Data Protection Best Practices

Version 2.0

January 27, 2025

Table of Contents

1 Introduction 2
2 Overview of Data Protection (DP) 2
2.1 Privacy and Data Protection 2
2.2 Information Assurance/Information Security and Data Protection 3
2.3 Storage and Data Protection 4
3 Managing Data 5
3.1 Data as an Asset 5
3.2 Data Lifecycle 6
3.3 Data Characteristics 8
4 Data Protection Drivers 10
4.1 Governance of IT 10
4.2 Compliance 12
4.3 Data Availability and Usability 12
5 Resiliency and Recovery 13
5.1 Metrics 13
5.2 Recovery Objectives and Capability 14
5.3 Data Protection (DP) Resilience and Recovery Technologies 16
5.4 Relevance of DP Technologies to the Data Lifecycle 17
6 DP Key Performance Indicators (KPIs) 18
6.1 Organizational KPIs 18
6.2 People KPIs 19
6.3 Physical KPIs 20
6.4 Technological KPIs 21
6.4.1 Resilience-oriented KPIs 21
6.4.2 Recovery-oriented KPIs 22
7 Storage Technologies Relevant to Data Protection 25
7.1 Resilience-oriented DP Technologies 25
7.1.1 RAID Storage 25
7.1.2 Mirroring 26
7.1.3 Erasure Coding 28

SNIA

7.2 Recovery-oriented DP Technologies 29
7.2.1 Cloning 29
7.2.2 Replication 30
7.2.3 Backups 32
7.2.4 Snapshots 38
8 Summary 39
9 Abbreviations 40
10 Acknowledgments 41
10.1 About the Authors 41
10.2 Reviewers and Contributors 42

List of Figures

Figure 1. Information Assurance – Interaction Between Security & Dependability 4
Figure 2. Data Lifecycle Based on DAMA-DMBOK2 6
Figure 3. Data Lifecycle Adapted for Data Protection 7
Figure 4. Model for Governance of IT 11
Figure 5 Quantification of reliability 13
Figure 6. Hypothetical Scenario – Multiple Disasters 15
Figure 7. Resilience and Recovery DP Technologies 17
Figure 8. Controller-based Mirroring 27
Figure 9. Storage Array Controller-based Mirroring 27
Figure 10. Fabric-based Mirroring 28

Executive Summary

“Data protection” can have different meanings, requirements, and solutions depending on the technological orientation and experience of the audience, be it with regards to privacy, security, or storage. This paper focuses on the storage aspect of data protection, covering topics such as the role of data protection throughout the data lifecycle, drivers for data protection, and contemporary data protection technologies. The paper provides useful descriptions of real-world requirements and scenarios, as well as recommendations and guidance.

1 Introduction

This paper is an update to the first version of the *SNIA Data Protection Best Practices* paper, published in 2017. As with the previous paper, the text is focused on storage-oriented data protection, which strives to safeguard important data from corruption, compromise, or loss and provides the capability to restore the data to a functional state should something render the data unusable.

New in this version of the paper is an exploration of issues and implications for data protection technologies commonly used in storage ecosystems as data traverses various lifecycle phases. It uses the DAMA International[1] data lifecycle model because it is a lifecycle model that is both simple and well suited to reflect modern data protection technologies and their usage.

This paper focuses on the more common data protection technologies which are grouped into either resilience-oriented or recovery-oriented categories. It also provides recommendations and best practices[2], along with example key performance indicators (KPIs).

2 Overview of Data Protection (DP)

The term "data protection" has different meanings, requirements, and solutions depending on the technological orientation and experience of the audience, be it with privacy, security, or storage. This overloaded term is used by the storage, information assurance/security, and privacy communities to describe very different requirements and technologies. Thus, context becomes important as ambiguities can lead to serious misunderstandings and lead to incidents having severe legal, operational, and/or financial consequences such as data breaches (with or without exfiltration), financial liability, and regulatory scrutiny, etc.

This section provides an overview of the general uses of DP, but much of this technical paper is focused on the storage aspect of DP.

2.1 Privacy and Data Protection

The terms data protection and data privacy are often used interchangeably, but there is an important difference between the two. The following ISO definitions and descriptions [6] can help highlight the differences:

- *Privacy* – freedom from intrusion into the private life or affairs of an individual when that intrusion results from undue or illegal gathering and use of data about that individual.
- *Privacy protection* – measures taken to ensure privacy. These measures include data protection and limitations on the gathering, combining, and processing of data about individuals.

[1] DAMA International (www.dama.org) is a not-for-profit, vendor-independent, global association of technical and business professionals dedicated to advancing the concepts and practices of information and data management. It promotes the understanding, development and practice of managing data and information as key enterprise assets to support the organization.

[2] Within the context of this paper, “best practices” is intended to mean proven techniques, leading practices, best methods, best techniques, etc. There is general recognition that such practices are not static and that they evolve over time to reflect changes in technology, the threat landscape, and legal/regulatory obligations.

- *Data protection* – implementation of administrative, technical, or physical measures to guard against unauthorized access to data.

In some jurisdictions, privacy is a right (i.e., freedom) of individuals to control or influence what information related to them may be collected and stored and by whom and to whom that information may be disclosed. The meaning of privacy can depend on prevailing laws and applicable regulations in a particular jurisdiction. In addition, these laws and regulations typically focus on specific types of data that must be protected (e.g., personal data or personally identifiable information).

Considering the regulatory perspective, data privacy is often focused on defining who has access to data while DP focuses on applying those restrictions. For example, data privacy defines the policies that DP tools and processes implement.

With the adoption of the European Union's (EU) General Data Protection Regulation[3] (GDPR), a set of data protection principles was introduced covering the lawful, fair, and transparent handling of personal information. Data handling involves the organization, collection, storage, structuring, use, consultation, combination, communication, restriction, destruction (eradication), and erasure of personal data. These data protection principles include purpose limitation, data minimization, storage period limitation, data quality and accuracy, data protection by design and by default, legal basis for processing, processing of special categories of personal data, measures to ensure data security, and the requirements in respect of onward transfers of data (e.g., accountability).

2.2 Information Assurance/Information Security and Data Protection

Information security, which covers more than just data, is typically defined in terms similar to NIST FIPS 199 [12]:

> *"protection of information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide confidentiality, integrity, and availability"*

The focus on confidentiality, integrity, and availability is consistently used within the context of security.

CNSSI No. 4009 [13] defines information assurance as:

> *"Measures that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation. These measures include providing for restoration of information systems by incorporating protection, detection, and reaction capabilities."*

Information assurance expands beyond confidentiality, integrity, and availability by incorporating information dependency elements (see Figure 1).

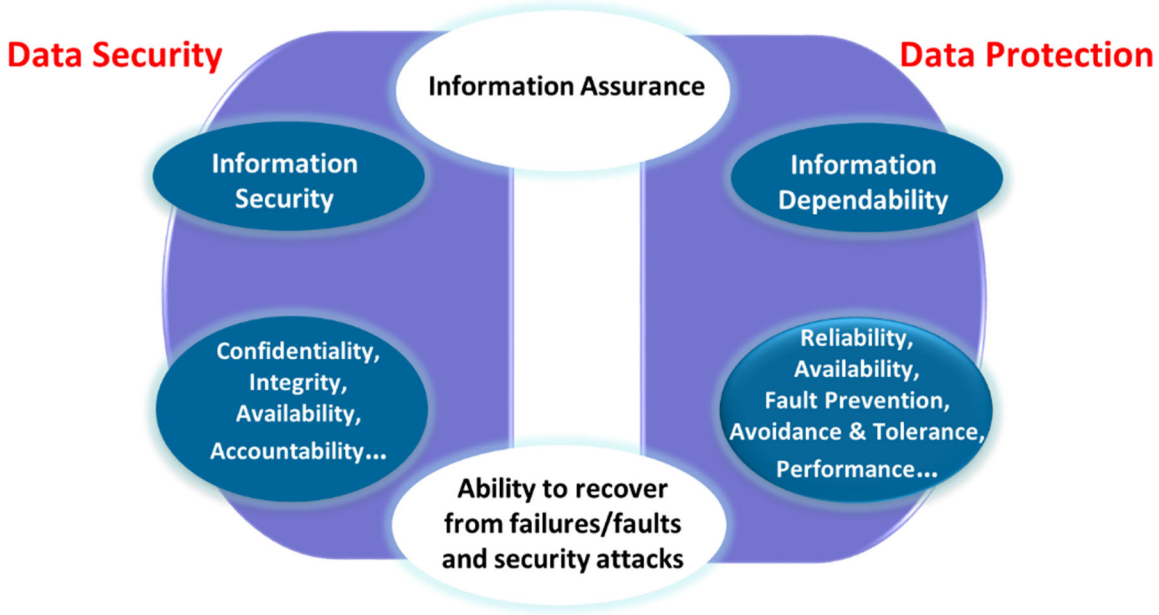
Dependability primarily focuses on how to quantitatively express the ability of a system to provide its specified services in the presence of failures, and is assessed through measures of:

- Reliability – probability that a system provides its services throughout a specified period.

[3] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

- Availability – fraction of time that a system can be used for its intended purpose within a specified period.
- Safety – probability that a system does not fail in such a way as to cause major damage to property, injury, or loss of life.
- Performability – quantitatively measures the performance level of a system in the presence of failures or faults.

Figure 1. Information Assurance – Interaction Between Security & Dependability[4]



It is worth noting that the dependability and security communities remain somewhat separated, but interactions between them are desirable. A simple and often cited difference between the two areas is that dependability focuses primarily on faults and errors in a system that are typically non-malicious in nature (primarily from the fault tolerance design area), while security focuses mainly on protection against maleficence.

2.3 Storage and Data Protection

From a storage perspective, DP is focused on safeguarding data from corruption, compromise, or loss, and on providing the capability to restore the data to a functional state should something happen that would otherwise render the data unusable.

