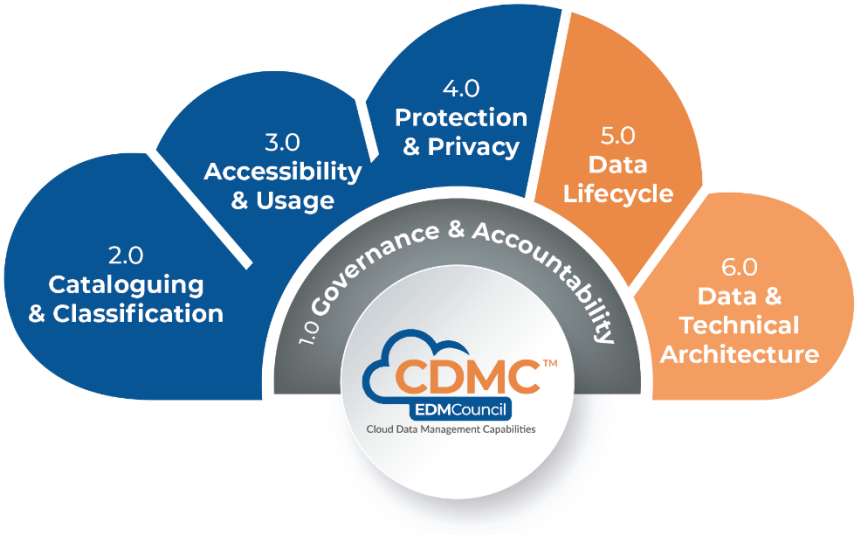
CDMC Key Controls and Automations

Version 1.1.1
September 2021

edmcouncil.org/page/CDMC

EDMCouncil



CDMC™ Terms of Use

This document is a constituent part of the Cloud Data Management Capabilities (CDMC™) model ("the Model") and is provided as a free license to any organization registered with EDM Council Inc. ("EDM Council") as a recipient ("Recipient") of the document. While this is a Free License available to both members and non-members of the EDM Council, acceptance of the CDMC Terms of Use is required to protect the Recipient's use of proprietary EDMC property and to notify the Recipient of future updates to the Model.

CDMC™ and all related materials are the sole property of EDM Council Inc. All rights, titles and interests therein are vested in the EDM Council. The Model and related material may be used freely by the Recipient for their own internal purposes. It may only be distributed beyond the Recipient's organization with prior written authorization of EDM Council. The Model may only be used by the Recipient for commercial purposes or external assessments if the Recipient's organization has entered into a separate licensing and Authorized Partner Agreement with EDM Council governing the terms for such use.

Please accept these CDMC™ Terms of Use by registering at
https://app.smartsheet.com/b/form/b3c66d25074f4422be037da82e64b65f

Feedback and Additional Information

Feedback on the document should be contributed via the Cloud Data Management Interest Community on EDMConnect: https://edmconnect.edmcouncil.org/clouddatamanagementinterestcommunity/home

For further information on the CDMC initiative please visit: https://edmcouncil.org/page/CDMC.

Any enquiries regarding EDM Council membership or CDMC Authorized Partnership should be directed to info@edmcouncil.org.

FOREWORD

The Cloud Data Management Capabilities (CDMC™) framework defines the capabilities necessary to manage and control data in the cloud effectively. Its creation represents an important milestone in the global adoption of industry best-practices for data management. The framework has been produced by the CDMC Work Group that was formed by the EDM Council in May 2020 with over 200 participants from over 70 organizations, including major consumers and providers of cloud services and technology in addition to leading advisory firms. The full framework will be published in September 2021.

This supplementary document is intended primarily for cloud service and technology providers. It summarizes and elaborates on the key controls required by organizations, equivalent to those implemented in their on-premises environments. It also highlights opportunities to support these controls with automation. Support of the controls and automation will streamline the adoption of cloud services.

The controls have been written with consideration of the many legal and regulatory requirements that exist and include the essential capabilities that a cloud service or technology provider may need to provide for an organization to use their services.

A specific organization may not need every control, but each of them could be necessary and applicable depending on the organization's requirements. As cloud service and technology providers can accommodate these automations and controls in alignment with the CDMC™ framework, the barrier to manage data in the cloud will continue to lower.

