Data Classification

Secure Cloud Adoption

June 2018

aws

[ Secure Cloud Adoption ]
Notices

This document is provided for informational purposes only. It represents AWS’s current product offerings and practices as of the date of issue of this document, which are subject to change without notice. Customers are responsible for making their own independent assessment of the information in this document and any use of AWS’s products or services, each of which is provided “as is” without warranty of any kind, whether express or implied. This document does not create any warranties, representations, contractual commitments, conditions or assurances from AWS, its affiliates, suppliers or licensors. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>1</td>
</tr>
<tr>
<td>What is data classification?</td>
<td>1</td>
</tr>
<tr>
<td>What is the value of data classification?</td>
<td>1</td>
</tr>
<tr>
<td>What are existing public sector models?</td>
<td>1</td>
</tr>
<tr>
<td>U.S.</td>
<td>2</td>
</tr>
<tr>
<td>United Kingdom (UK)</td>
<td>4</td>
</tr>
<tr>
<td>Customer considerations for implementing data classification schemes</td>
<td>4</td>
</tr>
<tr>
<td>Considerations in establishing a data classification framework</td>
<td>5</td>
</tr>
<tr>
<td>AWS Recommendations</td>
<td>6</td>
</tr>
<tr>
<td>Leveraging AWS Cloud to support data classification</td>
<td>7</td>
</tr>
<tr>
<td>Recommended resources</td>
<td>9</td>
</tr>
</tbody>
</table>
Purpose

This paper provides insight into classification schemes for public and private organizations to leverage when moving data to the cloud. It identifies practices and models currently implemented by global first movers and early adopters, examines how implementation of these schemes can simplify cloud adoption, and recommends practices to harmonize national requirements to internationally recognized standards and frameworks.

What is data classification?

Data classification is a foundational step in cybersecurity risk management. It involves identifying the types of data that are being processed and stored in an information system owned or operated by an organization. It also involves making a determination on the sensitivity of the data and the likely impact arising from compromise, loss, or misuse. The term “classification” as used in this document will imply the holistic approach of data categorization for confidentiality, integrity, and availability rather than the narrower scope of national security impact.

What is the value of data classification?

Data classification has been used for decades to help organizations safeguard sensitive or critical data with appropriate levels of protection. Regardless of whether data is processed or stored in traditional on-premises systems or the cloud, data classification is a starting point for maintaining the confidentiality (and potentially the integrity and availability) of data based on the data's risk impact level. For instance, data that is considered “restricted” should be treated with a higher standard of care than “unrestricted” data consumed by the general public. 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. Each data classification level should be associated with a baseline set of security controls that provide appropriate protection against vulnerabilities, threats, and risks commensurate with the designated protection level.

What are existing public sector models?

The United States (U.S.) and the United Kingdom (UK) have established data classification schemes for public sector data. Both governments use a three-tiered classification scheme with the majority of public sector data classified in the two lowest tiers. The city of Washington, D.C. has also established a data classification program, using a five-tiered classification scheme that was widely applauded by open data advocates, and may be a good model for other local governments. Data classification schemes have a short list of attributes and associated measures or criteria that help organizations determine the appropriate categorization level.
The U.S. government uses a three-tier classification scheme updated in Executive Order 13526 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" (FOUO) and "Controlled Unclassified Information" (CUI) that restrict disclosure to the public or unauthorized personnel.

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.

NIST employs the concept of a "high watermark" when categorizing a system, which means that the overall system is categorized at the highest level across confidentiality, integrity, and availability requirements. When using national security classifications (e.g. Confidential, Secret or Top Secret), the above categorizations are applied depending on the classification of the data that the system will process (Table 1). According to Fiscal

---

1. Executive Order 13926 of December 2009, sets forth a system for classifying, safeguarding, and declassifying national security information. This EO supersedes earlier executive orders.
2. The U.S. also employs a releasability caveat to their classification system for data authorized to be shared with other national governments through bi-lateral or multi-lateral information sharing agreements. Two examples include Japan (CONFIDENTIAL//RELTO JPN) and the North Atlantic Treaty Organization countries (NATO) (SECRET//RELTO NATO).
Year 2015 data, U.S. federal departments and agencies categorized 88 percent of their systems into the low and moderate categories. AWS has regions and services that are accredited to support all types of data categories and classifications.