[4] Figure is based on the "Information assurance: Interaction between security and dependability" figure in *Information Assurance – Dependability and Security in Networked Systems*, Qian, Joshi, Tipper, Krishnamurthy, 2008, New York, ISBN: 978-0-12-373566-9.

Storage-oriented DP can help an organization ensure that its data is not corrupted, is available for authorized use when needed, usable for its intended purposes, and is compliant with applicable legal or regulatory requirements (e.g., retention for a pre-determined period of time).

In addition to data availability and usability (see 4.3), DP can play a role with:

- Data immutability – capability of preventing any changes or modifications to data once it has been written or stored.
- Preservation and retention – the processes of ensuring data integrity, continued existence, and usability of stored data over a period.
- Data destruction/eradication – the process of erasing or removing data stored to writable media.

Storage technologies can be used to protect data by using disk, tape, or cloud backup to safely store copies of the data that can be used in the event of data loss or interruption of access.

The remainder of this paper primarily focuses on the storage-related aspects of DP as opposed to security or privacy perspectives.

3 Managing Data

3.1 Data as an Asset

Data is widely recognized as a personal and organizational asset, but data managers often struggle to manage it accordingly. This can result in liabilities and security risks as well as waste and missed opportunities.

Data has unique characteristics that make it different from other assets:

- Physical assets can be pointed to, touched, and handled but data is not physically tangible.
- Data is durable and does not wear out, though it can become stale or out-of-date.
- Data is easy to copy and transport.
- Data is not easy to reproduce if lost, destroyed, or corrupted.
- Data is not consumed when used, so it can be stolen without being gone (in the case of data breaches), unlike tangible assets.
- Data can be used simultaneously by multiple users such as people, applications, or other information systems.
- Data is dynamic and can be used for multiple purposes, and many uses of data generate still more data.

All these characteristics contribute to challenges related to managing data.

Further complicating the situation, the type or classification of data have different lifecycle management requirements. Data can be classified by:

- The function it serves (e.g., transactional data, reference data, primary data, metadata, etc.).
- Its content (e.g., data domains or subject areas).
- Its format.
- Its level of sensitivity and criticality.
- How and where it is stored and accessed.

Different types of data also have different requirements as they are associated with different risks and play different roles within an organization. These differing requirements mean that classification and control of data are of paramount importance.

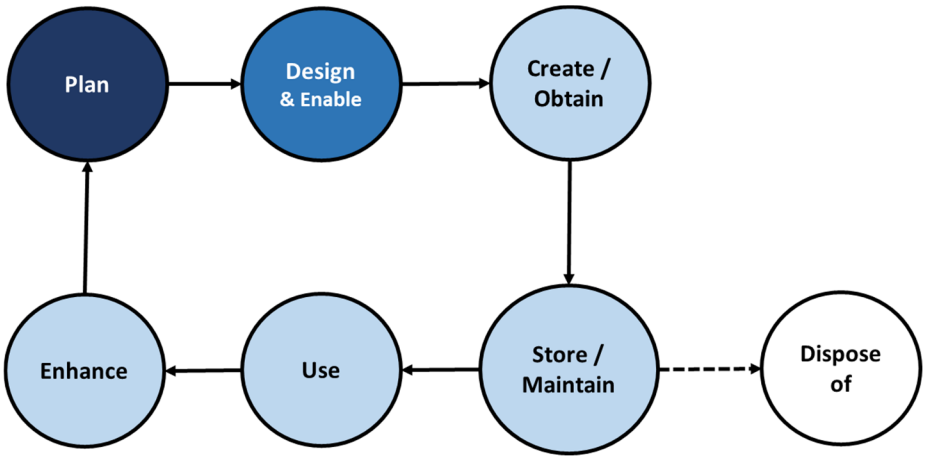
Data management is closely linked to technology management, so understanding the relationship between the two is important because decisions about technology can impact many facets of how data is managed. Data management focuses on ensuring that data is usable, secure, and trustworthy, whereas technology management focuses on building and maintaining infrastructure, systems, and applications that support data management. Successful data management requires sound decisions about technology, but managing technology is not the same as managing data.

3.2 Data Lifecycle

A lifecycle is the process of change and development of something throughout its useful life. Understanding and managing the lifecycle of a resource such as data is necessary to realize the full use and benefit of the resource. Lifecycles typically involve transitions through a series of phases or states. The transitions between states can be unplanned or carefully managed and policy-driven.

This paper uses the data lifecycle from DAMA International's Guide to the Data Management Body of Knowledge (DMBOK) [14]. The data lifecycle includes those processes that create or obtain data, that move, transform, store and enable it to be maintained and shared, that use or apply it, as well as that dispose of it. Figure 2 shows a slightly redrawn version[5] of the DAMA lifecycle.

Figure 2. Data Lifecycle Based on DAMA-DMBOK2



The "Plan" and "Design & Enable" phases Data Lifecycle are often associated with identifying objectives, planning information architectures, developing standards and definitions, as well as modeling, designing, and developing applications, databases, processes, organizations, etc. Many of these elements are addressed prior to a project going into production and are often difficult to modify after a data management system has entered production. Within the context of DP, these phases would identify

[5] "Dispose of" is shown to the right of "Store/Maintain" instead of below, as in the original diagram.

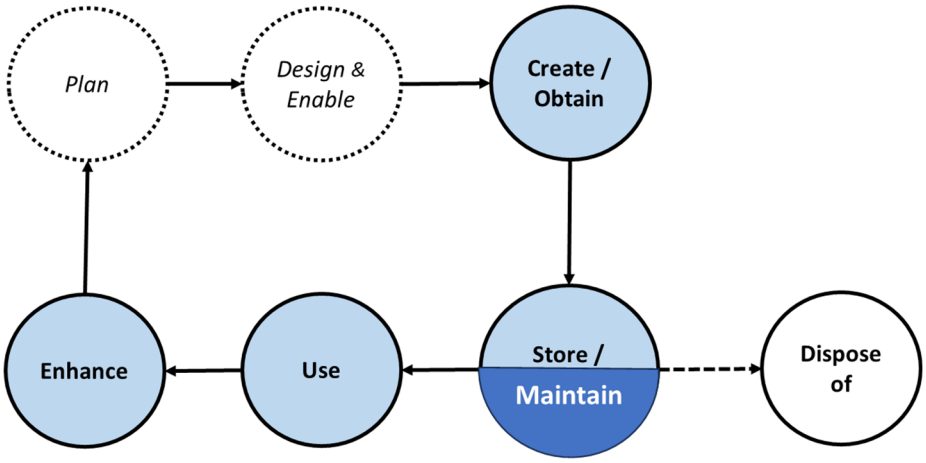
requirements and options along with initial implementations that are appropriate to the data. However, these phases are not likely to play a significant ongoing role. As such, they are not considered further in this paper, as depicted in Figure 3 below.

Another deviation from the DAMA lifecycle figure is shown in Figure 3 as part of the Store/Maintain phase: The split color coding is intended to show that there can be significant differences in this phase depending on whether the data is active or inactive.

Data is rarely static and can undergo periodic or continuous updating, cleansing, transformation, merging, enhancement, or aggregation, leading to iterations through states in its natural lifecycle. As data is used or enhanced, new data is often created, causing further iterations – the lifecycle of this newly created data typically develops its own separate data lifecycle – and so it is not necessarily reflected in the basic data lifecycle diagram of the data from which it was created.

Data and data management are intertwined with information technology (IT) and its management. Managing data requires an approach that ensures technology serves, rather than drives, an organization's data needs.

Figure 3. Data Lifecycle Adapted for Data Protection



The following is a summary of DAMA data lifecycle phases adapted for DP:

- *Create/Obtain* – This phase focuses on the generation, collection, acquisition, and capture of data.
- *Store/Maintain* – This phase focuses on storing activities associated with data storage, data quality and integrity, data classification, and data security, as well data maintenance activities such as data retention and digital preservation.
- *Use* – This phase focuses on viewing, processing, analyzing, organizing, editing/modifying/correcting, sharing, and interpreting data to convey meaningful information.
- *Enhance* – This phase focuses on identifying and acting on opportunities to get additional value from data. Data enhancement can also involve merging existing data with the most relevant, authoritative, externally-sourced data from third-parties. Such activities are also known as data enrichment or data append. Other enhancement activities can involve augmenting existing metadata with new knowledge or implementing new metadata requirements (e.g., enabling metadata-driven exchange of data in heterogeneous environments, leveraging ISO/IEC 11179 [2]

metadata registry standards). Generative artificial intelligence is also an example of an application that could be relevant for this phase.

- *Dispose of* – This phase focuses on the permanent, irreversible elimination (eradication) of data. A basic assumption is that all retention and/or preservation obligations have expired for data that is to be eradicated. In addition, all copies of the data are eradicated wherever it was stored (often referred to as undergoing data sanitization [9], which should not be confused with storage sanitization [15]).

As data transitions through its lifecycle, DP requirements at each stage can change significantly.

3.3 Data Characteristics

Data that is active (e.g., in use in production computing environments) is typically handled differently from inactive data. Active data is typically accessible without modification or reconstruction. Also, it is stored on directly accessible storage (e.g., internally or directly attached, on storage area networks, or on network attached storage) of computer systems. This storage is readily perceptible to the operating system and/or application software with which it was created and directly available to users. Inactive data is typically stored and maintained outside of the production computing environment (e.g., in archives and/or data repositories), and it is no longer required for everyday operations. The need to retain inactive data can be for its historical value, for audit/traceability purposes, and/or for legal or regulatory compliance obligations.

Data usability drives availability and accessibility requirements (see 4.3), whereas data trustworthiness drives confidentiality and integrity requirements. These requirements can change throughout the data lifecycle, so the specific measures used to meet the requirements can change as well.

