

How to classify data

by  **kriptos**[®] Advanced
Security

www.kriptos.io

Classifying your data

Kriptos - Locate critical information automatically using AI.

Background

To protect sensitive information properly it is important to have classified the data first. Classifying data can be tricky as the classification can be subjective depending on the person or Organization. A document can also change classification depending on the age and relevance of the document although the content remains the same. With this in mind we will attempt to explore ways to classify the confidentiality level of a document in a good and repeatable manner.

INDEX

Background	1
Levels of confidentiality	2
Public	3
Internal use	3
Restricted use	4
Confidential	4
The Bell Curve	5
How to classify data?	6
Important information	6
Evolution of a document	7
Aftermath	9
Where is my data?	9
Access Control	10
Protect or delete	11
Analysis	12
Consequences of classification	13

Levels of confidentiality



What are levels of confidentiality and how can we use them.

Let's look at what the Information Security Management System (ISMS) framework ISO27001 has to say:

ISO27002, A.8.2 Classification of information:

Goal: To ensure that information has a protection level in co-ordination with its criticality for the Organization.

A.8.2.1 Information should be classified according to legal requirements, value, criticality and sensitivity in relation to unauthorized disclosure or modification.

So, what does ISO27002 say about how many confidentiality levels there should be and how to classify the information. Well ISO27002 does not actually mention any number of classification levels nor does it mention names or how to classify. This is up to the Organization to decide based on what is correct and gives meaning to them. What it does say is that each level should be given a name that gives meaning within the context of the classification and that the whole Organization uses the same classification criterions. The results should say something about the value of the information assets and how sensitive and critical they are for the Organization.

As we can see ISO27002 does not say anything about how many levels there should be or how to define them. This is entirely up to each Organization to define based on their needs. We will however in this document operate with 4 levels of confidentiality as this is most common. The confidentiality levels are public, internal use, restricted use and confidential. As creating the confidentiality levels are up to each Organization, so is defining the levels. The definitions below are therefore not cut in stone but merely examples.



Public



Internal Use



Restricted



Confidential







Public



Information that may, should or could be available to the general public with no special access restrictions.

This information does not need any protection and can be shared with anyone without any consequences to the Organization. There is no restriction on who can access the information within the Organization.

Examples of public information are:

-  The Organization webpages
-  Information published on the Organization social media
-  Material for marketing or promoting the Organization
-  Contact information for the Organization
-  Forms used to gather information
-  Information flyers or product descriptions

NOTE: Just because someone has made their information public does not mean that you can download the information and treat it as public yourself.












Internal use



Information that is not open for everyone but restricted to people within the Organization. There are no laws or regulations requiring the added protection, but the information has some value to the Organization that fosters caution.

The information is usually restricted to people within the Organization, hence the name internal use. If the information should be shared with people outside the Organization or there is a leakage the damage to the Organization would be minimal.

Examples of internal use documents are:












-  Work documents
-  Organization charts or maps
-  Most meeting of minutes documents
-  Outdated Budgets and accounting papers and documents
-  Policies, procedures, processes and knowledge documents
-  Internal communication like copies of internal mail
-  Work orders
-  Business records
-  Public contracts
-  Internal information documents
-  Most inventory documents

Restricted use



Information that is required by law, regulations or agreements to have a higher degree of protection. This information is usually restricted to those who need to know and should rarely be shared with anyone outside the Organization. The information has value for the Organization and a leakage can affect the Organization in terms of lost business opportunities, loss of confidence with business partners or fines.

Examples of restricted use documents are:







-  Certain types of HR data
-  Strategic financial data
-  Security sensitive policies, processes and procedures
-  Security configuration documents
-  Project plans
-  Strategic plans
-  Contracts
-  Analysis
-  Security information
-  Personal customer data
-  Information shared in trust by business partners

Confidential



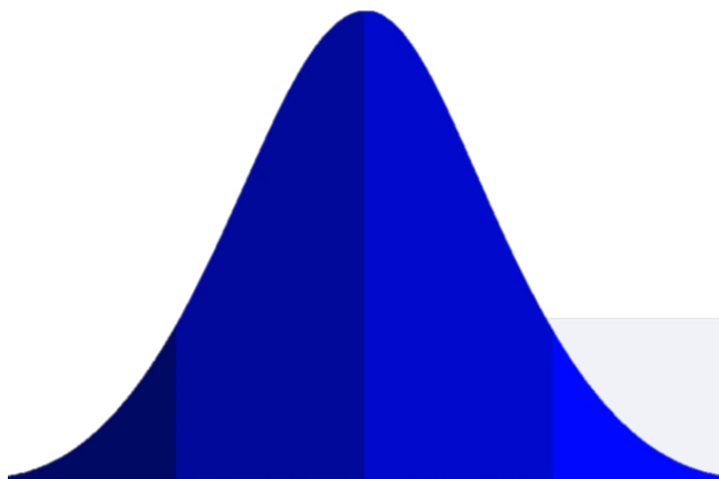
Information that is required by law, regulations or agreements to have a very high degree of protection. This information is usually restricted to a very few key personnel as the leakage of the information could have dire consequences for the Organization. Placement of data in this category should be done in cooperation with lawyers and information security personnel.