Table 1 — Aligning data classification to system security categorization

<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.

The city of Washington, D.C. implemented a new data policy in 2017 focused on being more transparent, while still protecting sensitive data. While Washington D.C. implemented a five tier model, these tiers can align with other widely-adopted three-tier classification schemes used in cloud accreditation regimes.

Level 0 — Open Data. Data readily available to the public on open government websites and datasets.

Level 1 — Public Data, Not Proactively Released. Data not protected from public disclosure or subject to withholding under any law, regulation, or contract. Publication of the data on the public Internet would have the potential to jeopardize the safety, privacy, or security of anyone identified in the information.

Level 2 — For District Government Use. Data that is not highly sensitive and may be distributed within the government without restriction by law, regulation, or contract. It is primarily daily government business operations data.

Level 3 — Confidential. Data protected from disclosure by law, regulation, or contract and that is either highly sensitive or is lawfully, regulatory, or contractually restricted from disclosure to other public bodies. This includes privacy-related data (e.g. personally identifiable information (PII), protected health information (PHI), payment card industry data security standard (PCI DSS), federal tax information (FTI), etc.)

Level 4 — Restricted Confidential. Data that unauthorized disclosure could potentially cause major damage or injury, including death to those identified in the information, or otherwise significantly impair the ability of the agency to perform its statutory functions.

1 https://octo.dc.gov/page/district-columbia-data-policy

United Kingdom (UK)

The UK government recently simplified its classification scheme by reducing the levels from six to three. They 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 well-being of the country or friendly nations).

The UK government has traditionally categorized approximately 90 percent of its data as “Official.” The UK uses a flexible, de-centralized 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. Most UK government agencies have determined that it is appropriate to use reputable, hyper-scale CSPs when running workloads with “Official” data.

Customer considerations for implementing data classification schemes

In addition to implementing a data classification scheme, it is equally important to determine data handling roles. ISO, NIST, and other standards place the responsibility of data classification on data owners, as they are the best positioned to determine the value, use, sensitivity, and criticality of their own data. Risk management obligations vary depending on the role of the parties that handle the data. In other words, data owners (i.e. controllers who generate and control content, such as agencies and ministries) and non-data owners (i.e. processors that handle data in order to provision services) should be subject to requirements appropriate for the roles they play. In the context of data classification, agencies work as the data owner and are responsible for classifying their data and determining the security accreditation that they expect their CSP to meet.

It is important to note that governments applying a blanket high classification level to all data (despite its true risk posture) do not reflect a risk-based, outcome-focused approach to security. Protecting data classified at higher levels requires a higher standard of care, which translates into the customer spending increased resources on securing, monitoring, measuring, remediating, and reporting risks. It is impractical to commit the significant

---


resources required to securely manage higher impact data for data that does not meet the requisite thresholds. Also, the additional controls placed on data at the lower classification levels can negatively affect the availability, completeness or timeliness of that data to the general workforce, customers, and/or constituents. Where risks can be managed so that data is handled at a lower classification level, governments will experience the most flexibility around how they use that data.

A risk-based approach means applying the security measures commensurate with the potential exposure and consequence of the loss of confidentiality, integrity, or availability of the data. Effective data classification schemes will apply the "high watermark" approach and classify data according to the highest measure that a particular data attribute meets. For example, in the U.S. model, if a data type has been categorized as "low" for confidentiality, "moderate" for integrity, and "low" for availability, the data type would be classified according to the highest factor- "moderate."