CDMC Component	CDMC Capability	Key Control and Automations
1. Governance & Accountability	Capability 1.1	(1) Data Control Compliance
	Capability 1.2	(2) Ownership Field
	Capability 1.3	(3) Authoritative Data Sources and Provisioning Points
	Capability 1.4	(4) Data Sovereignty and Cross-Border Movement
2. Cataloging & Classification	Capability 2.1	(5) Cataloging
	Capability 2.2	(6) Classification
3. Accessibility & Usage	Capability 3.1	(7) Entitlements and Access for Sensitive Data
	Capability 3.2	(8) Data Consumption Purpose
4. Protection & Privacy	Capability 4.1	(9) Security Controls
	Capability 4.2	(10) Data Protection Impact Assessments
5. Data Lifecycle	Capability 5.1	(11) Data Retention, Archiving and Purging
	Capability 5.2	(12) Data Quality Measurement
6. Data & Technical Architecture	Capability 6.1	(13) Cost Metrics
	Capability 6.2	(14) Data Lineage

Table 1: Overview of the key automations and controls and their positioning in the CDMC™ Framework

Acknowledgements

EDM Council would like to acknowledge the contribution of all participants in the CDMC initiative, and in particular the role of our CDMC Co-Chairs Oli Bage (LSEG) and Richard Perris (Morgan Stanley) and Project Executive Jubair Patel (Capco) in leading the initiative.

CONTENTS

INTRODUCTION

SCOPE OF CONTROLS

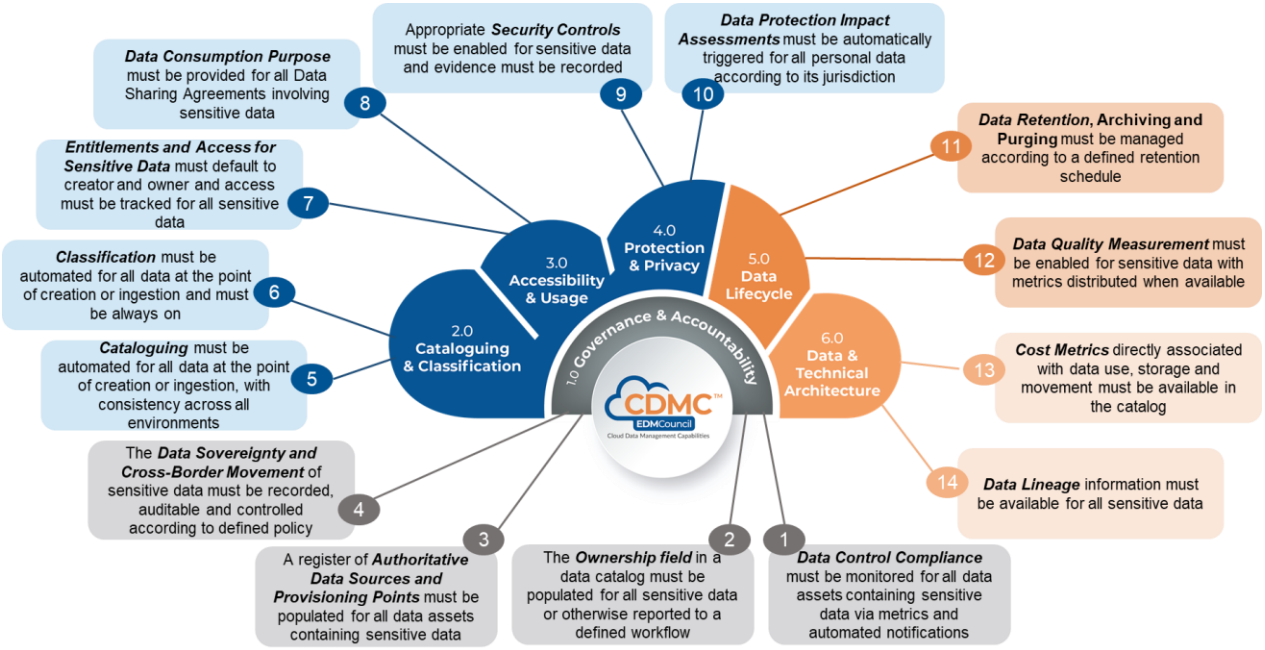
The framework addresses the control of data in cloud, multi-cloud and hybrid-cloud environments. Controls that address technology risks in other areas such as software development and service management are not within the scope of the document.