Throughout the data lifecycle (see 3.2), it is important to understand the data availability, data integrity, and data confidentiality characteristics of each phase. The following provides a high-level summary of these characteristics:

- *Create/Obtain* – Data in this lifecycle phase tends to be active data.
  - Confidentiality – High-value data and sensitive data are typically identified within this phase so that they can be protected appropriately (e.g., with encryption). Data ownership and access controls are often established during this phase, based on data criticality and/or sensitivity.
  - Integrity – Data in this phase often serves as a baseline (i.e., used as the basis for future integrity checks), so initial integrity calculations (e.g., checksums and/or hashes) used later for accuracy and completeness checks are often established and recorded.
  - Availability – This lifecycle phase is focused on data generation/ingestion, so the availability of required resources (e.g., networking, computational, and storage) to support these activities is critical to avoid delays in data processing, and data loss or corruption. In some instances, real-time data generation/ingestion activities can compete for shared resources, necessitating quality of services mechanisms.
- *Store/Maintain* – Data in this lifecycle phase can be active or inactive data.
  - Confidentiality – For both active and inactive data, this lifecycle phase focuses on implementing appropriate security controls. This includes protecting data at rest, in transit, and in use. Common controls used to maintain data confidentiality include encryption, access controls, and data masking (i.e., make it unreadable to unauthorized users).

  - Integrity – For both active and inactive data, this lifecycle phase focuses on protecting data from unauthorized modification, deletion, or insertion/addition. Common security techniques for maintaining integrity include using digital signatures, message authentication codes, and data hashing, as well as technology-oriented mechanisms (e.g., RAID and CRCs). For inactive data, there can also be issues of preservation (e.g., in archives), retention (e.g., legal or regulatory), authenticity, provenance, and usability (e.g., obsolete formats and technologies). Integrity is critical as there are not likely to be other copies of the inactive data.
  - Availability – For active data, this lifecycle phase focuses on ensuring that systems are highly available and can withstand infrastructure failures. Common techniques used to maintain availability include load balancing, redundancy, and disaster recovery planning as well as countermeasures to prevent denial of service and other cyber-attacks (e.g., ransomware). Inactive data is often accessed infrequently, so while availability requirements exist, a lower level of performance in accessing the data is acceptable.
- *Use* – Data in this lifecycle phase tends to be active data.
  - Confidentiality – The same requirements and characteristics as active data during the Store/Maintain lifecycle phase along with provisions for privacy (see 2.1) requirements (e.g., purpose limitation, data minimization, accuracy, etc.).
  - Integrity – The same requirements as active data during the Store/Maintain lifecycle phase.
  - Availability – The same requirements as active data during the Store/Maintain lifecycle phase.
- *Enhance* – Data in this lifecycle phase tends to be active data.
  - Confidentiality – The same requirements as active data during the Use lifecycle phase.
  - Integrity – The same requirements as active data during the Use lifecycle phase.
  - Availability – The same requirements as active data during the Use lifecycle phase.
- *Dispose of* – Data in this lifecycle phase can be active or inactive data, but in either case, the data ceases to exist (is eradicated) as an outcome of this lifecycle phase.
  - Confidentiality – The eradication process of sensitive and high-value data necessitates careful attention to avoid data breaches (e.g., proper data sanitization and/or storage sanitization).[9]
  - Integrity – Other than ensuring that the correct data is eradicated, integrity is no longer an issue.
  - Availability – Other than ensuring all retention and/or preservation obligations have expired for the data, and that all copies of that data are available to be eradicated, availability is no longer an issue.

Throughout the data lifecycle, DP is more likely to play an important role in ensuring data availability than play a role for integrity and confidentiality.

4 Data Protection Drivers

4.1 Governance of IT

ISO 37000 [16] describes governance of organizations as laying the foundation for the fulfilment of the purpose of the organization in an effective, responsible and ethical manner in line with stakeholder expectations. ISO/IEC 38500 [10] provides guidance on the governance implications of the use of IT, data and digital capabilities by an organization. This governance is applied broadly to the use of IT including emerging technology, data and digital capabilities.

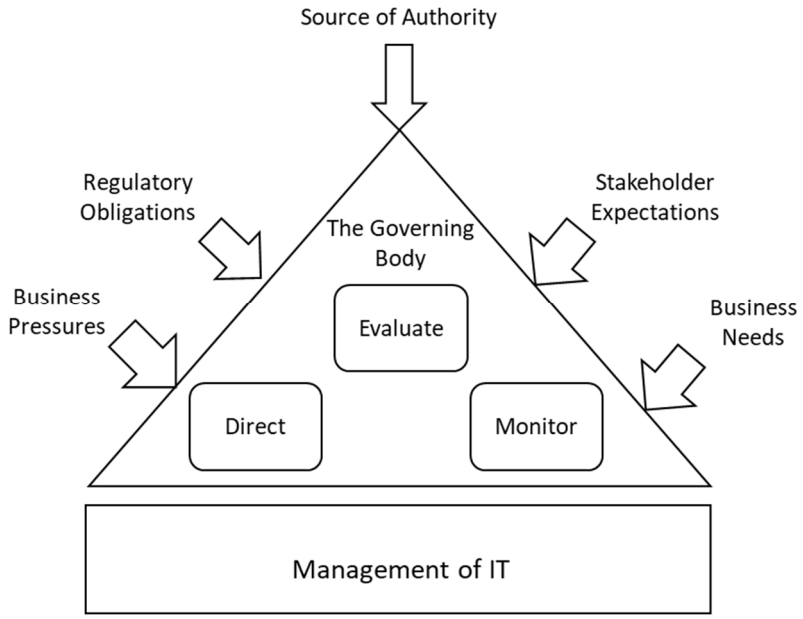
ISO/IEC 38500 also states that the whole organization operates within a governance framework. The framework for the governance of IT involves the strategies, policies, decision-making structures, and accountabilities through which the organization's use of IT operates today and plans for its future. Responsibility for specific aspects of IT may be delegated to management within the organization. However, accountability for the effective, efficient and acceptable use of IT by an organization remains with the governing body and cannot be delegated.

The governance of IT practice (see Figure 4), according to ISO/IEC 38500, is comprised of the following main tasks:

- *Engage Stakeholders* – relevant stakeholders (e.g., internal users of IT, employees, and developers, as well as external customers, suppliers, and governments) of the use of IT by the organization should be identified, consulted, and appropriately engaged; such engagement can help set appropriate policy and behavior so that the desired governance outcomes are understood, and its accountability obligations can be set.
- *Evaluate* – examining and making judgement on the current and future use of IT, including plans, proposals, and supply arrangements (whether internal, external, or both); consideration should be given to the external or internal pressures acting upon the organization, such as technological change, economic and social trends, regulatory obligations, legitimate stakeholder expectations, and political influences.
- *Direct* – assigning responsibility for, and directing preparation and implementation of strategies and policies to ensure that the use of IT meets business objectives; strategies should set the direction for investments in IT and what IT should achieve, and policies should establish acceptable behavior in the use of IT.
- *Monitor* – monitoring, through appropriate measurement systems, the performance of IT, which is in accordance with strategies, particularly with regard to business objectives as well as ensuring that IT conforms with external compliance obligations (whether regulatory, legislation, or contractual) and internal work practices.

In Figure 4, stakeholder engagement is shown as implied responses to business pressures, regulator obligations, source of authority, stakeholder expectations and business needs. In addition, Management of IT is shown underpinning the governance of IT; this management practice uses a management system of interrelated or interacting elements to establish management policies, objectives and processes to achieve those objectives. Additionally, there is a commitment of continual improvement to the management system.

Figure 4. Model for Governance of IT



This background information on the governance of IT is provide her because IT, and the data from which it unlocks value, is becoming increasingly effective and strategically significant to most organizations. This makes the governance of IT increasingly important for the organization – and stakeholders have high expectations of the outcomes of good governance of IT.

Gathering data and processing it into information to be used for decision making is the main use of IT. The decisions that utilize data can be made by people or machines both within the organization and outside it.

To ensure the appropriate use of data for decision making, the organization should ensure the data it uses:

a) Is appropriately classified so it can be disseminated, protected, and processed according to its classification.
b) Is delivered in the appropriate format, time, and quality for the decision-making process.
c) Respects the compliance obligations and restrictions placed on its use by others.
d) If disseminated to others, the rights over the data are respected.
e) The data is utilized ethically by the organization.

Recognizing that data is a valuable resource for decision making and taking the appropriate actions to manage and protect that resource can result in strategic use of data, responsible use of data, and ensuring requisite data quality. In addition, DP technologies can be critical to ensuring the ongoing integrity and availability of data.

An organization's governance maturity can play an important role in determining DP requirements. More mature governance practices are likely to result in more cost effective implementations of DP and well as protection of the high-value data assets.

4.2 Compliance

An organization's compliance obligations can also play an important role in determining DP requirements. These obligations can be legal in nature and take the form of contractual terms and conditions, preservation orders (e.g., electronic discovery), etc. Likewise, there can be statutory or regulatory requirements associated with certain types of data. In some cases, requirements can mandate the use of DP, and in other cases, there can be requirements imposed on the DP (e.g., confidentiality measures for personally identifiable information).

Compliance obligations can also originate from certification activities. For example, an organization seeking or maintaining ISO/IEC 27001 [7] certification is required to demonstrate certain DP measures.

4.3 Data Availability and Usability

Data protection goes beyond the notion of data availability, which is defined as the amount of time that data is accessible by authorized applications and users during those time periods when it is expected to be ready for use. Unacceptable performance can lower productivity levels such that access to applications and related data is effectively unavailable. Note that data security and compliance issues are also intimately involved in data availability, as the ultimate goal of DP is to eliminate data loss while mitigating vulnerabilities, costs, and downtime.

