IT needs a tool that can act upon the knowledge gained to provide proper:

- ✓ Levels of Manageability: Given the ongoing trend of the democratization of IT, more and more often, line-of-business employees are delegated to manage systems that were formerly under the umbrella of IT. This makes ease of use one of the most important areas of focus. If business managers need to set retention periods, blacklist/whitelist domains, and manage access controls, the solution needs to be intuitive.

- ✓ Access Controls: IT must ensure that only authorized users have access to information. This is where the multidimensional aspects come into play. Authorization considerations should be based on the user, devices used, geographic location, project, group, or department. It is rarely straightforward, given global security threats, highly mobile workers, and project-based business, so a solution that provides many dimensions for access is key. To minimize the risk associated with stolen credentials and malicious insiders, the least privilege security best practice whereby only those users who *must* have access to data assets are granted access is recommended. Organizations should also employ a "trust, but verify" approach to data access via audit trails. Lastly, access credentials can be temporal where some users will no longer need access to certain data assets; organizations should keep access control lists (ACLs) current.

- ✓ Archive: Archive is an underlying foundation for a strong governance solution. The solution must ensure data is stored in compliance with the above criteria, yet in a cost-effective manner and on the appropriate storage platform or repository at the appropriate time. Data is typically initially stored on a high-performance system, but as it ages, it may need to be retained for a long period of time, while staying accessible for e-discovery or business needs. Moving data off of tier-1 storage and into a long-term archive can significantly help drive down costs as data growth accelerates.

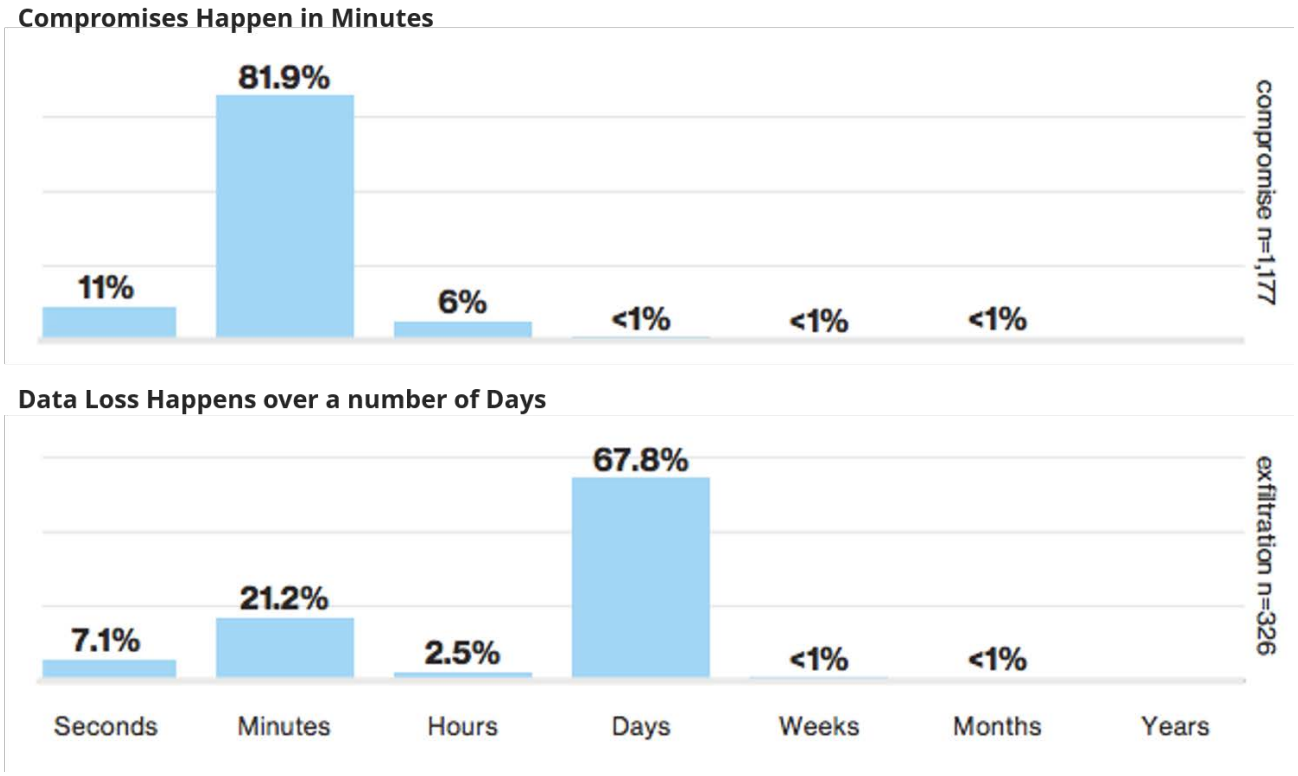
Of course, there are other concerns to be taken into consideration: retention periods and data residency, to ensure data is retained and accessible for the proper timeframe and in the geographic region(s) in line with regulatory requirements; encryption, to ensure data is not stored in clear text anywhere along the data path; and efficiency, to reduce data sprawl while aligning accessibility needs with media costs. IT is not (and never has been) a static environment. One of the biggest challenges regarding data governance is ensuring data governance policies are retained and enacted in a constantly changing environment. The data governance solution should be agnostic of the multiple content repositories so that it can take care of different aspects of enterprise businesses, such as when new content repositories are added as new apps are deployed, existing repositories are consolidated for any of a variety of reasons, or even if cloud migration changes the mix of content in the cloud and on-premises.