Examples of confidential use documents are:

-  Large amounts of personal data
-  Sensitive personal data such as health data, religious/political, sexual, criminal records etc.
-  Credit card information
-  Information with a high economic value
-  Financial data before informing the stock exchange
-  Intellectual property and research documents

The Bell Curve

The number of documents which should fall into each category or level should follow a bell curve where most documents fall into the categories of internal use or restricted. Unless your Organization makes a living publishing information, you probably want to be a little more cautious and use the



internal use category. Internal use still means that all, or at least most of the employees have access, but that it is not meant for the public. This also teaches employees to be careful and treat information as something that has a value without putting too many restrictions on it.

On the other end of the scale there should be more restricted documents than confidential. Confidential documents should by its nature be very valuable or secretive with access limited to few people. If there are many confidential documents, who are then these few people who can produce and process large amounts of confidential documents? Confidential data should stay secure in protected systems and databases and not in unprotected unstructured documents. An Organization that finds that they have as many or more confidential documents than restrictive is either in a very secretive business or they are doing something wrong.

HOW TO CLASSIFY DATA?



To better understand how to classify data it helps to go back to why you classify data or information. The goal of classifying data and information is so that you can better protect that which is important. So, when classifying data, you first need to understand what information is important to your Organization and why.

Important information

What is important information can vary from person to person and from Organization to Organization. It therefore helps to have some objective pillars to support your conclusion. As information can be important for many reasons, here are some helping pillars that raises the importance and criticality of the information:

- ☑ Regulated by laws and regulation
 - Personal information
 - Health Information
 - Genetical information
 - Credit card information
 - Information of importance to national security
- ☑ Trade secrets
- ☑ Insider information and Stock Exchange sensitive information
- ☑ Research and development data
- ☑ Information entrusted to you by a third party and where you are expected to key it secret.
- ☑ Information that could affect the security



When it comes to looking at importance of the information, it is valuable to also look at why it's important and the consequences of a leakage and loss of confidentiality. If a document is leaked but nobody finds it newsworthy, no fines will be handed out for regulation breach and no competitor cares, well then it is probably just internal use and not confidential. It does not matter if the person who created the document or who works with the information thinks it is critical and should be confidential. The classification should not be based on subjective evaluation on how important the information is for a person but how critical it is for the Organization. If a leakage has little or no consequence, the document is either public or internal. On the other side if a leak results in huge fines due to regulation breach, lots of negative media attention or loss of competitive advantage, it should be restricted or confidential.

Classification criteria should be easy to understand and yield consistent results. Even though the Organization uses tools to classify documents, the criteria for classifying should be simple and understood by all. Otherwise there will be little understanding for why a document has the classification level it has. Users who don't understand the classification of a document will complain or try to circumvent the protection measures put in place. Using complicated and fine-grained criteria to classify documents will only cause confusion and will not generate any added value. It is better to have a classification system that works but with a lower accuracy than a highly accurate system that fails because it is too complicated.

Evolution of a document

One of the reasons why classifying documents can be so difficult is because the same document can have many classifications during its lifecycle.