This concept of data availability can be further defined in terms of accessibility, integrity, and timeliness. Accessibility involves ensuring that the data is accessible at the right place, for the right uses, and in a timely manner. Integrity refers to the data being protected against unauthorized eradication or modification (i.e., is returned the "same" as it was stored, with no alteration or corruption to the information, whether intentional or accidental).

The concept of data usability is defined as the protected data should be usable for its intended purpose. Usability can require that steps be taken to provide data integrity, application consistency, versioning, availability, and acceptable performance.

5 Resiliency and Recovery

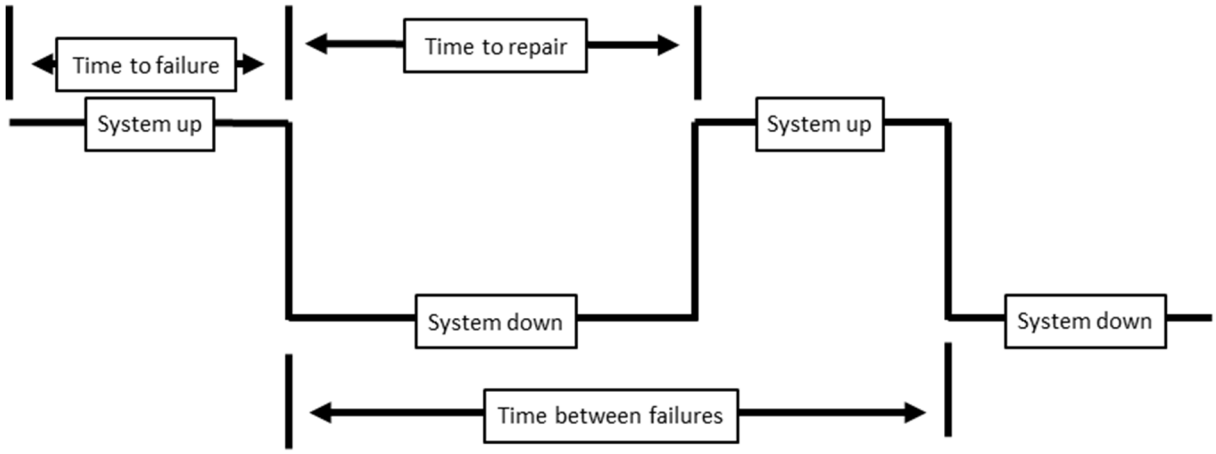
5.1 Metrics

The concept of resiliency is the ability of a storage subsystem to preserve data integrity and availability of access despite the unavailability of one or more of its storage devices.[6]

For storage, reliability is often considered the probability that a device performs its required function under stated conditions for a specific period and is quantified as:

- The MTBF (mean time between failures) for a product is the expected time, on average, between consecutive failures in a system or component and is sometimes thought of as the average time available for a system or component to perform its normal operations between failures (see Figure 5).
- The MTTR (mean time to repair) for a product is the expected or observed duration to return a malfunctioning system or component to normal operations and is sometimes thought of as the average time to repair a failed component.
- The MTTF (mean time to failure) for a product is the average time available for a system or component to perform its normal operations until it fails.

Figure 5 Quantification of reliability



The MTBF, MTTR, and MTTF metrics help an organization identify expected uptime – and by monitoring these statistics – the actual uptime of a system.

Not all system failures have impacts on data (e.g., corruption or loss of confidentiality), but they can have impacts on availability and usability (see 4.3). For example, a network outage that prevents users from

[6] SNIA Dictionary: (http://www.snia.org/education/dictionary)

accessing their data, but without harming the data, can later allow these users to access the data once the network outage is fixed.

5.2 Recovery Objectives and Capability

When considering system failures that result in data corruption or loss, the following DP factors can be important:

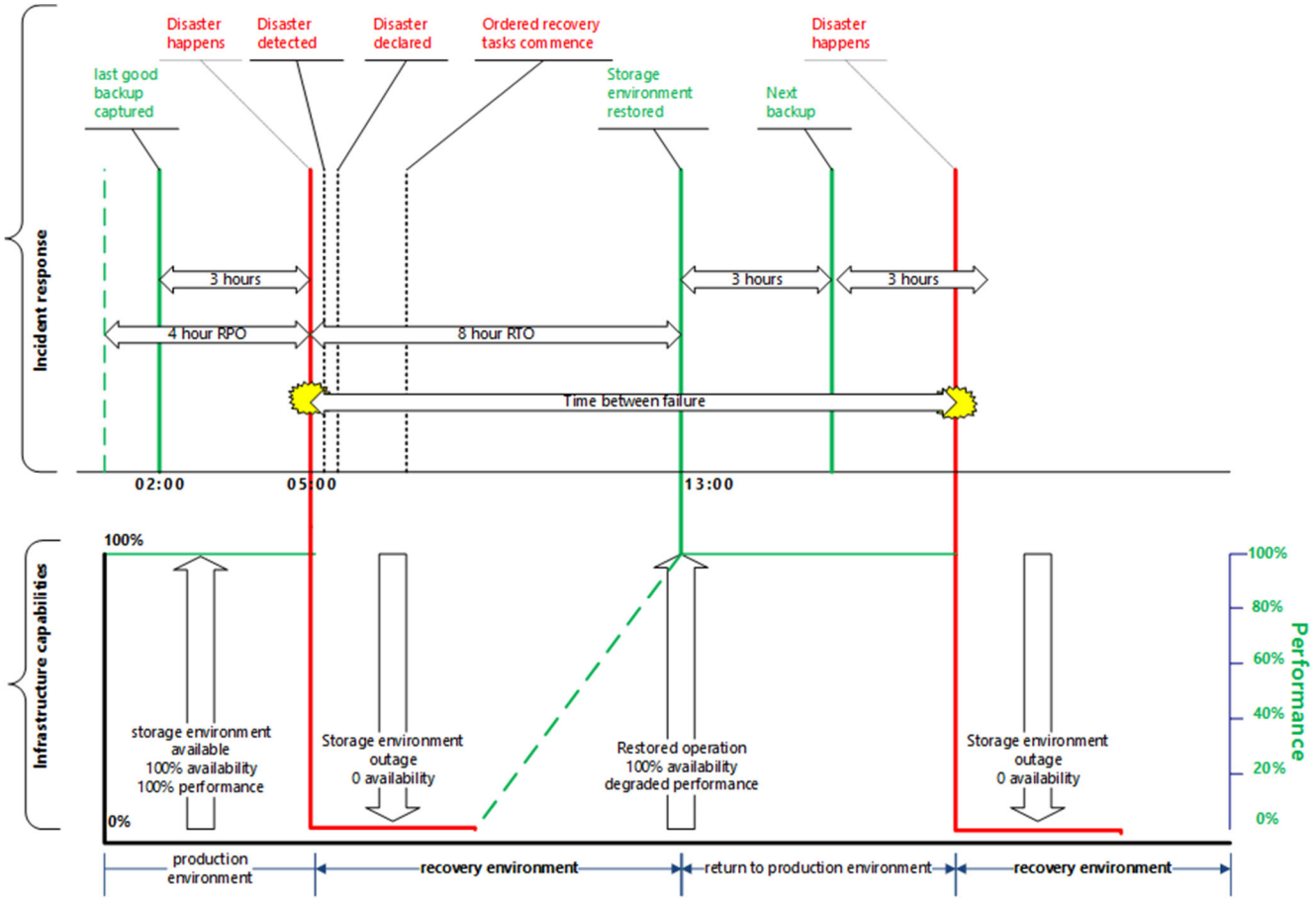
- Recovery Point Objective (RPO) is focused on data currency and represents the amount of data loss the business is willing to accept for each respective business process. Specified as time, RPO indicates the maximum amount of time data is permitted to exist before it is protected by a data recovery mechanism (e.g., backup). RPOs often drive the schedule for DP mechanisms.
- Recovery Time Objective (RTO) is developed by the business of the organization and is defined as the maximum permissible down time by the business (i.e., to avoid significant negative consequences). RTO is similar to MTTR in that both measure the time between the start of an outage and completion of recovery. However, MTTR is a mean value taken over several availability impacting events over a period of time, while RTO is an aspirational target for a single availability impacting event.
- Recovery Time Capability (RTC) is driven by IT of the organization and is defined as the shortest downtime that IT can support the recovery of services within budget. RTC can be the demonstrated amount of time in which systems, applications, and/or functions have been recovered during an exercise or actual event at the designated recovery/alternate location. An RTC that exceeds the organization's corresponding RTO is typically an indication of a risk to the organization.

RPO, RTO, and RTC are important factors in deciding what DP strategy the business needs to apply, and they should be a part of any organization's service level agreements with regard to DP.

A system that has a measured MTTR that exceeds the RTO should serve as a warning for the organization to consider adjusting either the expectations or protection mechanisms implemented by the system. This situation is a potential indication that each failure is likely to cause significant negative consequences (e.g., financial or contractual).

Figure 6 explores some of these concepts through the presentation of a hypothetical scenario for a fictitious organization wherein two disasters occur within a relatively short period of time. For the hypothetical scenario, details are provided for the incident response and for the infrastructure capabilities.

Figure 6. Hypothetical Scenario – Multiple Disasters



Consider the following in the Figure 6 hypothetical scenario:

- The time between failures was less than a day, which is likely be significantly less than the expected MTBF. Such a scenario would typically warrant further investigation by the organization.
- The organization has an RPO of four hours. The organization managed to capture good backups three hours or less prior to each of the disasters that rendered the storage unavailable. In both cases the RPO was met (i.e., data loss was within the acceptable range). However, there was a window between the time operations were restored after the first disaster (degraded performance) and when the second good backup was taken that the RPO was exceeded (i.e., seven hours).
- The organization was able to detect and respond to the first disaster in a timely manner, allowing the organization to begin the recovery within approximately two hours after the disaster. Such a response is critical to keeping the MTTR low.
- The organization has an RTO of eight hours. Following the first disaster, the organization was able to complete its disaster recovery activities such that operations were restored within four

hours with degraded performance and the production environment was fully recovered within eight hours (i.e., recovery time was within the acceptable range).

