DATA
CLASSIFICATION
CREDITS

Luis Almagro
Secretary General
Organization of American States (OAS)

OEA Technical Team
Farah Diva Urrutia
Alison August Treppel
Belisario Contreras
Kerry-Ann Barrett
Diego Subero
David Moreno
Mariana Cardona
Jaime Fuentes
Kadri Kaska
Elsa Neeme
Klaid Mägi
Lauri Luht

AWS Technical Team
Abby Daniell
Michael South
Andres Maz
Melanie Kaplan
Min Hyun
DATA
CLASSIFICATION
Introduction

Organizations, people, billions of connected devices generate, process and consume all sorts of data every day. Over 2.5 quintillion bytes of all kinds of new data\(^1\) is generated every day to be analyzed, processed, and stored. The range of data is also diverse, from ones and zeros coming from simple IoT devices that signal an on/off event (e.g. a motion detector sensor), to weather, traffic, financial transactions, health, and social media among others. Similarly, governments, which is the focus of this paper, generate, manage and store petabytes of data. The diversity of data leads to the question about the right policies that a government should follow to classify and store the data it holds. Governments’ answer to this question has been the development of Data Classification policies that are specific guidelines to government organizations about how different type of data should be classified and then secured, handled, stored, and processed based on its classification.

The first question often asked is “why don’t we just protect all data at the highest level and save time?” For governments, this is just not feasible financially and it negates some other benefits of properly classifying and labeling different types of data. First, the highest levels of data protection have additional costs, which results in the potential to spend more protecting data than what it is worth. Another aspect is that if all data is treated the same and not labeled accordingly, then it can be difficult to implement proper access controls and sensitive data may be accessible by people who don’t have an official reason to have access to that data. And lastly, there are efficiencies that can be gained in managing and reporting on data that are properly organized, grouped, secured, and labeled based on classification.

Data classification allows organizations to think about data based on sensitivity and business impact, which then helps the organization assess risks associated with different types of data. Reputable standards organizations, such as the International Standards Organization (ISO) and the National Institute of Standards and Technology (NIST), recommend data classification schemes so that information can be more effectively managed and secured according to its relative risk and criticality, advising against practices that treat all data equally. Despite the fact that each organization processes and classifies data according to their respective needs, regulations

---

and even capabilities, there continues to be an evolving need to establish a basic baseline of security controls that provide appropriate protection against vulnerabilities, threats, and risks commensurate with the designated protection level, notably in the public sector.

The benefits of an organization’s effective data classification are multiple. Not only is an organization able to improve its accessibility and organizational efficiency, but also an effective data classification ensures that information receives appropriate protection in accordance with its sensitivity, value, and criticality as well as the nature and degree of risks contained in undue disclosure, damage or destruction.

The aim of this Whitepaper is to provide guidance for the development of a data classification system for the purposes of ensuring access to and protection of information generated and processed by governments. It is important to highlight that a data classification policy is necessary independently of the type of infrastructure used by an organization such as onpremises or cloud or mobile. A Data Classification policy provides guidelines to organizations about the level of security and processes associated to store and manage different type of data. Additionally, the recommendations contained in this Whitepaper can be employed regardless of the type of organization, however they are primarily aimed to provide governmental entities that provide public services with key considerations for this process.

|Structure|

This Whitepaper seeks to provide guidance for the development of a data classification system for the purposes of ensuring access to and protection of information generated and processed by governments. The Whitepaper examines data classification approaches existing at the national and international level, in order to offer data classification as a functional tool, and means to avert potential risks such as under or overclassification of information.

The Whitepaper is divided into five sections, which provide an overview of the principles of data classification, as well as recommendations for its establishment. To illustrate some of the existing public sector models Section II analyses the experience of the United States, the United Kingdom and Argentina in their implementation and overall data classification regulations. The case studies from the US and the UK are particularly relevant given their level of rigor and sophistication. While the case of Argentina highlights the experience of a country in the Latin America and Caribbean region. Above all, the recommendations of this Whitepaper should be applied to the context and needs of your organization in the establishment of a data strategy.
The implementation of information management in general and data classification in particular varies by type of organization and may even vary depending on the individual organization. However, certain fundamental principles are common across governments, nongovernment organizations, and commercial organizations. The following is a refinement of six principles expressed by national (and regional) legal sources and international organizations’ instruments for information management. The principles below should be used as guidance rather than a single, standing benchmark in the construction and/or refinement of an information management and data classification strategy.