Many of the controls refer to being applicable to ***sensitive data***. Each organization will have a scheme for classifying their sensitive and important data and will determine the specific classifications to which the controls must be applied. Examples of classifications that may be in scope include:

- Personal Information (PI) / Sensitive Personal Data
- Personally Identifiable Information (PII)
- Client Identifiable Information
- Material Non-Public Information (MNPI)
- Specific Information Sensitivity Classifications (such as 'Highly Restricted' and 'Confidential')
- Critical Data Elements used for important business processes[1] (including regulatory reporting)
- Licensed data

KEY CONTROLS SUMMARY

The key controls are summarized in the following diagram:



Detail for each control is provided in the sections below. Words and phrases that are italicized and underlined are terms in the EDM Council Data Management Business Glossary.

[1] Important business processes and the threshold to be considered important are dependent on the maturity of the organization's data management and the extent of its data strategy.

CONTROL 1: DATA CONTROL COMPLIANCE	
Component	**1.0 Governance & Accountability**
Capability	**1.1 Cloud Data Management Business Cases are Defined and Governed**
Control Description	**Data Control Compliance** must be monitored for all *data assets* containing sensitive data via metrics and automated notifications. The metrics must be calculated from the extent of implementation of the CDMC Key Controls specified in subsequent sections.
Risks Addressed	An organization does not set or achieve its value and risk mitigation goals for cloud management. Data is uncontrolled and consequently is at risk of not being fit-for-purpose, late, missing, corrupted, leaked and in contravention of data sharing and retention legislation.
Drivers / Requirements	Organizations are required to demonstrate adequate control of data being created in or migrated to the cloud.
Legacy / On-Premises Challenges	Significant tranches of on-premises data do not have *data management* applied to them and consequently do not realize maximum value for the organization or can potentially pose an unquantified risk. When moving data to a new cloud environment, it is critical that organizations actively assess and apply the appropriate levels of *data management* to achieve their stated outcomes, apply controls to achieve this and measure compliance and value realization with those outcomes.
Automation Opportunities	• Where *evidence* of the existence of controls can be gathered automatically (including those controls referenced in subsequent sections of this document), the **Data Control Compliance metrics** may be calculated automatically. • Where the metrics fall below specified thresholds, alerts should be generated with automated notification to specified *stakeholders*.
Benefits	Cloud data is demonstrably controlled and supports the Cloud *Data management* business cases and risk mitigation requirements of the organization.
Summary	Organizations can demonstrate an awareness of the intended outcomes of cloud *data management* and focus on quantifiable value realization and risk mitigation.

CONTROL 2: OWNERSHIP FIELD	
Component	**1.0 Governance & Accountability**
Capability	**1.2 data ownership is Established for both Migrated and Cloud-generated Data**
Control Description	The **Ownership field** in a *data catalog* must be populated for all sensitive data or otherwise reported to a defined workflow.
Risks Addressed	Accountability for decisions on and control of sensitive data is not defined. Sensitive data is not effectively owned and consequently is at risk of not being fit for purpose, late, missing, corrupted, leaked and in contravention of data sharing and retention legislation.
Drivers / Requirements	Organizations have *policies* that require explicit ownership of data that is classified as sensitive.
Legacy / On-Premises Challenges	Significant amounts of legacy data do not have ownership recorded.
Automation Opportunities	The **Ownership field** in a *data catalog* must be populated "eventually" for sensitive data that is migrated to or generated within the cloud. • Automatically trigger workflows to enforce population when new *data assets* are created. • Provide the capability to automate workflows to review and update ownership periodically for sensitive data or when an owner leaves the organization or moves within the organization • Automatically trigger escalation workflows to address population gaps. • Implement ownership recommendations driven by the nature of data and ownership of similar data.
Benefits	Increased compliance with data ownership *policy*.
Summary	Infrastructure that supports the completion of data ownership information for sensitive data drives *policy* compliance.