While this hypothetical scenario is somewhat simplistic, it does help highlight some critical considerations. For example, the organization's backup strategy (routine, as well as post-disaster) helped mitigate data loss as a result of the two disasters. Also, the time between initial restoration of operations (at degraded performance levels) and full recovery of operations could represent a period of risk of unsatisfactory performance if production workloads commenced. Activities preceding the recovery effort, and those associated with the recovery effort, should be analyzed with an eye to reducing the time associated with each.

Ideally, organizations prefer to avoid failures that negatively impact data. Thus, they may implement a class of DP technologies that can mitigate component failures that under other circumstances would be a system failure.

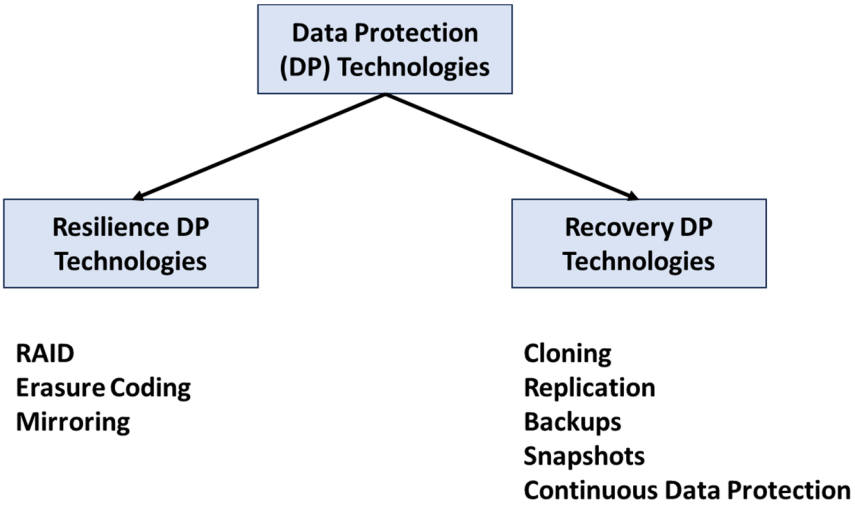
5.3 Data Protection (DP) Resilience and Recovery Technologies

When considering data protection technologies within storage systems and ecosystems, they generally fall into one of the following characterizations:

- Resilience DP technologies – Capabilities that allow for the failure of components without causing data integrity or data availability failures. The system or solution can continue providing its primary services, but potentially in a degraded manner. Another way to interpret the resilience of these systems or solutions is that they have an RTO and RPO of zero (i.e., undisrupted access to data regardless of component failure).
- Recovery DP technologies – Capabilities that facilitate the recovery of data and services in the shortest amount of time with the least amount of data loss or corruption after a failure or attack that renders the system or service non-operational. Recovery systems or solutions have a non-zero RTO or RPO.

This paper leverages these two characterizations as a way of grouping and describing DP technologies. Figure 7 shows the more common DP technologies for each of the characterizations.

Figure 7. Resilience and Recovery DP Technologies



5.4 Relevance of DP Technologies to the Data Lifecycle

Leveraging the lifecycle phases described in 3.2 and the data characterization in 3.3 (especially the active/inactive data details), the following summary describes the applicability of resilience and recovery DP technologies:

- *Create/Obtain* – During this phase, data can come from real-time sources and be transient or ephemeral in nature. In other situations, data can be amassed from different sources before being used (e.g., a data lake used to train a large language model). Another possible example is the creation of a new database that can take days or weeks to assemble.
  - Resilience DP Technologies – During this phase, resilience DP technologies are often used to guard against component failures that could corrupt or destroy data.
  - Recovery DP Technologies – For data that is not transient or ephemeral, recovery DP technologies such as snapshots and replication can be used to establish recovery points and/or roll-back options.
- *Store/Maintain* – During this phase, data is fully stored in the appropriate elements of the storage ecosystem, based on whether the data is active or inactive.
  - Resilience DP Technologies – A full range of resilience DP technologies are commonly used for active data. For inactive data, resilience DP technologies are less likely to be used.
  - Recovery DP Technologies – Any or all recovery DP technologies are likely to be used for active data. For inactive data, backups can be used.
- *Use* – During this phase, the focus is on active data and its usage. Generally, the data is read from where it is stored before or while it is used.
  - Resilience DP Technologies – No additional resilience DP technologies beyond those from the store/maintain phase (active data) are typically needed.
  - Recovery DP Technologies – No additional recovery DP technologies beyond those from the store/maintain phase (active data) are typically needed.

- *Enhance* – During this phase, active data can be enriched by merging with other data and/or enhanced (e.g., adding metadata). As an example, an existing large language model could be augmented with additional training data.
    - Resilience DP Technologies – No additional resilience DP technologies beyond those from the store/maintain phase (active data) are typically needed unless the enhanced data is stored separately from the production version of data.
    - Recovery DP Technologies – No additional recovery DP technologies beyond those from the store/maintain phase (active data) are typically needed unless the enhanced data is stored separately from the production version of data.
- *Dispose of* – During this phase, the focus is on eradicating data, so DP technologies are not used.
    - Resilience DP Technologies – Resilience DP technologies are not used for this phase.
    - Recovery DP Technologies – Recovery DP technologies are not used for this phase; however, it is important to ensure there are no residual copies of the data that is being eradicated (e.g., previous backups of data also need to be eradicated to ensure that the data is not recoverable).

6 DP Key Performance Indicators (KPIs)

Many organizations use key performance indicators (KPIs) to achieve strategic business objectives. Typically, KPIs are tailored to the specific needs of the organization. A common approach to developing KPIs involves identifying key strategic objectives, defining success, deciding on specific metrics, and crafting KPIs that are SMART[7] (specific, measurable, attainable, relevant, and time-bound).

This section proposes example KPIs that organizations can consider for their own DP activities. ISO/IEC 27002 DP oriented controls were used as inspiration for objectives and outcomes in the example KPIs.

ISO/IEC 27002:2022 organizes its controls based on themes: people (if they concern individual people), physical (if they concern physical objects), technological (if they concern technology), otherwise they are categorized as organizational. This paper uses a similar structure for data protection KPI materials.

6.1 Organizational KPIs

This section provides examples of organizational KPIs that are potentially relevant to data protection activities.

Title:	DP policies and procedures for data lifecycle
Objective:	IT policies and procedures address data lifecycle issues relevant to DP activities.
Success:	DP requirements are identified for all major categories of data assets.
KPI(s):	• All sources of data retention obligations (regulatory, legal holds, etc.) are identified and documented. • All approved data preservation approaches (e.g., archives and data repositories) and technologies (e.g., immutable storage) are identified and documented.

[7] Doran, G. T. (1981). "There's a S.M.A.R.T. way to write management's goals and objectives" (PDF). Management Review. 70 (11): 35–36. https://community.mis.temple.edu/mis0855002fall2015/files/2015/10/S.M.A.R.T-Way-Management-Review.pdf

	• All approved data eradication approaches (e.g., storage sanitization) are identified and documented.

Title:	DP policies and procedures for data security
Objective:	IT policies and procedures address data security issues relevant to DP activities.
Success:	Data security requirements are identified for all major categories of data assets.
KPI(s):	• Identified data classifications include provisions for sensitive data (e.g., personally identifiable information) and business critical data (e.g., intellectual property). • Confidentiality requirements specify approved encryption and access control mechanisms for data in-flight and data at-rest. • Integrity requirements address data accuracy, data completeness, and data consistency at a minimum. • Availability requirements specify the expected uptime (e.g., uptime / (uptime + downtime)).

Title:	DP policies and procedures for business continuity
Objective:	IT policies and procedures address business continuity issues relevant to DP activities.
Success:	Business continuity plans and requirements identify all major DP dependencies and expected outcomes.
KPI(s):	• Disaster recovery plans document expectations associated with DP activities/resources during catastrophic events. • Ransomware prevention and recovery plans are documented, including dependencies on production DP activities/resources. • RTO/RPO for business/mission critical data are identified and documented. • Business continuity plans are reviewed annually for DP content and DP elements are tested bi-annually.

Title:	DP compliance obligations
Objective:	Compliance obligations relevant to DP activities are identified and factored into procedures.
Success:	DP activities identify and document relevant statutory, regulatory, and legal requirements.
KPI(s):	• All mandated risk analysis and impact assessments include DP activities. • DP procedures address relevant data privacy requirements. • DP procedures include provisions to address data preservation solutions (e.g., cloud/non-cloud archives). • DP procedures include provisions to address data retention requirements (e.g., legal holds).

6.2 People KPIs

This section provides examples of people KPIs that are potentially relevant to data protection activities.

Title:	DP staff knowledge and skills
Objective:	Ensure DP personnel have appropriate knowledge and skills to address the ongoing operation activities and challenges.
Success:	DP personnel are adequately trained.
KPI(s):	• Job task analysis has been completed for each DP staff position. • Percentage of employees who have completed DP and security training.

Title:	DP staff supervised/managed
Objective:	Ensure DP personnel are able to perform the necessary DP activities in a manner that is beneficial to the organization.
Success:	DP personnel are sufficiently supervised/managed.
KPI(s):	• The unique staffing and facility requirements associated with the storage ecosystem are documented and reviewed bi-annually. • Number of DP-oriented incidents involving employee mistakes, accidents, or malicious behaviors.

Title:	DP staffing checks and balances
Objective:	Ensure there are sufficient checks and balances such that DP activities are not compromised.
Success:	There are no single points of failure in the DP staff due to skill gaps or workforce staffing.
KPI(s):	• Percentage of staff who can perform critical DP activities. • Requirements for segregation of duties (e.g., maker/checker) are documented and staffed appropriately.