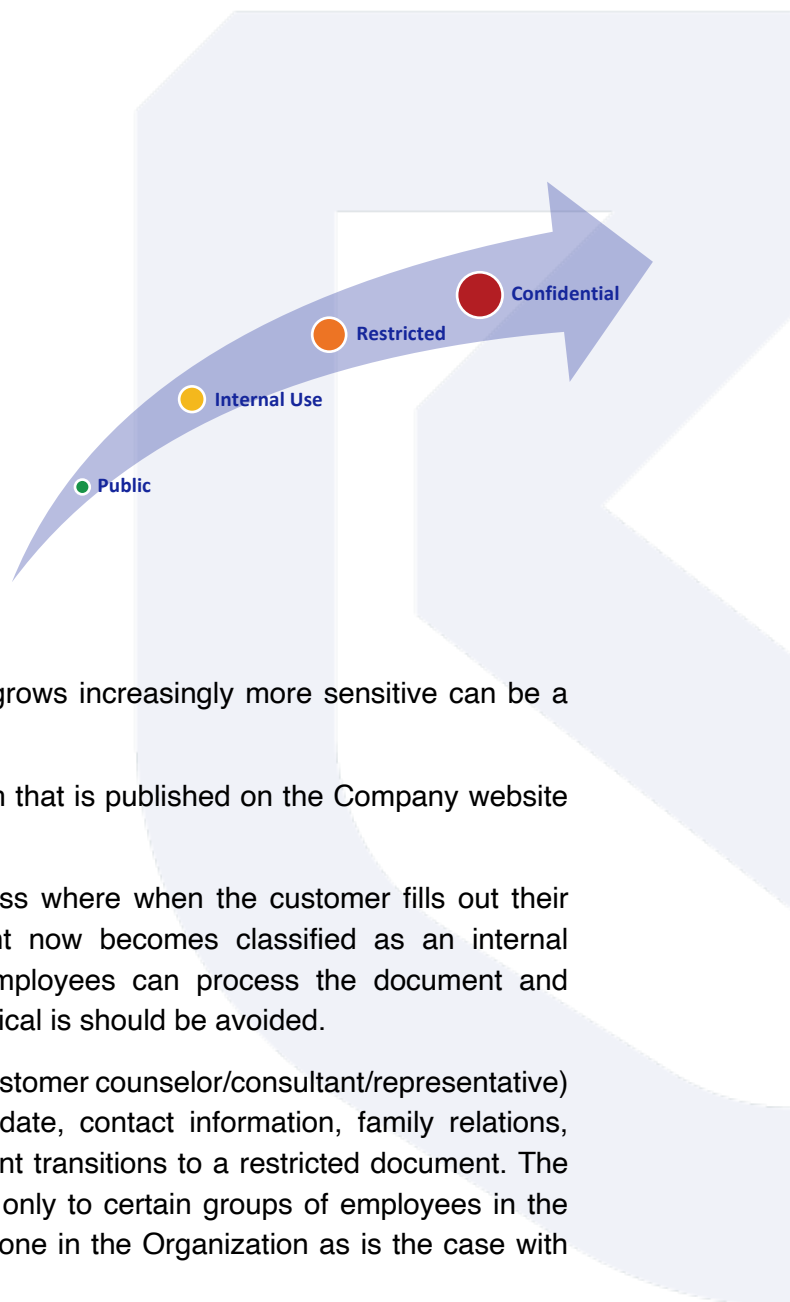
A document can start out as a public document, but as additional information is added to the document, it evolves and transitions through the categories internal use, restricted and may even reach confidential. The reverse is also possible as a confidential document is declassified due to age, relevance or loss of sensitivity.

An example of a public document that grows increasingly more sensitive can be a customer form.

Public: It starts out as a public form that is published on the Company website or given to customers to fill out.

Internal use: This starts the process where when the customer fills out their name and address. The document now becomes classified as an internal document. Only the company's employees can process the document and although a leakage would not be critical it should be avoided.

Restricted: As the customer (or a customer counselor/consultant/representative) fills out more information like birthdate, contact information, family relations, social security number, the document transitions to a restricted document. The document should now be available only to certain groups of employees in the Organization and not open to everyone in the Organization as is the case with internal use.



Confidential: The form is now completed with all necessary information filled out. This may also include financial statements, health data, credit card information and signatures, which brings the status to confidential. As a confidential document, access should be limited to business needs only, or upon request from the customer. The document would at this stage usually be stored or archived in a secure manner.

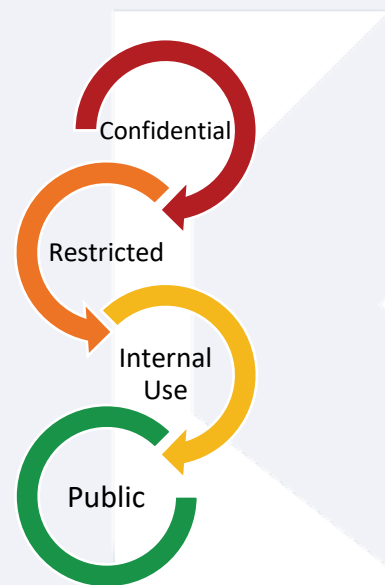
As we can see from this evolution the document increased its criticality level as more sensitive information was added. It is of course possible that the document would go directly from public to confidential as it would be filled out all at the same time. For documents where this is not the case, an Organization can for example do an analysis and see that documents older than one month that is still in the status internal or restricted probably stems from processes that was cancelled by the customer or for some other reason not completed.