Figure 1- Principles of data and information classification
1. **Openness, transparency, and societal values:** Classification should be used cautiously and in accordance with the sensitivity, value, and criticality of data. Access restrictions should only be chosen for cases where information disclosure may be detrimental to the legitimate interests and legal obligations of the organization itself, its staff, or third parties. In such cases, specified procedures should be strictly observed to ensure that the information is not compromised either purposely or inadvertently. The challenge will be to not overclassify for convenience or expediency; detrimental for transparency and public trust – and deprives stakeholders of ownership for their own risk management decisions.

2. **Content driven, technology neutral approach:** Information should be classified on the basis of its contents and the risks associated with the compromise of the content, regardless of its format, media, or origin. There should be no discrimination based on the format or media of the information – whether analogue (paper) or digital; stored in an information system, on storage media, on mobile devices, or in the Cloud. Likewise, the decision to classify information should depend on the content itself and not necessarily automatically derive from the source that information is based on, responds to, or is referring to). For example, reliance on public sources should not automatically determine that the aggregate information should be publicly releasable.

3. **Risk management approach:** Information should be afforded protection in accordance with the level of sensitivity, value, and criticality of the information; this is usually done in a graded approach based on levels corresponding to value and risk. A protection level circumscribes the set of measures to reduce risks to an acceptable level – i.e. the potential severity and likelihood – that information is compromised. In determining the level of sensitivity and value of the information, both the degree of potential damage of compromise (unauthorized disclosure, modification, or loss) as well as the potential value of the data should be taken into account.

4. **Proportionality:** Information shall be classified to an appropriate level which should be as low as possible, but as high as necessary.

5. **Clear roles and responsibilities:** with regard to data classification, policy and processes should be assigned for information security within the organization and upheld by management awareness and commitment to information security.

6. **Lifecycle approach:** As a part of an information management system, the classification system should have consideration for information throughout its lifecycle: from creation or receipt, storage, retrieval, modifications, transfer, copying, and transmission to destruction. Also, an organization’s information management/data processing policy should not be written in stone but regularly evaluated to ensure it corresponds to the needs of and expectations towards the organization.
What are existing public sector models?

Globalization has led a trend towards a convergence in data classification terminology. This convergence has notably been driven by the rigor of ICT industry standards (e.g. adherence to ISO/IEC, NIST definitions), the resulting regional political/legal developments (in particular in the European Union and its member states), but mainly the cross-domain interaction and interdependencies (e.g. a growing consideration of cybersecurity and data protection regulation for each other). Therefore, it is advisable to take such best practices into account when developing national definitions.

The United States (U.S.), the United Kingdom (U.K.), and Argentina have established data classification schemes for public sector data. Both the U.S. and U.K. governments use a three-tiered classification scheme with the majority of public sector data classified in the two lowest tiers. Argentina has been included as a case study as a regional example on implementation and challenges faced. The city of Washington, DC might also be a good model to highlight that has been widely applauded for the convergence of open data with data classification and without a national security component. Data classification schemes have a short list of attributes and associated measures or criteria that help organizations determine the appropriate categorization level.

|United States of America (U.S.)|

The U.S. government uses a three-tier classification scheme updated in Executive Order 135261 and based on potential impact to national security if it is disclosed (i.e. confidentiality):

1. **Confidential**—Information where unauthorized disclosure reasonably could be expected to cause damage to national security.

2. **Secret**—Information where unauthorized disclosure reasonably could be expected to cause serious damage to national security.

3. **Top Secret**—Information where unauthorized disclosure reasonably could be expected to cause exceptionally grave damage to national security.