6.3 Physical KPIs

This section provides examples of physical KPIs that are potentially relevant to data protection activities.

Title:	Survivability of DP capabilities
Objective:	The organization's DP capabilities remain functional during an environmental event affecting a single site/region.
Success:	DP solutions are sufficiently mobile/dispersed to meet RPO/RTO requirements during an incident.
KPI(s):	• Number of organization's critical/high-value data items not supported by multi-site DP solutions. • Frequency of testing of geofencing and/or alternate site DP solutions.

Title:	Isolation and independence of DP capabilities
Objective:	DP solutions associated with cyber recovery are sufficiently isolated and independent from production DP solutions.
Success:	Cyber recovery solutions are not impacted by major security events (e.g., ransomware attacks).
KPI(s):	• Frequency of recovery tests using the cyber recovery solution. • Number of security scans in a month that showed a lack of network isolation of cyber recovery solutions (i.e., a single failure in this regard could have severe consequences).

Title:	DP support for business continuity
Objective:	In support of the organization's business continuity management, there are provisions to securely store critical/high-value data remotely.
Success:	Critical/high-value data are routinely delivered and stored at appropriate remote facilities in appropriate forms.
KPI(s):	• Number of organization's critical/high-value data not stored on immutable media. • Number of organization's critical/high-value data not stored on media in a remote location or vault. • Measurement of time to retrieve data and make available to the DP solutions.

Title:	Physical security for DP solutions
Objective:	The organization's DP facilities (e.g., buildings, room, racks) housing critical/high-value data are secured against unauthorized access.
Success:	There are no incidents of unauthorized physical access to DP solutions.
KPI(s):	• All DP restricted areas are identified and documented, including requirements of surveillance tools appropriate to the relevant physical area. • Number of DP facilities not protected with detectors (e.g., motion, sound, and contact) that trigger an alarm when an intruder accesses physical premises.

6.4 Technological KPIs

This section provides examples of technological KPIs that are potentially relevant to data protection activities.

6.4.1 Resilience-oriented KPIs

From a DP perspective, resilience-oriented technologies are instrumental in mitigating component failures that could force a data recovery operation if such technologies are not used (i.e., RTO/RPO are zero). In this context, the components could be drives, software/firmware, and entire storage systems (e.g., storage array or network attached storage filer). This section provides examples of resilience-oriented KPIs.

Title:	Resilience-oriented requirements for failure mitigation
Objective:	Requirements for resilience DP technologies are identified and aligned with an RTO/RPO of zero.
Success:	Organization's storage infrastructure designs/architectures include resilience DP technologies to mitigate component failures.
KPI(s):	• All deployed resilience-oriented DP technologies (RAID, erasure coding, mirroring) are identified and documented. • Amount of organization's critical/high-value data not protected by resilience-oriented technologies.

Title:	Resilience-oriented component mitigations
Objective:	Resilience-oriented DP capabilities are sufficient to mitigate component failures.
Success:	Resilience-oriented DP technologies successfully mitigate component failures.
KPI(s):	• Monthly ratio of component failures resulting in outages versus total number of component failures.

6.4.2 Recovery-oriented KPIs

From a DP perspective, recovery-oriented technologies are instrumental in mitigating data loss or corruption after a failure or attack that renders the system or service non-operational. This section provides examples of recovery-oriented KPIs.

Title:	Recovery-oriented DP requirements for failure mitigation
Objective:	Requirements for recovery-oriented DP technologies are identified and aligned with the organization's RTO/RPO requirements.
Success:	Organization's storage infrastructure designs/architectures include recovery DP technologies to mitigate failures and attacks.
KPI(s):	• All deployed recovery-oriented DP technologies are identified and documented. • Percentage of organization's critical/high-value data protected by recovery-oriented technologies. • Number of recovery-oriented DP solutions that do not have current risk assessments and documented acceptance of the risk posed to organizational operations; RTO/RPO requirements associated with the protected data are included.

Title:	Backup solutions and data recovery
Objective:	Backup solutions provide sufficient protection (i.e., meets RTO/RPO requirements) for organization's critical/high-value data.
Success:	Data recoveries from backup solutions are within the organization's RTO/RPO limits.
KPI(s):	• Number of failed backups per month along with the number of root cause analyzes performed and resolutions of the underlying issues. • Number of monthly backups that fail to meet RPO requirements (e.g., time between backups for rapidly changing data exceeds the desired RPO). • Number of monthly recovery tests from backups; reports include information about the success rate of meeting the RTO/RPO requirements for the target data. • Measured impact of data rehydration (associated with data reduction mechanisms) on the throughput/performance of data recoveries from backups. • Percentage of data backups that are fully automated.

Title:	Cyber recovery solutions and data recovery
Objective:	Cyber recovery solutions provide sufficient protection (i.e., meets RTO/RPO requirements) for organization's critical/high-value data that could be subjected to ransomware and other destructive cyber-attacks.
Success:	Data recoveries from cyber recovery solutions are within the organization's RTO/RPO limits. Ransomware and other destructive cyber-attacks are handled within established RTO/RPO expectations.
KPI(s):	• Number of failed cyber recovery backups per month along with the number of root cause analyzes performed and resolutions of the underlying issues. • Number of monthly cyber recovery backups that fail to meet RPO requirements (e.g., time between backups for rapidly changing data exceeds the desired RPO). • Number of monthly recovery tests from cyber recovery backups (using staging or air-gapped environment); reports include information about the success rate of meeting the RTO/RPO requirements for the target data. • Number of monthly tests to verify the isolation of the cyber recovery solution (i.e., no dependencies on production data or regular backups) • Percentage of cyber recovery backups that are fully automated.

Title:	Replication solutions and data recovery
Objective:	Replication solutions provide sufficient protection (i.e., meets RTO/RPO requirements) for organization's critical/high-value data.
Success:	Data recoveries from replication solutions are within the organization's RTO/RPO limits.
KPI(s):	• Number of monthly tests to verify the replication approach (especially for business/mission critical data) is aligned with its associated reliability, fault-tolerance, or performance requirements.

Title:	Snapshot solutions and data recovery
Objective:	Snapshot solutions provide sufficient protection (i.e., meets RTO/RPO requirements) for organization's critical/high-value data.
Success:	Data recoveries from snapshot solutions are within the organization's RTO/RPO limits.
KPI(s):	• Number of monthly recovery tests from snapshots; reports include information about the success rate of meeting the RTO/RPO requirements for the target data. • Number of snapshots that are not protected from changes (e.g. read only).

Title:	Cloud DP solutions and data recovery
Objective:	Cloud DP solutions provide sufficient protection (i.e., meets RTO/RPO requirements) for the organization's critical/high-value data.
Success:	Data recoveries from cloud DP solutions are within organization's RTO/RPO limits.
KPI(s):	• Number of monthly recovery tests from cloud DP solutions; reports include information about the success rate of meeting the RTO/RPO requirements for the target data.

Title:	Securing recovery-oriented DP solutions
Objective:	Recovery-oriented DP systems have sufficient security capabilities to guard against attacks/unauthorized access.
Success:	Data sets and systems associated with the recovery-oriented DP solutions are appropriately secured.
KPI(s):	• Number of recovery-oriented DP solutions that do not employ in-flight or at-rest encryption (256-bit AES) for the organization's critical/high-value data; these exceptions necessitate a risk assessment and a documented risk acceptance. • Number of risk assessments/audits (annually) performed on recovery-oriented DP solutions. • Percentage of storage media, used in the recovery-oriented DP solutions, that are sanitized prior to reuse or disposal.

Title:	Managing and monitoring recovery-oriented DP solutions
Objective:	Recovery-oriented DP systems are appropriately configured, operated, and managed.
Success:	DP and recovery activities are conducted in accordance with the organization's RTO/RPO requirements.
KPI(s):	• Number of monthly data recoveries performed, including success/failure to comply with the organization's RTO/RPO requirements. • Delivery of weekly reports detailing the sources and amounts of data handled by the recovery-oriented DP solutions • Number of monthly data preservation overrides (e.g., legal holds) performed. • Amount of down-time for the recovery-oriented DP solutions as a result of maintenance and updates.

Title:	Recovery-oriented DP solutions and business continuity
Objective:	Recovery-oriented DP solutions that are integral to the organization's business continuity activities provide sufficient protection (i.e., meets RTO/RPO requirements) for organization's critical/high-value data.
Success:	Data recoveries from recovery-oriented DP solutions are within organization's business continuity RTO/RPO limits.
KPI(s):	• Number of business continuity tests (annually) that include recovery tests from recovery-oriented DP solutions; reports include information about the success/failure rate of meeting the business continuity RTO/RPO requirements for the target data.

7 Storage Technologies Relevant to Data Protection

This section identifies and describes common forms of storage-based DP technologies. These technologies are group based on the DP characteristics described in 5.3 and are further described in subsection on:

- Resilience-oriented technologies (see 7.1) where access, possibly degraded, to data is maintained even in the case of failures.
- Recovery-oriented technologies (see 7.2) where access to data has ceased and this access can only be reestablished after some form of data recovery.

7.1 Resilience-oriented DP Technologies

Resilience-oriented DP technologies are used to "protect" the data on the storage media itself from data corruption due to bit errors or media failures. This paper covers common forms of these technologies, which include RAID (redundant array of independent disks), mirroring, and erasure coding. RAID, mirroring, and erasure coding are methods of providing resilience blocks that use different error correction coding (ECC) methods. Note that ECC is used in many places in the storage infrastructure (e.g., transport, storage device, resilience blocks).

It is important to note, however, that resilience-oriented DP technologies should be part of an overall DP strategy as they do not inherently protect data from becoming compromised, deleted, or overwritten.

7.1.1 RAID Storage

RAID is an enabling technology that leverages multiple drives (spinning disk, solid state, or virtual) as part of a drive set that provides DP against drive failures, as well as media failures (e.g., unreadable sectors). RAID can also serve to improve storage system performance since input/output (I/O) requests can be serviced by multiple drives independently.