CONTROL 3: AUTHORITATIVE DATA SOURCES AND PROVISIONING POINTS	
Component	**1.0 Governance & Accountability**
Capability	**1.3 Data Sourcing and Consumption are Governed and Supported by Automation**
Control Description	A register of **Authoritative Data Sources and Provisioning Points** must be populated for all *data assets* containing sensitive data or otherwise must be reported to a defined workflow.
Risks Addressed	Architectural strategy for an organization is not fully defined. Authorized sources have not been defined or suitably controlled. Data is duplicative and/or contradictory, resulting in process breaks, architectural inefficiencies, increased cost of ownership and accentuating existing operational risks on all dependent business *processes*.
Drivers / Requirements	An important responsibility of a *data owner* is to designate the *authoritative data sources* and *provisioning points* of data for a specific scope of data. *Policy* controls require a *data asset* to be identified as authoritative or not when it is shared.
Legacy / On-Premises Challenges	Identification and remediation of the use of non-authoritative sources or copies of data require significant manual effort.
Automation Opportunities	• Automatically enforce the labeling of sources of data as authoritative or non-authoritative. • Control the consumption of sensitive data from sources that are non-authoritative. • Default the labeling of sources to non-authoritative until reviewed and updated by the *data owner*.
Benefits	Infrastructure that can run automated workflows to identify and retire non-authoritative data provides a cost savings opportunity to eliminate the manual effort involved in this work.
Summary	*Data assets* automatically tagged as authoritative or non-authoritative will greatly simplify *policy* compliance and eliminate manual costs of controlling data sourcing and consumption.

CONTROL 4: DATA SOVEREIGNTY AND CROSS-BORDER MOVEMENT

Component	**1.0 Governance & Accountability**
Capability	**1.4 Data Sovereignty and Cross-Border Data Movement are Managed**
Control Description	The **Data Sovereignty and Cross-Border Movement** of sensitive data must be recorded, auditable and controlled according to defined *policy*.
Risks Addressed	Data can be stored, accessed and processed across multiple physical locations in cloud environments, increasing the risk of breaches to jurisdictional laws, security and privacy rules, or regulation. Breaches can result in various penalties, including fines, reputational damage, legal action and removal of licenses.
Drivers / Requirements	The *data owner* should understand the jurisdictional implications of cross border data movement and any region-specific storage and usage rules for a particular *data set*. *Policy*-specified controls must be applied when establishing cross-border *data sharing agreements* to support requests to use data from a particular location.
Legacy / On-Premises Challenges	Maintaining data about the physical location of data stores and *processes* is a significant undertaking and applying rules consistently across multiple different technologies is prohibitive.
Automation Opportunities	• Automatically capture and expose the physical location of all storage, usage and *processing* infrastructure applying to a cataloged *data set* • Provide the ability to trigger cross border *data sharing agreement* workflows (for international data transfer and international data requests). • Automatically trigger regional storage, *processing* and usage constraints, with the ability to escalate to a *data owner* where required. • Automatically audit and allow workflow to be triggered when sensitive data is being accessed from a location without a *data sharing agreement*.
Benefits	Reducing the manual *processing* and audit of *data sharing agreements* will significantly reduce the cost and risk of data *processing* in the cloud.
Summary	Codifying and automatically applying jurisdictional *data management* rules and cross border sharing agreements will significantly reduce the risk of *processing* data in the cloud. This will increase the adoption of cloud services and reduce complexity in the day-to-day *processing* of data in the cloud.