Once you have insight into data, performing pre-production "what if" assessments of the business implications of a policy change becomes more realistic and reliable. Only then can the data management impact to governance policies from IT changes finally be understood, and business risk associated with making changes reduced.

The last piece of the puzzle is time sensitivity. In a constantly changing environment, IT and the business owner need to know if any level of risk is introduced, and they need to know quickly, in real time, in order to mitigate risk. Verizon's 2016 Data Breach Investigations Report (VDBIR) highlights both the speed at which a compromise can be successfully perpetrated, putting data assets at risk of exfiltration, and the lag in which it takes most organizations to detect and then

respond to an incident, a concept referred to as dwell time. The 2016 VDBIR notes that the vast majority of compromises happen in minutes (81.9%) and data loss in days (67.8%).[8] The gap is the first opportunity to prevent data loss. The second gap, dwell time, is typically measured in weeks and months, and represents the period in which the scope of a breach can widen to affect a broader set of data assets. That means having the ability to be notified of a possible governance issue in multiple ways as soon as it is detected so that it can be addressed in a timely manner is critical. And lines of business, like data, are not created equal; each has its own needs, so data governance needs to be real-time but customizable based on the line-of-business need.

Figure 4. The First Opportunity to Prevent Data Loss



Source: Verizon, 2017

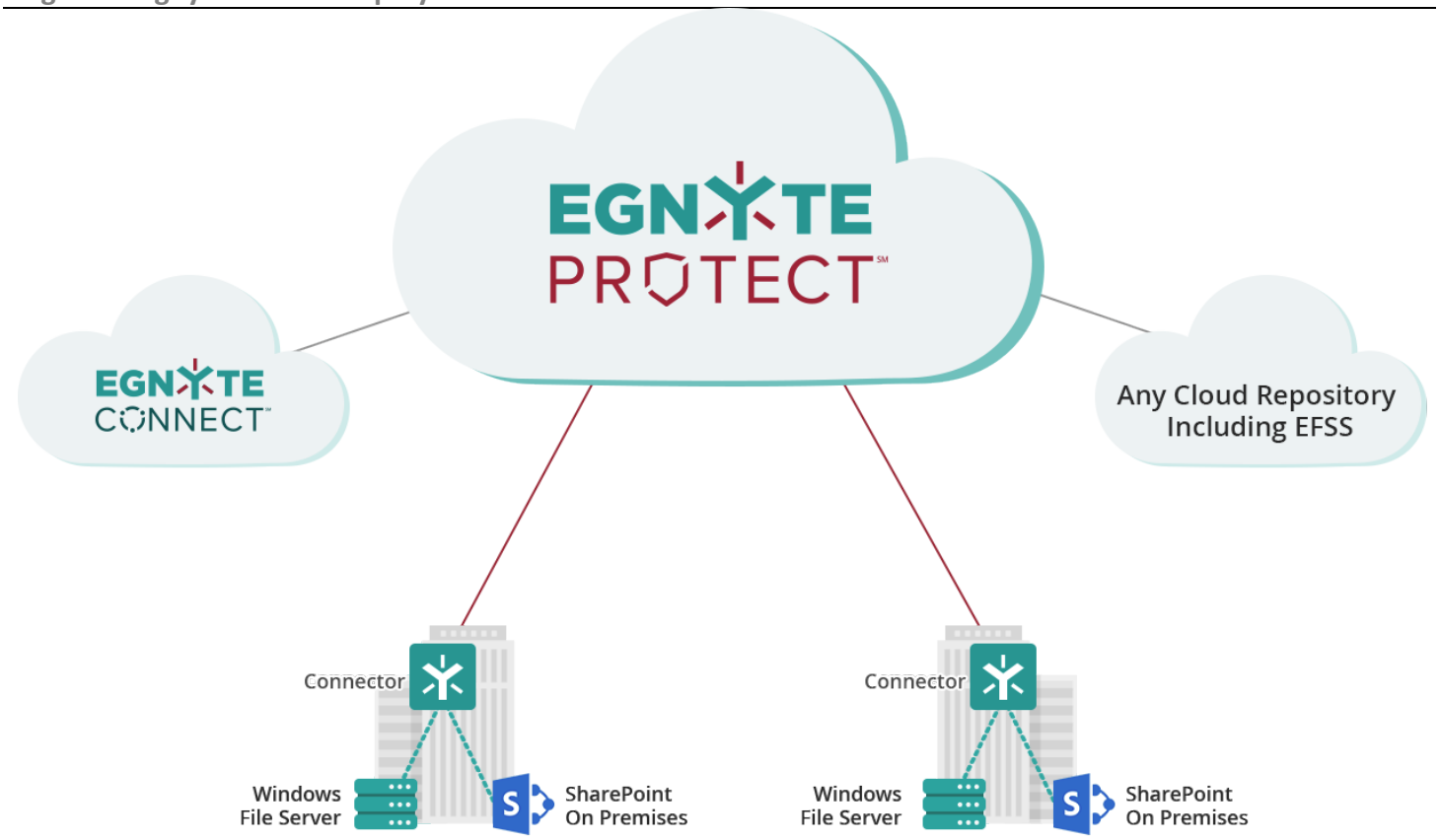
Solving the Data Governance Challenge with Egnyte Protect