Although RAID can be technically considered a form of erasure coding, generally, erasure coding refers to a very different approach to DP and is described in 7.1.3.

The industry has developed multiple RAID levels that focus on eliminating data recovery time and/or protection data integrity across failures. However, as technology matured, a limited number of RAID levels have dominated the market, those levels are:

Mirroring RAID

- RAID 1 (Two or more drives contain identical data)
- RAID 10 (Mirrored drive groups are striped for greater capacity)

Parity RAID

- RAID 5 Single parity (Enables recovery when any one disk out of a group of drives fails)
- RAID 6 Dual Parity (Enables recovery when any two disks out of a group of drives fails)
- Distributed Parity RAID (Enables recovery when one or more drives out of a group of drives fails)

Appropriate use of RAID levels includes making use of the RAID level that provides the right balance between cost, performance, and protection. For example, when wanting to achieve very high data availability, while maintaining the best possible performance, a RAID 10 configuration can be optimal. The downside, of course, is that RAID 10 doubles the cost of storing data, versus a RAID 0 (striping) configuration. Another consideration is the extended performance impacts that can be associated with RAID rebuilds (after a drive failure) when drive capacities are large.

7.1.2 Mirroring

Mirroring is a process whereby a group of two or more storage devices are configured to maintain identical copies of data (RAID 1 is one example of mirroring). This configuration of storage devices is known as a mirror group, and each of the storage devices are members of the mirror group. Every write operation is applied to *all* members of the mirror group and read operations can be fulfilled from *any* device member of within the mirror group.

The example configurations described here are simple examples to show the concepts of mirroring. Mirroring is typically used in more complex configurations that include additional redundancy techniques.

The primary benefit of mirroring is to avoid unplanned downtime due to hardware failure of a member of the mirror group. This protection against storage hardware failures is transparent to applications. Mirroring provides no protection against data being compromised, deleted, or overwritten.

Storage administrators must consider factors such as system performance, cost-effectiveness, and complexity when designing a mirroring-based solution.

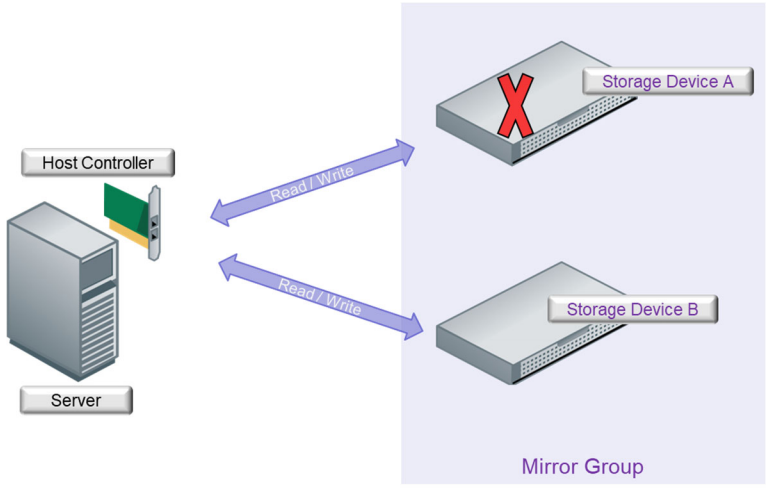
Software-based mirroring is a process provided by software in a host (e.g., by the operating system or third-party software) that consumes additional host processing resources instead of requiring additional hardware resources to implement the mirroring, but this does not eliminate the need for the additional storage required to support that mirroring.

Hardware-based mirroring is implemented in hardware (e.g., by a RAID controller) and is transparent to the host. There are typically multiple locations within a storage topology where hardware-based mirroring can be implemented.

- Host controller-based mirroring is depicted in Figure 8 and shows a configuration where more than one storage device in a mirror group is connected to the same host controller that performs the mirroring operation (i.e., host software performs writes to the host controller which writes the data to all the members of the mirror group; the host software performs reads to the host controller which then reads data from any member of the mirror group). Failure of a storage device in the mirror group does not impact access to the data (i.e., RTO and RPO are zero) since the host

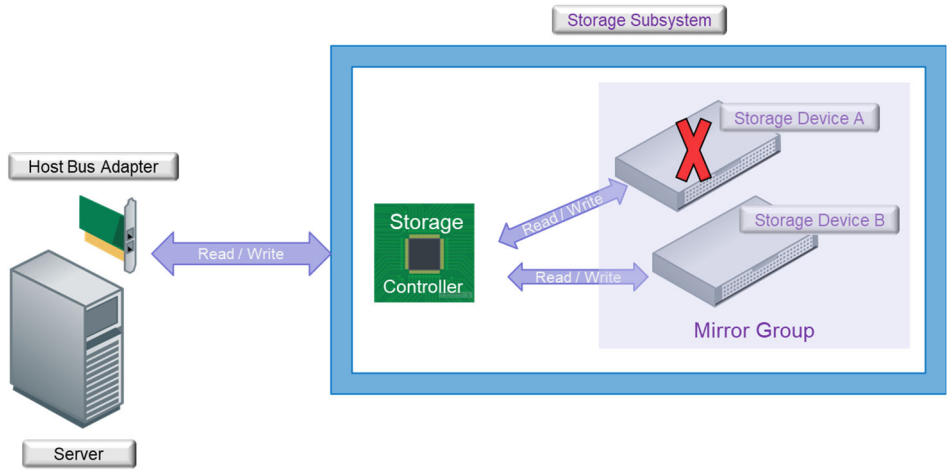
controller is able to read the data from another member of the mirror group. With this configuration, the host controller is a single point of failure.

Figure 8. Controller-based Mirroring



- Storage controller-based mirroring is depicted in Figure 9 and shows a configuration where the storage devices of a mirror group are connected to that controller in the storage subsystem. In this configuration, the host writes data to the storage controller which writes the data to the members of the mirror group. Failure of a storage device in the mirror group does not impact access to the data since the storage controller is able to read the data from another member of the mirror group. With this configuration, the storage controller is a single point of failure.

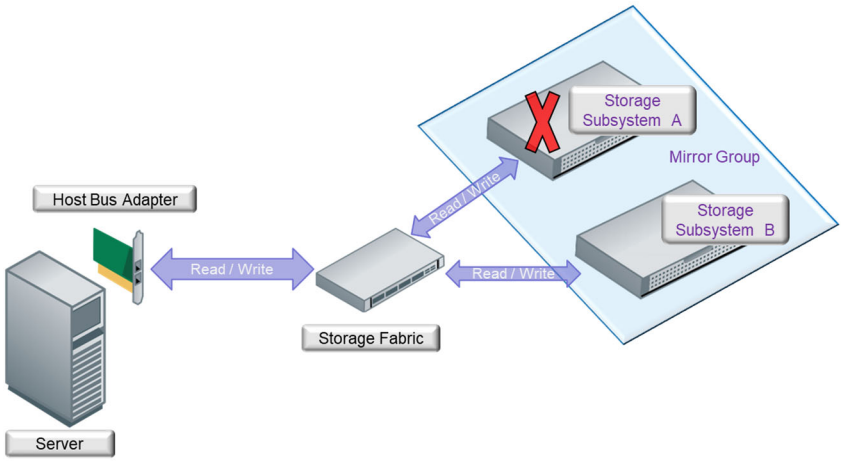
Figure 9. Storage Array Controller-based Mirroring



- Fabric-based mirroring is depicted in Figure 10 and shows a configuration where the storage fabric intercepts read and write commands from the host. For each write command the fabric issues write commands to all members of the mirror group. For each read command the fabric issues a read command to a member of the mirror group. Failure of one member of the mirror

group does not impact access to the data since the fabric is able to read the data from another member of the mirror group.

Figure 10. Fabric-based Mirroring



In summary, simple mirroring implementations write the same data to two or more identical disks can prove an effective way to eliminate a given drive as a single point of failure. More complex mirroring configurations, such as triple mirroring, offer greater resiliency, but at greater cost.

7.1.3 Erasure Coding

Erasure coding is an error correction technology that employs algorithmic “codes” to calculate parity data. Erasure coding mitigates the potential of data “erasure” (e.g., loss in transit, or loss at rest), providing data resiliency in a manner that is more efficient than mirroring. Erasure coding generally takes place in software above the hardware storage device layer. Erasure coding is suitable for large-scale, distributed storage systems that require a configurable balance between storage efficiency, performance, and rebuild speed.

Erasure coding is often used:

- To improve storage resiliency (e.g., spanning multiple storage devices, file systems, hosts)
- To increase storage utilization on large systems (e.g., where mirroring would be unfeasible and undesirable)
- For highly configurable storage efficiency and recovery performance (e.g., match the erasure coding algorithm to the error correction requirements of the workload)

Simple erasure coding algorithms are used for RAID parity calculations (e.g., RAID 5 and RAID 6), where a storage array must have some ability to self-heal in place.

Modern erasure coding algorithms provide highly configurable and scalable balances between available capacity, performance, redundancy, integrity validation, and rebuild time.

Erasure coding is often implemented using Reed-Solomon error correction algorithms.[8]

Erasure coding can be used for DP instead of RAID and is quite common as larger and larger multi-terabyte disk drives are being manufactured. For example, many of the “Object Storage” systems on the market today use some form of erasure coding rather than RAID. Also, erasure coding systems are typically software-based, as compared to traditional RAID 5 and 6 systems which often use specialized hardware to perform the necessary I/O processing.

