Protect and control your key information assets through information classification

Classify, label, protect, and audit (CLPA) your key information assets

Microsoft France

Published: October 2014 (Updated: March 2017)

Version: 2.0

Author: Philippe Beraud (Microsoft France)

Contributors/Reviewers: Jean-Yves Grasset (Microsoft France), Srinivasa Sivakumar (Microsoft India)

For the latest information on Microsoft Trustworthy Computing, please see

www.microsoft.com/twc

For the latest information on Azure Information Protection, please see

www.microsoft.com/aip

Copyright © 2017 [Microsoft Corporation](http://www.microsoft.com). All rights reserved.

Abstract: Due to increased regulation, the move the Cloud to save cost and bring agility as a result of the modernization of IT, the Consumerization of IT (CoIT) and the “Bring Your Own Device” (BYOD), the explosion of information with dispersed enterprise data, the Social Enterprise and its applications enabling new collaboration, etc., organization of all sizes are facing growing needs to protect and control sensitive information. Questions inevitably arise: How to appropriately identify sensitive information? And then in turn how apply the right level of control for maintaining the security and the privacy of the sensitive data?

In this context, this document discusses information classification as the foundation of the cross-organization effort, the suggested approach to sustain such an effort and thus to handle and manage information assets on that basis, as well as the Microsoft services, products, and technologies to consider in this context to implement a relevant information classification and enforcement infrastructure, which constitutes the starting point for an effective Information Protection (IP) system.

Table of Contents

[Feedback 1](#_Toc477267900)

[Introduction 2](#_Toc477267901)

[Objectives of this paper 4](#_Toc477267902)

[Non-objectives of this paper 5](#_Toc477267903)

[Organization of this paper 5](#_Toc477267904)

[About the audience 5](#_Toc477267905)

[Understanding current trends impacting information assets 6](#_Toc477267906)

[Modernization of IT 6](#_Toc477267907)

[Consumerization of IT 7](#_Toc477267908)

[Blocker questions regarding these trends 9](#_Toc477267909)

[An introduction to information classification 10](#_Toc477267910)

[Establishing an information classification process 11](#_Toc477267911)

[Defining a classification ontology and taxonomy 16](#_Toc477267912)

[Defining information assets ownership 23](#_Toc477267913)

[Deciding the allowed location(s) per classification level 24](#_Toc477267914)

[Defining the protection policies/profiles 26](#_Toc477267915)

[Classifying/labeling information assets 28](#_Toc477267916)

[Reclassifying/labeling information assets 28](#_Toc477267917)

[Managing information asset retention, recovery, and disposal 29](#_Toc477267918)

[Implementing a classification and enforcement infrastructure 30](#_Toc477267919)

[Promoting a culture of assets protection and awareness 32](#_Toc477267920)

[Managing current trends with information classification 33](#_Toc477267921)

[Modernization of IT 33](#_Toc477267922)

[Consumerization of IT 38](#_Toc477267923)

[Building a classification and enforcement infrastructure and beyond 40](#_Toc477267924)

[Classifying and labelling information at the time of creation or modification 40](#_Toc477267925)

[Protecting information with Rights Management 46](#_Toc477267926)

[Scanning for on-premises existing information 55](#_Toc477267927)

[Controlling information flows using a Data Loss Prevention (DLP) engine 64](#_Toc477267928)

[Monitoring the flow of sensitive data to cloud environments using CASB 75](#_Toc477267929)

[Benefiting from eDiscovery 78](#_Toc477267930)

[Integrating with other systems 82](#_Toc477267931)

# 

# Feedback

For any feedback or comment regarding this document, please send a mail to [AskIPteam@microsoft.com](mailto:AskIPteam@microsoft.com).

# Introduction

Every day, information workers use e-mail messages and collaboration solutions to exchange important and/or sensitive information such as financial reports and data, legal contracts, confidential product information, sales reports and projections, competitive analysis, research and patent information, customer records, employee information, medical records, etc.

Because people can now access their e-mail from just about anywhere, mailboxes have been transformed into repositories containing large amounts of potentially key and/or sensitive information assets. Likewise, collaboration and cloud storage solutions enable people to easily store and share information within the organization but also across partner organizations and beyond. Information leakage is a serious threat to organizations.

Moreover, according to a [2016 Verizon’s report](http://www.verizonenterprise.com/verizon-insights-lab/dbir/)[[1]](#footnote-2), hundred thousand (100K) confirmed security incidents occurred for 2016 with 3,140 confirmed data breaches.

All of the above can result in lost revenue, compromised ability to compete, unfairness in purchasing and hiring decisions, diminished customer confidence, and more. The above examples represent only a few of the real information assets management situations organizations must deal with today.

Furthermore, the ever-increasing dependency on and use of electronic data make information assets management more and more challenging especially in light of government and/or industry data-handling standards and regulations.

This overall information risks with the loss of sensitive information, the increased compliance obligations with data-handling standards and regulations, etc. demands effective Information Protection (IP) systems, which are not only secure but are also easy to apply, whether it’s about e-mail messages sent, documents accessed inside an organization or outside it to business partner organizations (e.g. suppliers and partners), customers, and public administration, or any other kind of information.

Organization of any size can benefit from an effective IP system in many of ways by helping to reduce:

* Violations of corporate policy and best practices.
* Non-compliance with government and industry regulations such as Health Insurance Portability and Accountability Act (HIPAA)[[2]](#footnote-3)/ Health Information Technology for Economic and Clinic Health (HITECH)[[3]](#footnote-4), Gramm-Leach-Bliley Act (GLBA)[[4]](#footnote-5), Sarbanes-Oxley (SOX)[[5]](#footnote-6), Canada's Personal Information Protection and Electronic Documents Act (PIPEDA)[[6]](#footnote-7), the forthcoming European Union (EU) General Data Protection Regulation (GDPR) (a.k.a. Regulation 2016/679)[[7]](#footnote-8) repealing the EU Data Protection Directive (a.k.a. Directive 95/46/EC), Japan's Personal Information Privacy Act (PIPA)[[8]](#footnote-9), etc. to just name a few.
* Loss of intellectual property and proprietary information.
* High-profile leaks of sensitive information.
* Damage to corporate brand image and reputation.

An IP system indeed provides capabilities to help:

* Identify and classify information assets against a previously defined taxonomy in order to identify key or sensitive information assets and further apply the right level of control for maintaining both the security and the privacy of the assets as follows.
* Safeguard the most important information assets at rest, in transit/motion, in use/process by applying controls based on the asset classification, and in accordance to the organization’s security and privacy standards.
* Protect/prevent classified information assets against information leak in accordance to the organization’s security and privacy standards.
* Enforce compliance and regulatory requirements for classified information assets in accordance to the organization’s security and privacy standards.

As illustrated above in the first bullet point, any effective IP system always grounds on a prior (and continuous) information classification effort. Such an effort provides organizations with the ability to not only identify the sensitive information but also to appropriately classify the information assets so that adequate controls and measures can be applied on that basis.

**Note** IP is also known as a different set of names: data leakage prevention, data loss protection, content filtering, enterprise rights management, etc. All of these categories aim to prevent an accidental and unauthorized distribution of sensitive information.

## Objectives of this paper

**This document introduces the information classification principles as a foundation for an Information Protection (IP) system.**

Thus, whilst information classification addresses a wide range of information management related needs, this document specifically focusses on the ones that help to identify key and/or sensitive information and consequently enable to handle and protect it appropriately:"

* **Electronic discovery**. Information classification can be used to identify documents that are relevant to a business transaction that has given rise to legal action.
* **Compliance audit**. Information classification can be used to find documents containing material that may be subject to the terms of a regulation.
* **Compliance and protection automation**. Information classification can be used to identify documents containing material that is subject to the terms of a regulation or a security requirement, and the classification labels can then be used as the basis for automating enforcement of protection policies that ensure compliance.
* **Access control**. Information classification can be used to identify confidential documents, and the classification labels can then be used as the basis for automating enforcement of an access control policy that ensures the documents are viewed only by users with a demonstrated need to know their contents.
* **Information flow control**. In a world of constantly shared information, data flows everywhere. Many of an organization's information flows are required to satisfy business goals, but sometimes data goes where it shouldn't. Information classification can be used to identify the types of information flowing within an organization and leaving its traditional perimeter. By knowing the business context of information, the destination to which it's heading and the level of protection it enjoys (or doesn't), security solutions can make decisions about whether to allow the communication."[[9]](#footnote-10)

In this context, this document covers the considerations that relate to the definition of an appropriate taxonomy, and subsequently the organization’s security and privacy standards to appropriately handle and protect (classified) information assets at rest, in transit/motion, or in use/process. From a practical standpoint, the document presents various approaches to apply and leverage classification on information assets.

The document finally considers the Microsoft services, products, and technologies that help to build in a holistic approach a relevant information classification infrastructure. This infrastructure will serve as the foundation of an effective IP system to sustain the organization’s security and privacy standards.

## Non-objectives of this paper

This document is intended as an overview document for information classification and how to implement a relevant information classification and enforcement infrastructure based on Microsoft services, products, and technologies. As such, it doesn’t provide neither in-depth description nor detailed step-by-step instructions on how to implement a specific covered feature or capability provided by the outlined Microsoft services, products, and technologies. Where necessary, it instead refers to more detailed documents, articles, and blog posts that describe a specific feature or capability.

Likewise, and as noticed in the previous section, aspects pertaining to information classification to address non-security related information management needs are not covered. This notably includes knowledge management and search as well as storage management.

Furthermore, although risk assessments are sometimes used by organizations as a starting point for information classification efforts, this document doesn’t discuss a process for a formal risk assessment. Organizations are strongly encouraged to consider identified risks that are specific to their business when developing an information classification process.

## Organization of this paper

To cover the aforementioned objectives, this document is organized by themes, which are covered in the following sections:

* Understanding current trends impacting information assets.
* An introduction to information classification.
* Managing current trends with information classification.
* Building a classification and enforcement infrastructure and beyond.

## About the audience

This document is intended for Chief Information Security Officers (CISO), Chief Risk Officers (CRO), Chief Privacy Officers (CPO), Chief Compliance Officers (CCO), Chief Digital/Data Officer (CDO)/Chief Digital Information Officer (CDIO), IT professionals, security specialists and system architects who are interested in understanding:

* The information classification principles,
* And how to ensure that the organization’s security and privacy standards are truly applied to protect the organization’s information assets that have to.

# Understanding current trends impacting information assets

There are many reasons why organizations of all sizes are currently facing growing demands to protect their information: increased regulation like the aforementioned GDPR that European Union will start to enforce in May 2018, explosion of information with dispersed enterprise data, mobility and, of course, social networking and popular collaboration tools.

We further discuss in this section two specific industry trends that are upon us or even already here:

1. Modernization of IT,
2. Consumerization of IT.

## Modernization of IT

Economic pressures are changing the way we live and do business:

* The global economic contraction faced by every enterprise and government,
* The productivity imperative to “do more with less”, with a better agility and time to market.

Organizations have no other choice than to become leaner, better focused, and more fit-to-purpose. Under such conditions, organizations need breakthrough changes to survive.

This renewal must apply to ALL systems of production and distribution - including IT - and the survivors will be the ones that specialize in what they do best and most efficiently.

As far as IT is concerned, the economic benefits from combined cloud innovations can lead to:

* The obvious: new ways of delivering and operating the IT infrastructure and systems,
* The profound: new business processes.

As the economy becomes progressively digital, organizations must decide how to refactor and redistribute business to accomplish the shift with the most efficiency. In this context, the cloud emerges as a major disruptive force shaping the nature of business and allowing dramatic global growth. **The cloud brings breakthrough change**, and thus represents a major opportunity and plays a determining role.

An increasing number of people recognize the benefits of locating some set of their IT services in the cloud to reduce infrastructure and ongoing datacenter operational costs, maximize availability, simplify management, and take advantage of a predictable pricing model provided that the resource consumption is also predictable, and the resource can be rapidly delivered to the market, can elastically scale in respect to the demand, and opens multichannel access for the business.

Many organizations have already begun to use both public and private cloud for their new applications. And many have already switched to using cloud services for generic IT roles (for example Office 365, Salesforce.com, etc.).

As part as the modernization of IT, cloud economics creates a federated IT for organizations with services but also (some of) their data in the cloud. Such a disruptive move inevitably raises a series of concerns within organizations.

Typically:

* The Chief Information Security Officer (CISO) and other decision makers worry about security.
* The Chief Privacy Officer (CPO) worries about (country-specific) privacy laws.
* The Chief Financial Officer (CPO) worries about SOX compliance.
* Etc.

**Note** To illustrate the above, see the series of articles [Legal issues in the Cloud - Part 1](http://www.cio.com.au/article/382624/legal_issues_cloud_-_part_1/)[[10]](#footnote-11), [Part 2](http://www.cio.com.au/article/382649/legal_issues_cloud_-_part_2/)[[11]](#footnote-12), [Part 3](http://www.cio.com.au/article/382652/legal_issues_cloud_-_part_3/)[[12]](#footnote-13), and [Part 4](http://www.cio.com.au/article/382657/legal_issues_cloud_-_part_4/)[[13]](#footnote-14).

**To embrace with confidence such a cloud journey, this implies to address the related challenges and provides an adequate answer at the organization level:**

* **Control**. With an increasingly (multi-sourced) cloud environment,organization workloads and the related information assets are then located on various services, platforms and providers/operators, outside any perimeter, beyond direct organizational control, which can have liability issues.
* **Security**. The move to the cloud plaids for clarity about control and security to assess and measure the introduced risks.
* **Vertical concerns.** Compliance is another area of challenge. Likewise, defense, financial services, and healthcare, etc. are some examples of sectors having specifics country-specific, industry-specific standards and requirements to conform to.
* **Privacy**. Compliance with country-specific laws that apply to, for example the aforementioned EU General Data Protection Regulation (a.k.a. Regulation 2016/679) repealing the EU Data Protection Directive (a.k.a. Directive 95/46/EC). (See [Forrester Brief: You Need An Action Plan For The GDPR](https://www.forrester.com/report/Brief+You+Need+An+Action+Plan+For+The+GDPR/-/E-RES136242)[[14]](#footnote-15)).
* **Data Sovereignty**. All the country-specific laws that apply in this space have to be observed.

## Consumerization of IT

Devices have become cheaper and more affordable over the last few years and unsurprisingly proliferate: netbooks, laptops, smartphones, slates and tablets. The same is true for both cellular and wireless networks that have become ubiquities. Social networks (Facebook, Google+, Yammer, etc.) are changing how people get information and communicate. People want content and services to work seamlessly across all these devices and environments. They are becoming connected all the time: at home, at work and everywhere in between, up to the point where personal and work communication can become indistinguishable.

As technology plays an increasingly important role in people’s personal lives, it has a profound effect on their expectations regarding the use of technology in their work lives. People have access to powerful and affordable PCs, laptops, and tablets, are using mobile devices more and more, expect “always on” connectivity and are connecting with each other in new ways using social networks. Ultimately, they have more choice, more options, and more flexibility in the technology they use every day and, as that technology spills over into their professional lives, the line between personal and professional time is blurring. People want to be able to choose what technology they use at work, and they increasingly want to use that same technology in all aspects of their lives. In fact, according to a study by Unisys (conducted by IDC), a full 95 percent of information workers use at least one self-purchased device at work.

"Consumerization of IT" (CoIT) is the current phenomenon whereby consumer technologies and consumer behavior are in various ways driving innovation for information technology within the organization. As people become more comfortable with technology innovation in their personal lives, they expect it in their professional lives.

Without any doubt, full-time employees (FTE), civil servants, contractors, etc. will demand access with anything anywhere:

* From any location: at work, at home, or mobile
* From any device (laptops, tablets, smartphones, etc.) regardless of the fact they’re managed/unmanaged, corporate/personally owned
* Etc.

**While CoIT has remarkable potential for improving collaboration and productivity, many organizations are grappling with the potential security, privacy and compliance risks of introducing consumer technologies into their environment**. **They’re particularly worried about data security and data leak.**



Figure 1 NEW normal environment

**Managing the CoIT challenge requires striking a balance between users’ expectations and the organization’s security, privacy, risk and compliance requirements.**

## Blocker questions regarding these trends

As highlighted throughout the previous short depictions of the current trends, they unsurprisingly impact the information assets and raise a series of blocker questions:

1. *What types of information and datasets exist across the organization, and how are they used in business processes?*
2. *What is the sensitivity of information stored and used* across the organization?
3. *What is the consequence of information loss or disclosure?*
4. *How is sensitive information discovered?*
5. *Are physical controls deployed?*
6. *How will information be destroyed?*
7. *How is the information protected in other organizations that handle it?*
8. Etc.

To only explicit a few of them.

**To tackle with confidence the above trends and unlock the above questions, organizations of all sizes should adequately leverage the information classification principles and thus subsequently ensure that IT security standards, and related rules and policies are applied to protect and control the organization key information assets along with their allowed location.**

**Note** The [Data Security And Privacy Playbook For 2017](https://www.forrester.com/The+Data+Security+And+Privacy+Playbook/-/E-PLA200)[[15]](#footnote-16) further elaborates on the above.

# An introduction to information classification

The very first step to protecting and controlling key or sensitive information assets - and thus for instance solving the information leakage issue – aims at having the ability to identify what information needs to be protected and how, depending on its nature and value.

Information data to be protected (in increasing order of complexity to identify) are typically:

* Personally identifiable information (PII) that can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual in context.
* Contextual/confidential information, e.g. unannounced financials, legal contracts, strategic plans, sales reports and projections, competitive analysis, etc.
* Intellectual property, e.g. trade secrets, research and patent information, etc.

Considering the above, **documents of any kind and email typically, form bulk of information assets to be protected, but not only. Data can indeed be either structured, semi-structured or unstructured.** Relational databases are typical example of structured data; likewise, XML files with specified XSD schema(s) are common examples of semi-structured data; emails and text files are common examples of unstructured data. Office and PDF documents may fall into the last two categories depending on their document formats.

**Depending on the nature of the data constitutes the information assets, the effort required to classify information is not the same in terms of magnitude.** In this respect, **classifying unstructured data is not an easy task** and could represent challenges.

In addition, the above data can exist in **three basic states:**

1. **Data at rest.** All[[16]](#footnote-17) data recorded on any storage media.
2. **Data in use/process.** All data not in an at rest state, that is on only one particular node in a network, for example, in resident memory, or swap, or processor cache or disk cache, etc. memory.
3. **Data in transit/motion.** All data being transferred between at least two nodes.

This must betaken in account in the classification approach.

“Any data classification effort should endeavor to understand the needs of the organization and be aware how the data is stored, processing capabilities, and how data is transmitted throughout the organization”[[17]](#footnote-18) inside it or outside it to business partner organizations (e.g. suppliers and partners), customers, and public administration.

Without classifying or categorizing data, organizations typically treat all data the same way, which rarely reflects the true differences in value among data sets in the organization. Data classification is a powerful tool that can help determine what data is appropriate to store and/or process in different computing architectures, like the cloud or on premises. Without performing data classification, organizations might under-estimate or over-estimate the value of data sets, resulting in inaccurate risk assessments and potentially mismanaging the associated risk.

Indeed, "classification facilitates understanding of data's value and characteristics. The value of data is relevant to each individual business and is determined by that business unit; an IT organization is unlikely to have any insight.”[[18]](#footnote-19)

## Establishing an information classification process

While organizations usually understand both the need for information classification and the requirement for conducting joined cross-functional effort, they’re usually stalled with the question on how initiate information classification: *where to begin?*

One effective and simple answer may consist in adopting the **PLAN–DO–CHECK–ADJUST (PDCA)** approach/cycle as a starting point.



Figure 1 PLAN-DO-CHECK-ADJUST cycle

PDCA is a well-known iterative four-step management method used in business for the control and continuous improvement of processes as notably popularized by the Microsoft Operations Framework (MOF) 4.0.

The multiple-cycles approach can be also found in the ISO/IEC 27001:2005 standard for implementing and improving security in the organization’s context.

**Note** Microsoft Operations Framework (MOF) 4.0 is concise guidance that helps organizations improve service quality while reducing costs, managing risks, and strengthening compliance. MOF defines the core processes, activities, and accountabilities required to plan, deliver, operate, and manage services throughout their lifecycles. The MOF guidance encompasses all of the activities and processes involved in managing such services: their conception, development, operation, maintenance, and—ultimately—their retirement. For additional information, see <http://www.microsoft.com/mof>.

### PLAN

This first **PLAN** step aims at establishing the objectives, all the references and materials, and processes necessary to deliver results in accordance with the expected output: an effective classification of the organization’s information assets, and in turn all the relevant security and privacy controls both in place and enforced.

Different people handle sensitive information assets for different purposes and in different ways throughout the organization. The flow of content across the organization varies from one business process to the next, from one geography to another, etc. Predicting where information assets might ultimately go inside or outside the network can be difficult. An organization therefore needs the support of staff from all departments - for example, IT, privacy/compliance, human resources, legal, marketing/communications, and business operations - to act on policies and remediate any incidents that are discovered: **Information classification has to be conducted as a joined cross-functional effort.**

This step is thus the time to define, invite, and involve people in the organization that are relevant for this effort:

* People with direct implications in Security, Privacy, Risk Management, Compliance, and Legal obligations have to be involved like the Chief Information Security Officer (CISO), the Chief Risk Officer (CRO), the Chief Privacy Officer (CPO) as well as the Chief Compliance Officer under the responsibility of the Chief Information Office (CIO). These people are well indicated to constitute an **Information Management Committee** within the organization.
* Other stakeholders, e.g. executives or key representatives from the business side of the organization, i.e. departments, business units, subsidiaries, etc., who have a strong understanding of the business along with a clear understanding of any compliance, industry or government regulations with which their data must comply, will need to be solicited as well. The same consideration applies to the IT professionals that provide the infrastructure and the services to sustain the (core) business.

A well-suited **Information Governance Board** can be formed to promote internal collaboration between business owners, Security, Privacy, Risk Management, Compliance, and Legal, and those who provide the infrastructure under the control and directions of the Information Management Committee.

Due to the sensitive nature of the information assets and the number of different teams and geographies in the organization that may need to be involved, we can’t stress enough how important it is to ensure that all stakeholders can provide input at an early stage, and work together to plan an effort that fulfills all key criteria for the organization.

This step is also the right time to conduct activities that will frame the classification effort (, and its cycles): defining the scope, building an appropriate inclusive ontology and taxonomy, defining the protection policies/profiles in this respect, and deciding the allowed location(s) as part of them.

The scope of the effort indeed needs to be clearly defined with, as a best practice, a preference to concentrate at a first glance on a limited number of applications/workloads, repositories, and type of data to create in the next step an inventory of these assets stored across the network. The best option as usual consists in focusing on achievable goals, and then in iterating on that basis when the first outcome reveals successful instead of embracing a larger scope and waiting for months or years without seeing any real outcome or improvement. This enable to smoothly and efficiently drive the organization beyond an initial implementation (or a proof of concept), i.e. one PDCA cycle, toward a complete operational implementation that is fully integrated with the day-to-day business operations at all appropriate points in the organization.

Furthermore, an inclusive ontology and taxonomy that includes the words and patterns is key for the rest of the effort and its success. The recommendation here is to start by building a relatively **simple and straightforward ontology and taxonomy that support the various (type of) data.**

**Such taxonomy scheme should take into account at least the sensitivity, the criticality, the legal requirements, and the (perceived) business value of the information assets**. **This constitutes the core properties or categories for information classification.**

**Note** The meaning of each category and example of possible values for each one are detailed in the next sections. On next PDCA cycles, the classification scheme could be easily enriched to add more values allowing to refine the organization’s specific requirements.

This activity represents an opportunity for an organization to develop a better understanding of its sensitive content according to its business and policy priorities. Thus, in order to help establishing a relevant scheme, the following non-exhaustive list of questions should be potentially addressed by the different stakeholders:

* *What information asset need to be classified?* Which means that, in the mass of (every day) produced information, what kind of data are sensitive and clearly important in the context of the organization and its (core) business?
* *How long is the information asset classified?* The sensitivity of an information asset is time dependent. For example, leak of financials results before expected announcement could be critical for the organization (stock value), whereas this information becomes unclassified once publicly disclosed.
* *Who can ask for it?* For example, if the data is considered as personal information, only the person and a restricted number of people from the Human Resources may have access to it.
* *How to get approval for viewing this data?* For certain type of information asset, it may exist a process to claim access to the data relying on the approval of the person designated as the owner or custodian. Consequently, the owner of each information or data must be known to be implied in controlling its access, classification and retention lifecycle.
* *How combination of information is handled?* Information that (closely) relates to a persona, and that could be later referred as to Personally Identifiable Information (PII) isn’t always private. For example, the last name and first name can be considered public if they can be easily found on social networks or on public Web sites on the Internet. Similarly, a pseudonym used to discuss on forums is a public information. However, this two-public information could become sensitive if they can be associated to a single person.

When suitable ontology and taxonomy are ready and validated by all the stakeholders through the above committee and board, the next activity consists in defining the set of controls – in the sense of the [ISO/IEC 27002:2013 standard](http://www.iso.org/iso/fr/home/store/catalogue_tc/catalogue_detail.htm?csnumber=54533)[[19]](#footnote-20) (formerly known as ISO/IEC 17799 standard), that need to be implemented to protect the information asset in accordance with its classification. This represents an opportunity to develop of a better understanding of how sensitive information assets are stored, used and where they travels on the network(s), in order to determine how to handle the myriad information assets protection situations that may arise.

The resulting outcome of this activity, named **protection policies/profiles**, describes, for each classification level (specific category value or set of values from different categories), the list of mandatory controls to apply as well as a very important data property: its location (regardless of its (possible) state(s) (data at rest, data in use/process, and data in transit/motion)).

In a world where cloud computing - a key element that grounds the Modernization of IT - is (perceived as) an issue from a security standpoint, and more specifically in terms of control, **the location of an information asset naturally goes into the equation**. Same kind of consideration applies to the mobility, i.e. one of the underlying aspect of the Consumerization of IT, and we thus end up with the same conclusion.

Therefore, it seems both legitimate and beneficial to include the information asset location as a new variable. This allows the organization, e.g. the above Information Management Committee and Information Governance Board, to discuss and decide the allowed location(s) of an information assets per classification level, and then reflect the decision in the protection policy that relates to this classification level.