Additionally, while not an actual classification, the U.S. also uses the term “unclassified data” to refer to any data that is not classified under...
the official three classification levels. Even with unclassified data there are some caveats for sensitive information, such as “For Official Use Only” (FOOU) and “Controlled Unclassified Information” (CUI) that restrict disclosure to the public or unauthorized personnel. And this does not account for the various data protection laws based on more narrow types of data such as individual tax data, criminal data, credit card data, healthcare data, and others.

Due to the narrow focus of the U.S. classification system, which does not directly consider data integrity and availability in its classification levels factors that should be required when assessing information protection requirements - NIST developed a three tiered categorization scheme based on the potential impact to the confidentiality, integrity, and availability of information and information systems applicable to an organization’s mission. Most of the data processed and stored by public sector organizations can be categorized into the following:

- **Low**— limited adverse effect on organization operations, organization assets, or individuals.
- **Moderate**— serious adverse effect on organization operations, organization assets, or individuals.
- **High**— severe or catastrophic adverse effect on organization operations, organization assets, or individuals.

<table>
<thead>
<tr>
<th>Data Classification</th>
<th>System Security Categorization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclassified</td>
<td>Low to High</td>
</tr>
<tr>
<td>Confidential</td>
<td>Moderate to High</td>
</tr>
<tr>
<td>Secret</td>
<td>Moderate to High</td>
</tr>
<tr>
<td>Top Secret</td>
<td>High</td>
</tr>
</tbody>
</table>

For many other national, provincial, state, and local governments, this dual system of classification and categorization may be too complex and not necessary to meet information assurance needs. In these situations, a simpler option may be to merge the two concepts into the single term “classification” addressing national security (if applicable) and the importance of all three pillars of information assurance – Confidentiality, Integrity, and Availability – to the organization’s mission and business. For this reason, use of the word “classification” in this document will imply the holistic approach of categorization for confidentiality, integrity, and availability rather than the narrower scope of national security impact.
United Kingdom (U.K.)

The U.K. government recently simplified its classification scheme by reducing the levels from six to three. These are:

1. **Official**— Routine business operations and services, some of which could have damaging consequences if lost, stolen, or published in the media, but none of which is subject to a heightened threat profile.

2. **Secret**— Very sensitive information that justifies heightened protective measures to defend against determined and highly capable threat actors (e.g., compromise could significantly damage military capabilities, international relations, or the investigation of serious organized crime).

3. **Top secret**— Most sensitive information requiring the highest levels of protection from the most serious threats (e.g., compromise could cause widespread loss of life or could threaten the security or economic wellbeing of the country or friendly nations).

The U.K. government has traditionally categorized approximately 90 percent of its data as “Official.”³ The U.K. uses a flexible, decentralized accreditation approach where individual agencies determine the cloud services suitable for “Official” data based on a cloud service provider’s (CSP’s) security assurance against 14 cloud security principles.⁴ Most UK government agencies have determined that it is appropriate to use reputable, hyperscale CSPs when running workloads with “Official” data.

The U.K. government laid out various considerations for all information security when stored using the Cloud:

1. **Official**— All information and assets categorized as Official is suitable for different GCloud⁵ services. Nevertheless, it is required that all risk owners to have a full understanding any GCloud accreditation. All Information Communication Technology (ICT) services must continue to follow the risk management process as set out in the UK government’s Information Assurance Standards, in addition to follow standard architectural approaches which must be hosted within the UK.

2. **Secret**— All ICT services dealing or storing Secret information must be accredited as appropriate in accordance to the secret threat model. Specific design patterns or advice should come from the National Technical Authority for Information Assurance (CESG). A preliminary risk assessment and implications of enabling functionality. Information exchange outside the Secret tier will be highly constrained and managed using shared accredited capability.

3. **Top Secret**— ICT systems designed must be accredited as appropriate in order to hold Top Secret materials. Customized architectural advice may be necessary.