An example of a document that is declassified may be financial statement reports.

Confidential: Before the financial statement reports are sent to the Stock Exchange, they are considered confidential. A leakage can cause fines, negative press and suspicions of insider trading.

Public: Once the reports have been sent to the stock exchange, they are considered public. The content of the document is exactly the same, the only difference is the date is now after the official release of the numbers.

With documents that evolve and transitions from one classification level to another it is important to consider the ongoing business impact of classification. A document that is sensitive enough to be classified as confidential due to its content should maybe be kept as internal or restricted until the process is finished. Otherwise the employees may lose access to the document, or the Organization may have to operate with giving many employees access to confidential documents.



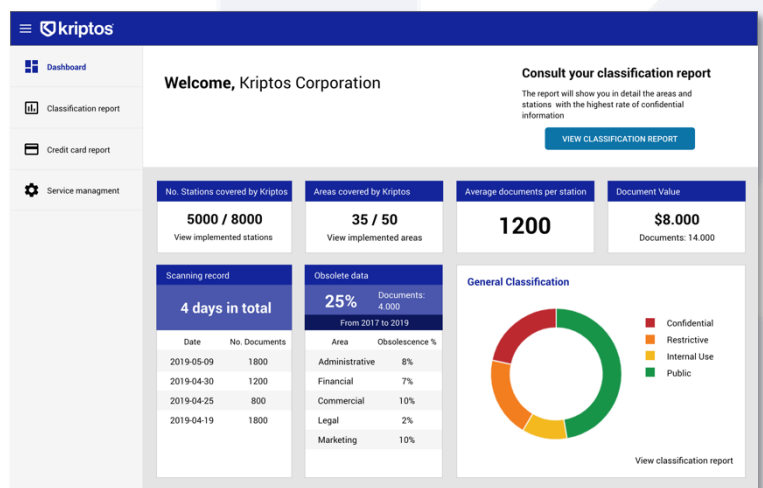
AFTERMATH

All your documents have been classified. Now what?

Classifying the documents was only the beginning. As referenced in ISO27002 A.8.2 **the goal is to protect the information according to its criticality**. Once you have classified all documents you can start to protect it accordingly. When the information is structured data located in a database accessed through a system, you can apply the necessary encryption and access controls. This is possible because you know where the data is, and you already have a system in place to provide you with role-based access controls. **When the data is in the form of thousands of unstructured documents, then what can you do.**

Where is my data?

To evaluate and reduce risk related to confidential documents, it is important to know where they are located. Are they located on an internal file server, spread around on different user machines or on a cloud server service like OneDrive, Google Drive or SharePoint? Getting an overview over where your data is located is a first step in reducing risk. Getting an overview over where the data is also means getting an overview of how many documents you are responsible for.



How well can you protect the documents at their current location? A leakage can come in many different ways:

- 🔒 A deliberate leak by one of the users
- 🔒 An unprotected server connected to the net
- 🔒 A document forwarded in a mail
- 🔒 A lost laptop/USB memory stick
- 🔒 An anonymous link to a folder on a cloud server
- 🔒 A computer hacked via phishing
- 🔒 A ransomware copying all files on a machine/server before encrypting them

The more sensitive/critical a document is, the less it should be floating around on file servers, desktops and laptops where identity and access control is difficult. Removing

confidential documents from the endpoints of “risky” users should therefore be a priority. These users should maybe not have access to confidential documents in the first place. By moving the documents to either a restricted file server or a cloud solution like SharePoint means that access control can be exercised to a greater degree. On SharePoint you can have one copy of the file that can be accessed by all who needs access instead of 10 copies on 10 different machines. You can easily remove access when it is no longer needed. You can log who has accessed the file, when they accessed it, if they created a link or downloaded the file. One file that you can control access to and log activities on is a lot safer than 10 copies of a file that you have little or no control over.

NB: Some regulations like GDPR prohibits you from storing personal data of a European outside of Europe without a documented guarantee of protection. This means you must be careful about where the physical location of your server is. This includes the physical location of your cloud servers like OneDrive and SharePoint.

Access Control

So now you know what data you have, at least in terms of classification and where it is located. Next step is to start exercise access control. Unstructured data usually means elevated vulnerability to data leakage. The reason for this is the lack of access control. If the data is structured in a database, you can easily find it. You know what data you have, where you have it and maybe with a fine-grained role-based access control where read and write access is given according to need. With unstructured documents this is different as there can be many copies of the same document located on different locations.