As a result of the modernization of IT and the CoIT, the possible locations could in a non-exhaustive manner:

* The on-premises, a private cloud, a community cloud, a hosting environment, public cloud(s), etc.
* Managed devices, known devices or unmanaged devices.

### DO

Once the classification taxonomy is defined, the protection policies are agreed upon, the information assets scope is well established, etc., this second step **DO** deals with implementing the plan, executing the defined process and, of course, implementing and rolling out the classification and enforcement infrastructure as needed for that purpose.

Depending on the initial scope for the classification effort, an inventory of the information assets is created. The information assets need to be examined and assigned values for each classification category.

Starting this step with information assets discovery enables the organization to understand the magnitude of the sprawl of sensitive information assets that has accumulated over time. This effort aids greatly in estimating and focusing subsequent efforts, and related PDCA cycles.

It is also wise to approach an information assets-inventory activity with a furthered narrow initial scope that scans for a single class of content (i.e. one value within a classification category) or a limited number of classes (i.e. a few values within a few classification categories) – for example, "Company confidential" and "High Business Impact" (see later in this document). Limiting the class of content for discovery initially enables IT and Compliance executives to keep tighter controls over discovery and remediation.

Consequently, and per the protection policies, controls applying to the assets are inferable. For example, if the targeted workload (application) processes data classified as "Company confidential" and "High Business Impact", the associated protection policy should dictate than the data must be encrypted, accessible only for a restricted group of people, stored on-premises, and that corresponding patterns must be included in the Data Loss Prevention (DLP) system (if any) to detect and prevent any leakage.

In parallel, either already deployed or new technical solutions have to be chosen to implement the controls assigned for protecting information assets based on their classification. One should note that controls can also correspond or refer to (manual or automated) processes.

In addition to classification, and controls implementation, **access control** to the targeted information asset must be set to grant access rights only to people with the correct accreditation. One must determine who owns the data, which means who is producing the information, is able to classify it and to decide who will have access to it.

### Check

The third **CHECK** step mainly consists in studying the actual results from the classification data metering, reports, and logs (measured and collected in the **DO** step above) and compare against the expected classification results (targets or goals from the **PLAN** step) to ascertain any differences. This supposes to review, assess, and validate the various data metrics/reports to ensure that the tools and methods in place are effectively addressing the classification of the information assets regardless of their specific nature and the protection policies/profiles. This allows to highlight the risks and the business value at risk.

A special emphasis should be put on looking for deviation in implementation from the plan and also looking for the appropriateness and completeness of the plan to enable the execution.

On that basis, the classification results are communicated to the organization, the full-time employees, the civil servants, etc.

### Adjust

This fourth and last **ADJUST** step is the opportunity to:

* Request/Implement corrective actions on significant differences between actual and planned classification results: information assets reclassification, revision of the underlying methodology to adopt, etc.
* Analyze the differences in information assets that require revision in terms of classification, protection policies/profiles discrepancies, etc. to determine their root causes.
* Determine where to apply changes that will include improvement(s) of the whole information classification process.

Such a step enables to streamline the classification effort and to create a comprehensive phased approach that:

* Includes clear articulation and enforcement of IT governance, thorough engagement of business owners to prioritize information risks, and service-oriented operational processes.
* Allows to expand the coverage over the time, address new information risk(s).
* Enables continuous improvements by implementing additional safeguards and capabilities.

**Note** The classification effort should also drive the creation of an Information life cycle management framework, an incident management system for breach, etc. These considerations are outside the scope of the current document.

The next sections further depict activities of the above **PLAN–DO–CHECK–ADJUST (PDCA)** steps.

## Defining a classification ontology and taxonomy

Organizations need to find ways to categorize their information assets that make sense for their business, organizational, conformance, and legal requirements. According to the [ISO/IEC 17799:2005 standard](http://www.iso.org/iso/catalogue_detail?csnumber=39612)[[20]](#footnote-21), “information should be classified in terms of its value, legal requirements, sensitivity, and criticality to the organization”.

To be in coherence with this vision, four classification categories can typically be defined (as previous suggested) to initiate the effort:

1. The **Sensitivity** category, which is mainly correlated to the information **confidentiality**,
2. The **Privacy** category that will be mapped to **criticality**,
3. The **Compliance** category embracing the **legal requirements** topic.
4. The **Business Impact** categorycorresponding to the **value** of the asset for the organization,

One can argue that the above correspondence is not totally accurate: a critical data may not be related to a personally identifiable information (PII) (see later in this document) but to a business value, as well as the sensitivity may not concern only the data confidentiality. That is true and a different choice of classification categories or naming can be used to reflect a slightly different vision based on the already organization’s existing conventions.

This said, the information classification approach should solve a compromise between two divergent objectives: a) offer a limited number of categories and values to remain manageable and understandable by non-IT people who product the business information, but also b) reflect/include all the complexity of data characteristics to offer the proper protection.

The following sections describe each of the four suggested classification categories. Examples of values that can be used to characterize the information are provided as part of the description.

### Defining the sensitivity levels that addresses your needs for information protection

Information classification starts by analyzing information sensitivity to determine into which of predefined sensitivity level it falls to be later in a position to apply the right control(s) according the security standards.

One can consider that the information **sensitivity refers primarily to confidentiality**.

The sensitivity comes from the military and public sectors where information is classified based on the notion of secrecy and protection against divulgation of crucial information that could damage military capabilities, international relation or homeland security. For example, the [(UK) Government Security Classification Policy document](https://www.gov.uk/government/publications/government-security-classifications)[[21]](#footnote-22) (dated 5 March 2014) recommends to classify assets into three types, OFFICIAL, SECRET and TOP SECRET, assuming that “ALL information [...] has intrinsic value and requires an appropriate degree of protection”. It is interesting to point out that information with no label is by default considered as OFFICIAL, but that a sort of new classification OFFICIAL-SENSITIVE must be used to mark “a little bit more sensitive” assets to reinforce the access control based on the “need to know‟ principle.

Concerning the private sector, companies generally define at the bare minimum three sensitivity global levels or values:

* **Confidential** (also labelled “Company confidential” or “restricted”). Information assets falling into this level typically include data where disclosure to unauthorized parties could **cause severe or catastrophic harm** to one or more individuals like the information asset owner, the organization, and/or relying parties. Access to such assets are frequently strictly controlled and limited for use on a "need to know[[22]](#footnote-23)" basis only.  This might include in a non-exhaustive manner:
  + Business material, such as unannounced financials, trade secrets, documents or data that is unique or specific intellectual property.
  + Legal data, including potential attorney-privileged material.
  + Any information that can be used to directly or indirectly authenticate or authorize valuable transactions in the context of the organization, including private cryptography keys, username password pairs, or other identification sequences such as private biometric key files.
* **General** (also labelled “internal use only”, “medium” or “sensitive”). Information assets falling into this level include data where unauthorized or improperly disclosure, loss or destruction may **cause limited** **harm** to individual(s), the organization, and/or relying parties due to identity/brand damage, operational disruptions. Access to such assets are frequently limited for use by only those who have a legitimate business need for access.

This might include in a non-exhaustive manner:

* + Emails, most of which can be deleted or distributed without causing a crisis (excluding mailboxes or emails from individuals who are identified in the confidential classification).
  + Documents and files that do not include confidential data.

Such a level indeed typically includes anything that is not tagged/labelled "Confidential".

* **Public** (also labelled “low”, “unrestricted”, or “non-critical”). Information assets classified to this level include data where unauthorized or improperly disclosure could result **in none to very limited harm** to the organization, individual(s), or relying parties. These assets are typically intended to be widely published or disseminated and have minimal consequences if improperly disclosed. Non‑critical information is not subject to protection or data handling procedures.

This might include:

* + Organization’s public web sites, public keys, and other publically available information.
  + Commonly shared internal information, including operating procedures, policies and interoffice memorandums.
  + Companywide announcements and information that all employees, contingent staff, and those under Non-Disclosure Agreement (NDA) have been approved to read.

Two additional global levels are generally typically added to the above to enhance both the granularity in terms of sensitivity and the distinction between corporate vs. personal information, i.e. on one hand **Highly confidential** (also labelled “high” or “restricted”) for any information that must be considered more than confidential, and on other hand **Non-Business** (also labelled “personal”).

The [ISO/IEC 27001:2013 standard](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=54534)[[23]](#footnote-24) considers the following values: "Company confidential data", "client confidential data", "proprietary”, “unclassified”, and "public”.

**Whatever the terminology model is for the Sensitivity category, organizations usually adhere to a scheme with three to five distinct global standardized and approachable levels.**

**Sub-labels can further categorize the information per key departments within the organization and/or per (secret) internal projects, acquisition plans, etc.**

The resulting so-called Sensitivity category is sometimes referred as to **Enterprise Classification**.

### Handling the levels of the privacy category

As a reference, the [ISO/IEC 29100:2011 standard](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=45123)[[24]](#footnote-25) first defines a PII principal as a "natural person to whom the personally identifiable information (PII) relates" and Personally identifiable information (PII) "any information that can be used to identify the PII principal to whom such information relates, or is or might be directly or indirectly linked to a PII principal".

Beyond this formal definition, one can define two privacy global levels to take in account the sensitivity factor and the difference of impact that could be important considering privacy regulations:

* **Personally identifiable information (PII)**. PII represents information that can be used to uniquely identify, contact, or locate a single person or can be used with other sources to uniquely identify a single individual. Examples include but are not limited to name, email addresses, phone numbers, credit card information and real-time location data.
* **Highly-Sensitive personally identifiable information (HSPII)**. HSPII corresponds to a subset of PII considered to be so important to the individual that it deserves special protection.

This includes data that could:

* 1. Be used to discriminate (e.g., race, ethnic origin, religious or philosophical beliefs, political opinions, trade union memberships, sexual lifestyle, physical or mental health),
  2. Facilitate identity theft (e.g., mother’s maiden name),
  3. Or permit access to a user's account (e.g., passwords or PINs).

This might notably include:

* + Government-provisioned identification credentials (e.g., passport, social security, or driver’s license numbers).
  + Financial transaction authorization data (e.g., credit card number, expiration date, and card ID).
  + Financial profiles (e.g., consumer credit reports or personal income statements).
  + Medical profiles (e.g., medical record numbers or biometric identifiers).

Furthermore, identifying and labeling PII information becomes mandatory as it falls under privacy (country specific) requirements for example, the Data Protection Act (1998), the EU Data Protection Directive (a.k.a. Directive 95/46/EC) repealed by the EU General Data Protection Regulation (GDPR) (a.k.a. Regulation 2016/679), CA SB1386, etc.

In addition, emerging standards like the [ISO/IEC 27018:2014 standard](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=61498)[[25]](#footnote-26) establish guidelines to protect PII in a cloud computing context: identifying PII from a client side perspective becomes a necessity before envisaging to migrate data outside the organization and put it under the responsibility of a cloud provider.

**The organization must consequently identify any privacy (country specific) requirements that apply to its information assets and the above global categories in consequence and/or provide sub-categories to embrace the range of situations to cover.**

The resulting category is sometimes referred as to **Privacy Classification**.

### Identifying regulatory and compliance requirements

Depending on the organization’s business, information assets may fall under (country specific) strict compliance and/or regulatory handling requirements, e.g., SOX, GLBA, PCI DSS, HIPAA/HITECH, etc.

Table 1 Some Compliance and/or Regulatory Standards

| **Cloud Security Certification** | **Description** |
| --- | --- |
| Sarbanes–Oxley Act (SOX) | The Sarbanes–Oxley Act of 2002, also commonly called Sarbox or SOX, is a U.S. federal law that set new or enhanced standards for all U.S. public company boards, management and public accounting firms. “As a result of SOX, top management must individually certify the accuracy of financial information. In addition, penalties for fraudulent financial activity are much more severe. Also, SOX increased the oversight role of boards of directors and the independence of the outside auditors who review the accuracy of corporate financial statements."[[26]](#footnote-27) |
| Gramm–Leach–Bliley Act (GLBA) | GLBA requires financial institutions to put processes in place to protect their clients' nonpublic personal information. GLBA enforces policies to protect information from foreseeable threats in security and data integrity. |
| Health Insurance Portability and Accountability Act (HIPAA)/ Health Information technology for Economic and Clinic Health (HITECH) | HIPAA and the HITECH Act are U.S. federal laws that apply to healthcare companies, including most doctors’ offices, hospitals, and health insurers. They establish requirements for the use, disclosure, and safeguarding of individually identifiable health information. In other words, HIPAA/HITECH impose on covered healthcare organizations security, privacy, and reporting requirements regarding the processing of electronic protected health information. |
| Payment Card Industry (PCI) Data Security Standards (DSS) | The PCI DSS is an information security standard designed to prevent fraud through increased controls around credit card data. PCI certification is required for all organizations that store, process or transmit payment cardholder data. |

**The organization must identify any country specific regulatory and compliance requirements that apply to its information assets.**

**Important note** Even if the above listed compliance and/or regulatory standards could appear to be US-centric, they in fact potentially apply in various regions regardless of the country.

The resulting category is sometimes referred as to **Government Regulations & Industry Classification**.

### Defining the business impact scale for the organization (optional)

The definition of the sensitivity levels for information protection, the privacy levels in accordance to the privacy (country specific) requirements that apply, and the (country specific) regulatory and compliance requirements allows in turn the organization to optionally scale the business impacts for the organization.

The Business Impact category is reflecting the **value** of the information **from a business point of view**. It must be clearly distinguished from the sensitivity category where the notion of business is not present. To understand how to determine the level of Business Impact for an asset, you have to answer to the following question: *what would be the impact on my business if this information is leaking outside the organization?*

For an organization, this category may become **the most visible classification category** and the reference for data owners to assign a classification level.

**Important note** For their (full-time) employees or civil servants depending of the considered sector, the Sensitivity category is probably easier to be accustomed with.One should remember that the information asset owners or their delegates have the responsibility to classify their assets per the classification rules.

The Business impact category must be seen as an optional **higher level classification** that will include information assets eventually already classified in the three previous categories (Sensitivity, Privacy, and Compliance).

Three levels are generally considered relevant to appropriately scale the business impact for the organization as follows:

1. **High Business Impact (HBI)**. Information assets classified to this level typically include data where disclosure to unauthorized parties could cause **severe or catastrophic harm** to one or more individuals like the information asset owner, the organization, and/or relying parties. Access to such assets are frequently strictly controlled and limited for use on a "need to know" basis only.

Examples of HBI data are:

* Authentication credentials or private keys theft that could be used to carry out fraudulent transactions,
* Leakage of HSPII or PII under stringent regulations that could lead to a substantial penalty,
* Public divulgation of confidential data with highly sensitive strategic content that could be harmful for the future of the enterprise or impair seriously the company reputation,
* Critical intellectual property theft that could be capable of conferring an immediate, competitive advantage to an unauthorized party, etc.

1. **Moderate Business Impact (MBI)**. Information assets classified to this level include data where unauthorized or improperly disclosure, loss or destruction may cause **limited harm** to individual(s), the organization, and/or relying parties due to identity/brand damage, operational disruptions, and or legal/regulatory liability. Access to such assets are frequently limited for use by only those who have a legitimate business need for access.

Examples of MBI data are:

* Intellectual property (source code or trade secret) theft that could adversely affect a limited competitive advantage,
* Disclosure of an unreleased product roadmap,
* Leakage of not sensible PII (some personal information like Social Security Number (SSN) are considered as HBI).

1. **Low Business Impact (LBI)**. Information assets classified to this level include data where unauthorized or improperly disclosure could result **in none to very limited harm** to the organization, individual(s), or relying parties. These assets are typically intended to be widely published or disseminated and have minimal consequences if improperly disclosed.

Examples of LBI data are:

* Publicly accessible Web pages,
* Public cryptographic keys,
* PII information considered as public, i.e. than can be found on social networks, blogs, public directories, etc.
* Commonly shared (internal) information, internal emails, including operating procedures, policies,
* Companywide announcements and information that all employees, contingent staff have read access,
* Source code shared under an open source license.

**Note** The LBI category regroups different kind of information: some is labelled Public in the Sensitive category whereas other data is categorized PII or company internal.

The final decision for the classification level is under the **responsibility of data owners** who have discretion to elevate certain assets to the highest level where warranted.

The below diagram depicts the 3 core categories, e.g. Sensitivity, Privacy, and Compliance, along with on the above level, the optional **Business Impact Classification** category.



Figure 2 Information Classification

To summarize the benefits of the suggested approach, the optional "Business Impact" category:

* Takes the business value as privileged indicator, what should make sense for most organizations.
* Offers a three-levels classification easily understandable by data owners.

As outlined above, **(full-time) employees, civil servants, etc. who may play the role of data owners and will then have finally the decision for classifying information may find more understandable the three/five-levels Enterprise classification**.

* Extends the well-known three levels classification (used by Defense or public entities based on information sensitivity to include compliance and privacy considerations.
* Offers the capability to label information more precisely with the aim of to put in place appropriate controls.

The "Business Impact" category is sometimes referred as to **Asset Classification**.

## Defining information assets ownership

In addition to defining a classification ontology and taxonomy, it’s also important for organizations to establish a clear custodial chain of ownership for all information assets in the various business units/departments of the organization.

The following identifies different information asset ownership roles in information classification efforts:

* **Information asset owner**. This role corresponds to the original creator of the information asset, and who can delegate ownership and assign a custodian. When a file is created, the owner should be able to assign a label to classify the information, which means that they have a responsibility to understand what needs to be classified in accordance to their organization’s information classification taxonomy. For instance, in terms of sensitivity, all of an information asset owner’s data can be auto-classified as for “General” unless they are responsible for owning or creating “Confidential” data types. Frequently, the owner’s role will change after the information asset is classified. For example, the owner might create a database of classified information and relinquish their rights to the data custodian.

**Note** Information asset owners often use a mixture of services, devices, and media, some of which are personal and some of which belong to their organization. A clear organizational policy can help ensure that usage of devices such as laptops and smart devices is in accordance with information classification guidelines.

* **Information** **asset custodian**. This role is assigned by the asset owner (or their delegate) to manage the information asset according to agreements with the asset owner or in accordance with applicable policy requirements. Ideally, the custodian role can be implemented in an automated system. An asset custodian ensures that necessary access controls are provided and is responsible for managing and protecting assets delegated to their care. The responsibilities of the asset custodian could include in a non-exhaustive manner:
  + Protecting the information asset in accordance with the asset owner’s direction or in agreement with the asset owner.
  + Ensuring that protection policies/profiles are complied with (see section § Defining the protection policies/profiles later in this document).
  + Informing asset owners of any changes to agreed-upon controls and/or protection procedures prior to those changes taking effect.
  + Reporting to the asset owner about changes to or removal of the asset custodian’s responsibilities.
* **IT professional**. This role represents a user who is responsible for ensuring that integrity is maintained, but they are not an information asset owner, custodian, or user. In fact, many IT professional/administrator roles provide data container management services without having access to the information asset. Such roles include backup and restoration of the information asset data, maintaining records of the information assets, and choosing, acquiring, and operating the devices and storage that house the information assets, etc.
* **Information asset user**. This role includes anyone who is granted access to data or a file. Access assignment is often delegated by the owner to the asset custodian.

The following table identifies the respective rights of the above information asset ownership roles. This doesn’t represent an exhaustive list, but merely a representative sample.

Table 2 List of roles and rights in information classification efforts

| **Role** | **Create** | **Modify/Delete** | **Delegate** | **Read** | **Archive/Restore** |
| --- | --- | --- | --- | --- | --- |
| Owner | ✓ | ✓ | ✓ | ✓ | ✓ |
| Custodian |  |  | ✓ |  |  |
| IT professional |  |  |  |  | ✓ |
| User |  | ✓ |  | ✓ |  |

## Deciding the allowed location(s) per classification level

As already noticed before, in the era of the transition to the cloud with the modernization of IT, organizations have to make tough decisions and evaluate the pro and cons of moving to the cloud for (part of) each application or workload.

The challenge mostly consists - for the security, privacy, risk, and compliance decision makers - in finding the right balance between benefits resulting from the cloud economics (like cost-reduction or faster time-to-market, which offers immediate competitive advantage) versus the risk of hosting or processing data outside the direct control of organization whilst respecting/enforcing all the (country-specific) privacy and compliance regulations that applies.



Figure 3 Decision framework for the cloud

The classification categories and their values greatly help in expressing the rules that prohibit or allow under certain conditions the processing or hosting of information in the cloud.

Let’s take an example of a decision tree that illustrates the interest of considering the different classification categories to decide what kind of information could migrate to the cloud.



Figure 4 Decision Tree for migrating information to the cloud

The first test considers the **Privacy** level, which eliminates immediately HSPII. In the case of non-PII, the **Business Impact** value is evaluated and any HBI information is also denied. On the contrary, any Low Business Impact data is authorized. Considering the left-hand branch of the synoptic, PII or MBI data are submitted to approval, which means that the decision could be based on other criteria like regulations constraints and a positive decision could be associated with requirements like implementing controls (encryption, audit, etc.).

Such a decision tree can also ensure that information is stored and removed on best practices that pertain to the allowed location(s).

The next step consists in translating the decision framework in:

1. The protection policies/profiles, see next section.
2. The retention, recovery, and disposal policies, see section § Managing information asset retention, recovery, and disposal later in this document.

The same kind of considerations also apply for CoIT. Likewise, a decision framework can be built in respect of devices.

## Defining the protection policies/profiles

As implicitly suggested so far, information assets are protected based on their classification. In other words, unless the information is classified correctly, it cannot be (automatically) protected.

The classification of an information asset indeed dictates the minimum-security controls that must be utilized when handling, storing, processing information, and/or transporting information asset.

Each protection policy/profile contains a set of rules defining the minimum requirements in terms of security controls for protecting the confidentiality, integrity, and availability (CIA) (triad) of information assets. For example, as far as confidentiality is concerned, regarding the three aforementioned states for data:

* **Data at rest** can be regarded as "secure" if and only if data is protected by strong encryption (where “strong encryption” is defined as “encryption requiring a computationally infeasible amount of time to brute force attack”) AND the encryption key is a) not present on the media itself b) not present on the node associated with the media; and c) is of sufficient length and randomness to be functionally immune to a dictionary attack.
* **Data in use/process** can be regarded as "secure" if and only if a) access to the memory is rigorously controlled (the process that accessed the data off of the storage media and read the data into memory is the only process that has access to the memory, and no other process can either access the data in memory, or man-in-the-middle the data while it passes through I/O), and b) regardless of how the process terminates (either by successful completion, or killing of the process, or shutdown of the computer), the data cannot be retrieved from any location other than the original at rest state, requiring re-authorization.
* **Data in transit/motion** can be regarded as “secure” if and only if a) both hosts are capable of protecting the data in the previous two states and b) the communication between the two hosts is identified, authenticated, authorized, and private, meaning no third host can eavesdrop on the communication between the two hosts.

In addition, the location of an information asset can further constraint the set of security controls to apply.

Consequently, for each classification level, and then for each location allowed in this respects, a related protection policy/profile should be developed by the organization’s IT.

In a phase approach as allowed as per PLAN–DO–CHECK–ADJUST (PDCA) approach, one can start by focusing on the confidentiality.

The security controls should deal with the following subjects in a non-exhaustive manner:

* Authentication of internal users.
* Authentication of external users.
* Authorization and access permissions.
* Back up, archival, and retention.
* Disposal of non-electronic media (paper and whiteboards).
* Disposal of electronic media (hard drives, USB drives, DVD-ROMs, etc.).
* Encryption – data confidentiality at rest.
* Encryption – data confidentiality in transit/motion.
* Event logging.
* Physical transport.

**Important note**  The above list encompasses two notions: protection and access control.

**Note**  The National Institute of Standards and Technology (NIST) provides a comprehensive catalog of security and privacy controls through the [Special Publication 800-53 Rev. 4](http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf) document[[27]](#footnote-28). The SANS Institute provides a list[[28]](#footnote-29) of the critical controls. This constitutes a subset of the aforementioned catalog.

Furthermore, the organization’s IT may delegate to a subset of business users the task to relevantly define the rules, since IT professionals may lack the context/information to successfully perform it.

The protection policy/profiles with their rules for the various classification levels make up the organization’s information security and privacy policies.

Whereas IT professionals are in charge of deploying and managing the security controls that relates to the rules, information owners and information custodians in the various business units/departments have the responsibility to ensure that all organization’s information assets are classified according to the organization’s ontology and taxonomy.

This leads us to the topic on how to classify information assets, and then protect information assets based on their prior classification.

## Classifying/labeling information assets

To fit potentially different needs and process, information classification can be realized in various ways and methodologies:

* **Manual classification**. The information classification is specified manually by the content creator, the business owner (if distinct), or a trained expert for the specifics and characteristics of the business. Templates of documents can be used for default settings if applicable.
* **Automatic classification**. Based on content or other characteristics, the information classification is (either) done automatically (or recommended to an end-user who in turn enforces it (or not)). The centerpiece resides in the classification/content-analysis engine that evaluate information assets by using a variety of techniques. This requires modelling or formulate a set of rules that map search results onto the classification levels as defined in the classification taxonomy. For that purpose, these techniques might include at the core searching for specific keywords, terms, phrases, identifying patterns via for instance the use of regular expressions, and analyzing the context in which a match is detected, etc.

Such an approach usually imposes to find the right balance between efficiency and accuracy. This typically requires some tuning and optimization to maximize overall performance and scalability while achieving an acceptable percentage of both false positives and negatives. This includes creating a test body of information assets with known classification level that you can use to evaluate the regular expression's accuracy.

* **Application-based**. Using certain applications by default sets the labels such as the ones defined by the Sensitivity category. For example, data from customer relationship management (CRM) software, HR, and health record management tools are “Confidential” by default.
* **Location-based**. Information asset location/ can help identify data sensitivity. For example, data that is stored by an HR or financial department is more likely to be confidential in nature. So, the information classification is directly based on the storage (e.g. the folder, the collaboration library, the database, etc.) where the information is created/stored. This might be driven by the business (owner) that makes the storage available.

Management considerations apply to all above classification methodologies. These considerations need to include details about who, what, where, when, and why an information asset would be used, accessed, changed, or deleted.

All information asset management must be done with an understanding of how an organization and/or a business unit views its risks. Additional considerations for information classification include the introduction of new applications and tools, and managing change after a classification method is implemented.