Considerations in establishing a data classification framework

Policy makers could consider adopting a three-tiered data classification scheme. In particular, NIST’s data classification scheme has been widely recognized in sector-specific, national and international certifications. In fact, governments such as the Philippines and Indonesia are evaluating and adopting data classification schemes that apply similar principles as the US and UK models. However, governments are best positioned to develop their own classification schemes based on organizational and risk management needs. Government agencies should select the appropriate cloud deployment model according to their specific needs, the type of data they handle, and assessed risk (refer to table below). Depending on the classification of the data, agencies will need to apply the relevant security controls (e.g., encryption) within their cloud environment.6

When assessing risk and determining security controls, it is important to understand how the commercial cloud works and how various services may differ from on-premises environments, the differences in implementation of controls, and that there may be alternate controls to consider as compared to traditional IT implementation. When agencies have fully evaluated the commercial cloud with the numerous security benefits available (e.g., improved availability and resiliency, improved visibility and automation, and continually audited infrastructure), agencies may find that the vast majority of their workloads can be deployed in the cloud with due regard to a particular classification, similar to what the US and UK have done. Globally, we are seeing governments increasingly leverage the native security benefits of commercial cloud and meeting their security and compliance requirements through appropriate data classification and implementation of security controls.

AWS Recommendations

In most cases, AWS recommends a three-tier data classification approach (Table 2). We have found this works well for governments and government organizations as well as commercial customers. The naming convention for the three tiers can be tailored to each. However, for organizations that have more complex data environments, additional tiers can be added. This may include a tier for open public data, separate and below from the “Unclassified” tier, or a tier for extremely sensitive and strategic data above the standard “Secret” tier.

Table 2 — Example Three-tier classification scheme

<table>
<thead>
<tr>
<th>Data Classification</th>
<th>System Security Categorization</th>
<th>Cloud Deployment Model Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclassified</td>
<td>Low to High</td>
<td>Accredited public cloud</td>
</tr>
<tr>
<td>Official</td>
<td>Moderate to High</td>
<td>Accredited public cloud</td>
</tr>
<tr>
<td>Secret and above</td>
<td>Moderate to High</td>
<td>Accredited private/hybrid/community cloud/ tightly-controlled public cloud</td>
</tr>
</tbody>
</table>

Phase I — Inventory. The very first step is to conduct an inventory of the various data types that exist in the organization. It doesn’t need to be minutely detailed, but enough to gain a broad understanding of what type of data exists, where within the organization it exists, how it is used, and if any of it is governed by a compliance regulation or policy. Once the inventory is complete, group the data types into one of the data classification tiers the organization has adopted. For example, general business correspondence may be assigned to Tier 1 for Unclassified data, sensitive contract information and privacy information may be assigned to Tier 2 for Official data, and national security or intellectual property may be assigned to Tier 3 for Secret data.

Phase II — Risk Assessment and Classify. We recommend that organizations conduct a risk assessment for each broad data type and assign a level of potential risk (Low, Moderate, or High) to each security objective - confidentiality, integrity and availability – with an associated risk matrix. The organization can then assign overall risk to the data (and supporting information systems) based on the high-watermark level across the three factors. Once classified, security controls should align with the risk to the type of information, ensuring to include any required compliance requirements for the specified data (e.g., PCI). This is where data labeling and handling guidance and procedures are developed and implemented as part of the administrative and technical controls. It is important to keep the classification system as simple as possible and to not over-classify data as a risk avoidance technique as it may incur additional costs when implementing controls and limit availability of data and capability opportunities.

Phase III — Security Assessment and System Authorization. A quality assurance process to assess the implementation of security controls is needed to ensure that the proper controls were selected and implemented for the specified system and data, that the controls work as designed, and that all (if any) compliance requirements are met. It is recommended for a senior IT leader (CIO or CISO) or organizational business leader, who is responsible for the system and data, review the security assessment and formally authorize the system for use in the organization.
Phase IV — Continuous Monitoring. In addition to ever-present external threats, normal system operations, updates, and changes to the environment can potentially affect security controls. It is critical that automated and manual means of monitoring the security of systems are included in the system’s development and operations to ensure that changes don’t adversely affect the security of a system and take it out of compliance.

Data Residency Consideration. AWS encourages customers to assess their data classification approach and hone in on which data needs to stay within their country or region, and why. By doing so, customers may find that their data, potentially even sensitive and critical data, may be stored and/or replicated elsewhere if there is no particular legal or policy geographical requirement. This can further reduce risk of loss in the event of a disaster and provide access to technologies and capabilities that may not be available in their area.