CONTROL 5: CATALOGING	
Component	**2.0 Cataloging & Classification**
Capability	**2.1 Data Catalogs are Implemented, Used, and Interoperable**
Control Description	**Cataloging** must be automated for all data at the point of creation or ingestion, with consistency across all environments.
Risks Addressed	The existence, type and context of data are not identified, resulting in the inability of all other controls to be applied that are dependent on the data scope. Data is uncontrolled and consequently is at risk of not being fit for purpose, late, missing, corrupted, leaked and in contravention of data sharing and retention legislation.
Drivers / Requirements	Organizations must ensure the necessary controls are in place for large or complex workloads that involve sensitive data such as client identifiers and transactional details. Knowledge of all data that exists is foundational to ensuring that all sensitive data has been identified.
Legacy / On-Premises Challenges	Organizations cannot scan and catalog the significant variety of *data assets* that exist in legacy on-premises environments. Without comprehensive catalogs of all existing data, organizations cannot be confident that all sensitive data within their *data assets* have been identified.
Automation Opportunities	• Ensure that catalog entries are generated for all data migrated to or created in the cloud. • Ensure catalog entries are generated for data in development, test and production environments and for both online and archived data. • Generate *evidence* of the comprehensiveness of the *data catalog*. • Implement APIs and support open data *standards* for *metadata* sharing and catalog interoperability. (Refer to the *CDMC Information Model*).
Benefits	An organization can guarantee that all data has been cataloged and can use this as the foundation on which to automate and enforce controls based on the *metadata* in the catalog.
Summary	This is the infrastructure describing what data exists, to see how much there is and how many different types there are. It is the foundation of all the other controls.

CONTROL 6: CLASSIFICATION	
Component	**2.0 Cataloging & Classification**
Capability	**2.2 Data Classifications are Defined and Used**
Control Description	***Classification*** must be automated for all data at the point of creation or ingestion and must be always on. • *Personally Identifiable Information* auto-discovery • *information sensitivity classification* auto-discovery • Material Non-Public Information (MNPI) auto-discovery • Client identifiable information auto-discovery • Organization-defined *classification* auto-discovery
Risks Addressed	Sensitive data is not classified, resulting in the inability of all other controls to be applied that are dependent on the *classification*. Data is uncontrolled and consequently is at risk of not being fit for purpose, late, missing, corrupted, leaked and in contravention of data sharing and retention legislation.
Drivers / Requirements	*Information sensitivity classification* (ISC) is required by most organizations' information security *policies*. An organization is required to know whether data is highly restricted (HR), classified (C), internal use only (IUO), or public (P), and if it is sensitive. Knowing whether data is sensitive is the foundation of most other controls in the framework. This requires certainty that all data has been cataloged and certainty that the sensitivity of the data has been determined.
Legacy / On-Premises Challenges	The variety of *data assets* in legacy environments impacts the ability to ensure that all data has been identified. Sensitive data may exist in *data assets* that have not been identified. *Classification* of *data assets* is often manual and can be both error-prone and expensive. Even where assets are identified, there may be gaps or errors in the *classification*. The proliferation of copies of data in legacy environments can lead to *classifications* in data sources not being carried through to copies of the data.
Automation Opportunities	• Apply *classification processing* to all data migrated to or created in the cloud. • Use automated *data classification* to identify the *classification* that applies. • Support organization-specified *classification* schemes. • Default *classifications* to the highest level until explicitly reviewed and changed.
Benefits	The operations team that is responsible for classifying data is expensive. Auto-*classification* can significantly streamline and reduce the amount of manual effort required to perform this function.
Summary	Auto-*classification* of data provides confidence that all sensitive data has been identified and can be controlled.

CONTROL 7: ENTITLEMENTS AND ACCESS FOR SENSITIVE DATA	
Component	**3.0 Accessibility & Usage**
Capability	**3.1 Data Entitlements are Managed, Enforced, and Tracked**
Control Description	1. **Entitlements and Access for Sensitive Data** must default to creator and owner until explicitly and authoritatively granted. 2. Access must be tracked for all sensitive data.
Risks Addressed	Access to data is not sufficiently controlled to those who should be authorized. This could result in data leakage, reputational damage, regulatory censure, criminal manipulation of business *processes*, or data corruption. Data is uncontrolled and consequently is at risk of not being fit for purpose, late, missing, corrupted, leaked and in contravention of data sharing and retention legislation.
Drivers / Requirements	Once the auto-classifier has identified sensitive *data assets*, enhanced controls should be placed on those *data assets*, including how *entitlements* are granted. The users that have access to data and how frequently they access it needs to be tracked.
Legacy / On-Premises Challenges	It is difficult to track which *data consumers* are using which *data assets* unless tracking is turned on and is consistent across all the data in the catalog.
Automation Opportunities	• Automate the defaulting of *entitlements* to restrict access to the creator and owner until explicitly and authoritatively granted to others • Automatically track which users have access to which data and how frequently they access it and store that information in a *data catalog*. • Provide all *data owners* access to the usage tracking tool • Hold *entitlements* as *metadata* to enable their use by any tool used to access the data.
Benefits	Tracking of data consumption enables consumption-based allocation of costs. Automation can reduce the cost of performing these allocations manually.
Summary	**Entitlements and access for sensitive data** at a minimum should be automated to default to being restricted to just the creator and owner of the data until they grant permissions to other people. Once other people have access to that data, monitoring should be in place to track who is using it and how frequently they are accessing it. Costs can then be correctly allocated.