## Reclassifying/labeling information assets

Reclassifying or changing the classification state of an information asset needs to be done when a user or system determines that the information asset’s importance or risk profile has changed. This effort is important for ensuring that the classification status continues to be current and valid for the organization and/or its business units/departments:

**Note** Most content that is not classified manually can be classified automatically or based on usage by an information asset custodian or an information asset owner (see section § Defining information assets ownership).

* **Manual reclassification**. Ideally, this effort would ensure that the details of a change are captured and audited. The most likely reason for manual reclassification would be for reasons of sensitivity, or for a requirement to review data that was originally misclassified etc.

Because this document considers information classification and potentially moving them to the cloud (see section § Managing current trends with information classification later in this document), manual reclassification efforts would require attention on a case-by-case basis and a risk management review would be ideal to address classification requirements. Generally, such an effort would consider the organization’s policy about what needs to be classified, the default classification state (all data and files being sensitive but not “Confidential”), and take exceptions for high-risk data.

* **Automatic reclassification**. This uses the same general rule as manual classification, the exception being that automated solutions can ensure that rules are followed and applied as needed. Information classification can be done as part of an information classification enforcement policy, which can be enforced when data is stored, in use, and in transit using authorization technology.
* **Application-based**. See previous section.
* **Location-based**. Ibid.

## Managing information asset retention, recovery, and disposal

Information asset recovery and disposal, such as information reclassification as per previous section, is an essential aspect of managing information assets for organization. The principles for information asset recovery and disposal would be defined by an information asset retention policy and enforced in the same manner as information reclassification.

Such an effort would be performed by the information asset custodian and IT professional roles as a collaborative task. (See section § Defining information assets ownership.)

Failure to have an information asset retention policy could mean data loss or failure to comply with regulatory and legal discovery requirements. Most organizations that do not have a clearly defined information asset retention policy tend to use a default **keep everything** retention policy. However, such a retention policy has additional risks in cloud services scenarios.

However, such a retention policy has additional risks in cloud-based scenarios as part of a modernization of IT effort, see section § Modernization of IT later in this document. For example, a data retention policy for cloud service providers can be considered as **for the duration of the subscription** (as long as the cloud service is paid for, the information asset is retained). Such a pay-for-retention agreement may not address corporate or regulatory retention policies.

Deciding the allowed location(s) per classification level (see eponym section) can ensure that information assets are stored and removed in accordance to the organization’s decision framework for the cloud (see section § Deciding the allowed location(s) per classification level).

In addition, an archival policy can be created to formalize an understanding about what data should be disposed of and when.

Data retention policy should address the required regulatory and compliance requirements, as well as corporate legal retention requirements. Classified data might provoke questions about retention duration and exceptions for data that has been stored with a provider. Such questions are more likely for data that has not been classified correctly.

## Implementing a classification and enforcement infrastructure

Beyond the ability to effectively and relevantly classify and label information assets, information assets MUST then be protected according to the applicable protection policie(s)/profile(s) requirements based on its defined classification (see below).  **The higher the value of the information, the tighter should be the security controls.**

Aside applying access control measures, this also imposes to consider how prevent data loss, how to conduct electronic discovery (e-discovery) process, etc. This is the purpose of the next sections.

### Preventing data loss

As outlined in the introduction, today’s working environment provides many “opportunities” for information data leakage or loss to occur inside an organization or outside the organization.

With the externalization and rationalization of IT, the growing use of increasingly powerful computers and devices, the introduction of extensive connectivity through networks and the Internet, the need to exchange information with business partner organizations (e.g. suppliers and partners), customers, as well as public administration, along with the various collaboration and cloud storage solutions, **organizations are not islands and have to address the threats of theft and mishandling information assets.**

In addition to these threats – accidental leakage/loss occurs more often than malicious insider –, a growing list of regulatory requirements adds on top of the ongoing task of protecting digital files and information. For example, the financial, government, healthcare, and legal sectors are increasingly taxed by the need to better protect digital information assets due to emerging regulatory standards such as HIPAA and the GLBA in the financial services market.

All of the above demand systems in place designed to help preventing such data leakage/loss.

According to [Wikipedia](http://en.wikipedia.org/wiki/Data_Loss_Prevention)[[29]](#footnote-30), “Data Loss Prevention (DLP) is a computer security term referring to systems that **identify, monitor, and protect** **data in use** (e.g., endpoint actions), **data in motion** (e.g., network actions), and **data at rest** (e.g., data storage) through deep content inspection, contextual security analysis of transaction (attributes of originator, data object, medium, timing, recipient/destination, etc.), and with a centralized management framework. The systems are designed to detect and prevent the unauthorized use and transmission of confidential information."

Such systems must thus be based on ensuring compliance with the aforementioned handling standards and the related set of classification rules, which make up the organization’s information security and privacy policies.

Depending on their scope of applicability, e.g. endpoint actions, network actions, and data storage to rephrase the above definition, they mainly consist of filters on email and internet access, control of devices storage, etc. They put up barriers to stop sensitive and confidential information from leaving the organization and thus allowing to monitor access to that information from within the company.

Although no form of information will ever be completely risk-free from unauthorized use and no single approach will shield data from misuse in all cases, the best defense is a comprehensive multi-layered solution for safeguarding information. Consequently, such systems usually can operate in conjunction with [Information Rights Management](https://www.prot-on.com/information-rights-management) (IRM) solutions, a.k.a. Enterprise Digital Rights Management (e-DRM) solutions in order to better control how data is used and distributed beyond the use of simple access control.

In addition to the above DLP capability, an [Information Rights Management](https://www.prot-on.com/information-rights-management) (IRM) solution should indeed help protect an organization's records and documents on the organization’s intranet, as well as from being shared with unauthorized users. It should help ensuring that data is protected and tamper-resistant. When necessary, information should expire based on time requirements, even when that information is sent over the internet to other individuals.

### Conducting electronic discovery (e-discovery) process

As already noticed, with the explosive growth compliance requirements both inside and outside organizations, compliance has become everyone’s responsibility. Neither the IT department nor the legal and compliance departments can keep tabs on all of the information that is exchanged in the ordinary course of business. Organizations need tools that enable self-service and automated compliance wherever possible. Enabling legal teams to search, hold and export the right information without intervention from IT is cost saving for the organization.

This is the role devoted to the electronic discovery (e-discovery), which "is the **identification, preservation, collection, preparation, review and production of electronically stored information associated with legal and government proceedings**"[[30]](#footnote-31)

E-discovery enables compliance officers to perform the discovery process in-place where data is not duplicated into separate repositories.

In this regard, the [Electronic Discovery Reference Model (EDRM)](http://www.edrm.net/)[[31]](#footnote-32) provides a set of guidelines and processes for conducting e-discovery for customers and providers, which was developed by a group of industry experts in a number of working projects. EDRM focuses on reducing the cost and complexity of e-Discovery through the development of standards such as the EDRM XML schema to attempt to provide a clear and repeatable process for e-Discovery solutions. "The goal of the EDRM XML schema is to provide a standard file format that enables an efficient, effective method for transferring data sets from one party to another and from one system to another in pursuit of electronic discovery (e-discovery) related activities."[[32]](#footnote-33)

As such, e-discovery solutions optimally operate on data where it lives, and preserve minimum amount of data needed. Since content is held in-place, teams can respond quickly by accessing data in its native format (without any loss of fidelity often associated with copying data to separate archives). Then, teams have an easy way to package the result by exporting it according to the EDRM XML schema so that it can be imported for example into a review tool.

## Promoting a culture of assets protection and awareness

Technologies to apply security controls and (security and privacy) policies alone will not necessarily protect an organization. The organization must continuously evangelize the importance of protecting key and sensitive information assets, and must provide both information and training on appropriate ways to reach this objective.

Establishing who within the organization has ownership over information assets is indeed just the first step in promoting an attentive and vigilant culture of information assets security.

The organization must continually educate users. An organization should thus provide dedicated resources allowing employees to:

* Understand that they’re responsible for classification and compliance.
* Easily find up-to-date definitions of the practices that align to the organization’s information security and privacy policies.
* Understand these definitions.
* Integrate them on a day-to-day basis with adequate implementation procedures, and easy to use tool as well. For example, if a custom office toolbar can be developed to classify every document or email that gets authored, and if the tool suggests the users to classify before sending or saving and add the classification into the document metadata – search engines and DLP engines would know what do be done.

Providing training and ongoing oversight is just as important as the technical safeguards and solutions that the organization implements. Gamification could be a suitable approach.

# Managing current trends with information classification

The beginning of this paper has discussed two specifics trends, e.g. the Modernization of IT and the CoIT, where it becomes an imperative to adequately answer specifics challenges, to find the right balance between some benefits, expectations, etc. and the organization’s security, privacy and compliance requirements.

## Modernization of IT

The information classification as envisaged so far is not enough even if the established decision framework along with the related allowed location(s) greatly help in clarifying the risk(s) taken by the organization with the transition to the cloud for certain (type or part of) workloads, and their information assets.

### Updating the IT strategy

The information classification enables to update the whole IT strategy and sourcing strategy based on (type of) workloads as per [ISO/IEC 17788 standard](http://www.iso.org/iso/catalogue_detail.htm?csnumber=60544)[[33]](#footnote-34):

* Infrastructure as a Service (IaaS),
* Platform as a Service (PaaS),
* Software as a Service (SaaS),
* And Data Storage as a Service (DSaaS).

Service providers can then be prepositioned regarding the above type of workloads.

### Mapping the classification requirements with cloud security requirements

The sourcing strategy should then map the workloads (applications) and associated information assets classification requirements with the cloud security certifications and compliance standards like ISO/IEC 27001, SSAE 16 SOC 1 and SOC 2, HIPPA/HITECH, PCI DSS 1, FedRAMP/FISMA, EU Model Clauses, FERPA, PIPEDA, etc. per type of workloads.

The following table list the top certifications and compliance standards.

Table 2 Top Certifications and Compliance Standards

| **Cloud Security Certification** | **Description** |
| --- | --- |
| ISO/IEC 27001:2005 | ISO 27001 is one of the best security benchmarks available in the world which defines a rigorous set of physical, logical, process and management controls to be implemented in a (public cloud) service. |
| Statement on Standards for Attestation Engagements No. 16 (SSAE 16) SOC 1 and SOC 2 | The Service Organization Control (SOC) 1 Type 2 audit report attests to the design and operating effectiveness of service’s controls. The SOC 2 Type 2 audit included a further examination of service’s controls related to security, availability, and confidentiality. The service is audited annually to ensure that security controls are maintained.  Audits are conducted in accordance with the Statement on Standards for Attestation Engagements (SSAE) No. 16 put forth by the Auditing Standards Board (ASB) of the American Institute of Certified Public Accountants (AICPA) and International Standard on Assurance Engagements (ISAE) 3402 put forth by the International Auditing and Assurance Standards Board (IAASB). SSAE 16 is an enhancement to the former standard for Reporting on Controls at a Service Organization, i.e. the SAS70.  In addition, the SOC 2 Type 2 audit included an examination of the Cloud Controls Matrix (CCM) from the Cloud Security Alliance (CSA) (see below). |
| Health Insurance Portability and Accountability Act (HIPAA)/ Health Information technology for Economic and Clinic Health (HITECH) | HIPAA and the HITECH Act are U.S. federal laws that apply to healthcare companies, including most doctors’ offices, hospitals, and health insurers. They establish requirements for the use, disclosure, and safeguarding of individually identifiable health information. In other words, HIPAA/HITECH impose on covered healthcare organizations security, privacy, and reporting requirements regarding the processing of electronic protected health information. A service provider can provide physical, administrative, and technical safeguards to help customer’s organizations comply with HIPAA.  In many circumstances, for a covered healthcare organization to use a cloud service, the service provider must agree in a written agreement, e.g. the Business Associate Agreement (BAA), to adhere to certain security and privacy provisions set forth in HIPAA and the HITECH Act.  Microsoft has published a [HIPAA white paper](http://smb.blob.core.windows.net/smbproduction/Content/Microsoft_Cloud_Healthcare_HIPAA_Security_Privacy.pdf)[[34]](#footnote-35) which provides details about our approach to HIPAA and the HITECH Act. |
| Payment Card Industry (PCI) Data Security Standards (DSS) Level 1 | The PCI DSS is an information security standard designed to prevent fraud through increased controls around credit card data. PCI certification is required for all organizations that store, process or transmit payment cardholder data. Customer’s organizations can reduce the complexity of their PCI DSS certification by using compliant services in the cloud. The compliancy with Level 1 is indeed as verified by an independent Qualified Security Assessor (QSA), allowing merchants to establish a secure cardholder environment and to achieve their own certification. |
| Federal Risk and Authorization Management Program (FedRAMP)/ Federal Information Security Management Act (FISMA) | FedRAMP is a mandatory U.S. government program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud services. Prior to FedRAMP, service providers were required to undergo FISMA assessments by individual federal agencies. FISMA required U.S. federal agencies to develop, document, and implement controls to secure their information and information systems. A service provider has the ability follow security and privacy processes relating to FedRAMP/FISMA. |
| Gramm–Leach–Bliley Act (GLBA) | GLBA requires financial institutions to put processes in place to protect their clients' nonpublic personal information. GLBA enforces policies to protect information from foreseeable threats in security and data integrity. Organizations subject to GLBA can possibly use a service provider and still comply with GLBA requirements. |
| European Union (EU) General Data Protection Regulation (GDPR) | The EU General Data Protection Regulation (GDPR) (a.k.a. Regulation 2016/679) repealing below EU Data Protection Directive imposes new rules on companies, government agencies, non-profits, and other organizations that offer goods and services to people in the European Union (EU), or that collect and analyze data tied to EU residents. This European privacy law is due to take effect in May 2018, and will require big changes, and potentially significant investments, by organizations all over the world (including Microsoft and our customers). The GDPR applies no matter where you are located. See next section § A zoom on the GDPR. |
| European Union (EU) Model Clauses | The EU Data Protection Directive (a.k.a. Directive 95/46/EC), a key instrument of EU privacy and human rights law, requires organization in the EU to legitimize the transfer of personal data outside of the EU. The EU model clauses are recognized as a preferred method for legitimizing the transfer of personal data outside the EU for cloud computing environments. Offering the EU model clauses for a service provider involves investing and building the operational controls and processes required to meet the exacting requirements of the EU model clauses. Unless a service provider is willing to agree to the EU model clauses, a customer’s organization might lack confidence that it can comply with the EU Data Protection Directive's requirements for the transfer of personal data from the EU to jurisdictions that do not provide “adequate protection" for personal data. |
| EU-U.S. Privacy Shield | The EU-U.S. Privacy Shield Framework is set forth by the U.S. Department of Commerce regarding the collection, use, and retention of personal information transferred from the European Union to the United States. It is designed to provide companies on both sides of the Atlantic with a mechanism to comply with data protection requirements when transferring personal data from the European Union to the United States in support of transatlantic commerce. |
| Family Educational Rights and Privacy Act (FERPA) | FERPA imposes requirements on U.S. educational organizations regarding the use or disclosure of student education records, including email and attachments. A service provider can agree to use and disclosure restrictions imposed by FERPA that limit its use of student education records, including agreeing to not scan emails or documents for advertising purposes |
| Canadian Personal Information Protection and Electronic Documents Act (PIPEDA) | The Canadian Personal Information Protection and Electronic Documents Act pertains to how private sector organizations collect, use, and disclose personal information in the course of commercial business. |

Even though control and security represent a shared responsibility between the organization and the public service providers, clear information and status regarding each envisaged service provider’s security practices and its cloud security certifications should be obtained and assessed based on the previous mapping.

**Note**  The [Microsoft Trust Center](https://www.microsoft.com/en-us/TrustCenter/default.aspx)[[35]](#footnote-36) offers detailed security, privacy, and compliance information for all Microsoft cloud services: Azure, Office 365, Dynamics 365, etc. This is the place where Microsoft share our commitments and information on trust-related topics. These places aim at helping organization to address a wide range of international, country, and industry-specific regulatory requirements.

### A zoom on the GDPR

The European Union (EU) General Data Protection Regulation (GDPR) (a.k.a. [Regulation 2016/679](http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679)[[36]](#footnote-37)) is a new European privacy law[[37]](#footnote-38) published in May 2016, which applies to the protection of personal data for all European citizens. This regulation will come into force in May 2018 and, unlike a directive, will be immediately applicable in the legislation of the 28 member states, leaving little time for organizations, i.e. companies and service providers, to comply with.

Contrary to what one might think, all organizations, whether they are European or not, are concerned when they recover, process or store private data from EU nationals. For example, search engines, cloud service providers, e-commerce companies, applications that garner user data for targeting advertising or bulk processing are directly impacted. Fines for non-compliance are far from symbolic because they can reach **up to 20 M€ or 4% of the overall turnover of the organization**, by choosing the **higher of the two**, which can represent several billion for global companies.

To meet the requirements of the regulation, personal data must be clearly identified to be protected accordingly, in function of the risk. The flow of data, i.e. their end-to-end life cycle, from the collection, processing, storage, possibly sharing and deletion should be described precisely to prove that the necessary controls have been implemented to avoid any leakage of personal information.

Finally, the obligation to consider the protection of personal data from the design of the service/system is one of the key concepts of the regulation, requiring that the reflection on the controls to be implemented are designed from the beginning in correlation with the associated risks.

For 'high risk' personal data, the determination of risk must be made through the conduct of a Data Protection Impact Assessment (DPIA), allowing both to adapt the protection to the sensitivity of the information and demonstrate the approach during an audit.

The technical protection of the data will be based, not surprisingly, on encryption but also on the principle of pseudonymizing, widely quoted in the regulation, which makes it possible to eliminate the nominative character of the data using pseudonyms, see [ISO/TS 25237:2008 Health informatics – Pseudonymization](https://www.iso.org/standard/42807.html)[[38]](#footnote-39).

To be able to move towards compliance with the GDPR, companies will have not only to set up an organization around data security by using the role of Data Protection Officer (DPO) as defined in the Regulation, but also to rely on technical tools and solutions.

It can be seen that the various aspects concerning the discovery, classification and tagging/labeling of data, and protection by encryption either on-premises or in the cloud, resonate in this new context, and the overall approach of Azure Information Protection (AIP) facilitates the coverage of many technical aspects described in the requirements of the regulation. (See later in this document).

The European Union will start to enforce the GDPR in May 2018.

**Note**  Microsoft cloud services can help you stay focused on your core business while efficiently becoming compliant with the GDPR. For more information, see [Microsoft.com/GDPR](https://www.microsoft.com/gdpr)[[39]](#footnote-40).

### Creating a roadmap alignment

This previous activity allows to conduct a risk analysis and create a roadmap alignment based on information classification and cloud security alignment of the initially selected service providers. The Cloud Controls Matrix (CCM)[[40]](#footnote-41) and the [Security, Trust and Assurance Registry (STAR)](https://cloudsecurityalliance.org/star/)[[41]](#footnote-42) from the [Cloud Security Alliance (CSA)](https://cloudsecurityalliance.org/)[[42]](#footnote-43) can be leveraged to assess how an offering fulfills the security, compliance, and risk management requirements as defined by the CCM and what extend.

**Note**  The CCM is a meta-framework of cloud-specific security controls, mapped to leading standards, best practices and regulations in order to provide organizations with the needed structure, detail and clarity relating to information security tailored to cloud computing so that they can confidently evaluate an offering from a service provider.

**Note**  The STAR program is a publicly accessible registry designed to recognize the varying assurance requirements and maturity levels of service providers and consumers, and is used by customers, service providers, industries and governments around the world. STAR consists of 3 levels of assurance, which currently cover 4 unique offerings: a) Self-Assessment, b) STAR Attestation, c) STAR Certification, and d) STAR Continuous Monitoring. All offerings are based upon the succinct yet comprehensive list of cloud-centric control objectives in the CCM.

**Note** Microsoft Azure, Microsoft Intune, and Microsoft Power BI have obtained above STAR Certification, which involves a rigorous independent third-party assessment of a cloud provider’s security posture.

### Entering the new era of hybrid cloud deployment

The roadmap alignment enables the organization to enter with confidence, minimized risk, and compliance a new era of “hybrid cloud deployment” in which progressively more IT functionalities are hosted by cloud service providers at the same time that organizations retain IT information assets running in on-premises datacenters or in a private cloud.

Hybrid deployment requires that organizations increase their core infrastructure operational maturity in two main areas:

1. Improving the management of their on-premises IT landscape and traditional systems (servers, storage devices, network components, and other datacenter assets) by evolving towards a private cloud, i.e. an optimized and more automated way to provision and operate on-premises services.
2. Enhancing their service offerings by utilizing suitable public cloud services for lower-cost add-ons or replacements for existing IT assets. This implies as developed above that the cloud services are aligned with the strategy roadmap of the organization. Such a dynamic is actually twofold:
3. When suitable and inexpensive "pay-as-you-go" (PAYG) alternatives exist for generic components: organizations can retire their expensive in-house systems/applications/services (or parts of them), benefiting from cost reduction, agility, scalability, etc.
4. Organizations can cut the costs of new systems, applications or services by building them as cloud services and using other specialized cloud services as inexpensive building blocks that reduce their own labor.

The combination and interaction of the two approaches (private vs. public) driven by the information classification outcomes provides the best strategy for the Hybrid Era and meets:

* Meets the demands and the expectations of the businesses.
* Fulfills the security, privacy, risk, and compliance requirements of the organization.

**Note**  The [Enabling Hybrid Cloud Today with Microsoft Technologies whitepaper](http://www.microsoft.com/en-us/download/details.aspx?id=39052)[[43]](#footnote-44) discusses how Microsoft can help your organization achieve a successful hybrid cloud strategy and present the enabling technologies from on-premises, cross-premises, and off-premises implementation of (parts of) the services. Several typical patterns and common scenarios are illustrated throughout this paper.

## Consumerization of IT

The information classification as envisaged so far is also not enough here. However, it can have a great contribution to help building a foundational CoIT framework that will server to classify the apps based on the information (assets classification). Like the modernization of IT, an alignment roadmap can be in turn created using a CoIT maturity model across people, process, and technology.

The following figure illustrates the outcome in terms of a CoIT maturity model driven by information classification.



Figure 5 Example of CoIT maturity model

# Building a classification and enforcement infrastructure and beyond

Organizations of all sizes are challenged to protect a growing amount of valuable information assets against careless mishandling and malicious use. The increasing impacts of information theft and the emergence of new regulatory requirements to protect data emphasize the need for better protection of them.

In this shared context, this section depicts how to build in a holistic approach a classification and enforcement infrastructure on-premises, in the cloud, or in hybrid environment with Microsoft services, products, and technologies.

Such an infrastructure constitutes the starting point for an Information Protection (IP) system that can effectively help organizations of all sizes to enforce the defined protection policies/profiles, and thus apply the right level of control for maintaining the security (confidentiality and integrity) of their key and/or sensitive information assets throughout the complete data lifecycle – from creation to storage on-premises and in cloud services to sharing internally or externally to monitoring the distribution of files and finally responding to unexpected activities.

## Classifying and labelling information at the time of creation or modification

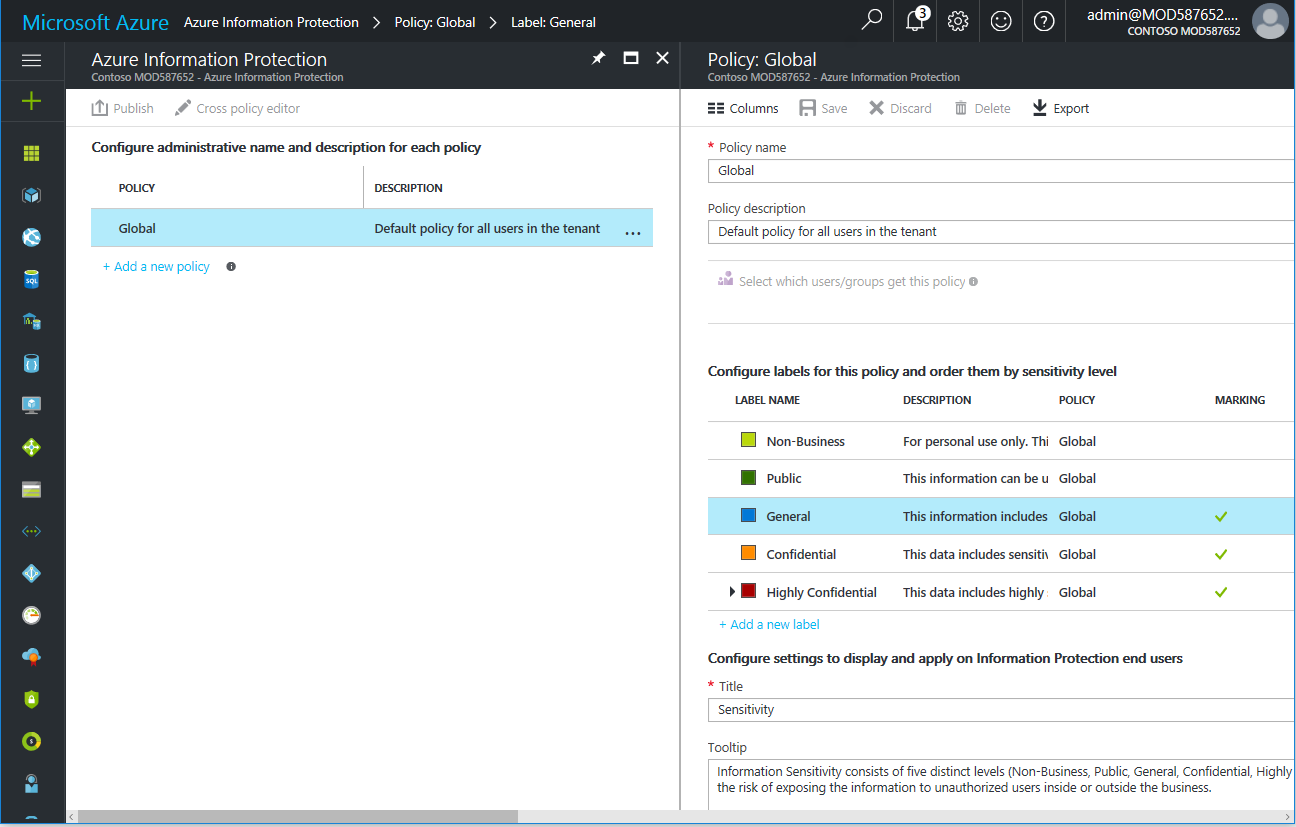
Azure Information Protection[[44]](#footnote-45) (sometimes abbreviated to AIP) is a cloud-based solution that helps an organization classify, and label its documents and emails, and in turn can apply Rights Management protection (encryption, authentication and use rights) to protect sensitive information (see next section).