Leveraging AWS Cloud to support data classification

Cloud service providers (CSPs), such as AWS, provide a standardized, utility-based service that is self-provisioned by customers. CSPs do not have visibility into the type of data customers run in the cloud, which means CSPs do not distinguish, for example, personal data from other customer data when providing cloud services. It is the customer’s responsibility to classify their data and implement appropriate controls within their cloud environment (e.g., encryption). However, the security controls CSPs implement within their infrastructure and their service offerings can be used by customers to meet the most sensitive data requirements.

AWS services are “content agnostic” and offer the same high level of security to all customers, regardless of the type of content being stored. AWS adopts a high security bar across all services. These services are then queued for certification against international security and compliance “gold” standards, which translates to customers benefiting from elevated levels of protection for customer data processed and stored in the cloud. The risk events and threat vectors of greatest concern are largely accounted for through foundational cyber hygiene disciplines (e.g., patching and configuring systems), which CSPs can demonstrate through widely adopted, internationally-recognized security certifications such as ISO 27001,7 PCI DSS,8 and Service Organization Controls (SOC)9. In evaluating CSPs, customers should leverage these existing CSP certification frameworks so that the customer can appropriately determine whether a CSP (and services within the CSP’s service offerings) can support their data classification security requirements. We encourage organizations to implement a policy

7 ISO 27001/27002 is a widely-adopted global security standard that sets out requirements and best practices for a systematic approach to managing company and customer information that’s based on periodic risk assessments appropriate to ever-changing threat scenarios.
8 The Payment Card Industry Data Security Standard (also known as PCI DSS) is a proprietary information security standard administered by the PCI Security Standards Council (https://www.pcisecuritystandards.org/), which was founded by American Express, Discover Financial Services, JCB International, MasterCard Worldwide and Visa Inc. PCI DSS applies to all entities that store, process or transmit cardholder data (CHD) and/or sensitive authentication data (SAD) including merchants, processors, acquirers, issuers, and service providers.
9 Service Organization Controls reports (SOC 1, 2, 3) are intended to meet a broad range of financial auditing requirements for U.S. and international auditing bodies. The audit for this report is conducted in accordance with the International Standards for Assurance Engagements No. 3402 (ISAE 3402) and the American Institute of Certified Public Accountants (AICPA): AT B01 (formerly SSAE 16).
identifying which existing national, international, or sector-specific cloud certifications and attestations are acceptable for each level in the data classification scheme to streamline accreditation and accelerate migrating workloads to the cloud.

AWS offers several services and features that can facilitate an organization's implementation of a data classification scheme. For example, Amazon Macie can help you inventory and classify your sensitive and business-critical data stored in the cloud. Amazon Macie uses machine learning to automate the process of discovering, classifying, labeling, and applying protection rules to data stored in AWS. This helps you better understand where sensitive information is stored and how it's being accessed, including user authentications and access patterns.

Other AWS services and features that can support data classification include, but are not limited to:

- Identity and Access Management (IAM) for managing user credentials, setting permissions, and authorizing access.
- Encryption in transit and at rest to protect sensitive data
- AWS KMS or AWS CloudHSM for encryption key Management with AWS-generated keys or bring your own key (BYOK) with FIPS 140-2 validation
- AWS CloudTrail for extensive logging to track who, what, and when data was created, accessed, copied/moved, modified, and deleted
- AWS Systems Manager to view and manage service operations like patching along with AWS Inspector to conduct vulnerability scans
- AWS GuardDuty for intelligent threat detection supporting continuous monitoring requirements
- AWS Config to manage configuration changes and implement governance rules
- AWS Web Application Firewall (WAF) and AWS Shield to protect web applications from common attack vectors (e.g., SQL Injection, Cross-Site Scripting, and DDoS).

To review the entire list of AWS security services, visit https://aws.amazon.com/products/security/

10 Some classification schemes may include a fourth tier, "Public data," information which has been or is intended to be disclosed to the public, therefore disclosure carries no risk.
Recommended resources

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