---


⁵ The G-Cloud framework is an agreement between the UK government and suppliers who provide cloud-based services.
As early as 2004, the government of Argentina began to outline the requirements to form and implement a nationwide data protection strategy. This initial strategy encompassed creating a model of security policy, forming a committee on information security, establishing its functions as well as designating a coordinator to oversee the work of the committee. The policy was formalized in 2005 when the National Office of Technologies (ONTI), the Argentine entity responsible for the transformation and implementation of technology solutions in the public sector, enacted the Policy on the Security of Information Mode Decreto N°378, which was later updated and amended in 2014, based on a series of recommendations obtained from its 2013 review becoming Disposición 1/2015.

The policy designates best practices for the protection and management of assets as part of its risk management. The proprietors of the data and information are responsible for classifying the information based on the degree of sensitivity, document and update the classification of information, and define which users should have access to the information based on their functions and roles. Within the classification, the policy should be based on the following three factors: confidentiality, integrity and availability. Each of the three factors has a scale of 0 to 3, which then determine the degree of protection that it should receive. The scale in which it is divided is as follows:

- **Low Criticality**: Information is classified as public. The information is commonly known and used by any person or employee. Its unauthorized modification can be easily repaired, and does not compromise the operations of the Organization.

- **Medium Criticality**: Information is classified as reserved or for internal use. The information may be known or used by some employees of the organization and some external delegated authorities, its use could cause slight risks or losses for the Agency, the National Public Sector or third parties. Its unauthorized modification can be repaired, although it could cause slight losses for the public agency or associated third parties. Its permanent loss or of a day could cause significant damages of the operations in the organization.

- **High Criticality**: Information is classified as confidential or secret. This information can only be known by a group or very small group of employees, usually the top management of the organization, and its unauthorized disclosure or use could cause serious losses to the Public Sector or other associated third parties. The permanent loss could cause serious damage to the organization.

Recommendations for establishing a Data Classification system

Existing data classification systems recognize different levels of sensitivity, value, and criticality of the information as well as varied levels of severity and likelihood of compromise. Unless security levels for certain data are prescribed by law (e.g. for national security information or privacy) and/or require alignment with regional or international commitments, the definition of security levels is at the discretion of the particular organization. This does not necessarily mean that compliance with legal requirement necessitates a myriad of separate classification categories. Where the risk and required protections are equivalent, it is feasible to accommodate the required protections under one classification level.

The following section offers an outline of the primary phases of the data classification process. It does not substitute the systematic implementation of information security standards or legal requirements arising out of specific instruments, but intends instead to offer an overall understanding of the main steps needed to develop and implement a data classification system. There are four main steps:

**Audit**

- **Inventory of data assets**
  The first step to data classification within an organization is to carry out a data inventory, or a ‘data audit’. This activity should provide a broad understanding of the types of data and information processed within the organization, their value, sensitivity and criticality.

  This step also involves the identification of legal requirements that apply; and an audit of existing organizational or administrative policies and procedures for data management, including existing organizational roles and responsibilities in data processing.

- **Risk assessment**
  After the definition of data classification policies, the data classification system can be deployed. The next step is to conduct a risk assessment for the types of data processed which identifies and quantifies risks for severity and likelihood, and prioritizes risks against criteria for risk acceptance and objectives relevant to the organization. The result of this exercise should guide and determine the selection of appropriate technical and organizational measures as well as priorities for risk management. Risk assessments
should be periodic – recognizing that the technological and threat environment as well as security practices continuously evolve over time – and, preferably, comparable.\(^6\)

Risk assessment is the task of the data controller – in certain instances backed up by legal requirements, as discussed in the previous section.\(^7\) Applicable law may mandate that the controller be able to demonstrate that processing complies with established requirements and constraints (e.g. the GDPR does so with regard to personal data processing). See Annex I for some considerations on risk factors that could be considered when undertaking this process.

**Defining protection levels and their application**

Appropriate protection requirements – grouped by classification categories – should then be defined for each type of information asset.