**Note**  For more information, see articles [What is Azure Information Protection?](https://docs.microsoft.com/en-us/information-protection/understand-explore/what-is-information-protection)[[45]](#footnote-46) and [Frequently asked questions about classification and labeling in Azure Information Protection](https://docs.microsoft.com/en-us/information-protection/get-started/faqs-infoprotect)[[46]](#footnote-47).

**Note**  Azure Information Protection has evolved from a long history of established technologies from Microsoft that implement Rights Management protection. Because of this evolution, you might know this solution by one of its previous names, or you might see references to these names in documentation, the UI, and log files. For more information, see article [Azure Information Protection - also known as ...](https://docs.microsoft.com/en-us/information-protection/understand-explore/aka)[[47]](#footnote-48).

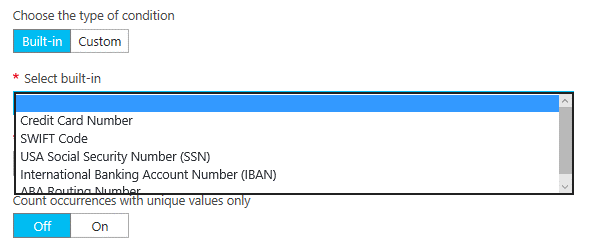
For that purpose, Azure Information Protection allow you to define classification and protection policies that specify the way different types of information assets should be classified, labelled and optionally protected for sensitive information.

These classification and protection policies are defined and managed over the time through the through Azure Information Protection service in the Azure portal[[48]](#footnote-49). **As a starting point, IT professionals are provided with a default global policy with a set of default sensitivity labels, which they can view or modify as needed to fit their own organization’s needs and requirements**.

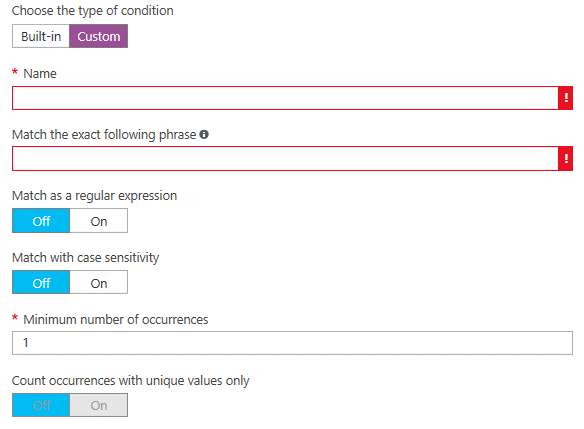


These default sensitivity labels are available with options to define custom labels based on your business needs and requirements as previously discussed as part of this paper. These labels come with predefined rules that govern how data is labelled, the actions to take and enforce based on classification, such as visual marking (headers, footers, watermarking), and protection (Rights Management template to apply, see next section), the conditions to meet to automatically apply the label, etc.

Such conditions enable to look for patterns in data, such as words, phrases or expression. For that purpose, you can select a built-in condition.



You can alternatively specify a custom condition to meet your own requirements. This allows you to find matches as a regular expression if you want to.



Over use of automatic classification can frustrate end-users. So, you can rely on recommendations in lieu of.

**Note** For more information, see article [Configuring Azure Information Protection policy](https://docs.microsoft.com/en-us/information-protection/deploy-use/configure-policy)[[49]](#footnote-50).

The following table lists the default top level labels for the default global policy along with the related actions.

Table 3 Default labels for Sensitivity

| **Label** | **Description** | **Action** |
| --- | --- | --- |
| Non-Business | **For personal use only**. This data will not be monitored by the business. Personal information must not include any business-related information. | Remove protection |
| Public | This business information is specifically prepared and approved for public consumption. **It** **can be used by everyone inside or outside the business**. Examples are product datasheets, whitepapers, etc. Content Is not encrypted, and cannot be tracked or revoked. | Remove protection |
| General | This business information is NOT meant for public consumption. This includes a wide spectrum of **internal business data that can be shared with internal employees, business guest and external partners as needed**. Examples are company policies and most internal communications. Content is not encrypted, and cannot be tracked or revoked. | Remove protection  Document footer:  Sensitivity: General |
| Confidential | This business information includes **sensitive business information which could cause harm if over-shared**. Exposing this data to unauthorized users may cause damage to the business. Examples are employee information, individual customer projects or contracts and sales account data. Recipients are trusted and get full delegation rights. Content is encrypted. Information owners can track and revoke content. | Footer –  “Sensitivity: Confidential” |
| Highly Confidential | This business information includes **highly sensitive information which would certainly cause business harm if over-shared**. Examples are personal identification information, customer records, source code, and pre-announced financial reports. Recipients do NOT get delegation rights and rights are enforced. Information owners can track and revoke content. | Footer –  “Sensitivity: Highly Confidential” |

Interestingly enough, the above default global policy or the changes that you configured for the global policy are seamlessly downloaded by Office client applications (Office 365 ProPlus, Office 2016, and Office 2013 SP1) on Windows devices.

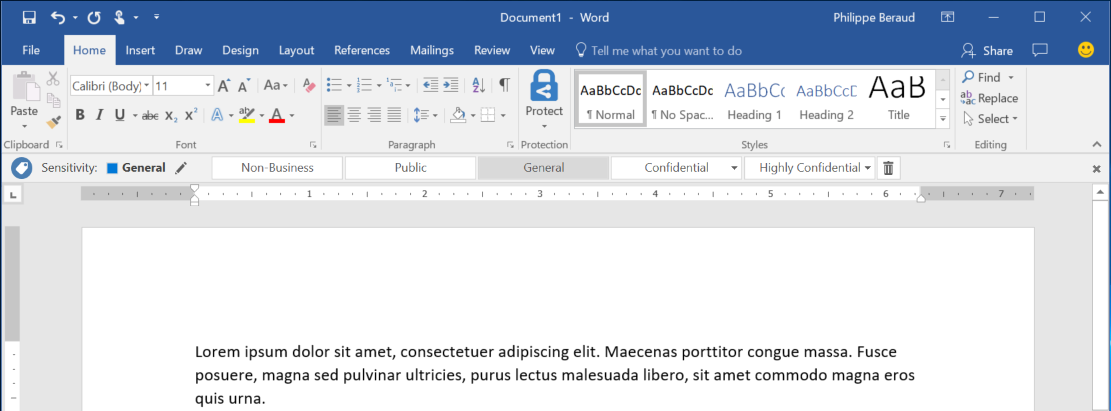


Figure 7 Policy download from Azure Information Protection

This simply implies to make that happens a prior deployment - on each Windows device - of the new Azure Information Protection client for Windows[[50]](#footnote-51), i.e. a [free, downloadable client](https://www.microsoft.com/en-us/download/details.aspx?id=53018)[[51]](#footnote-52) for organizations. (No administrative rights are required to install it on Windows devices.)

**Note**  For more information, see article [Download and install the Azure Information Protection client](https://docs.microsoft.com/en-us/information-protection/rms-client/install-client-app)[[52]](#footnote-53).

This client indeed provides among other benefits a plug-in for Office client applications. Information labeling is then available through a new Information Protection bar for Office client applications.



**Note**  To help you get started, see article [Quick start tutorial for Azure Information Protection](https://docs.microsoft.com/rights-management/information-protection/infoprotect-quick-start-tutorial)[[53]](#footnote-54).

All users get the labels and related settings from the global policy. **Sub-labels** can be created for key departments, projects, acquisition plans, etc. to accommodate “special” needs and requirements in how sensitive information assets are handled in those related specific entities and/or contexts. For instance, you create a sub-label “Finance” for “Highly confidential”, it’s trivial for someone to classify a content as “Highly Confidential \ Finance”.

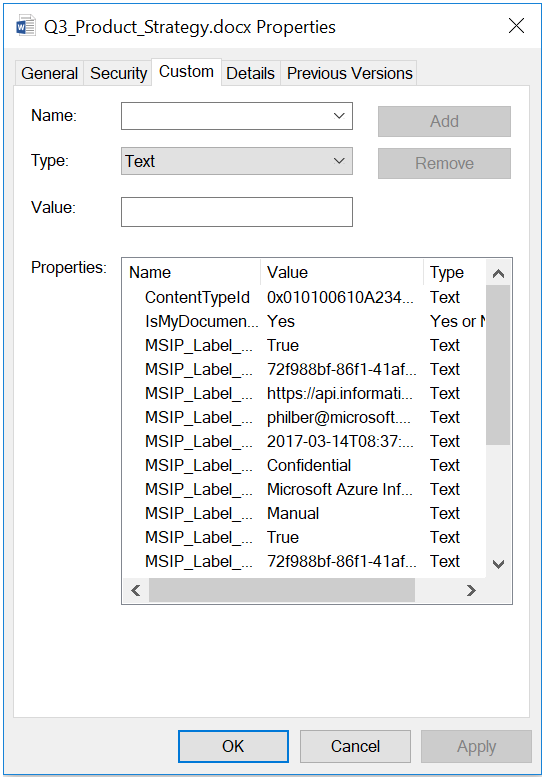
**Important note** When you use sub-labels, you ought to not configure visual markings, protection, and conditions at the primary label. When you use sub-levels, you should configure these setting on the sub-label only. If you configure these settings on the primary label and its sub-label, the settings at the sub-label take precedence.

Moreover, if you want to supplement these for specific roles, employees and groups of employees such as teams, business units/departments or projects, by not only having different labels and settings, offering specialty behaviors, etc. but also by controlling who can see what sub-labels, and to offer specialty behaviors, you must create a **scoped policy** tailored/configured for those specialized users and teams. A scope policy is layered on top of the above global policy and is available only to users that are member of the specified security group.

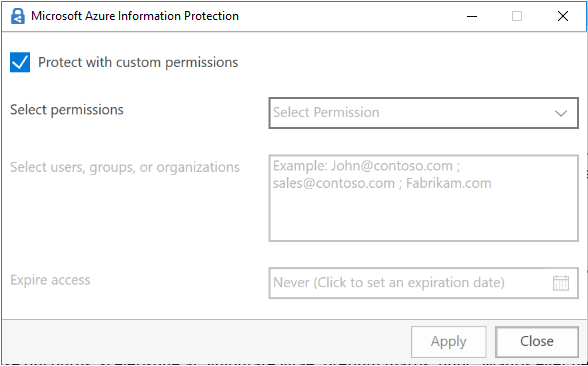
**Note**  For more information, see article [How to configure the Azure Information Protection policy for specific users by using scoped policies](https://docs.microsoft.com/en-us/information-protection/deploy-use/configure-policy-scope)[[54]](#footnote-55).

Information can be classified at the time of creation based on content either automatically or by users as per related classification and protection policy. For user-driven classification, users can select the sensitivity label applicable to the document. A classification can be recommended as per policy as mentioned above. It this cases, users are prompted to make an informed classification decision.

Classification and labeling information are then embedded to the document and defined actions are enforced. As one of the main outcomes, a label is added as multiple meta data entries, i.e. a set of keys and values, to the file (within the files and in the file system). Label’s data entries are in clear text so that other systems such as a DLP engine can read them, and take actions accordingly.



if the applied label is associated with a Rights Management template in the classification and protection policy to enforce, the document is also protected (encryption, authentication, and usage rights). See next section. Custom protection is available through the **Protect** menu of the Office client applications.



File labeling and visual content markings are persistent – same is true for protection if any -, traveling with the data throughout its lifecycle, so that it’s detectable and controlled at all times – regardless of where it’s stored or with whom it’s shared – internally or externally.

Eventually, based on policies and the related rules, one should outline that users can be empowered to override a classification and optionally be required to provide a justification. All the related users activity is stored into the Windows event logs.

The local event logs can in turn be forwarded via an event forwarder to a centralized Security Information and Event Management (SIEM) system: only the required events would be forwarded to feed and customize a central report.

**Note**  For more information, see article [Azure Information Protection client files and client usage logging](https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-files-and-logging)[[55]](#footnote-56).

Considering the above, and after information assets are classified and labelled, finding and implementing ways to protect sensitive information becomes an integral part of any Information Protection (IP) system deployment strategy.

**Protecting sensitive information requires additional attention to how information is stored and transmitted in conventional architectures as well as in the cloud**. This is the purpose of the next sections.

## Protecting information with Rights Management

Beyond allowing to classify (automatically based on preset rules) emails and documents, such as financial reports, product specifications, customer data, etc., add markers to content like custom headers, footers, and watermarks, Azure Information Protection also provides organizations with the ability to protect these information assets using encryption (, authentication and use rights) if a Rights Management template has been specified in the classification and protection polici(es) for the label(s) applied to adequately classify the information assets.

These capabilities as a whole enable organizations to have a greater end-to-end control over their sensitive and proprietary information assets. In this context, Azure Information Protection plays an important role in securing organization's information assets.

**Note**  For more information, see article [The role of Azure Information Protection in securing data](https://docs.microsoft.com/en-us/enterprise-mobility-security/solutions/azure-information-protection-securing-data)[[56]](#footnote-57).

### Azure RMS

**Azure Rights Management (Azure RMS) is the protection technology used by Azure Information Protection for encryption to safeguard organization's sensitive and proprietary information** such as confidential emails and documents with Rights Management.

**Note**  For more information, see article [What is Azure Rights Management?](https://docs.microsoft.com/en-us/information-protection/understand-explore/what-is-azure-rms)[[57]](#footnote-58).

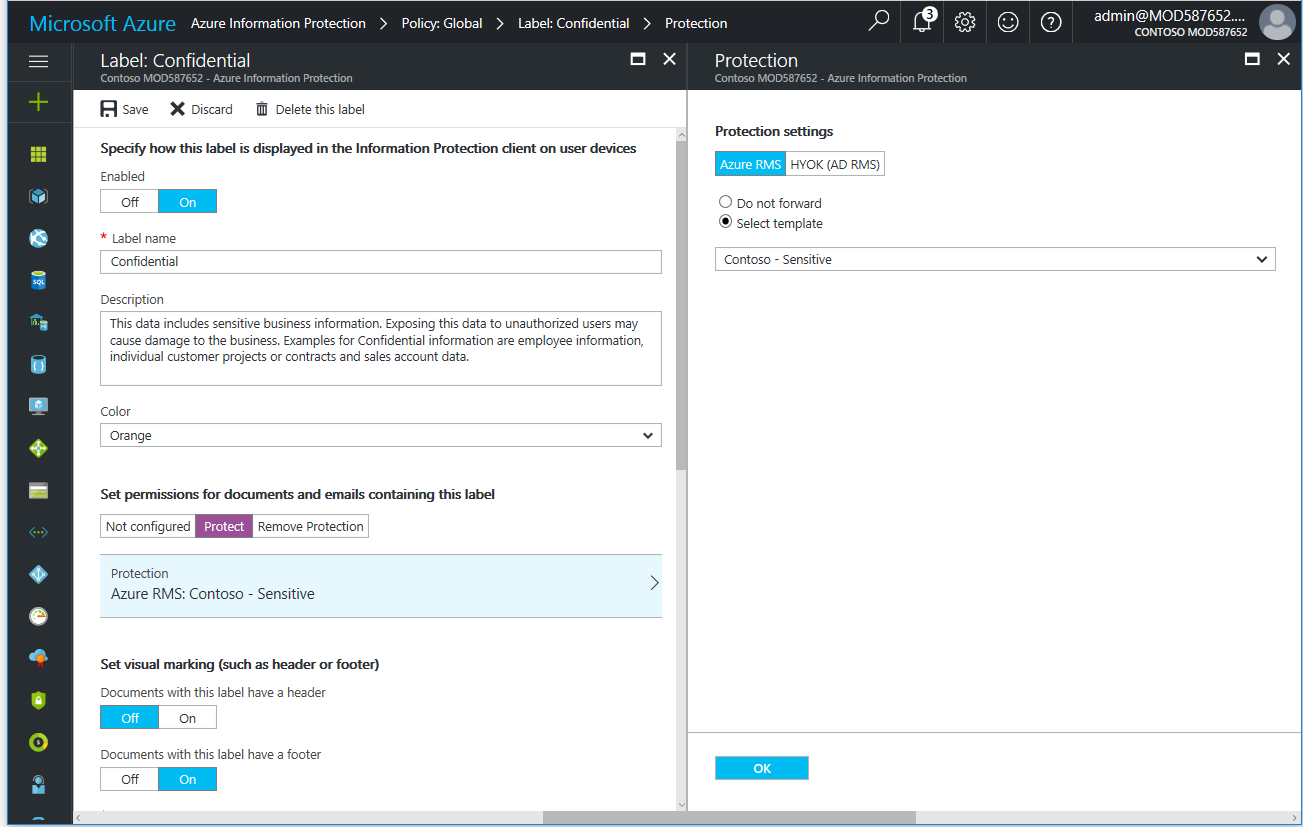
Azure RMS provides protection above what is available for documents in the typical IT environment – usually perimeter protection only – by protecting the emails or files themselves. This protection results in the file or the mail being encrypted and having persistent usage rights policies enforced by the RMS-enlightened applications and (online) services (see below). **The protection travels with data: data is protected at all times, regardless of where it’s stored or with whom it’s shared.**

The Azure RMS protection technology indeed allows to:

* Use RSA 2048-bit keys for public key cryptography and SHA-256 for signing operations.
* Encrypt the emails and files to a specific set of recipients both inside and outside their organization.
* Apply specific set of usage rights to restrict the usability of the email or file as defined by the Rights Management template being applied (in accordance to the label set as).
* Decrypt content based on the user’s identity and authorization in the usage rights policy resulting from the enforcement of the specified Rights Management template.

**Note** When protecting sensitive information with the help of Azure RMS, it benefits from multiple layers of security and governance technologies, operational practices, and compliance policies that combine to enforce data privacy and integrity at a very granular level. For more information, see white paper [Azure RMS Security Evaluation Guide](https://thermsguy.blob.core.windows.net/slides/AzureRMS-SecurityEvaluation-v1.docx)[[58]](#footnote-59). This paper describes security capabilities in Azure RMS, including mechanisms for encryption, management, and access control that you can leverage for managing and protecting your sensitive data.

Manually or automatically applying a label with Azure Information Protection results in eventually seamlessly applying a Rights Management template that defines a usage rights policy. As already outlined, the Azure Information Protection classification and protection policies indeed allows to optionally specify for each label a Rights Management template to apply thus allowing **standard default usage rights policies for many common business scenarios**.



The usage rights policies vary from ‘View only’ to full rights including the ability to unprotect content - usage rights can be accompanied by conditions, such as when those rights expire.

**Note** Azure RMS allows to create simple and flexible usage policies to enforce protection - customized rights policy templates provide a quick and easy solution for you to apply policies in accordance to its classification, and thus to apply the correct level of protection and restrict access to the intended people.

The Azure RMS protection technology is designed to protect all file types (whatever they are and not only Office or PDF files) and protect files anywhere - when a file is saved to a location, the protection stays with the file, even if it is copied to storage that is not under the control of IT, such as a cloud storage service.

The usage policies remain with the protected information, no matter where it goes, even in transport, rather than the rights merely residing on an organization’s corporate network. Moreover, all commonly used devices, not just Windows computers are supported.

Azure RMS enables to audit and monitor usage of the protected files, even after these files leave your organization’s boundaries. Thanks to the authentication part of the Rights Management protection, protected documents (that have been protected via the Azure Information Protection client and/or Office client applications) can be tracked to show who has opened the protected document and from which geographical location (IP address-based). So, users, i.e. the authors/owners of the data, as well as IT professionals have access to detailed tracking and reporting information in the document tracking portal[[59]](#footnote-60) to see what’s happening with shared information to gain more control over it, and to eventually revoke access to it in the event of unexpected activities.

**Note**  For more information, see articles [Configuring and using document tracking for Azure Information Protection](https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-document-tracking)[[60]](#footnote-61) and [Track and revoke your documents when you use Azure Information Protection](https://docs.microsoft.com/en-us/information-protection/rms-client/client-track-revoke)[[61]](#footnote-62).

**Important note**  You should be aware of the potential legal and political implications with document tracking. As described in the first of the above articles, tracking can disable if needed.

Azure RMS is built to work in conjunction with RMS-enlightened applications, i.e. applications that are enhanced to consume or publish Rights Management protected content. Azure RMS has tight integration with Microsoft Office applications and services, and extends support for other applications by using the new Azure Information Protection client for Windows (that now replaces the previously available RMS sharing application).

**Note**  The existing RMS sharing application is still available on the Microsoft download center[[62]](#footnote-63) and will be supported for a period of 12 months with support ending January 31, 2018.

The Rights Management Software Development Kits (SDK) provides your internal developers and software vendors with APIs to write custom RMS-enlightened applications that support Azure RMS such as [Foxit PDF Security Suite](http://www.foxitsoftware.com/landingpage/rms.php)[[63]](#footnote-64), [SECUDE](http://secude.com/)[[64]](#footnote-65), etc. The Rights Management SDKs provided with Azure Information Protection for protection purposes are available on the most important devices and platforms: Windows and Mac OS/X computers, Windows Phone devices, iOS, and Android devices. Applications in other environments can interact with Azure RMS by utilizing the service’s RESTful APIs directly. (see section § Integrating with other systems.)

Note For more information, see articles [How applications support the Azure Rights Management service](https://docs.microsoft.com/en-us/information-protection/understand-explore/applications-support)[[65]](#footnote-66) and [Applications that support Azure Rights Management data protection](https://docs.microsoft.com/en-us/information-protection/get-started/requirements-applications)[[66]](#footnote-67).

As a highly scalable and available cloud-based solution run by Microsoft in data centers strategically located around the world, and that can be subscribed and activated in a couple of minutes, Azure RMS enables to:

* Benefit from rights management for those organizations who choose to subscribe to an Office 365 Enterprise plan: Exchange Online, SharePoint Online, OneDrive for Business, and Microsoft Office 365 ProPlus.
* Support for existing on-premises Exchange Server and SharePoint Server deployments, and Windows Server running File Classification Infrastructure (FCI) in addition to working seamlessly with Office 365 with the use of the RMS connector or the new cmdlets installed by the Azure Information Protection client (see section § Leveraging an on-premises File Classification Infrastructure (FCI) for existing data later in this document).



Figure 8 Azure RMS integration with Office workloads on-premises and in the cloud

Azure RMS also enables organizations of any size to minimize the effort required to implement an effective protection to sustain the collaboration inside and outside of the organization’s boundaries. For business-to-business (B2B) collaboration, there is no need to explicitly configure trusts with other organizations (or individuals) before you can share protected content with them. You fully benefit from Azure Active Directory (Azure AD) as a “trust fabric” that saves organizations a lot of time, work and complexity.  Essentially, in terms of directory federations, organizations federate once with Azure AD which then operates as a claims broker between them and all their external partners who have also federated with Azure AD. (Azure AD is the directory behind Microsoft Online Services subscriptions like Office 365, Dynamics 365, Intune, etc.). You can run federation service (ADFS, etc.) on-premises for authentication or let Azure AD do the work.



Figure 9 Azure AD as a “trust fabric”

Consequently, secure collaboration is enhanced and extended to organizations and individuals using computers and mobile devices.

Moreover, Azure RMS protects your organization's documents and emails by using a tenant key that protects the documents and emails. All service artifacts (per-document/mail encryption keys) are cryptographically chained to that cryptographic key.

**Note** Azure RMS uses industry-standard cryptography in FIPS 140-2 validated modules. Azure RMS services are certified for ISO/IEC 27001:2005, has SOC 2 SSAE 16/ISAE 3402 attestations, and compliant with EU Model Clause. For more information about these external certifications, see [Microsoft Trust Center](https://www.microsoft.com/en-us/TrustCenter/default.aspx)[[67]](#footnote-68).

The tenant key can be managed by Microsoft (the default), or managed by you with the “Bring Your Own Key” (BYOK) solution by storing your tenant key in Hardware Security Modules (HSMs) in Azure Key Vault (AKV)[[68]](#footnote-69), thus allowing you to maintain control of your tenant key, and therefore your protected information assets, including ensuring that Microsoft cannot see or extract your tenant key.

**Note** For more information, see article [Planning and implementing your Azure Information Protection tenant key](https://docs.microsoft.com/en-us/information-protection/plan-design/plan-implement-tenant-key)[[69]](#footnote-70) and white paper [Bring Your Own Key (BYOK) with Azure Key Vault for Office 365 and Azure](http://download.microsoft.com/download/F/6/3/F63C9623-053F-44DD-BFA8-C11FA9EA4B61/Bring-Your-Own-Key-with-Azure-Key-Vault-for-Office-365-and-Azure.docx)[[70]](#footnote-71).

Azure RMS supports auditing and rich logging capabilities that you, and IT professionals can use to analyze for business insights, monitor for abuse, (if you have an information leak) perform forensic analysis, and reason over the data for compliance and regulatory purposes.

**Note**  For more information, see article [Logging and analyzing usage of the Azure Rights Management service](https://docs.microsoft.com/en-us/information-protection/deploy-use/log-analyze-usage)[[71]](#footnote-72) and/or whitepaper [Get Usage Logs from Azure RMS](http://download.microsoft.com/download/F/6/3/F63C9623-053F-44DD-BFA8-C11FA9EA4B61/Get-Usage-Logs-from-Azure-RMS.docx)[[72]](#footnote-73).

Likewise, and in addition with BYOK, you can configure the Azure Key Vault service to monitor when and how your key vault assigned to the Azure Information Protection/Azure RMS service is accessed and by whom.

**Note** When Key Vault logging is activated, logs will be stored in an Azure storage account that you provide. To turn on logging and get these logs, follow the instructions outlined in the article [Azure Key Vault Logging](https://docs.microsoft.com/en-us/azure/key-vault/key-vault-logging)[[73]](#footnote-74). The logs you receive from the Azure Key Vault will contain every transaction performed with your tenant key.