Fortunately, some products coming to market help IT organizations address these issues. One such product is Egnyte Protect, a comprehensive data governance solution. Egnyte Protect builds on Egnyte's extensive data management, security, and analytics experience within its own platform and extends that analysis capability across any content repository and application data repositories, such as SharePoint, EFSS, and file servers, whether they are on-premises or in the cloud, providing unified governance across multiple repositories. Such coverage is essential for today's hybrid cloud enterprises.

Egnyte Protect provides content analytics and visibility for discovering and classifying information. It uses metadata analysis, content classification, and custom tagging to help IT organizations meet data governance needs in hybrid cloud environments.

[8] Source: Verizon, *2016 Data Breach Investigations Report*.

Figure 5. Egnyte Protect Deployment Architecture



Source: Egnyte, 2017

Leveraging Egnyte Protect and the content analytics it provides, IT can scan across content repositories to find where your sensitive content is and classify that data, allowing users to centrally enforce access policies and maximize control and security. It allows users to set policies across a heterogeneous file storage environment, such as setting, monitoring, and reporting on file access controls, and only retaining data for the appropriate amount of time relative to regulatory requirements.

Egnyte Protect is an enterprise-grade SaaS hybrid platform for content governance that is built on Egnyte's mature expertise, content intelligence, and a smart content orchestration platform, providing a comprehensive analytics engine for unstructured data, with near real-time file analysis and full visibility at your fingertips. But perhaps one of the most compelling features is that it provides comprehensive support for multiple data sources—whether you have applications on-premises or in the cloud, whether data is in an enterprise content management (ECM) application, like Microsoft SharePoint, and whether you're using Egnyte Connect. The visibility it provides allows IT to choose the right tools, such as multiple ECM providers meeting different departmental needs, while offering visibility for governance across the entire corpus of corporate file data.

The Bigger Truth

It was difficult enough to manage data growth when IT was only run within the four walls of the data center. Cloud, mobility, competitive pressure, and the resulting increasing speed of business mean that IT applications are now running in many places in addition to the data center, and can be spun up quickly in support of line-of-business needs. But data governance mandates persist and even change as the legal landscape changes to deal with advances in technology.

To deal with the constantly changing landscape, IT needs visibility into sensitive data assets, including where that data resides. But visibility alone is not enough. Based on the understanding of the environments IT gains from better data asset visibility, IT needs to be able to classify data to ensure it is managed in accordance with corporate policies and data governance requirements throughout its lifecycle. And it needs the control levers to ensure data is appropriately protected across the lifecycle. That means multidimensional analysis and control over who has permission to access that data based on a comprehensive set of criteria. It means understanding who is actually accessing and modifying data, how long that data needs to be retained, and how fast it must be retrieved. And it includes not just retention, but also destruction. And it needs to have tools that notify the right people immediately upon detection of a breach, via multiple channels to make sure the breach can be acted upon quickly and the damage can be contained.

When evaluating data governance solutions, ask your vendor the questions in Table 1.

Table 1. Evaluating Solutions

Questions to Ask	Y/N
Does the data governance solution provide a level of visibility into my data assets that allows me to see who can access the data, who is sharing the data, and where the data is being shared?	☐
Does the data classification system automatically generate tags to reduce the risk of misclassification due to human error? Is it based on simple, basic file metadata or can it apply custom tags to obtain a deeper understanding of all my content?	☐
Does it have the reporting and auditing capability to identify abnormal data usage patterns, report them in a timely manner, and provide infrastructure optimization recommendations?	☐
Can it provide insight into and analysis of any content repository (cloud or on-premises) regardless of location?	☐

This list is just a start, but it covers the basics of visibility, classification, action, and reporting. It is important to evaluate solutions on common criteria and understand how comprehensive the solutions you evaluate are.

The biggest challenge with managing and classifying data has been having a single tool that provides visibility across silos of application data, on-premises and in the cloud, and allows you to act on that visibility, analyze patterns, and report any associated risk. The good news is that solutions are evolving and new solutions, like Egnyte Protect, mean that data management and security technology is finally rising to the challenge.

Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides actionable insight and intelligence to the global IT community.

 www.esg-global.com

 contact@esg-global.com

 P. 508.482.0188