CONTROL 8: DATA CONSUMPTION PURPOSE	
Component	**3.0 Accessibility & Usage**
Capability	**3.2 Ethical Access, Use, & Outcomes of Data are Managed**
Control Description	**Data Consumption Purpose** must be provided for all *data sharing agreements* involving sensitive data. The purpose must specify the type of data required and include country or legal entity scope for complex international organizations.
Risks Addressed	Data is shared or used in an uncontrolled manner with the result that the producer is not aware of how it is being used and cannot ensure it is fit for the intended purpose. Data is not shared in compliance with the ethical, legislative, regulatory and *policy* framework where the organization operates.
Drivers / Requirements	There are emerging ethical-use frameworks and *guidelines* that include specifications for what should happen when the use of data changes.
Legacy / On-Premises Challenges	It is difficult for human capabilities to recognize when the use of data has changed into a new kind of *processing* that could be protected under some regulatory or legal basis without specific authorization.
Automation Opportunities	• Record data access tracking information for sensitive data. • Enforce the capture of purpose, for example, integrated with *model* governance. • Provide alerts to the *data owner* or data governance teams when there is an additional use case for existing user access to sensitive data. • Recognize when specific technologies are employed (e.g., Machine Learning) and leverage usage and cost tracking to highlight potential new use cases.
Benefits	Streamlined ethical data accountability for data that is accessed for new purposes.
Summary	A *data sharing agreement* between a consumer and the authoritative source expresses the intent to use the data for a specific purpose. Automated tracking and monitoring of **data consumption purpose** can alert *data owners* and data governance teams when there is new or changed use.

CONTROL 9: SECURITY CONTROLS	
Component	**4.0 Protection & Privacy**
Capability	**4.1 Data is Secured, and Controls are Evidenced**
Control Description	1. Appropriate **Security Controls** must be enabled for sensitive data. 2. Security control *evidence* must be recorded in the *data catalog* for all sensitive data.
Risks Addressed	Data is not contained within the parameters determined by the legislative, regulatory or *policy* framework where the organization operates. Data loss or breaches of privacy requirements resulting in reputational damage, regulatory fines and legal action.
Drivers / Requirements	The sensitivity level of the data dictates what level of *encryption*, obfuscation and data loss prevention should be enforced. The requirements for **Security Controls** and Data Loss Prevention become increasingly more stringent as the sensitivity level of the data increases.
Legacy / On-Premises Challenges	It is difficult to ensure that *encryption* is always on for sensitive data.
Automation Opportunities	• Provide **security controls** capabilities including *encryption*, masking, obfuscation and *tokenization* that are turned on automatically based on the sensitivity of a *data set*. • Automate recording of the application of security controls.
Benefits	*Evidence* that the appropriate level of *encryption* is on and has been consistently applied is easy to produce. During a security audit, a *data owner* has a list of their data and how much of it is sensitive. Every piece of sensitive data can provide *evidence* that the data is encrypted, and there is a data loss prevention regime in place for all the compute environments it resides. Having security control *evidence* to deliver through the catalog rather than performing a forensic cyber review is a cost savings opportunity. A full-time team of employees typically handles this work.
Summary	Automation that enforces and records the appropriate *encryption* level based on a data asset's sensitivity level ensures security compliance and reduces manual effort to provide *evidence* of the controls.