**Regardless of the way the tenant key is managed, the information that you protect with Azure RMS (and Azure Information Protection) is never sent to the cloud; the protected documents and emails are not stored in Azure unless you explicitly store them there or use another cloud service that stores them in Azure.**

### Hold-Your-Own-Key (HYOK)

The Hold-Your-Own-Key (HYOK) capability of Azure Information Protection also allows - with some restrictions to be aware of – to use [Active Directory Right Management Services (AD RMS)](https://technet.microsoft.com/en-us/library/74272acc-0f2d-4dc2-876f-15b156a0b4e0.aspx)[[74]](#footnote-75) to protect highly confidential information when you do not want to have the organization’s root key stored in the cloud even in a hardware security module (HSM) as it is with Azure RMS.

Note For more information about the HYOK capability and its restrictions, see blog post [Azure Information Protection with HYOK (Hold Your Own Key)](https://blogs.technet.microsoft.com/enterprisemobility/2016/08/10/azure-information-protection-with-hyok-hold-your-own-key/)[[75]](#footnote-76) and article [Hold your own key (HYOK) requirements and restrictions for AD RMS protection](https://docs.microsoft.com/en-us/information-protection/deploy-use/configure-adrms-restrictions)[[76]](#footnote-77).

First shipped in Windows Server 2003 timeframe, AD RMS is, in the latest releases of Windows Server, a server role initially designed for organizations that need to safeguard sensitive information and apply protection with the data. If Azure RMS is the protection technology for Azure Information Protection, for Office 365 services and B2B collaboration scenarios thanks to the Azure AD as a “trust fabric”, AD RMS provides on-premises Rights Management protection.

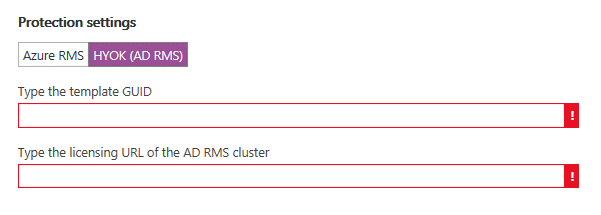
**Note** For more information, see article [Active Directory Rights Management Services Overview](http://go.microsoft.com/fwlink/?LinkId=84726)[[77]](#footnote-78).



Figure 10 AD RMS and Office workloads on-premises

**This protection technology can be used with Azure Information Protection and might be suitable for a very small percentage of documents and emails that must be protected by an on-premises key.**

In lieu of defining an Azure RMS-based Rights Management template for the considered label in the Azure Information Protection classification and protection policy, you will have to specify both the (GUID of) the AD RMS Rights Management template to use in this context and the (licensing) URL where you can reach on the internal corporate network the on-premises AD RMS infrastructure and. As implicitly outlined, this requires having a line of sight on the on-premises AD RMS infrastructure.



**Note** For more information, see article [How to configure a label for Rights Management protection](https://docs.microsoft.com/en-us/information-protection/deploy-use/configure-policy-protection)[[78]](#footnote-79).

In terms of user experience (UX), users aren’t aware when a label uses AD RMS protection rather than Azure RMS protection. So, the aforementioned scoped policies of Azure Information Protection represent a good way to ensure that only the specific set(s) of users who need to use the HYOK capability see labels that are configured for AD RMS protection.

Protected documents or emails can be shared only with partners that you have defined with explicit point-to-point trusts: you do not benefit from the Azure AD as a “trust fabric”.

**Note** AD RMS on-premises requires that trusts must be explicitly defined in a direct point-to-point relationship between two organizations by using either trusted user domains (TUDs) or federated trusts that you create by using Active Directory Federation Services (AD FS) server role. For more information, see articles [Trusted User Domain](https://technet.microsoft.com/en-us/library/dd983944(WS.10).aspx)[[79]](#footnote-80) and [Deploying Active Directory Rights Management Services with Active Directory Federation Services](https://technet.microsoft.com/en-us/library/dn758110(v=ws.11).aspx)[[80]](#footnote-81).



Figure 11 HYOK capability to protect highly confidential information

Considering the above, and as guidance, the HYOK capability should be used only for documents or emails that match all the following criteria:

* The content has the highest classification in your organization and access is restricted to just a few people that “need to know”.
* The content will never be shared outside the organization.
* The content will only be consumed on the internal corporate network.
* The content does not need to be consumed on Mac OS computers or mobile devices.

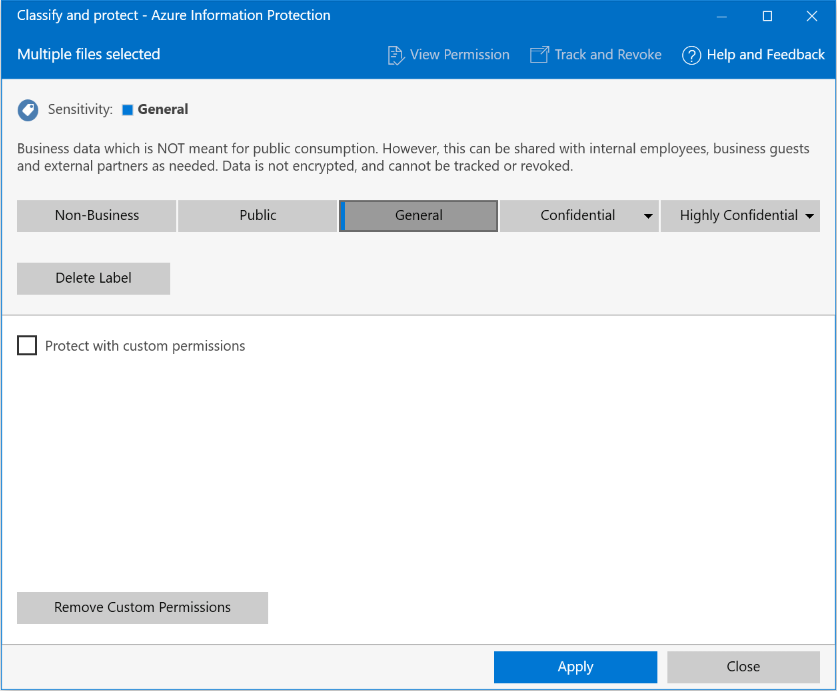
## Scanning for on-premises existing information

In today’s working environments, data centers, and corporate networks, the need for high levels of data storage keeps growing exponentially. The huge number of files combined with the increased regulations and data leakage is a cause of increasing risk concerns for organizations. These files are very frequently kept in storage because of a lack of control and proper classification of their data. Organizations need to get insight into their files to help them manage their data more effectively, and mitigate risks.

Thus, file classification is a key part of performing the organization of the stored files.

### Using the Azure Information client wizard

The aforementioned Azure Information Protection client for Windows provides among other benefits a wizard for information labeling and protection through the Windows File Explorer to classify and protect multiple files. So, information can be classified based on context and source by users. (You can also apply customs permissions to files if preferred or required.)



File protection (encryption, authentication, and use rights) is based on Azure RMS as previously discussed. As such, the wizard supports generic protection as well as native protection, which means that file types other than Office documents can be protected. You can notably apply any label to PDF files (also labels that do not apply protection).

**Note** For more information, see article [File types supported by the Azure Information Protection client](https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-file-types)[[81]](#footnote-82) from the Azure Information Protection client admin guide.

Custom protection can be defined to sustain collaboration with the ability to notably protect a file for:

* A group of users at an organization, e.g. finance@contosos.com
* Any user at a specified organization, e.g. [anyuser]@contoso.com)

The former enables **group collaboration** where, as an illustration, two organizations can collaborate effectively with each other without having to know precisely who is in the group, for example legal teams needing to work on briefs, project teams working on a joint effort. By simply being a member of the group, adequate permissions are granted to users. This requires that the group exists in Azure AD (either as a native group or a group that have been sync thanks to Azure AD Connect as illustrated in Figure 8). Azure AD as a “trust fabric” will do the rest so that it just works.

Likewise, the latter enables **organization collaboration** with content to be protected to all users within a specified organization, for example any user who works at Contoso. Such a feature is of particular interest for business-to-business (B2B) scenarios.

Unlike the group collaboration that requires no additional configuration beyond being member of an Azure AD group, the organization collaboration must be enabled by IT professionals using cmdlets of the updated Azure RMS PowerShell module[[82]](#footnote-83).

Following is an example on how to create a suitable Rights Management definition (with New-AadrmRightsDefinition) and template (with Add-AadrmTemplate) for the collaboration between the Contoso and Fabrikam organizations:

$names = @{}

$names[1033] = "Contoso-Fabrikam Confidential"

$descriptions = @{}

$descriptions[1033] = "This content is confidential for all employees in Contoso and Fabrikam organization"

$r1 = New-AadrmRightsDefinition -DomainName contoso.com -Rights "VIEW","EXPORT"

$r2 = New-AadrmRightsDefinition -DomainName fabrikam.com -Rights "VIEW", "EXPORT"

Add-AadrmTemplate -Names $names -Descriptions $Descriptions -LicenseValidityDuration 5 -RightsDefinitions $r1, $r2 -Status Published

**Note** For more information, see blog post [Azure Information Protection December update moves to general availability](https://blogs.technet.microsoft.com/enterprisemobility/2017/02/08/azure-information-protection-december-update-moves-to-general-availability/)[[83]](#footnote-84) and article [Administering the Azure Rights Management service by using Windows PowerShell](https://docs.microsoft.com/en-us/information-protection/deploy-use/administer-powershell)[[84]](#footnote-85).

### Using PowerShell for a file-location based solution

The Azure Information Protection client for Windows also provides a set of PowerShell cmdlets so that you can manage the client by running commands that you can put into scripts for automation for labeling and protection of files in folders on file servers and network shares that are accessible through SMB/CIFS, e.g. \\server\finance.

So, information can be classified based on context and source automatically.

This PowerShell module[[85]](#footnote-86) replaces the RMS Protection Tool and the AD RMS Bulk Protection Tool. It notably includes all the Rights Management cmdlets from the RMS Protection Tool. It also includes two new cmdlets that use Azure Information Protection for tagging/labeling, i.e. Set-AIPFileLabel and Get-AIPFileStatus.

The former cmdlet allows to set or remove an Azure Information Protection label for a file. Protection is automatically applied or removed when labels are configured for Rights Management protection in the Azure Information Protection policy. When the command runs successfully, any existing label or protection is replaced.

The latter gets the Azure Information Protection label and protection information for a specified file or files in folders and network shares.

**Note** For more information, see article [Using PowerShell with the Azure Information Protection client](https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-powershell)[[86]](#footnote-87).

Moreover, these cmdlets support generic protection as well as native protection, which means that file types other than Office documents can be protected, including PDF files.

**Note** For more information, see article [File types supported by the Azure Information Protection client](https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-file-types)[[87]](#footnote-88) from the Azure Information Protection client admin guide.

These cmdlets help to implement a file-location based solution by letting you automatically protect all files in a folder on a file server running Windows Server. However, it may be required to (automatically) classify files meet a specific criterion in terms of format and content as previously covered.

### Leveraging an on-premises File Classification Infrastructure (FCI) for existing data

The Windows Server File Classification Infrastructure (FCI) technology introduces an extensible built-in solution for file classification and management allowing IT professionals to classify file and apply policy based on classification, thus helping ensure that your information assets as files are identifiable and secure, e.g. a key requirement of the GDPR – regardless of where they’re stored and how they’re shared.

**Note** FCI is controlled and exposed through the File Server Resource Manager (FSRM). FSRM is a feature of the File Services role in Windows Server 2008 R2 SP1 and above. It can be installed as part of the File Services role, using Server Manager. For more information on the FCI technology, see white paper [File Classification Infrastructure](http://download.microsoft.com/download/D/D/3/DD3FB42C-D3C7-47C0-9431-40D80256FB0A/FCI_TDM_WP_May_09.pdf)[[88]](#footnote-89), [videos on Channel 9](http://channel9.msdn.com/tags/FCI/)[[89]](#footnote-90) and of course the [Server Storage at Microsoft Team Blog’s](https://blogs.technet.microsoft.com/filecab/tag/file-classification-infrastructure-fci/)[[90]](#footnote-91) post’s on FCI.

The infrastructure can be leveraged as a centerpiece for new content classification, file tagging/labeling, and by solutions and products spanning compliance, data loss prevention, backup and archival, etc.

**Note** For an illustration on how Microsoft IT has leveraged the FCI technology internally to create a solution to automatically classify, manage, and protect sensitive data, including personally identifiable information and financial information, see technical case study [Microsoft IT Uses File Classification Infrastructure to Help Secure Personally Identifiable Information](https://www.microsoft.com/itshowcase/Article/Content/278/Microsoft-IT-Uses-File-Classification-Infrastructure-to-Help-Secure-Personally-Identifiable-Information)[[91]](#footnote-92).

FCI’s out-of-the-box functionality includes the ability to:

* Define classification properties, which correspond to or are aligned with the various classification levels of the defined classification taxonomy,
* Automatically classify files based on folder’s location and content - by scanning content automatically at the file level for key words, terms and patterns -,
* Apply file management tasks such as applying right management protection, file expiration and custom commands based on classification,
* Finally produce reports that show the distribution of a classification level (a.k.a. property) on a file server.

FCI provides by default the following ontology and taxonomy to label and thus classify the files. With a closer look at the **Confidentiality**, **Personally Identifiable Information**, **Compliancy**, and **Impact** classification properties below, one can easily do the link with the previously suggested **Sensitivity**, **Privacy**, **Compliance** and **Business Impact** categories and see how to turn organization’s classification categories into technology.

Table 3 FCI ontology

| **Areas** | **Properties/Categories** | **Values** |
| --- | --- | --- |
| Information Security | Confidentiality | High; Moderate; Low |
|  | Required Clearance | Restricted; Internal Use; Public |
| Information Privacy | Personally Identifiable Information | High; Moderate; Low; Public; Not PII |
|  | Protected Health Information | High; Moderate; Low |
| Legal | Compliancy | SOX; PCI; HIPAA/HITECH; NIST SP 800-53; NIST SP 800-122; EU-U.S. Safe Harbor Framework; GLBA; ITAR; PIPEDA; EU Data Protection Directive; Japanese Personal Information Privacy Act |
|  | Discoverability | Privileged; Hold |
|  | Immutable | Yes/No |
|  | Intellectual Property | Copyright; Trade Secret; Parent Application Document; Patent Supporting Document |
| Organizational | Impact | High; Moderate; Low |
|  | Department | Engineering; Legal; Human Resources… |
|  | Project | <Project> |
|  | Personal Use | Yes/No |
| Record Management | Retention | Long-term; Mid-term; Short-term; Indefinite |
|  | Retention Start Date | <Date Value> |

The above properties enable to establish the baseline configuration for the classification with FCI. They of course have to be adapted to reflect the adopted taxonomy scheme in the organization.

Once the baseline configuration is ready, classification with FCI can be conducted. Classification includes:

* **Manual classification**. An end user can manually classify a file using the file properties interface built into the Microsoft Office system file, and the FCI technology will recognize these properties.
* **Automatic classification**. Using automatic classification rules, the FCI technology can classify files according to the folder in which the file is located or based on the contents of the file thanks to regular expressions.
* **Application-based and IT scripts**. Using an API, applications and IT scripts can set a classification level to files.

The FCI pipeline is extensible by third-party plug-ins at the point where files are classified (classifier plug-in) and the point where properties get read/stored for files (property storage module plug-in).

* **File classification (Classifier plug-in)**. FCI provides a set of classification mechanisms by default. When more custom logic is required to classify files, a classifier plug-in can be created. These plug-ins represent a primary developer extensibility points for file classification. Products and technologies that classify data can then hook into the FCI automatic classification rules by providing a classification plug-in.

**Note The** PowerShell **host classifier, IFilter based classifier, text-based classifier, managed content classifier are some of the examples provided by the Microsoft SDK. For more information, see article** [How to create a custom file classification mechanism](http://msdn.microsoft.com/en-us/library/jj552967(v=vs.85).aspx)[[92]](#footnote-93)**.**

* **File property storage (Property storage module plug-in)**. A few modules are provided by the system, including a plug-in to store properties in Office files. The method for storing and reading classification properties is also extensible, so that, if properties need to be stored within other file types, a custom storage plug-in can be created. For example, a third-party video file plug-in can provide a way to store and extract properties from a video file. In another example, classification properties can be stored in a database or in the cloud. Property storage modules are queried when FCI reads properties for a file as well as when it writes properties for a file. Each plug-in can indicate which file extensions it can support.

**Note For more information, see article** [How to create a custom storage plug-in](http://msdn.microsoft.com/en-us/library/jj552962(v=vs.85).aspx)[[93]](#footnote-94)**.**

In addition, the custom file management tasks can be extended by applications and custom scripts. They run on a scheduled basis and invoke a custom script or application based on a condition. Custom scripts can be provided by IT professionals to apply automatic data management to files based on their classification.

**Note For more information, see article** [How to configure a file management task](http://msdn.microsoft.com/en-us/library/jj552961(v=vs.85).aspx)[[94]](#footnote-95)**.**

#### Introducing the Data Classification Toolkit

The Data Classification Toolkit[[95]](#footnote-96) is a Solution Accelerator designed to help enable an organization to identify, classify, and protect data on their file servers based on the FCI technology, and to easily configure default central access policy across multiple servers. The out-of-the-box classification and rule examples help organizations to build and deploy their policies to protect critical information on the file servers in their environment.

The toolkit provides support for configuring data compliance on file server deployments for Windows Server 2012 R2, as well as for mixed deployments of Windows Server 2012 R2, Windows Server 2012, and Windows Server 2008 R2 SP1, to help automate the file classification process, and thus making file management more efficient in your organization.

In addition to configuring the file classification infrastructure, it allows to provision and standardize central access policy across the file servers in a forest. The toolkit enhances the user experience for IT professionals through a scenario based user interface that enables to configure classification information and apply default access policies on your file servers. For that purpose, the toolkit adds a user interface to the existing Windows PowerShell experience, including a Classification Wizard that you can use to manage file classifications.

**Note** The toolkit also provides tools to provision user and device claim values based on Active Directory Domain Services (AD DS) resources to help simplify configuring Dynamic Access Control (DAC) (see next section hereafter).

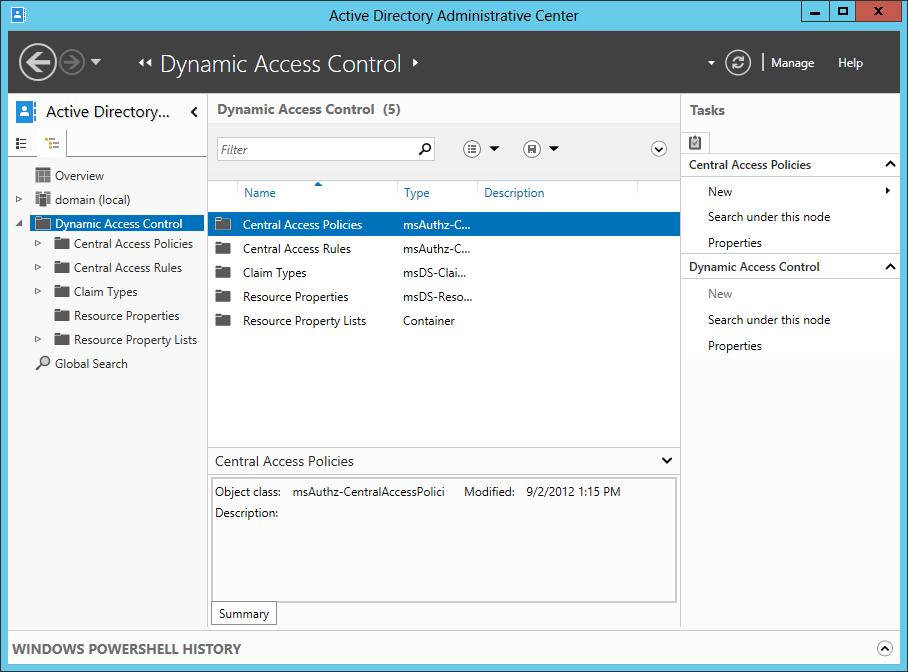
#### Leveraging Dynamic Access Control within FCI

The FCI technology in Windows Server allows classifying files by assigning metadata labels manually or thanks to automated rules. In essence, FCI helps in identifying and tagging/labeling certain type of information.

We all agree that simply classifying data intrinsically offers limited benefits without defining the related handling standard and its security controls for protecting the information assets at the right level. In other word, the protection level along with the associated security controls are generally both specified and applied in accordance of the asset’s classification level.

Choice of protection with FCI includes and is not limited to Dynamic Access Control (DAC) or RMS protection (see section § Protecting information with Rights Management).

First introduced with Windows Server 2012, DAC is a technology that enables to configure who has access to which file resources based on claim values after a successful authentication (claims-based authentication). Interestingly enough, the outcomes of the classification with FCI can be leveraged with DAC to express access rules and thus applying the right protection level taking into account the information access context.



As such, you can define claims-based access rules that not only leverage the information classification results as claims but also might include in a non-exhaustive manner claims that convey additional information about:

* The user status such as full-time employee (FTE)/civil servant, intern or contractor,
* The user’s department,
* The strength of authentication,
* The type of devices being used such as managed or personal,
* The type of network corporate network vs. Internet,
* Etc.

Thus, the file server can have an associated access policy that applies to any HBI file (File.Impact = High) and grants a read and write access to the considered file if the user belongs to the same department at the file’s department (User.Department == File.Department) and their device is ‘managed’ (Device.Managed == True):

*Applies to: @File.Impact = High*

*Allow | Read, Write | if (@User.Department == @File.Department) AND (@Device.Managed == True)*

In other words, an organization can enforce an access rule saying that access to finance’s HBI data is strictly restricted to users from the finance department using an IT managed devices.

A Claims Wizard helps to manage central access policy on the file servers in your organization.

The first key benefit of DAC is that it extends traditional access control mechanisms based on security group to claim-based access control. A claim can be any trusted attribute stored on Active Directory and not only membership of security group. Such an approach enables new scenarios by granting access based on different attributes in AD DS, such as a user's department, manager, location, role or title, as well as how files are classified. The aforementioned Claims Wizard can be used to build claim values in AD DS.

An organization doesn't need to upgrade all of its file servers to Windows Server 2012 or above like Windows Server 2016 for leveraging DAC: one domain controller running Windows Server 2012 or above is at least required.

**Note** There is NO specific requirement for the forest functional level.

**Note** If you want to leverage device claims, domain-joined Windows 8 client and above are required.

DAC constitutes a new way to enable access to sensitive information assets primary stored on-premises while enforcing the suitable security controls on who can access them, particularly from a risk perspective and in BYOD context.

**Note** For information on DAC, see article [Dynamic Access Control: Scenario Overview](https://technet.microsoft.com/windows-server-docs/identity/solution-guides/dynamic-access-control--scenario-overview)[[96]](#footnote-97).

#### Leveraging Azure Information Protection within FCI

By leveraging Azure Information Protection, FCI can automatically apply adequate right management protection to the files to prevent them from leaking beyond their intended audience. This feature is particularly important when information is flowing into hybrid environment where files can be stored in many places such as USB drives, mobiles devices or cloud services.

Starting with Windows Server 2012, there is a built-in file management task (**Protect files with RMS**) that can apply information protection for sensitive documents after a file is identified as being sensitive. This task allows to run custom command like a PowerShell script where the PowerShell cmdlets of the Azure Information Protection client for Windows can be used.

As already outlined, this PowerShell module replaces the RMS Protection Tool and the AD RMS Bulk Protection Tool. It notably includes all the Rights Management cmdlets from the RMS Protection Tool.

Moreover, and as already outlined, these cmdlets support generic protection as well as native protection, which means that file types other than Office documents can be protected.

**Note** For more information, see article [File types supported by the Azure Information Protection client](https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-file-types)[[97]](#footnote-98) from the Azure Information Protection client admin guide.

For that reason, such an approach should be preferred the one that leverages the RMS connector for FCI[[98]](#footnote-99) to protect files with Azure RMS.

**Note** The ability to automatically encrypt files based on their classification is indeed done through the Rights Management SDK 2.1, more specifically by using the File API. By default, native protection is enabled for Microsoft Office files and encryption is blocked for all other file types. To change this behavior and enable encryption on PDF files and all other files, you must modify the registry as explained in the File API Configuration page[[99]](#footnote-100).

**Note** For more information, see article [RMS protection with Windows Server File Classification Infrastructure (FCI)](https://docs.microsoft.com/en-us/information-protection/rms-client/configure-fci)[[100]](#footnote-101).

This approach allows classify and protect information based on content, context and source automatically.

## Controlling information flows using a Data Loss Prevention (DLP) engine

As outlined in the introduction of this document, leakage or loss of data represents a growing risk and concern for many organizations today – because of regulations, breaches of trust or loss of business-critical information.

In such a context, organizations must ensure that data is properly handled during regular business processes, preventing inappropriate access or sharing. They must consequently organize and classify their content to assign security controls as part of their compliance and security practice. Current classification systems place an enormous responsibility on the user and can interfere with their ability to complete their job.