Here are some ways to restrict access to documents:

- 🔒 Password protect sensitive documents
- 🔒 Place sensitive documents on cloud solutions like SharePoint or Google Drive and limit access according to Organizational unit, project participation or based on individual need
- 🔒 Use Active Directory to limit users who have access to file servers or folders on a server.
- 🔒 Segregate network based on location, department or field.
- 🔒 Use a DLP solution to prevent sensitive documents from being copied or transferred from their secure location to a less secure location, like a laptop or mailbox.

Protect or delete



Let's be honest, how many people go through old reports, project status reports, spreadsheets, minutes of meetings, old guidelines, forms etc.? Very few! Why is that? Well we usually have too much to do to indulge ourselves with reading through old data. The unstructured documents are seldom more than a snapshot in time. The updated data is where the Master data is located which is usually in structured database. When looking for an interesting document to read you would look for the one that states final version, and not one of those 20 early draft versions. Since storage is so "cheap" we tend to keep everything whether we it or not. This means that there are a lot of documents that could just be deleted because they are already obsolete. That an Organization has 5000 documents per user would make sense if that Organization was in the business of making documents. Most companies are not in the document making business and so they do not need that much unstructured information.

When you see an "old" document, here are some questions that should be asked:

- Is this version the latest official version or some earlier draft version?
- How many duplicate copies exist of this document?
- Is the information still correct?
- Is the information still relevant?
- Is the information used to gather insight or support business decisions?
- Are there any legal requirement for storing the information?
- What is the retention policy (if any) for this kind of document?
- Would this document show up as "noise" when searching for an important document?

Confidential and restricted information must still be protected even though the document is no longer useful, so why spend resources protecting it. Save only those documents that are important to the business and those documents you are legally obliged to keep. Move the important documents to a safe location and delete the rest. Set up an automatic retention policy that ensures that documents are removed before growing old. Deleting old documents can come with many benefits like:

- Reduced risk for a leakage with reduced documents to protect.
- Less documents to protect.
- Reduced diskspace. Maybe you don't need to upgrade or buy additional capacity.
- Reduced diskspace means reduced energy need (reduced CO2 footprint).
- Reduced backup need.
- Less documents to classify.
- Easier to find the important documents that you really need.

With thousands of new documents created every day, removing or deleting old and outdated documents that no longer brings value not only makes perfect sense, it is necessary. This is especially true if those same documents carry a risk if they are leaked.

Analysis



Using an automatic data classification tool like Kriptos, gives you the possibility to perform several analyses to gain insight into risk related to data classification. With Kriptos, all new documents will be classified the next day. You can then either use the analysis reports that comes from Kriptos or transfer data into Excel or a SIEM tool like Splunk. Here are some examples of data analysis that you can perform:

- 🔍 Which Organizational department have the most confidential documents.
- 🔍 Which Organizational department produces the most new confidential documents.
- 🔍 What types of documents are confidential, excel spreadsheets, PDF reports, minutes of meeting documents, forms filled out by clients or other?
- 🔍 Have there been any spikes in production of confidential documents?
- 🔍 Are you finding copies of confidential documents laying around or are they limited to the safe location where they should be?
- 🔍 Who are the top risk users who have access to or who are sharing confidential documents?
- 🔍 Are users who should not have access to confidential information storing confidential documents on their laptops or on their area of OneDrive?
- 🔍 Is there an increase/decrease in production/storage of confidential documents among users who have resigned? They might be trying to copy confidential information, or they have already copied it and are now deleting the files before they leave.
- 🔍 Estimate the value of your unstructured information.
- 🔍 Find out how many documents you have which are liable to GDPR/HIIPA/PCI DSS etc.?
- 🔍 Combine knowledge of classification data with security log data to discover breaches or security incidents.

Consequences of classification

Classifying a document does not make it secure or protected. It is what you do with the document or the knowledge afterwards that decides how well protected the information is. Protecting the information will also come with a price. Limiting access through identity management and access control or installing a DLP to prevent leakage is a great way to protect the information. This can be good if the documents need to be protected but not so good if the documents are classified wrongly. By being too strict with the classification you are limiting access to information that should be available to a larger group of people. This can create annoyance with users who can no longer do their job. If the users are prevented from doing their job, they will look for other creative ways of circumventing the processes and controls. If a type of document is classified confidential but users need to share the document, they can:

- ❑ Create a new document and password protect it to prevent it from being classified
- ❑ Create a new document in a file format that is not being classified.
- ❑ ZIP file the document to prevent it from being classified.
- ❑ Encrypt the document to prevent it from being classified.

Classifying a document according to its confidentiality and then protecting it is a fine balance between securing sensitive information and allowing business processes to flow.