CONTROL 10: DATA PROTECTION IMPACT ASSESSMENTS	
Component	**4.0 Protection & Privacy**
Capability	**4.2 A Data Privacy Framework is Defined and Operational**
Control Description	***Data Protection Impact Assessments*** (DPIAs) must be automatically triggered for all *personal data* according to its jurisdiction.
Risks Addressed	Data is not secured to an appropriate level for the nature and content of that *data set*. This results in either data being secured at greater cost and inconvenience than required or data loss or breaches of privacy requirements resulting in reputational damage, regulatory fines and legal action.
Drivers / Requirements	If a *data set* is classified as containing personal information, an organization needs to be able to demonstrate that it has performed a *data protection impact assessment* on it in certain jurisdictions.
Legacy / On-Premises Challenges	It is a very expensive workflow to initiate and complete a *data protection impact assessment* for the *data assets* classified as containing personal information. Identifying the DPIAs that need to be performed can be challenging, and completing those DPIAs can be very expensive.
Automation Opportunities	• Automatically initiate ***Data Protection Impact Assessments*** based on factors such as the geography of the data infrastructure, *classification* of the data or the specified consumption purpose.
Benefits	*Evidence* that all privacy requirements have been met for sensitive data is easy to produce since DPIAs are automatically initiated. Cost savings opportunities arise from more efficient identification of the need for DPIAs.
Summary	Automatically enforcing a DPIA on data that is classified as personal ensures *policy* compliance and reduces manual labor costs for that function.

CONTROL 11: DATA RETENTION, ARCHIVING AND PURGING	
Component	**5.0 Data Lifecycle**
Capability	**5.1 The Data Lifecycle is Planned and Managed**
Control Description	**Data Retention, Archiving, and Purging** must be managed according to a defined retention schedule.
Risks Addressed	Data is not removed in line with the legislative, regulatory or *policy* requirements of the organization's environment, leading to increased cost of storage, reputational damage, regulatory fines, and legal action.
Drivers / Requirements	Organizations have a master retention schedule that determines how long data needs to be retained in each jurisdiction it was created based on its *classification*.
Legacy / On-Premises Challenges	Organizations will have huge repositories of historical data, often retained to support the requirements of potential future audits. *Data sets* in different jurisdictions will have different retention schedules. It is difficult to comply with these requirements manually since different applicable legal requirements can modify the retention schedule.
Automation Opportunities	• Automate **data retention, archiving and purging** *processing* based on the data's jurisdiction, purpose and *classification* and according to a defined retention schedule. • Collect and provide *evidence* of the data retention, archiving and purging plan and execution.
Benefits	Automatically retaining, archiving, or purging data based on its *classification* and association retention schedule will reduce the manual effort required to perform this function and ensure *policy* compliance.
Summary	Organizations with this automation and control can provide the necessary *evidence* to verify that their data is being retained, archived or *purged* based on the retention schedule of its *classification*.

CONTROL 12: DATA QUALITY MEASUREMENT	
Component	**5.0 Data Lifecycle**
Capability	**5.2 Data Quality is Managed**
Control Description	**Data Quality Measurement** must be enabled for sensitive data with metrics distributed when available.
Risks Addressed	Data is not consistently fit for the organization's purposes, resulting in the inability to provide expected customer service, process breaks, the inability to demonstrate risk management, inefficiencies, and a lack of trust in the data and decisions based on flawed information.
Drivers / Requirements	*Data quality* metrics will enable *data owners* and *data consumers* to determine if data is fit-for-purpose. That information needs to be visible to both owners and *data consumers*.
Legacy / On-Premises Challenges	The limited application of *data quality* management in many legacy environments results in a lack of transparency on the quality of data and an inability for *data consumers* to determine if its fit-for-purpose. *Data owners* may not be aware of *data quality* issues.
Automation Opportunities	• Automatically deliver *data quality* metrics to *data owners* and *data consumers*. • Make *data quality* metrics available in the *data catalog*. • Automatically alert *data owners* to *data quality* issues.
Benefits	*Data consumers* can determine if data is fit-for-purpose. *Data owners* are aware of *data quality* issues and can drive their prioritization and remediation.
Summary	Providing clarity on *data quality* and support to ensure data is fit-for-purpose will help *data owners* address *data quality* issues.