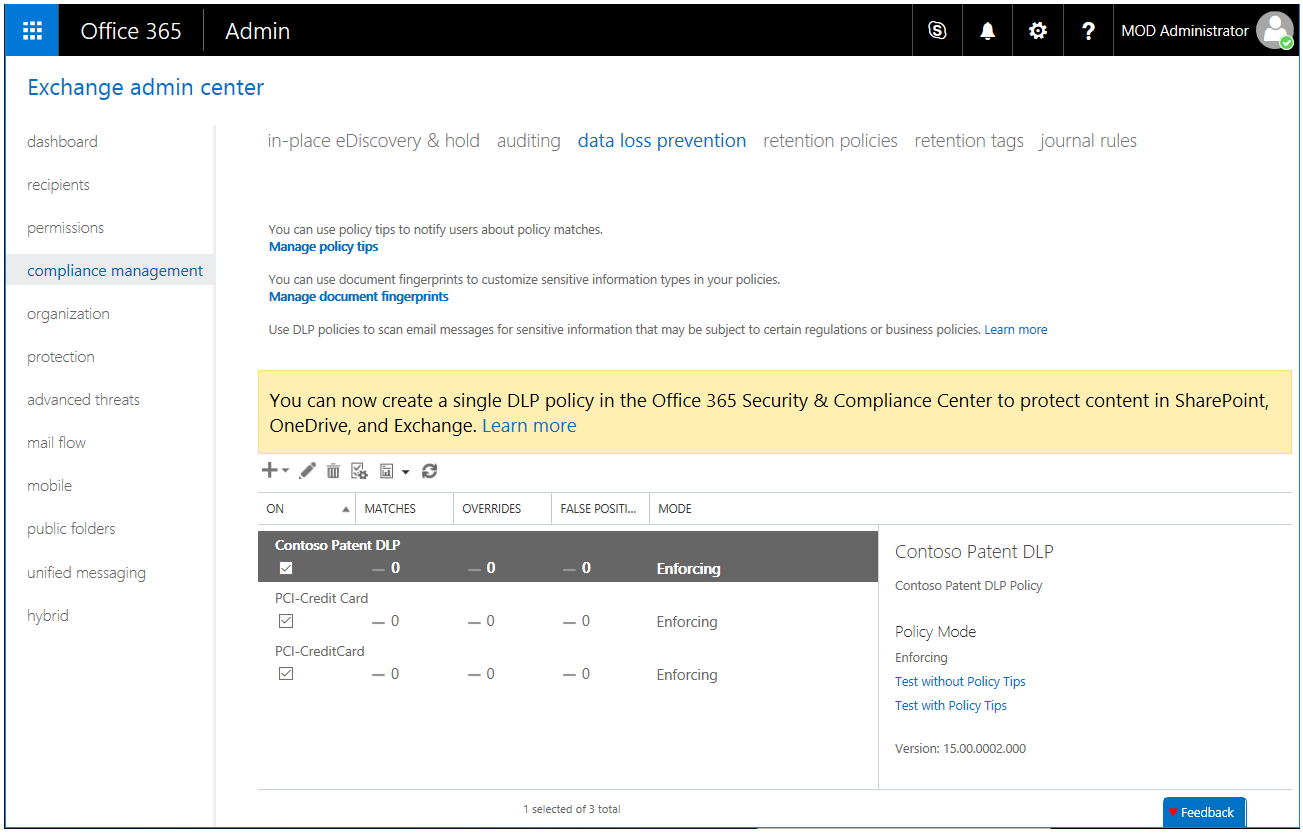
Data loss prevention (DLP) technologies are an important issue for enterprise messaging systems because of the today’s extensive use of email - inside an organization or outside the organization to business partner organizations - for business-critical communication that includes sensitive or confidential data. The same considerations apply to collaboration systems. DLP technologies can help ensure that these systems do not transmit information that has been classified as “Confidential”. They can read the labels and take actions accordingly.

Organizations can take advantage of DLP features in existing products to help prevent data loss. Let’s consider what Exchange Online, Exchange Server, SharePoint Online and OneDrive for Business, SharePoint Server, etc. provide.

### DLP features in Exchange Online and Exchange Server 2013/2016

#### DLP features via Exchange transport rules

In order to help protecting sensitive information, enforcing compliance requirements, and managing its use in email, without hindering the productivity of employees, DLP features in both Exchange Online and Exchange Server 2013/2016 make managing sensitive data easier than ever before with the ability to identify, monitor, and protect sensitive data through deep content analysis.



As such, DLP features provide a range controls that can detect sensitive data in email before it is sent and automatically block, hold, notify the sender or apply usage rights restriction.

DLP policies are simple packages that you create in the Exchange admin center and then activate to filter email messages. You can create a DLP policy, but choose to not activate it. This allows to test your policies without affecting mail flow.

DLP policies use the full power of existing transport rules that provide mail flow control capabilities for IT professionals. The basic goal, when creating a transport rule, is to have Exchange Online and Exchange Server inspect e-mail messages sent to and received by the users and trigger actions against that e-mail message based on conditions.

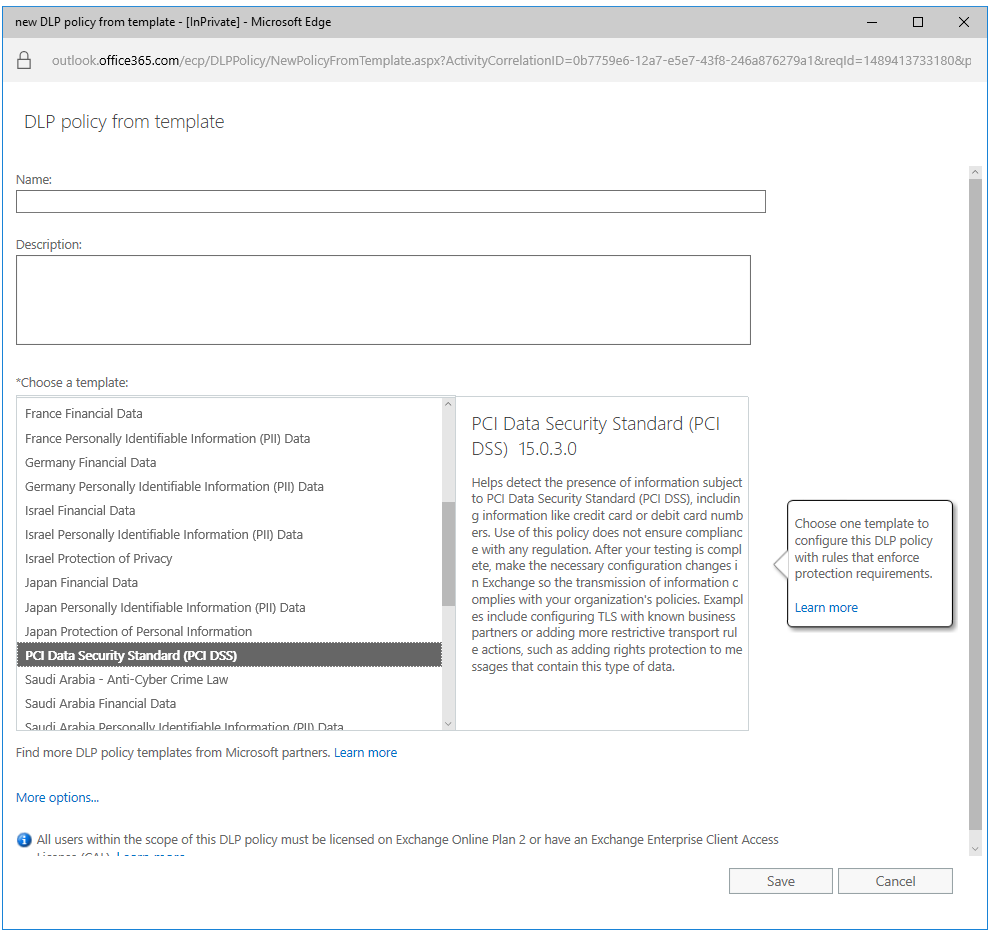
Exchange transport rules use a set of conditions, actions, and exceptions:

* **Conditions** identify specific criteria within an e-mail message and determine why the rule is triggered.
* **Actions** are applied to e-mail messages that match these conditions.
* **Exceptions** identify e-mail messages to which a transport rule action shouldn't be applied, even if the message matches a transport rule condition.

In fact, a number of types of transport rules have been created in Exchange Online and Exchange Server in order to accomplish DLP capability. One important feature of transport rules is an innovative approach to classifying sensitive information that can be incorporated into the mail flow processing. This DLP feature performs deep content analysis through keyword matches, dictionary matches, regular expression evaluation, and other content examination to detect content that violates organizational DLP policies.

**Note** For more information about transport rules, see articles [Mail flow or transport rules](https://technet.microsoft.com/en-us/library/dd351127(v=exchg.150).aspx)[[101]](#footnote-102) and [Integrating Sensitive Information Rules with Transport Rules](http://technet.microsoft.com/en-us/library/jj150583(v=exchg.150).aspx)[[102]](#footnote-103).

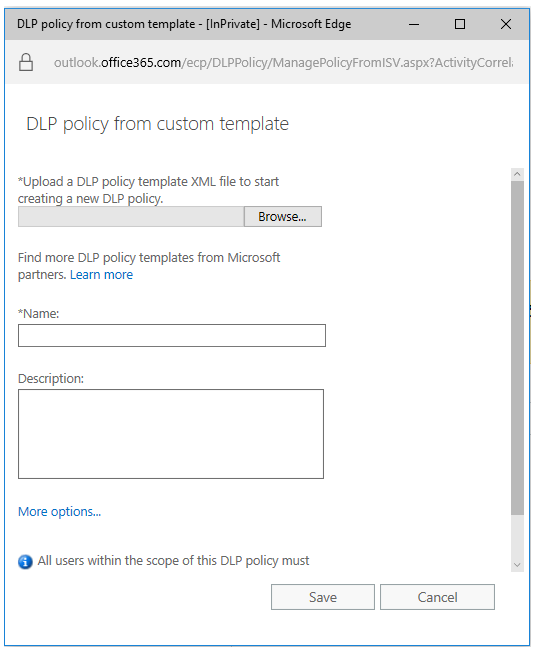
Built-in templates for a DLP policy based on regulatory standards such as Personally Identifiable Information (PII) and Payment Card Industry Data Security Standard (PCI-DSS)[[103]](#footnote-104) are offered by default as illustrated hereafter:



This set of DLP policies is extensible to support other policies important to your business. IT professionals can easily create DLP policies in the Exchange admin center (EAC). For example, a DLP policy built for a financial institution would take action on emails that include credit card information.

**Note** IT Professionals can also manage the organization’s DLP policies by using Exchange Management Shell cmdlets. For more information about policy and compliance cmdlets, see article [Policy and compliance cmdlets](http://technet.microsoft.com/en-us/library/dd298082(v=exchg.150).aspx)[[104]](#footnote-105).

Fundamentally, a DLP policy with Exchange transport rules is an .xml document (i.e. a block of configuration) that will determine what content should be detected and what is the response (action) when that content is detected.



Such a DLP policy can include one or more rules that in turn have condition(s), action(s), and exception(s), and thus uses the full power of Exchange transport rules.

In terms of condition to identity sensitive information in the context of this document, the Document Fingerprinting available in Exchange Online and Exchange Server 2013 Service Pack 1 (SP1) helps you detect sensitive information in standard forms that are used throughout the organization.

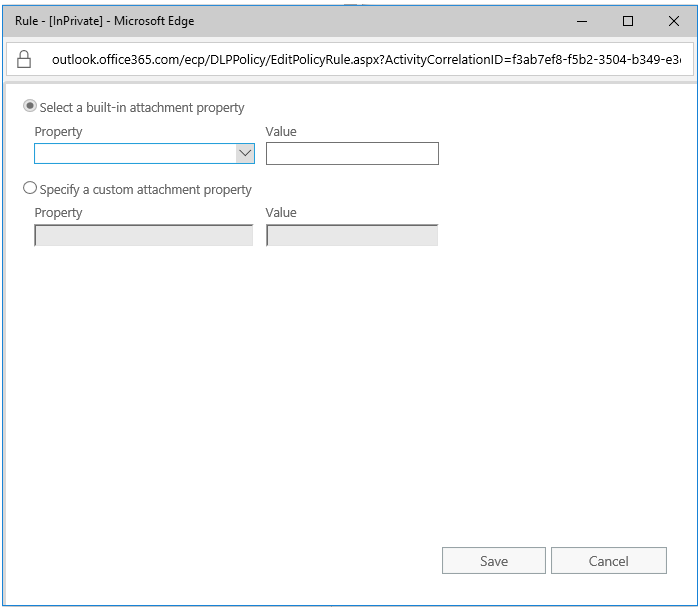
**Note** For more information about document fingerprinting, see article [Document Fingerprinting](http://technet.microsoft.com/en-us/library/dn635176(v=exchg.150).aspx)[[105]](#footnote-106).

Document Fingerprinting is a DLP feature that converts a standard form into a sensitive information type, which you can use to define transport rules and DLP policies. For example, you can create a document fingerprint based on a blank patent template and then create a DLP policy that detects and blocks all outgoing patent templates with sensitive content filled in.

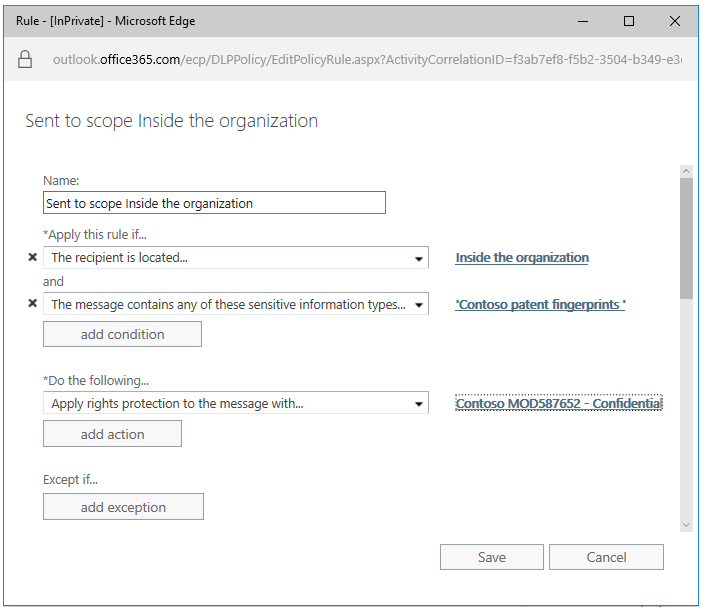
As far as the forms are concerned, Document Fingerprinting supports the same file types as the ones that are supported in transport rules.

**Note** For more information about the supported file types, see article [Use mail flow rules to inspect message attachments](http://technet.microsoft.com/en-us/library/jj919236(v=exchg.150).aspx)[[106]](#footnote-107).

Furthermore, for the attachments if any, you can also leverage the document properties as a condition and thus leverage any label’s data entries previously set by Azure Information Protection, FCI, and/or any other systems.



Upon identifying sensitive information, the DLP policy can automatically take action such as applying Rights Management protection (see section § Protecting information with Rights Management), appending a disclaimer, generating an audit log, sending the message for moderation, or preventing a message from being sent.



In addition to the customizable DLP policies themselves, DLP works with a feature called Policy tips that informs users of a potential policy violation before it occurs, i.e. even before they send an offending message. Policy tips are similar to MailTips, and can be configured to present a brief note in the Outlook 2013/2016 client that provides information about possible policy violations to a person creating a message. Policy tips help educate users about what sensitive data has been found in the email and, more generally, about related organization’s security and privacy policies. This ongoing education helps users to manage data appropriately and avoid sending sensitive data to unauthorized users.

In Exchange Online and Exchange 2013 SP1 and above, Policy tips are also displayed in Outlook Web App (OWA) and OWA for Devices.

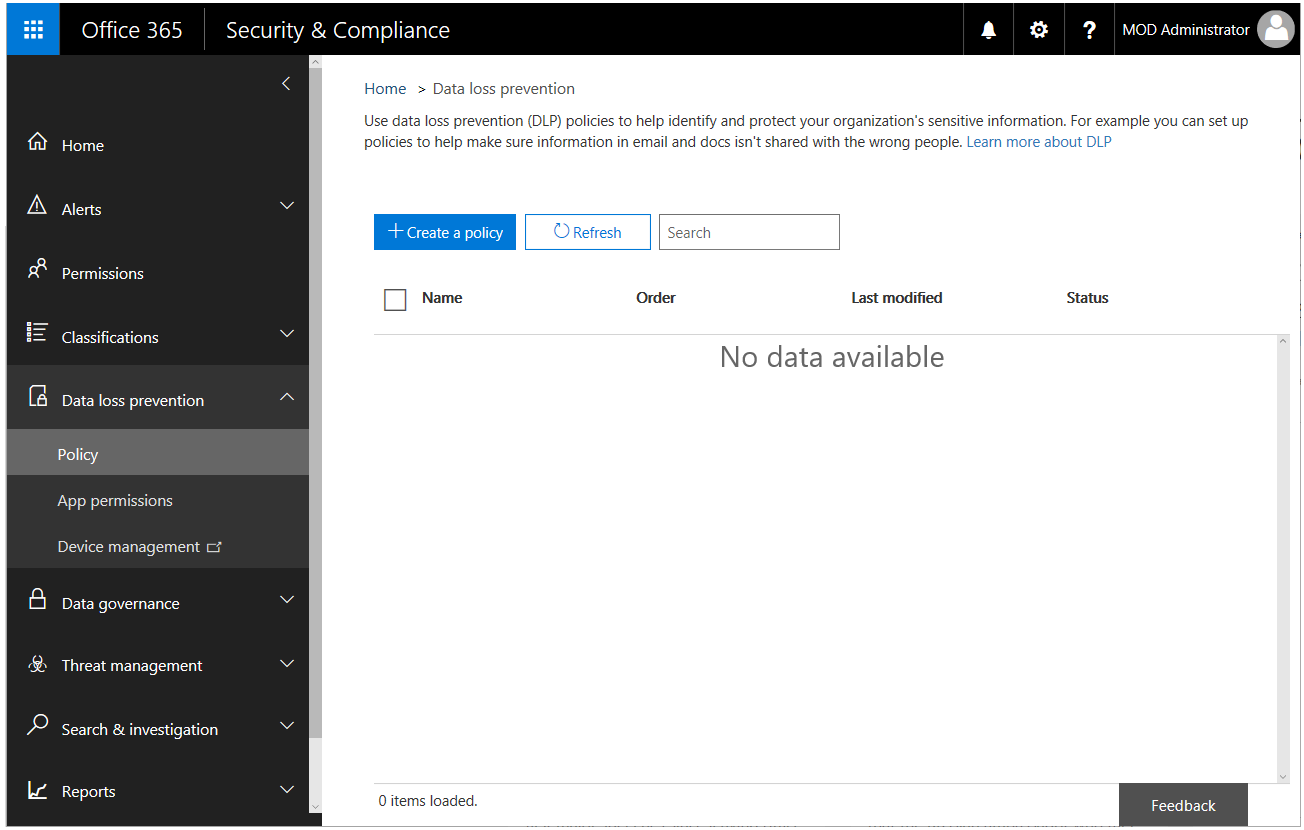
**Note** For more information, see article [Policy Tips](http://technet.microsoft.com/en-us/library/jj150512(v=exchg.150).aspx)[[107]](#footnote-108).

The DLP technology is part of the compliance management features provided in Exchange Online and Exchange Server 2013/2016. In terms of compliancy, we would like to take this paper to shortly describe the InPlace eDiscovery and Hold feature. (see section § Benefiting from eDiscovery later in this document).

**Note** For more information, see article [Messaging Policy and Compliance](http://technet.microsoft.com/en-us/library/aa998599(v=exchg.150).aspx)[[108]](#footnote-109).

#### DLP features in the Office 365 Security & Compliance Center

The new (unified) Office 365 Security & Compliance Center now also allows you to create DLP policies.

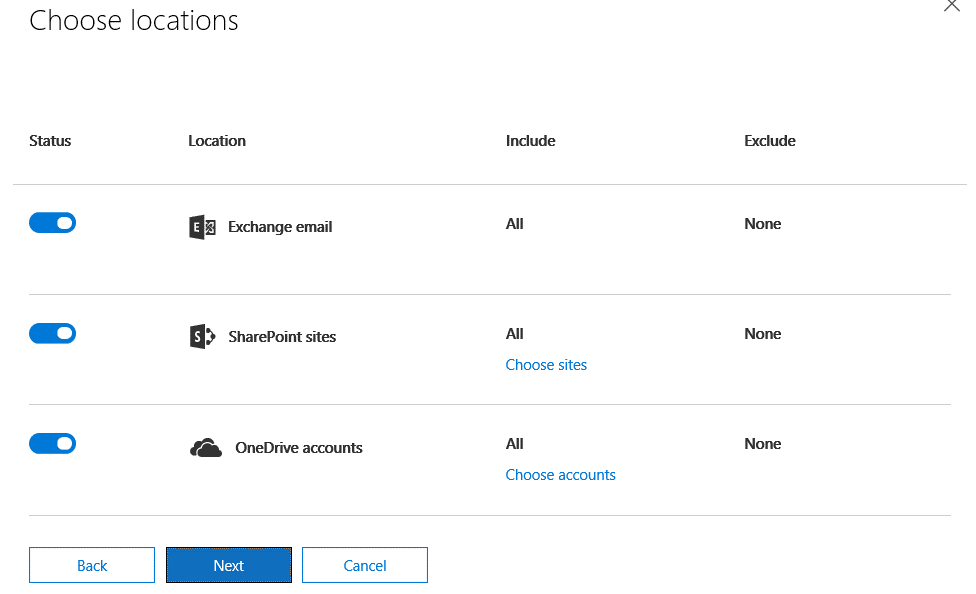


**Note** For more information, see blog post [Unifying Data Loss Prevention in Office 365](https://blogs.office.com/2017/01/09/unifying-data-loss-prevention-in-office-365/)[[109]](#footnote-110).

Like the DLP policies with Exchange transport rules, these DLP policies can automatically detect sensitive content - thanks to ready-to-use policy templates that address common compliance requirements, such as helping you to protect sensitive information subject -, and take actions.

However, unlike the above DLP policies, actions do not allow to apply usage rights restriction, and thus to integrate with Azure RMS in terms of protection.

These DLP policies are however deployed to all the locations included in the policy, i.e. Exchange Online but also SharePoint Online, and/or OneDrive for Business sites as discussed in the next section.



**Note** For more information, see articles [Overview of data loss prevention policies](https://support.office.com/en-us/article/Overview-of-data-loss-prevention-policies-1966b2a7-d1e2-4d92-ab61-42efbb137f5e)[[110]](#footnote-111) and [What the DLP policy templates include](https://support.office.com/en-us/article/What-the-DLP-policy-templates-include-c2e588d3-8f4f-4937-a286-8c399f28953a)[[111]](#footnote-112).

If the policy includes Exchange Online, the policy is synced there and enforced in exactly the same way as a DLP policy created in the Exchange admin center. If DLP policies have been created in the Exchange admin center, those policies will continue to work side by side with any policies for email that you create in the Office 365 Security & Compliance Center, but rules created in the Exchange admin center take precedence. All Exchange transport rules are processed first, and then the DLP rules from the Office 365 Security & Compliance Center are processed.

Conversely, policy tips can work either with DLP policies and Exchange transport rules created in the Exchange admin center, or with DLP policies created in the Office 365 Security & Compliance Center, but not both. This is because these policies are stored in different locations, but policy tips can draw only from a single location.

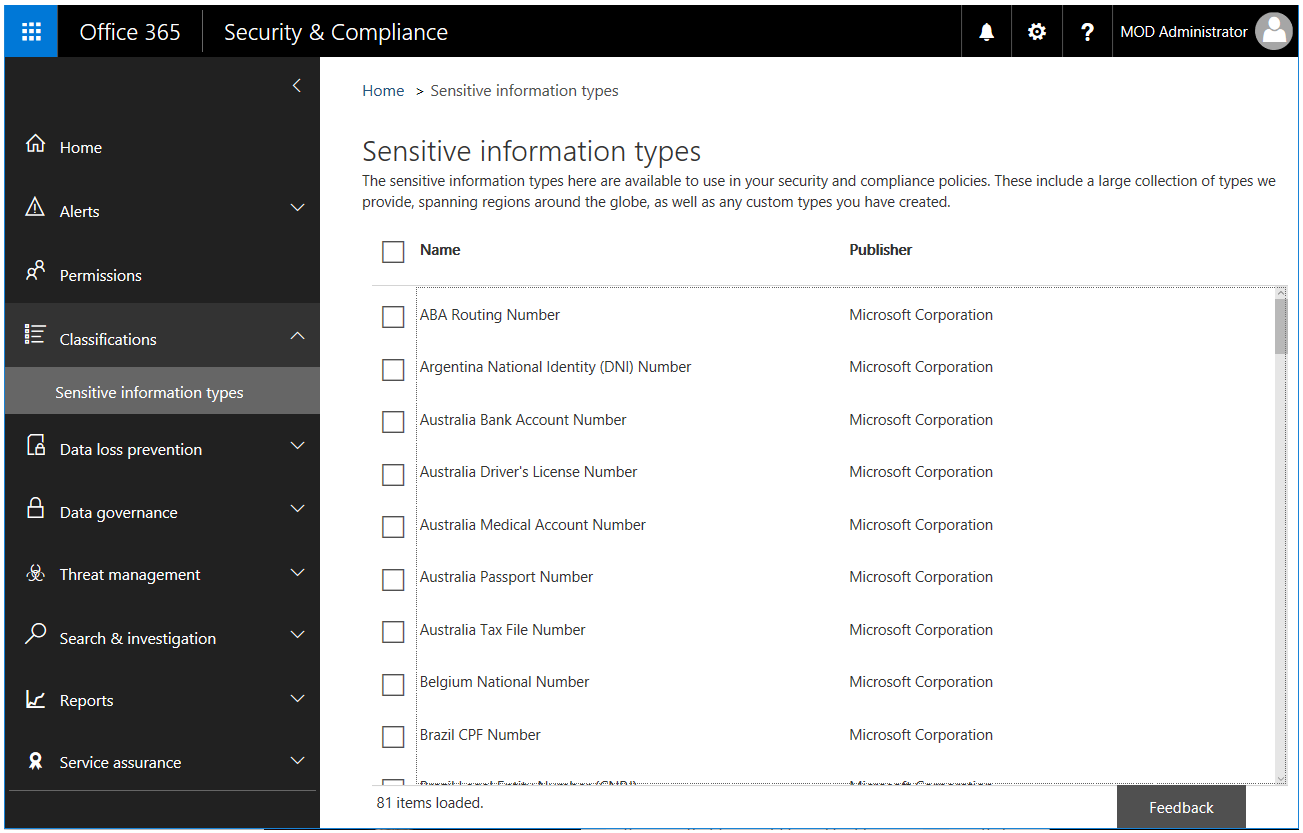
If you’ve configured policy tips in the Exchange admin center, any policy tips that you configure in the Office 365 Security & Compliance Center won’t appear to users in Outlook Web App (OWA) and Outlook 2013/2016 until you turn off the tips in the Exchange admin center. This ensures that your current Exchange transport rules will continue to work until you choose to switch over to the Office 365 Security & Compliance Center.