The amount of data classification levels should be optimal for the organization’s needs. An overly nuanced approach is difficult to manage, may result in inconsistently protected data and increased risk, and confuse data controllers and processors. An overly simplified model presents the risk of over or under classification. A three tiered approach tends to meet both information security standards (ISO, NIST, national standards) as well as, in most cases, legal compliance expectations.

**Determining data management roles**

The next step is defining the roles and responsibilities of the organization and staff with regard to information classification and protection. Together with the roles, the risk management obligations appropriate to each role should be defined. The aim is to ‘translate’ the above into organizational routines by means of policies and procedures. This is also a good phase to revise and update existing internal regulations as a part of this process.

Ultimately, it is the organization as the ‘data controller’ who is responsible for compliance and shall be able to demonstrate such compliance (accountability).

**Implementation**

**Classification**

Based on the risk assessment, the risk level is assigned, considering each security objective (confidentiality, integrity, and availability) individually. An overall classification to the data is assigned according the highest value across the three factors.\(^8\) Some systems also recognize a combined level (‘high confidentiality, moderate integrity, low availability’).\(^9\)

**Considerations for Emerging Technologies – Cloud, Mobile, and IoT**

A risk based approach should be adopted for all technical evaluations and implementations, whether they be traditional on premise equipment, mobile devices, in the Cloud, or with Internet of Things (IoT) devices. Strategies for adopting emerging technologies should be influenced by the organizational risk strategy, but also provide feedback to update the organization’s risk strategy as new capabilities become available. An evaluation of the data assets, risk levels, and the confidentiality, integrity and availability requirements should give the organization an understanding for their risk tolerance as well as the acceptable combinations of deployment, service models, and locations that emerging technologies can offer.

---

\(^6\) ISO 27000:2018; ISO/IEC 27005 provides information security risk management guidance, including advice on risk assessment, risk treatment, risk acceptance, risk reporting, risk monitoring and risk review. Examples of risk assessment methodologies are included as well.

\(^7\) See, e.g., GDPR preamble section 75.


For example, one of the most misunderstood emerging technologies today is the “Cloud”. Where governments and organizations lack a data classification program, risk management processes, and focus on legacy technical controls rather than security objectives of the data, there tends to be fear, uncertainty, and doubt (FUD). This FUD impedes the organization from adopting emerging tech and benefiting from new capabilities, performance, and cost efficiencies.

A ‘staged migration’ approach can be helpful with regard to adopting emerging technologies. In that case, assets and services are initially assigned ‘macro categories’ (e.g. non sensitive and non critical, medium sensitive, and medium critical etc.) and a detailed classification is assigned per each asset and service as it is migrated to the cloud.¹⁰

| Monitoring |

- **Supervision and quality assurance**
  An appropriate entity should be assigned for supervision, advice and consulting, as well as revision of classification decisions – e.g. Chief Information Officer (CIO), Chief Data Officer (CDO, or Chief Information Security Officer (CISO) with dedicated responsibility for data classification, data risk decisions, and required protection measures. That entity should also be empowered to ensure quality assurance for the implementation of security controls, the suitability and adequacy of existing controls for meeting the desired security objectives, and any compliance requirements.

- **Continuous improvement and monitoring**
  After the data assets have been classified, the security procedures need to be implemented with a view of constant monitoring and assessment in order to continue meeting risk management and compliance requirements. In order to continue meeting the policy’s security objectives, it is advisable to develop security standards and implementation guides based on current technical and non technical capabilities, which can be updated to adopt new innovations more readily without having to update the policy.

| Review |

- **Periodic review and adjustment**
  Beyond the continuous monitoring and assessment, periodic systematic reviews enable adjustments to data access and review of classified data. A reclassification and revision methodology can ensure that security measures are applied that are suited to the current technology and threat/risk environment, but also the changing value and sensitivity of the classified data. Classified information should be reviewed regularly to prevent legacy information lingering, which is costly to store and manage. It is advisable to likewise review classification policies and procedures on a periodic basis.