CONTROL 13: COST METRICS	
Component	**6.0 Data & Technical Architecture**
Capability	**6.1 Technical Design Principles are Established and Applied**
Control Description	**Cost Metrics** directly associated with data use, storage, and movement must be available in the catalog.
Risks Addressed	Costs are not managed, detrimentally impacting the commercial viability of the organization.
Drivers / Requirements	As the cloud changes the cost paradigm from Capex to Opex, organizations require additional visibility on where data movement, storage and usage costs are incurred. Poor data architectural choices concerning data placement can incur additional costs through ingress or egress costs. For example, extra compute costs will be incurred when running data warehouse workloads on OLTP infrastructure.
Legacy / On-Premises Challenges	Limited need to manage data *processing* or storage costs at a *data asset* level. There is no line-item costing on the assets in a *data catalog*, so organizations cannot run a cost-analysis to understand where their *data management* costs are specifically being incurred.
Automation Opportunities	• Automatically track data assets' movement, storage, and usage costs and make this information available via the *data catalog*. • Support automated *policy*-driven cost management and optimization of data *processing*.
Benefits	*Data owners* would be able to understand who is using what data, the frequency of that access and the cost incurred to provide that data.
Summary	The financial operations infrastructure of *cloud service providers* is robust enough to identify accounts and operations that are incurring costs and associating those costs to specific *data assets* as line items in the *data catalog*.

CONTROL 14: DATA LINEAGE	
Component	**6.0 Data & Technical Architecture**
Capability	**6.2 Data Provenance and Lineage are Understood**
Control Description	***Data lineage*** information must be available for all sensitive data. This must at a minimum include the source from which the data was ingested or in which it was created in a cloud environment.
Risks Addressed	Data cannot be determined as having originated from an authoritative source resulting in a lack of trust of the data, inability to meet regulatory requirements, and inefficiencies in the organization's system architecture.
Drivers / Requirements	Organizations need to trust data being used and confirm that it is being sourced in a controlled manner. Regulated organizations produce lineage information as *evidence* that the information on regulatory reports has been taken from an authoritative source for that type of data. Consumers of sensitive data must be able to *evidence* sourcing of data from an authoritative source, for example, by showing lineage from the authoritative source or providing the provenance of the data from a supplier.
Legacy / On-Premises Challenges	Lineage information is produced manually by tracing the flow of data through systems from source to consumption. The cost of this approach and the consequences of producing incorrect data can be significant.
Automation Opportunities	• Record ingestion source of all data of specific *classifications* migrated to the cloud. • Record source-to-target lineage of all movement of data of specific *classifications* within the cloud environment. • Record destination lineage of all data of specific *classifications* egressing from the cloud (whether to on-premises or another cloud).
Benefits	Easy to produce *evidence* of the ***data lineage*** for regulatory reports. Major financial organizations incur significant costs producing this information manually and retrospectively.
Summary	Automatically tracking lineage information for data that feed regulatory reports would streamline the reports' data and eliminate cost by replacing the manual labor required to produce that information.

ADDITIONAL DOCUMENTATION

This document is a constituent part of the CDMC™ framework focusing on the key controls for effective management of data risk in cloud, multi-cloud and hybrid environments. This section provides a summary of additional parts of the overall framework.

CDMC Framework

Full documentation of the 6 components, 14 capabilities and 37 sub-capabilities of the CDMC framework, along with the 14 controls presented in this document. This 150+ page document details the objectives of each sub-capability and presents best practice advice written from both the data practitioner and cloud service and technology provider perspectives. A set of questions, artifacts and scoring guidance for each sub-capability provide the basis for organizations to perform capability assessments.

Reference: CDMC Framework Version 1.1 – published September 2021

CDMC Controls Testing Procedures

Specifications of tests of the 14 key controls within the framework to form the basis of certification of cloud products and services against the framework.

Reference: CDMC Controls Testing Procedures V1.1 – to be published Q4 2021

CDMC Information Model

An ontology that draws on and combines related open frameworks and standards to describe the information required to support cloud data management. This provides a foundation for interoperability of data catalogs and automation of controls across cloud service and technology providers.

Reference: CDMC Information Model Version 1.1 – to be published Q4 2021

Data Management Business Glossary

A standard set of over 150 data management terms, with definitions and commentary for each.

Reference: https://www.dcamportal.org/glossary/