### DLP features in SharePoint Online and OneDrive for Business

#### DLP features via Enterprise Search

DLP features are also present in SharePoint Online and OneDrive for Business into the existing Enterprise Search capability, thereby allowing to search for sensitive content with simple or complex queries and crawl a variety of sources, including team sites and users’ OneDrive for Business folders, keeping content in place and enabling you to search in real time.

A wide range of sensitive information types from different industry segments and geographies (81 at the time of this writing), many of which you may already be using to search for sensitive content in email (see previous section). Likewise, these sensitive information types are detected based on pattern matching and are easy to set up.



**Note** For more information, see blog post [What the sensitive information types look for](https://support.office.com/en-us/article/What-the-sensitive-information-types-look-for-fd505979-76be-4d9f-b459-abef3fc9e86b)[[112]](#footnote-113).

The identified offending documents can be reviewed inline in real time and/or exported for further review to inspect them, check for false positive, then take manual actions such as adjusting sharing permissions, removing data on shared sites, etc.

#### DLP features in the Office 365 Security & Compliance Center

Additional capabilities that go beyond simply discovering and reviewing sensitive content are provided with the Office 365 Security & Compliance Center. The Office 365 Security & Compliance Center indeed now allows you to create DLP policies as outlined in the previous section.

These DLP policies can automatically detect sensitive content and take actions like blocking access to the content, sending a notification to the primary site collection administrator, etc. These DLP policies can be applied to a set of locations, i.e. SharePoint Online and OneDrive for Business sites and Exchange Online.

**Note** For more information, see articles [Overview of data loss prevention policies](https://support.office.com/en-us/article/Overview-of-data-loss-prevention-policies-1966b2a7-d1e2-4d92-ab61-42efbb137f5e)[[113]](#footnote-114) and [What the DLP policy templates include](https://support.office.com/en-us/article/What-the-DLP-policy-templates-include-c2e588d3-8f4f-4937-a286-8c399f28953a)[[114]](#footnote-115).

For organizations that have a process in place to identify and classify sensitive information by using the labels in Azure Information Protection (see section § Classifying and labelling information at the time of creation), the classification properties in FCI (see section § Leveraging an on-premises File Classification Infrastructure (FCI) for existing data), or the document properties applied by a third-party system, you can create a DLP policy that recognizes the properties that have been applied to documents by Azure Information Protection (AIP), FCI, or other system, by defining rules that use the condition **Document properties contain any of these values** with the appropriate properties’ value.

So the DLP policy can then be enforced on Office documents with specific Azure Information Protection label property, specific FCI property, or other property values to take appropriate property-based actions such as blocking access to those files or sending an email notification. (Before you can use an Azure Information Protection property, a FCI property or other property in a DLP policy, you need to create a managed property in the SharePoint admin center.)

In this way, DLP in SharePoint Online integrates with Azure Information Protection, FCI, or other systems if any and can help protect Office documents uploaded or shared to Office 365 from Windows Server–based file servers on-premises.

**Note** For more information, see blog post [Create a DLP policy to protect documents with FCI or other properties](https://support.office.com/en-us/article/Create-a-DLP-policy-to-protect-documents-with-FCI-or-other-properties-1b9e3c6c-4308-4a20-b11e-c37b8013e177)[[115]](#footnote-116).

As of this writing, the above condition **Document properties contain any of these values** is temporarily not available in the UI of the Office 365 Security & Compliance Center. So, such a DLP policy should be created using the Office 365 Security & Compliance Center cmdlets from Windows PowerShell.

The New\Set\Get-DlpCompliancePolicy cmdlets allows to work with a DLP policy and the ContentPropertyContainsWords parameter to add the condition **Document properties contain any of these values**.

**Note** For more information, see blog post [Office 365 Security & Compliance Center cmdlets](https://technet.microsoft.com/en-us/library/mt587093(v=exchg.160).aspx)[[116]](#footnote-117).

## Monitoring the flow of sensitive data to cloud environments using CASB

As sensitive information assets travels to cloud-based applications as part of the modernization of IT effort most organizations have already initiated, it becomes critical to monitor how information is being used, shared, and distributed.

In this context, Cloud Access Security Brokers (CASB) are on-premises, or cloud-based security policy enforcement points, placed between the organizations on-premises IT infrastructure and the cloud-based applications they consume typically as part of the modernization of their It as previously covered.

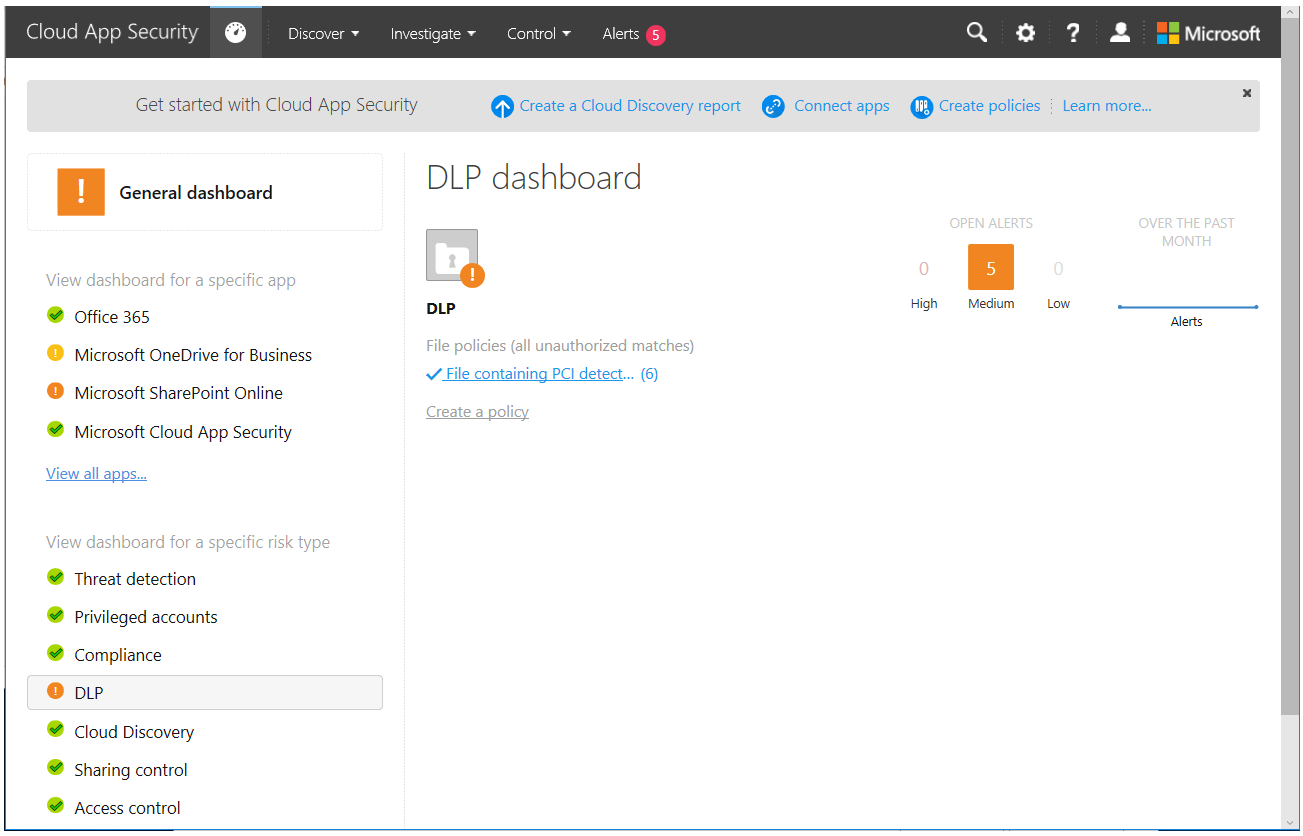
This approach should not be confused with that of a virtual private network (VPN); CASBs are designed to provide a transparent layer to cloud-based applications. CASBs are designed to work seamlessly and transparently between users and the cloud-based applications. They are placed into the data flow between user devices and cloud-based applications.

As such, CASBs can provide a means to manage and secure information that has been classified as “Confidential” by encrypting the data in transit as well as data at rest. You can indeed via the configuration of policies look for sensitive information in cloud-based applications and take actions or get alerted in case of abnormal behaviors.

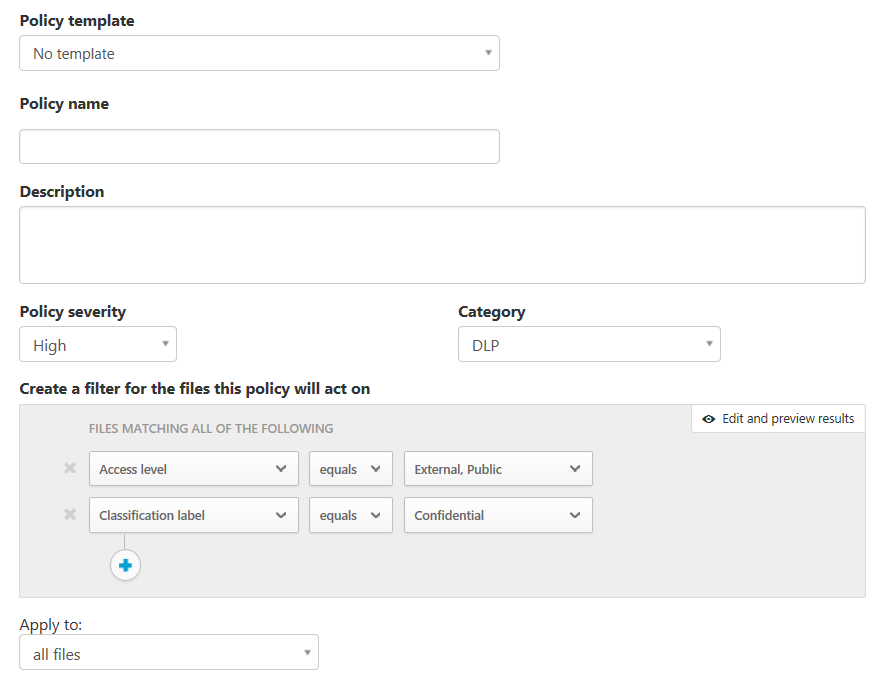
To further illustrate the holistic approach based on Microsoft solutions, let’s introduce Cloud App Security (CAS)[[117]](#footnote-118), a critical component of the Microsoft Cloud Security stack. It's a comprehensive solution that can help your organization as you move to take full advantage of the promise of cloud-based applications, but keep you in control, through improved visibility into activity.

**Note**  For more information, see blog post [Why you need a cloud access security broker in addition to your firewall](https://blogs.technet.microsoft.com/enterprisemobility/2016/12/20/why-you-need-a-cloud-access-security-broker-in-addition-to-your-firewall/)[[118]](#footnote-119) and article [What is Cloud App Security?](https://docs.microsoft.com/en-us/cloud-app-security/what-is-cloud-app-security)[[119]](#footnote-120).

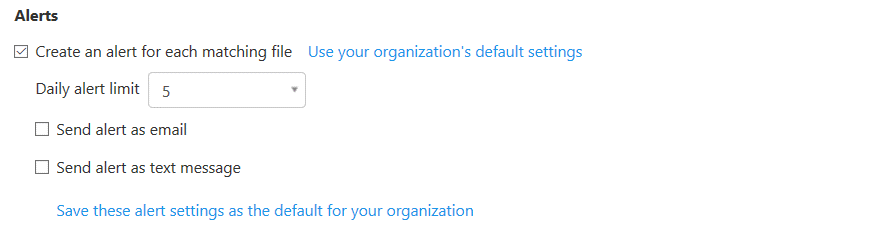
It also helps increase the protection of critical data across cloud applications and notably integrates for that purpose DLP features as illustrated hereafter.



Policies can be defined to leverage any label previously set by Azure Information Protection, FCI, and/or any other systems to reflect their level of sensitivity to the business.



An alert can then be created.



And the governance further allows to put the matching file(s) in quarantine, to restrict sharing for them, etc.

**Note** For more information on how the integration with Azure Information Protection and Cloud App Security works as well as technical guidance on how IT professionals can gain visibility and control over sensitive information in cloud-based applications with Cloud App Security, see blog post [Azure Information Protection and Cloud App Security integration: Extend control over your data to the cloud](https://blogs.technet.microsoft.com/enterprisemobility/2016/11/07/azure-information-protection-and-cloud-app-security-integration-extend-control-over-your-data-to-the-cloud/)[[120]](#footnote-121) and article [Azure Information Protection integration](https://docs.microsoft.com/en-us/cloud-app-security/azip-integration)[[121]](#footnote-122).

## Benefiting from eDiscovery

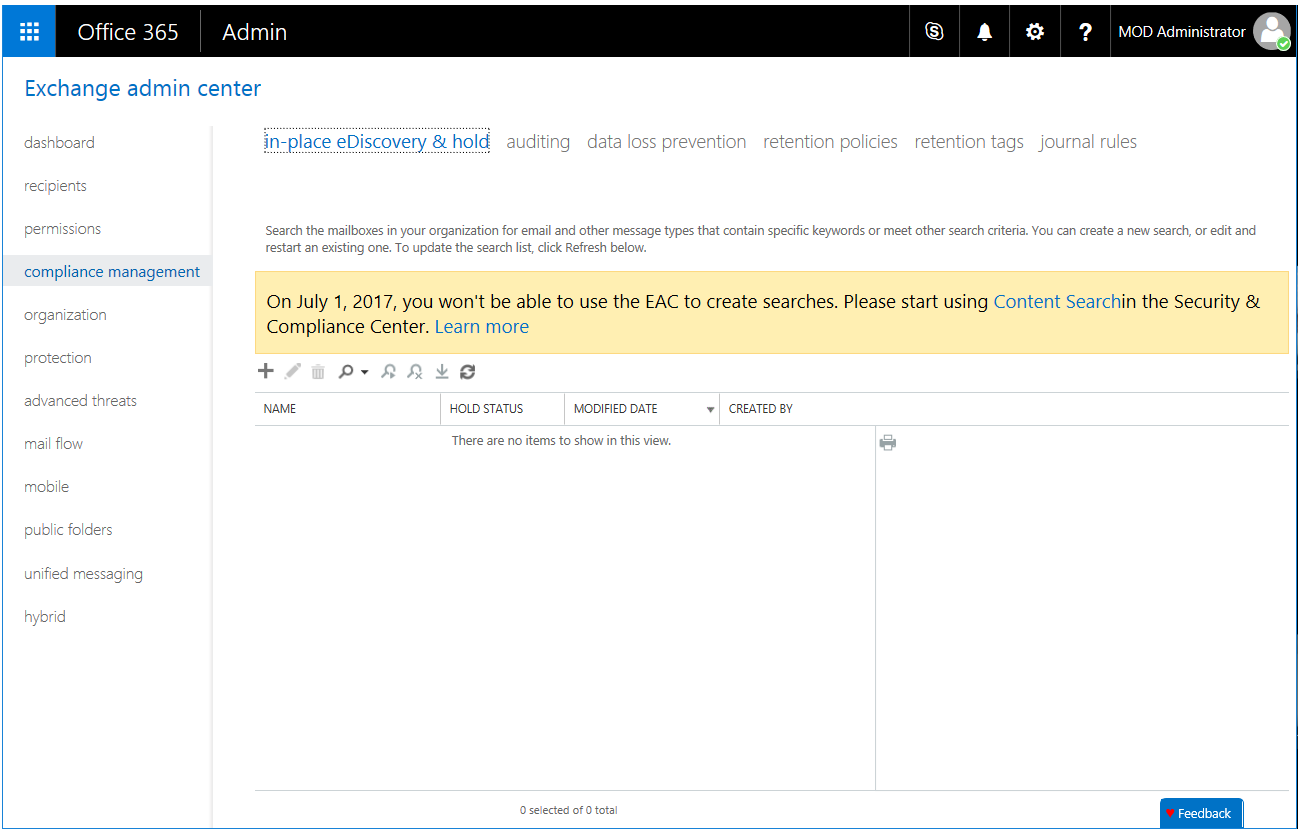
Electronic Discovery, or eDiscovery, is the discovery of content in electronic format for litigation or investigation. This typically requires identifying contents - that are relevant to a particular subject - spread across laptops, email servers, file servers, and many other sources.

If your organization adheres to legal discovery requirements (related to organizational policy, compliance, or lawsuits), you can use eDiscovery in Office 365 to search for content in Exchange Online mailboxes, Office 365 Groups, Microsoft Teams, SharePoint Online and sites, and Skype for Business conversations.

**Note** For more information, see article [eDiscovery in Office 365](https://support.office.com/en-us/article/eDiscovery-in-Office-365-143b3ab8-8cb0-4036-a5fc-6536d837bfce)[[122]](#footnote-123)

### Searching mailboxes

If you only need to search mailboxes, in-Place eDiscovery in the Exchange admin center (EAC) of Exchange Online and Exchange Server 2013/2016 can help you perform e-discovery searches for relevant content within mailboxes.



**Important note** As mentioned on the above screenshot, on July 1, 2017, you'll no longer be able to create In-Place eDiscovery searches in Exchange Online. To create eDiscovery searches, you ought to start using the Content Search page in the Office 365 Security & Compliance Center. See section § Searching mailboxes and sites in the same eDiscovery search later in this document.

You will still be able to modify existing In-Place eDiscovery searches. In Exchange hybrid deployments, searches run from your on-premises organization aren't affected by this change with Exchange Server 2013/2016.

**Note** For more information, see article [In-Place eDiscovery](https://technet.microsoft.com/en-us/library/dd298021(v=exchg.150).aspx)[[123]](#footnote-124)

Thanks to this feature, compliance officers can perform the discovery process in-place as data is not duplicated into separate repositories. This addresses pre-processing stages of e-discovery, including information management, identification, preservation, and collection. As such, tools operate on data where it lives, and preserve minimum amount of data needed. Since content is held in-place, teams can respond quickly by accessing data in its native format – without any loss of fidelity often associated with copying data to separate archives –. Then, IT professionals and compliance officers have an easy way to package the result by exporting it according to the aforementioned EDRM XML specification so that it can be imported for example into a review tool.

**Note** In order to be able to decrypt any IRM protected content within an organization (see later in this document), IT professionals and compliance officers must be added as super user in Microsoft Rights Management service.

In-Place e-Discovery can be used in an Exchange hybrid environment to search cloud-based and on-premises mailboxes in the same search.

This capability resonates with the site mailbox, another feature of Exchange Online and Exchange Server 2013/2016. A site mailbox is a shared inbox in Exchange Online and Exchange Server 2013/2016 that all the members of a SharePoint Online and SharePoint Server 2013/2016 site, i.e. the individuals listed in the Owners and Members groups of the site can access (security groups or distribution lists are not supported). It’s accessible from the site in which it is created. The e-mail address of the site mailbox is generated automatically from the name of the site.

With site mailboxes, Exchange Online and Exchange Server 2013/2016 work with SharePoint Online and SharePoint Server 2013/2016 to give users more ways to collaborate while keeping data safe.

**Note** To support such an integrated scenario on-premises, the servers must be able to request resources from each other in a secure way. This leverages the Server-to-Server (S2S) authentication, a feature of Exchange Server 2013/2016, SharePoint Server 2013/2016, and Lync Server 2013/Skype for Business Server 2016 that allows a server to request resources of another server on behalf of a user. This feature uses the industry standard Open Authorization (OAuth) 2.0 protocol.

Beyond the site itself, site mailboxes are surfaced in Outlook 2013/2016 and give you easy access to the email and documents for the projects you care about. It's listed in the Folder Explorer in Outlook 2013/2016, letting you store emails or documents into the shared project space simply by dragging the email, document, or attachment into the site mailbox. (Site mailboxes are not available in Outlook Web App).

Users view site mailbox emails just as they would any other Exchange message, while SharePoint enables versioning and coauthoring of documents.

**Note** For more information, see blog post [Site Mailboxes in the new Office](https://blogs.technet.microsoft.com/exchange/2012/08/22/site-mailboxes-in-the-new-office/)[[124]](#footnote-125).

Site mailboxes can be searched using the Exchange eDiscovery Center where the built-in e-Discovery functionality makes it easy to find needed information across held, archived, and current e-mail messages. As a consequence, e-mail messages and documents stored in site mailboxes can be put on legal hold. Additionally, site mailboxes adhere to the lifecycle policies applied to the SharePoint site with which they are associated, enabling automated retention and archiving of the entire site mailbox. Site mailboxes allow users to naturally work together – while compliance policies are applied behind the scenes.

Beyond the site mailboxes, Exchange Online and Exchange Server 2013/2016 offer federated search capability and integration with SharePoint Online and Microsoft SharePoint 2013/2016.

### Searching sites

The eDiscovery Center in SharePoint Online and SharePoint Server 2013/2016 can help you manage e-discovery cases and related searches.

The eDiscovery Center is a SharePoint site collection used to perform electronic discovery actions. In an eDiscovery Center, you can create cases, which are SharePoint sites that allow you to identify, hold, search, and export content from SharePoint Online and SharePoint Server 2013/2016 sites, searchable file shares indexed by SharePoint, content in Exchange mailboxes, and archived Lync 2013/Skype for Business 2016 content

**Important note** On July 1, 2017, you'll no longer be able to create new eDiscovery cases in SharePoint Online. To create eDiscovery cases, you ought to start using the Content Search page in the Office 365 Security & Compliance Center. See section § Searching mailboxes and sites in the same eDiscovery search later in this document.

You will still be able to modify existing eDiscovery cases in SharePoint Online.

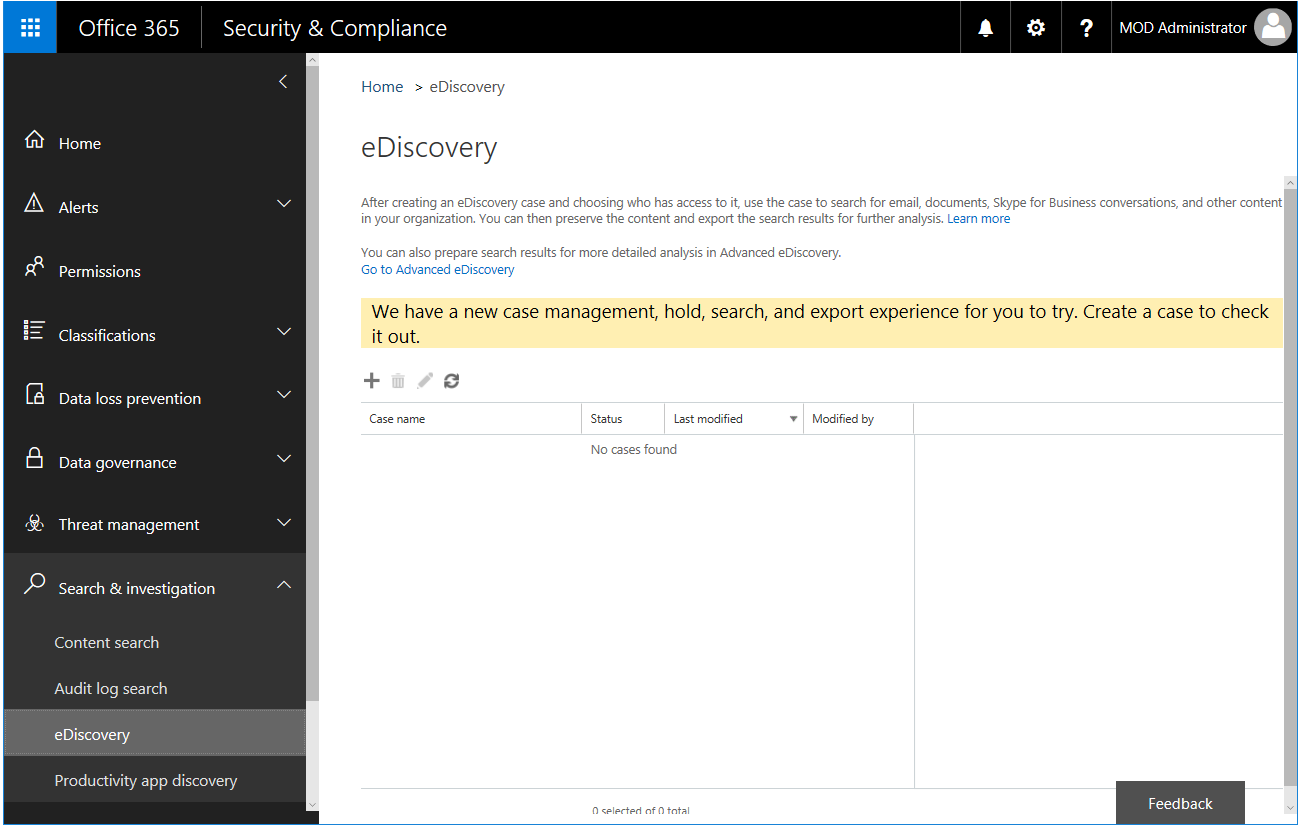
**Note** For more information, see article [Plan and manage cases in the eDiscovery Center](https://support.office.com/en-us/article/Plan-and-manage-cases-in-the-eDiscovery-Center-d955aeb8-0d48-4291-a8e2-f3b84f17943f?ui=en-US&rs=en-US&ad=US)[[125]](#footnote-126).

### Searching mailboxes and sites in the same eDiscovery search

If you need to search mailboxes and sites in the same eDiscovery search for Office 365, you can create an eDiscovery case and an associated Content Search in the Office 365 Security & Compliance Center. You can then identify, hold, and export content found in mailboxes and sites.

**Note** For more information, see article [Manage eDiscovery cases in the Office 365 Security & Compliance Center](https://support.office.com/en-us/article/Manage-eDiscovery-cases-in-the-Office-365-Security-Compliance-Center-edea80d6-20a7-40fb-b8c4-5e8c8395f6da)[[126]](#footnote-127).

The Office 365 Security & Compliance Center is one place to manage compliance across Office 365 for your organization.



You can define central policies that apply across your data in Office 365, such as preserve policies that keep content in Exchange Online and in SharePoint Online indefinitely or for a set time period.

Links to existing Exchange Online and SharePoint Online compliance features bring together the compliance capabilities across Office 365.

The eDiscovery case cmdlets alternatively be used from Windows PowerShell.

**Note** For more information, see blog post [Office 365 Security & Compliance Center cmdlets](https://technet.microsoft.com/en-us/library/mt587093(v=exchg.160).aspx)[[127]](#footnote-128).