---

Recommended resources

https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680084826

Data Classification for Cloud Readiness. Microsoft, April 2017.
https://gallery.technet.microsoft.com/Data-Classification-for-51252f03

Data Classification: Secure Cloud Adoption. AWS, June 2018.

https://fam.state.gov/fam/05fam/05fam0480.html; see 5 FAM 482.5 for classification categories.

Good Practice Guide for Securely Deploying Governmental Clouds. ENISA, 2013,


https://www.bsi.bund.de/EN/Topics/ITGrundschutz/itgrundschutz_node.html


NIST Special Publication 800-60 Rev. 1 (Volume 1, Volume 2), Guide for Mapping Types of Information and Information Systems to Security Categories


### Annex I. Risk Scenarios

The following categorization summarizes risk scenarios commonly recognized in the instruments considered in the previous sections of this Whitepaper. Rather than a predetermined catalogue of risks, it can offer guidance for developing a data classification system for the purposes of managing risks arising from breaches of information security (i.e. confidentiality, integrity or availability).

<table>
<thead>
<tr>
<th>Risks to person</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Effect of compromise on the physical safety and security of an individual,</td>
<td>- Effect of compromise on the physical safety and security of an individual, including direct or indirect threat to life or health, regardless of the individual’s relation with the organization (staff or third party);</td>
</tr>
<tr>
<td>including direct or indirect threat to life or health, regardless of the</td>
<td>- Effect of compromise to individual non material rights (where the result could be loss or breach of privacy, discrimination, damage to reputation, or another significant social disadvantage, or where a data subject might be deprived of their rights and freedoms or prevented from exercising control over their personal data);</td>
</tr>
<tr>
<td>individual’s relation with the organization (staff or third party);</td>
<td></td>
</tr>
<tr>
<td>- Effect of compromise to individual non material rights (where the result</td>
<td></td>
</tr>
<tr>
<td>could be loss or breach of privacy, discrimination, damage to reputation, or</td>
<td></td>
</tr>
<tr>
<td>another significant social disadvantage, or where a data subject might be</td>
<td></td>
</tr>
<tr>
<td>deprived of their rights and freedoms or prevented from exercising control</td>
<td></td>
</tr>
<tr>
<td>over their personal data);</td>
<td></td>
</tr>
<tr>
<td>- Effect of compromise to individual material rights and interests (where the</td>
<td></td>
</tr>
<tr>
<td>result could be e.g. identity theft or fraud, financial loss, or another</td>
<td></td>
</tr>
<tr>
<td>significant economic disadvantage);</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Risks to an organization’s operations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Effect of compromise on the effective operation and administration of the</td>
<td>- Effect of compromise on the effective operation and administration of the organization and its processes;</td>
</tr>
<tr>
<td>organization and its processes;</td>
<td></td>
</tr>
<tr>
<td>- Effect of compromise to free and independent internal decision making process and (internal) investigations;</td>
<td></td>
</tr>
</tbody>
</table>
| Risks to an organization’s assets or business interest | -Risk of financial loss to the organization; effect of compromise on the financial interests of the organization or those of other parties involved;  
-Effect of compromise on the partners of the organization, including over information exchanged with third parties under an expectation of confidentiality;  
-Effect of compromise of information covered by legal privilege;  
-Effect of compromise to the organization in commercial or political negotiations;  
-Risk to the organization’s reputation, stability, or security; |
|-----------------------------------------------|
| Risk to national security, public order, or foreign relations | -Effect of compromise on national security and defence capability (incl. technological and economic matters relating to national security) or prejudice security operations or activities;  
-Effect of compromise on the exercise of foreign relations (incl. foreign government information);  
-Effect of compromise on public order and the operation of security authorities  
-Effect of compromise of information concerning vulnerabilities or capabilities of systems, installations, infrastructures, projects, plans, or protection services relating to national security;  
-Effect of compromise on infrastructure and protection of information;  
-Effect of compromise to administrative or judicial interests, including an investigation or trial;  
-Effect of compromise on public trust of the organization and its operations. |

**Sources and examples:**
GDPR, ISO/IEC, NIST, ICC, national security law  
(US, Estonia and NATO/EU countries).
DATA
CLASSIFICATION