## Integrating with other systems

As covered so far, Azure Information Protection delivers a holistic, agile, comprehensive, and flexible Information Protection (IP) platform for today’s businesses to cope with current industry trends such modernization of IT and consumerization of IT.

It aims at providing a flexible infrastructure that can be leveraged to meet the most rigorous protection and compliance requirements.

Integration is virtually possible with any system to enable the protection of content and/or the consumption of protected content from it, to augment it to fulfill some industry verticals’ requirements, etc.

Examples of type of integration are Line of Business (LoB) application, document management systems, DLP systems, etc. Some integrations exist today natively, some involve partners in the Azure Information Protection partners’ ecosystem, some others will involve independent software vendor (ISV), developers, etc. to modify their code.

The [Azure Information Protection Developer's Guide](https://docs.microsoft.com/en-us/information-protection/develop/developers-guide)[[128]](#footnote-129) will orient you to tools for extending and integrating with Azure Information Protection to provide information protection. The intent of this guide thus aims at allowing independent software vendor (ISV), developers, etc. who want to leverage the Rights Management capabilities to build different types of applications for a range of supported devices and platforms.

It features the following Software Development Kits (SDK) and their libraries that are provided with Azure Information Protection for protection purposes:

* **Rights Management SDK 4.2**[[129]](#footnote-130). A simplified, next-generation API that enables a lightweight development experience in upgrading device apps with information protection via the Azure RMS protection technology in Azure Information Protection.
* **Rights Management SDK 2.1**[[130]](#footnote-131). A platform that enables developers to build applications that leverage Rights Management capabilities to provide information protection both with Azure Information Protection (Azure RMS) and AD RMS on-premises. The RMS SDK 2.1 handles complex security practices such as key management, encryption and decryption processing and offers a simplified API for easy application development.

**Important note** Above Azure Information Protection SDKs only have as of this writing the Rights Management component. The classification and labelling are under development.

A GitHub repository provides a series of samples[[131]](#footnote-132) that can inspire or provide directions on the integration path.

**This concludes our guided tour on how to build an information classification and protection enforcement infrastructure on-premises, in the cloud, or in hybrid environment with Microsoft services, products, and technologies.**

We hope that you’re now better equipped with a clear understanding of what an information classification effort can provide to your organization and how Microsoft solutions can help to apply the required security controls for maintaining the security (confidentiality and integrity) of the key and/or sensitive information assets of your organization.

**Note** For an overview of the Azure Information Protection, see [online documentation](https://docs.microsoft.com/en-us/information-protection/)[[132]](#footnote-133), the [series of whitepapers](https://aka.ms/aippapers)[[133]](#footnote-134) to which the current document belong as well as the posts on the [Enterprise Mobility + Security (EMS) Team blog](https://blogs.technet.microsoft.com/enterprisemobility/?product=azure-information-protection)[[134]](#footnote-135).

|  |
| --- |
| The information contained in this document represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of publication.  This white paper is for informational purposes only. Microsoft makes no warranties, express or implied, in this document.  Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in, or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.  Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.  © 2017 Microsoft Corporation. All rights reserved.  The example companies, organizations, products, domain names, e-mail addresses, logos, people, places, and events depicted herein are fictitious. No association with any real company, organization, product, domain name, e-mail address, logo, person, place, or event is intended or should be inferred.  Microsoft, list Microsoft trademarks used in your white paper alphabetically are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.  The names of actual companies and products mentioned herein may be the trademarks of their respective owners. |

1. Verizon 2016 Data Breach Investigations Report: http://www.verizonenterprise.com/verizon-insights-lab/dbir/ [↑](#footnote-ref-2)
2. Passed in 1996, HIPAA relates to healthcare coverage and, for example, how companies may use medical information. [↑](#footnote-ref-3)
3. Enacted in 2009 to promote and expand the adoption of health information technology. [↑](#footnote-ref-4)
4. Gramm-Leach-Bliley, also known as the Financial Services Modernization Act, was passed in 1999. [↑](#footnote-ref-5)
5. The Sarbanes–Oxley Act of 2002, also known as the 'Public Company Accounting Reform and Investor Protection Act' (in the Senate) and 'Corporate and Auditing Accountability and Responsibility Act' (in the House), is a United States federal law that set new or enhanced standards for all U.S. public company boards, management and public accounting firms. [↑](#footnote-ref-6)
6. Passed in 2000, and reviewed every 5 years, PIPEDA is a Canadian law relating to data privacy that governs how private sector organizations collect, use and disclose personal information in the course of commercial business. [↑](#footnote-ref-7)
7. Passed in 2016, EU GDPR entered into force on 24 May 2016 and is due to apply from 25 May 2018. [↑](#footnote-ref-8)
8. Passed in 2003, PIPA spells out duties of the national and local government for handling personal information and measures for protecting personal information. It also sets out obligations of businesses that handle personal information. [↑](#footnote-ref-9)
9. Gartner, “Information Classification: An Essential Security Thing You're (Still) Not Doing,” September 2012 [↑](#footnote-ref-10)
10. Legal issues in the Cloud - Part 1: http://www.cio.com.au/article/382624/legal\_issues\_cloud\_-\_part\_1/ [↑](#footnote-ref-11)
11. Legal issues in the Cloud - Part 2: http://www.cio.com.au/article/382649/legal\_issues\_cloud\_-\_part\_2/ [↑](#footnote-ref-12)
12. Legal issues in the Cloud - Part 3: http://www.cio.com.au/article/382652/legal\_issues\_cloud\_-\_part\_3/ [↑](#footnote-ref-13)
13. Legal issues in the Cloud - Part 4: http://www.cio.com.au/article/382657/legal\_issues\_cloud\_-\_part\_4/ [↑](#footnote-ref-14)
14. Forrester Research Inc, Brief: You Need An Action Plan For The GDPR: https://www.forrester.com/report/Brief+You+Need+An+Action+Plan+For+The+GDPR/-/E-RES136242 [↑](#footnote-ref-15)
15. Forrester Research Inc, The Data Security And Privacy Playbook For 2017: https://www.forrester.com/The+Data+Security+And+Privacy+Playbook/-/E-PLA200 [↑](#footnote-ref-16)
16. This comprises archival and retention data. [↑](#footnote-ref-17)
17. Data Classification for Cloud Readiness: http://download.microsoft.com/download/0/A/3/0A3BE969-85C5-4DD2-83B6-366AA71D1FE3/Data-Classification-for-Cloud-Readiness.pdf [↑](#footnote-ref-18)
18. Gartner, “ILM: The Art of Data Placement," December 2009 [↑](#footnote-ref-19)
19. ISO/IEC 27002:2013 Information technology -- Security techniques -- Information security management systems – Code of Practice for Information Security Controls: http://www.iso.org/iso/fr/home/store/catalogue\_tc/catalogue\_detail.htm?csnumber=54533 [↑](#footnote-ref-20)
20. ISO/IEC 17799:2005: Information technology -- Security techniques -- Code of practice for information security management: http://www.iso.org/iso/catalogue\_detail?csnumber=39612 [↑](#footnote-ref-21)
21. (UK) Government Security Classification Policy: https://www.gov.uk/government/publications/government-security-classifications [↑](#footnote-ref-22)
22. The need-to-know principle claims that “Access to sensitive information must be no wider than necessary for the efficient conduct of an organisation’s business and limited to those with a business need and the appropriate personnel security control.” Source: (UK) Government Security Classification Policy. [↑](#footnote-ref-23)
23. ISO/IEC 27001:2013 Information technology -- Security techniques -- Information security management systems -- Requirements: http://www.iso.org/iso/home/store/catalogue\_tc/catalogue\_detail.htm?csnumber=54534 [↑](#footnote-ref-24)
24. ISO/IEC 29100:2011 Information technology -- Security techniques -- Privacy framework http://www.iso.org/iso/home/store/catalogue\_tc/catalogue\_detail.htm?csnumber=45123 [↑](#footnote-ref-25)
25. ISO/IEC 27018:2014 Information technology -- Security techniques -- Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors http://www.iso.org/iso/home/store/catalogue\_tc/catalogue\_detail.htm?csnumber=61498 [↑](#footnote-ref-26)
26. Sarbanes–Oxley Act: http://en.wikipedia.org/wiki/Sarbanes%E2%80%93Oxley\_Act [↑](#footnote-ref-27)
27. NIST Special Publication 800-53 Revision 4 Security and Privacy Controls for Federal Information Systems and Organizations: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf [↑](#footnote-ref-28)
28. Critical Controls for Effective Cyber Defense: http://www.sans.org/critical-security-controls/ [↑](#footnote-ref-29)
29. Data Loss Prevention: http://en.wikipedia.org/wiki/Data\_Loss\_Prevention [↑](#footnote-ref-30)
30. Gartner Says E-Discovery Software Marketplace is Set to Continue High-Growth Pace: http://www.gartner.com/it/page.jsp?id=1257113 [↑](#footnote-ref-31)
31. Electronic Discovery Reference Model: http://www.edrm.net/ [↑](#footnote-ref-32)
32. EDRM XML White Paper: http://www.edrm.net/projects/xml/white-paper [↑](#footnote-ref-33)
33. ISO/IEC 17788 Information technology -- Cloud computing -- Overview and vocabulary: http://www.iso.org/iso/catalogue\_detail.htm?csnumber=60544 [↑](#footnote-ref-34)
34. Addressing HIPAA Security and Privacy Requirements in the Microsoft Cloud: http://smb.blob.core.windows.net/smbproduction/Content/Microsoft\_Cloud\_Healthcare\_HIPAA\_Security\_Privacy.pdf [↑](#footnote-ref-35)
35. Microsoft Trust Center: https://www.microsoft.com/en-us/TrustCenter/default.aspx [↑](#footnote-ref-36)
36. Regulation (EU) 2016/679: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679 [↑](#footnote-ref-37)
37. Reform of EU data protection rules: http://ec.europa.eu/justice/data-protection/reform/index\_en.htm [↑](#footnote-ref-38)
38. ISO/TS 25237:2008 Health informatics – Pseudonymization: https://www.iso.org/standard/42807.html [↑](#footnote-ref-39)
39. Microsoft.com/GDPR: https://www.microsoft.com/gdpr [↑](#footnote-ref-40)
40. Cloud Controls Matrix Working Group: https://cloudsecurityalliance.org/group/cloud-controls-matrix// [↑](#footnote-ref-41)
41. CSA Security, Trust and Assurance Registry (STAR): https://cloudsecurityalliance.org/star/ [↑](#footnote-ref-42)
42. Cloud Security Alliance: https://cloudsecurityalliance.org/ [↑](#footnote-ref-43)
43. Enabling Hybrid Cloud Today with Microsoft Technologies whitepaper: http://www.microsoft.com/en-us/download/details.aspx?id=39052 [↑](#footnote-ref-44)
44. Azure Information Protection: https://www.microsoft.com/en-us/cloud-platform/azure-information-protection [↑](#footnote-ref-45)
45. What is Azure Information Protection?: https://docs.microsoft.com/en-us/information-protection/understand-explore/what-is-information-protection [↑](#footnote-ref-46)
46. Frequently asked questions about classification and labeling in Azure Information Protection: https://docs.microsoft.com/en-us/information-protection/get-started/faqs-infoprotect [↑](#footnote-ref-47)
47. Azure Information Protection - also known as ...: https://docs.microsoft.com/en-us/information-protection/understand-explore/aka [↑](#footnote-ref-48)
48. Azure Information Protection service in the Azure portal: https://portal.azure.com/signin/index/?Microsoft\_Azure\_InformationProtection=true [↑](#footnote-ref-49)
49. Configuring Azure Information Protection policy: https://docs.microsoft.com/en-us/information-protection/deploy-use/configure-policy [↑](#footnote-ref-50)
50. Azure Information Protection client for Windows: https://docs.microsoft.com/en-us/information-protection/rms-client/aip-client [↑](#footnote-ref-51)
51. Azure Information Protection: https://www.microsoft.com/en-us/download/details.aspx?id=53018 [↑](#footnote-ref-52)
52. Download and install the Azure Information Protection client: https://docs.microsoft.com/en-us/information-protection/rms-client/install-client-app [↑](#footnote-ref-53)
53. Quick start tutorial for Azure Information Protection: https://docs.microsoft.com/rights-management/information-protection/infoprotect-quick-start-tutorial [↑](#footnote-ref-54)
54. How to configure the Azure Information Protection policy for specific users by using scoped policies: https://docs.microsoft.com/en-us/information-protection/deploy-use/configure-policy-scope [↑](#footnote-ref-55)
55. Azure Information Protection client files and client usage logging: https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-files-and-logging [↑](#footnote-ref-56)
56. The role of Azure Information Protection in securing data: https://docs.microsoft.com/en-us/enterprise-mobility-security/solutions/azure-information-protection-securing-data [↑](#footnote-ref-57)
57. What is Azure Rights Management: https://docs.microsoft.com/en-us/information-protection/understand-explore/what-is-azure-rms [↑](#footnote-ref-58)
58. Azure RMS Security Evaluation Guide: https://thermsguy.blob.core.windows.net/slides/AzureRMS-SecurityEvaluation-v1.docx [↑](#footnote-ref-59)
59. Document tracking portal: https://track.azurerms.com/ [↑](#footnote-ref-60)
60. Configuring and using document tracking for Azure Information Protection: https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-document-tracking [↑](#footnote-ref-61)
61. Track and revoke your documents when you use Azure Information Protection: https://docs.microsoft.com/en-us/information-protection/rms-client/client-track-revoke [↑](#footnote-ref-62)
62. Microsoft Rights Management sharing application for Windows: https://www.microsoft.com/en-us/download/details.aspx?id=40857 [↑](#footnote-ref-63)
63. Foxit PDF Security Suite – Rights Management for PDF: http://www.foxitsoftware.com/landingpage/rms.php [↑](#footnote-ref-64)
64. SECUDE End-to-End Information Security for SAP: http://secude.com/ [↑](#footnote-ref-65)
65. How applications support the Azure Rights Management service: https://docs.microsoft.com/en-us/information-protection/understand-explore/applications-support [↑](#footnote-ref-66)
66. Applications that support Azure Rights Management data protection: https://docs.microsoft.com/en-us/information-protection/get-started/requirements-applications [↑](#footnote-ref-67)
67. Microsoft Trust Center: https://www.microsoft.com/en-us/TrustCenter/default.aspx [↑](#footnote-ref-68)
68. Azure Key Vault: https://azure.microsoft.com/en-us/services/key-vault/ [↑](#footnote-ref-69)
69. Planning and implementing your Azure Information Protection tenant key: https://docs.microsoft.com/en-us/information-protection/plan-design/plan-implement-tenant-key [↑](#footnote-ref-70)
70. Bring Your Own Key (BYOK) with Azure Key Vault for Office 365 and Azure: http://download.microsoft.com/download/F/6/3/F63C9623-053F-44DD-BFA8-C11FA9EA4B61/Bring-Your-Own-Key-with-Azure-Key-Vault-for-Office-365-and-Azure.docx [↑](#footnote-ref-71)
71. Logging and analyzing usage of the Azure Rights Management service: https://docs.microsoft.com/en-us/information-protection/deploy-use/log-analyze-usage [↑](#footnote-ref-72)
72. Get Usage Logs from Azure RMS: http://download.microsoft.com/download/F/6/3/F63C9623-053F-44DD-BFA8-C11FA9EA4B61/Get-Usage-Logs-from-Azure-RMS.docx [↑](#footnote-ref-73)
73. Azure Key Vault Logging: https://docs.microsoft.com/en-us/azure/key-vault/key-vault-logging [↑](#footnote-ref-74)
74. Active Directory Rights Management Services Overview: https://technet.microsoft.com/en-us/library/74272acc-0f2d-4dc2-876f-15b156a0b4e0.aspx [↑](#footnote-ref-75)
75. Azure Information Protection with HYOK (Hold Your Own Key): https://blogs.technet.microsoft.com/enterprisemobility/2016/08/10/azure-information-protection-with-hyok-hold-your-own-key/ [↑](#footnote-ref-76)
76. Hold your own key (HYOK) requirements and restrictions for AD RMS protection: https://docs.microsoft.com/en-us/information-protection/deploy-use/configure-adrms-restrictions [↑](#footnote-ref-77)
77. Active Directory Rights Management Services Overview: http://go.microsoft.com/fwlink/?LinkId=84726 [↑](#footnote-ref-78)
78. How to configure a label for Rights Management protection: https://docs.microsoft.com/en-us/information-protection/deploy-use/configure-policy-protection [↑](#footnote-ref-79)
79. Trusted User Domain: https://technet.microsoft.com/en-us/library/dd983944(WS.10).aspx [↑](#footnote-ref-80)
80. Deploying Active Directory Rights Management Services with Active Directory Federation Services: https://technet.microsoft.com/en-us/library/dn758110(v=ws.11).aspx [↑](#footnote-ref-81)
81. File types supported by the Azure Information Protection client: https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-file-types [↑](#footnote-ref-82)
82. Azure Rights Management Administration Tool: https://www.microsoft.com/en-us/download/details.aspx?id=30339 [↑](#footnote-ref-83)
83. Azure Information Protection December update moves to general availability: https://blogs.technet.microsoft.com/enterprisemobility/2017/02/08/azure-information-protection-december-update-moves-to-general-availability/ [↑](#footnote-ref-84)
84. Administering the Azure Rights Management service by using Windows PowerShell: https://docs.microsoft.com/en-us/information-protection/deploy-use/administer-powershell [↑](#footnote-ref-85)
85. AzureInformationProtection Module: https://docs.microsoft.com/en-us/powershell/azureinformationprotection/vlatest/aip [↑](#footnote-ref-86)
86. Using PowerShell with the Azure Information Protection client: https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-powershell [↑](#footnote-ref-87)
87. File types supported by the Azure Information Protection client: https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-file-types [↑](#footnote-ref-88)
88. Windows Server 2008 R2 File Classification Infrastructure Technical White Paper: http://download.microsoft.com/download/D/D/3/DD3FB42C-D3C7-47C0-9431-40D80256FB0A/FCI\_TDM\_WP\_May\_09.pdf [↑](#footnote-ref-89)
89. FCI videos on Channel 9: http://channel9.msdn.com/tags/FCI/ [↑](#footnote-ref-90)
90. The Server Storage at Microsoft Team: https://blogs.technet.microsoft.com/filecab/tag/file-classification-infrastructure-fci/ [↑](#footnote-ref-91)
91. Microsoft IT Uses File Classification Infrastructure to Help Secure Personally Identifiable Information: https://www.microsoft.com/itshowcase/Article/Content/278/Microsoft-IT-Uses-File-Classification-Infrastructure-to-Help-Secure-Personally-Identifiable-Information [↑](#footnote-ref-92)
92. **How to create a custom file classification mechanism**: http://msdn.microsoft.com/en-us/library/jj552967(v=vs.85).aspx [↑](#footnote-ref-93)
93. **How to create a custom storage plug-in**: http://msdn.microsoft.com/en-us/library/jj552962(v=vs.85).aspx [↑](#footnote-ref-94)
94. **How to configure a file management task**: http://msdn.microsoft.com/en-us/library/jj552961(v=vs.85).aspx [↑](#footnote-ref-95)
95. Microsoft Data Classification Toolkit: http://www.microsoft.com/en-us/download/details.aspx?id=27123 [↑](#footnote-ref-96)
96. Dynamic Access Control: Scenario Overview: https://technet.microsoft.com/windows-server-docs/identity/solution-guides/dynamic-access-control--scenario-overview [↑](#footnote-ref-97)
97. File types supported by the Azure Information Protection client: https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-file-types [↑](#footnote-ref-98)
98. Deploying the Azure Rights Management connector: https://docs.microsoft.com/en-us/information-protection/deploy-use/deploy-rms-connector [↑](#footnote-ref-99)
99. FILE API CONFIGURATION: https://docs.microsoft.com/en-us/information-protection/develop/file-api-configuration [↑](#footnote-ref-100)
100. Using PowerShell with the Azure Information Protection client: https://docs.microsoft.com/en-us/information-protection/rms-client/client-admin-guide-powershell [↑](#footnote-ref-101)
101. Mail flow or transport rules: https://technet.microsoft.com/en-us/library/dd351127(v=exchg.150).aspx [↑](#footnote-ref-102)
102. Integrating Sensitive Information Rules with Transport Rules: http://technet.microsoft.com/en-us/library/jj150583(v=exchg.150).aspx [↑](#footnote-ref-103)
103. PCI Data Security Standards: http://www.pcicomplianceguide.org/ [↑](#footnote-ref-104)
104. Policy and compliance cmdlets: http://technet.microsoft.com/en-us/library/dd298082(v=exchg.150).aspx [↑](#footnote-ref-105)
105. Document Fingerprinting: http://technet.microsoft.com/en-us/library/dn635176(v=exchg.150).aspx [↑](#footnote-ref-106)
106. Use mail flow rules to inspect message attachments: http://technet.microsoft.com/en-us/library/jj919236(v=exchg.150).aspx [↑](#footnote-ref-107)
107. Policy Tips: http://technet.microsoft.com/en-us/library/jj150512(v=exchg.150).aspx [↑](#footnote-ref-108)
108. Messaging Policy and Compliance: http://technet.microsoft.com/en-us/library/aa998599(v=exchg.150).aspx [↑](#footnote-ref-109)
109. Create a DLP policy to protect documents with FCI or other properties: https://support.office.com/en-us/article/Create-a-DLP-policy-to-protect-documents-with-FCI-or-other-properties-1b9e3c6c-4308-4a20-b11e-c37b8013e177 [↑](#footnote-ref-110)
110. Overview of data loss prevention policies: https://support.office.com/en-us/article/Overview-of-data-loss-prevention-policies-1966b2a7-d1e2-4d92-ab61-42efbb137f5e [↑](#footnote-ref-111)
111. What the DLP policy templates include: https://support.office.com/en-us/article/What-the-DLP-policy-templates-include-c2e588d3-8f4f-4937-a286-8c399f28953a [↑](#footnote-ref-112)
112. What the sensitive information types look for: https://support.office.com/en-us/article/What-the-sensitive-information-types-look-for-fd505979-76be-4d9f-b459-abef3fc9e86b [↑](#footnote-ref-113)
113. Overview of data loss prevention policies: https://support.office.com/en-us/article/Overview-of-data-loss-prevention-policies-1966b2a7-d1e2-4d92-ab61-42efbb137f5e [↑](#footnote-ref-114)
114. What the DLP policy templates include: https://support.office.com/en-us/article/What-the-DLP-policy-templates-include-c2e588d3-8f4f-4937-a286-8c399f28953a [↑](#footnote-ref-115)
115. Create a DLP policy to protect documents with FCI or other properties: https://support.office.com/en-us/article/Create-a-DLP-policy-to-protect-documents-with-FCI-or-other-properties-1b9e3c6c-4308-4a20-b11e-c37b8013e177 [↑](#footnote-ref-116)
116. Office 365 Security & Compliance Center cmdlets: https://technet.microsoft.com/en-us/library/mt587093(v=exchg.160).aspx [↑](#footnote-ref-117)
117. Cloud App Security: https://www.microsoft.com/en-us/cloud-platform/cloud-app-security [↑](#footnote-ref-118)
118. Why you need a cloud access security broker in addition to your firewall: https://blogs.technet.microsoft.com/enterprisemobility/2016/12/20/why-you-need-a-cloud-access-security-broker-in-addition-to-your-firewall/ [↑](#footnote-ref-119)
119. What is Cloud App Security?: https://docs.microsoft.com/en-us/cloud-app-security/what-is-cloud-app-security [↑](#footnote-ref-120)
120. Azure Information Protection and Cloud App Security integration: Extend control over your data to the cloud: https://blogs.technet.microsoft.com/enterprisemobility/2016/11/07/azure-information-protection-and-cloud-app-security-integration-extend-control-over-your-data-to-the-cloud/ [↑](#footnote-ref-121)
121. Azure Information Protection integration: https://docs.microsoft.com/en-us/cloud-app-security/azip-integration [↑](#footnote-ref-122)
122. eDiscovery in Office 365: https://support.office.com/en-us/article/eDiscovery-in-Office-365-143b3ab8-8cb0-4036-a5fc-6536d837bfce [↑](#footnote-ref-123)
123. In-Place eDiscovery: http://technet.microsoft.com/en-us/library/dd298021(v=exchg.150).aspx [↑](#footnote-ref-124)
124. Site Mailboxes in the new Office: https://blogs.technet.microsoft.com/exchange/2012/08/22/site-mailboxes-in-the-new-office/ [↑](#footnote-ref-125)
125. Plan and manage cases in the eDiscovery Center: https://support.office.com/en-us/article/Plan-and-manage-cases-in-the-eDiscovery-Center-d955aeb8-0d48-4291-a8e2-f3b84f17943f [↑](#footnote-ref-126)
126. Manage eDiscovery cases in the Office 365 Security & Compliance Center: https://support.office.com/en-us/article/Manage-eDiscovery-cases-in-the-Office-365-Security-Compliance-Center-edea80d6-20a7-40fb-b8c4-5e8c8395f6da [↑](#footnote-ref-127)
127. Office 365 Security & Compliance Center cmdlets: https://technet.microsoft.com/en-us/library/mt587093(v=exchg.160).aspx [↑](#footnote-ref-128)
128. Azure Information Protection Developer's Guide: https://docs.microsoft.com/en-us/information-protection/develop/developers-guide [↑](#footnote-ref-129)
129. Rights Management SDK 4.2: https://docs.microsoft.com/en-us/information-protection/develop/active-directory-rights-management-services-multi-platform-thin-client-sdk-portal [↑](#footnote-ref-130)
130. Rights Management SDK 2.1: https://docs.microsoft.com/en-us/information-protection/develop/microsoft-information-protection-and-control-client-portal [↑](#footnote-ref-131)
131. Azure Information Protection samples on GitHub: https://github.com/Azure-Samples/Azure-Information-Protection-Samples [↑](#footnote-ref-132)
132. Azure Information Protection Documentation: https://docs.microsoft.com/en-us/information-protection/ [↑](#footnote-ref-133)
133. Microsoft Azure Information Protection whitepapers: https://aka.ms/aippapers [↑](#footnote-ref-134)
134. Enterprise Mobility + Security (EMS) Team blog: https://blogs.technet.microsoft.com/enterprisemobility/?product=azure-information-protection [↑](#footnote-ref-135)