Erasure coding parses incoming data into multiple component chunks. Then, somewhat like a parity calculation, expands each chunk with some additional information and creates parity chunks, creating a slightly redundant but more resilient superset of data. With a mathematical algorithm, the system can use these expanded blocks to recreate the original data set, even with missing or corrupted chunks. This allows the storage system to still deliver data, even after multiple drive or node failures. There is little overhead for “reads” when using erasure coding, except when there are drive failures, since the expansion and parity calculations are implemented during “writes”. Most erasure coding schemes allow the user to configure the level of resiliency, essentially by increasing the amount of parity data generated for each chunk. There are also different levels at which erasure coding can be applied: at the array level, at the node level (for scale-out architectures) or at the system level – which can affect how much processing overhead it consumes.

Erasure coding can be combined with data distribution or dispersion to improve resiliency and eliminate the need to make dedicated copies for off-site storage. This process essentially spreads data chunks across multiple nodes or systems, usually in different physical locations. However, using a distributed architecture where data chunks are spread between different physical locations can create a latency problem, since network bandwidth quickly becomes the limiting factor when blocks are pulled across the Wide Area Network (WAN). Some object storage systems combine erasure coding and replication, using erasure coding at the local system level and copying data between geographic locations to alleviate latency.

As storage devices scale in capacity faster than their host interconnect bandwidth does, rebuild times using standard RAID algorithms by necessity get longer (can increase from multiple minutes to multiple hours). Because of this, the need for erasure coding is becoming much more important to sustain multiple drive failures without having to suffer the prolonged period of “read/recovery” time per drive, as seen with many RAID-protected storage systems.

7.2 Recovery-oriented DP Technologies

7.2.1 Cloning

Cloning is a process of taking an asynchronous offline full copy of the source dataset. Once the clone process has completed, the link between the source dataset and the clone can be severed, resulting in the clone becoming a full copy of the original dataset at the time of the link was severed (i.e., at the completion of the cloning process). Once completed, the clone is a read/write copy of the dataset. A clone is typically used for development and or quality assurance to provide developers with a replica of the dataset for testing on real live data. While it is typically not used for resiliency or recovery, it can be used as a full backup, though typically taken less frequently than snapshots. A good, uncorrupted point

[8] https://en.wikipedia.org/wiki/Reed%E2%80%93Solomon_error_correction

in time clone can accelerate the recovery of the dataset after a corruption of the original occurs (whether because of random failure or malware such as ransomware).

Cloning can also be used to reduce the impact of taking a backup (which usually requires all application access to be stopped). If a dataset is cloned, then the backup can be taken of a completed clone while applications continue to have read/write access to the main dataset.

7.2.2 Replication

7.2.2.1 General

Replication is a process by which a copy of the data is created from an array or a storage frame to another array that can be local or remote. Replication is different than cloning in that it works to keep the source and destination data copies in sync (either in real time (synchronously) or after the fact (asynchronously) on a regular cadence). Replication can be accomplished in three types of implementations:

1. Host based replication
2. Array/storage-based replication
3. Fabric based replication

Replication is a continuous process (not a one-off point-in time-copy) of writing data to two or more targets in a replication group and having each target member capable of being used for read operations. Each member of the replication group could be “local”, or a different storage array in the same data center, or it could be “remote” (in another building on campus or across town, or in a different geographic region, or even in another country).

There are two types of replications: “synchronous” and “asynchronous”. Synchronous replication and asynchronous replication are both continuous processes, which means that I/O are not dated and transmitted (or cached). With synchronous replication, no further write I/Os are allowed until the remote array acknowledges that it has successfully written the current I/O. With asynchronous replication, subsequent write I/Os are not held off. With asynchronous replication one or more write I/Os that haven’t yet been confirmed to have been written at the remote site can be lost (effectively in transit) in the case of a failure, and thus there could be a nonzero RPO in case of a disaster. The distance between local and remote sites can be so far that the latency for acknowledgement required by synchronous replication is so long that the resultant write throughput performance penalty is so unacceptable that synchronous replication cannot be used.

Remote replication is often used in conjunction with drive-based backups, however data replication does not require a “restore” of the data (as backup does), before the data can be used/accessed. In other words, assuming the appropriate resources are in place at the replication target location, the replicated data can be used (nearly) instantly (depending on the specific vendor implementation).

Whether replication is used as part of the data loss/availability procedures greatly depends upon the importance (criticality) of the data to the business, along with its availability requirements. Requirements for availability of the data set(s), are often depicted as RPO and RTO requirements. For example, for a banking customer’s transaction a logging data set is likely to require a “zero” RPO, and an RTO of a few seconds (or less). In this case, even a minute of disruption can pose an unacceptable risk and can lead to devastating consequences for the business. Data “backup” is not powerful enough to prevent such data loss/availability at this level. So, as the criticality of the data rises, more sophisticated and costly geographically dispersed replication solutions are often deployed.

End-users typically deploy three primary categories of replication:

(1) Complete replication of data between two or more sites

(2) Caching of frequently used data at remote sites with the complete data sitting at a home site

(3) Hybrid complete replication with caching

Complete replication is often used to provide continued access to data in the case of a disaster and the "copies" of the data are often in geographically dispersed locations. In addition, complete replication is utilized in standard production environments to provide faster access to the data by localizing the data access for the user or application that is requesting the data. Replication can even be used within the same data center or in another data center within close proximity simply to maintain more than one copy of the data and/or distribute the access load. One example of complete replication would be for a financial institution such as a credit card company where complete data copies must reside in differing geographical locations to account for natural disasters, terrorism, etc.

The third type of topology used for replication is a hybrid approach of the previous two described topologies: a combination of complete replication between multiple, geographically dispersed sites with smaller, cache sites fanning out from each of the complete sites. This can be thought of as a "multi-hop" topology that is often used in very large organizations where requirements necessitate multiple copies of data (often for regulatory reasons) and fast access to data at many different sites.

The appropriate use of replication will depend on the defined business requirements of the respective data sets – based on the importance of the data to the business. For example, for financial data that is critical to the organization, the type of replication used can be complete replication of the critical data sets to another geographically dispersed data center, to protect against disasters, acts of terrorism, etc.

Some considerations when using replication are:

- Replication can assist in complying with business continuity requirements, which often include ensuring that the data is stored in a remote location, at a sufficient distance to escape any damage from a disaster at the main site.[8]
- Replication can involve significant resources and adds more complexity to the environment. This is important because some data is not necessarily “critical”, and therefore non-critical data should not be replicated the same way required for “critical” data.

Best practices for replication include using the appropriate implementations of each, for meeting the specified business requirements, specifically in the areas of business continuity management planning.

7.2.2.2 Host-based replication

The concept of host-based replication is the process of copying the data that was managed and controlled by the host computers. Host-based replication can be computationally demanding and have a significant impact on the overall solution cost. However, with the ever-declining processor cost, host-based replication has become widely accepted.

Finally, in a replicated environment, both replication source and destination stay in sync either in real time (synchronous replication) and on controlled intervals (asynchronous replication) as required, depending on how critical the application is.

7.2.2.3 Array/storage-based replication

With the enhancement in storage controller technology, the industry adopted the concept of using the extended capabilities of a storage controller to replicate the changes to the stored data to a member of the replication group, on-site or off-site. The introduction of array-based replication was a game changer where all replication overhead was moved off the host computer to the storage frame, making access to storage replication more affordable. While still in use today, almost all storage vendors offer array-based replication – in most cases at no additional cost to the storage solution.

7.2.2.4 Fabric-based replication

Fabric-based replication was supposed to offer an attractive alternative to both host-based and array-based replication where all storage intelligence takes place at the storage transport fabric layer. While in very limited use today, fabric-based replication has proven to be cumbersome and has not become a popular solution.

7.2.3 Backups

7.2.3.1 General

A “backup” or “backup copy” is defined by the SNIA as a collection of data, often stored on non-volatile storage media for purposes of recovery, in case the original copy of data is lost or becomes inaccessible[9]. A backup process aims to preserve data in case of equipment failure, cyberattack, natural disaster, or other data loss events. Backups are relevant to individual users, work groups, and enterprises to protect valuable data assets.

While backup and recovery processes are fundamental components of IT operations, several misconceptions can undermine their effectiveness. By addressing the followings misconceptions and adopting best practices, organizations can enhance their DP strategies and ensure they are well-prepared to recover from data loss incidents:

- *Assumption of Success in Automated Backups* – Due to the repetitive and automated nature of backup processes, backup jobs, which can take anywhere from a few hours to entire weekends to complete, are generally left unattended. Administrators often assume that all backup jobs have been successful if no immediate errors are reported; however, a small number of unsuccessful backup jobs among tens of thousands can pose a significant risk, if not corrected.
- *Misplaced Priority on Backup Speed Over Restore Speed* – Backup performance (i.e., amount of time to complete the backup) is often assumed to be the most critical criterion for a backup solution. While completing backups within the designated backup window and minimizing disruptions to IT operations is important, the real priority often needs to be the restore speed, to ensure minimal downtime and quick recovery in the event of data loss.
- *Misunderstanding the Impact of Data Deduplication* – While data deduplication can improve backup performance by reducing data volume and associated writes, it introduces complexities during data restoration because deduplication requires data to be rehydrated, or fully reconstructed, during the restore process. Such complexity can become a significant challenge during large-scale restore operations, in major disaster recovery incidents or ransomware recovery.

[9] SNIA Dictionary: (http://www.snia.org/education/dictionary)

- *Overlooked Importance of Backup Validation* – Failure to regularly validate backups, and/or assuming that the successful completion of a backup job equates to a reliable restore, can expose an organization to the risk of discovering corrupt or incomplete backups only when they need them the most, during a data loss event.
- *Neglecting Backup Security* – In the face of increasing cyber threats, securing backup data is often an overlooked aspect. Backup data can be a target for ransomware attacks or unauthorized access and tampering, so implementing strong security measures, such as encryption and access controls, is crucial.
- *Ignoring the Cost of Downtime* – Organizations sometimes underestimate the financial impact of downtime caused by data loss. While investing in robust backup and recovery solutions can seem costly, the cost of extended downtime can far exceed these initial investments. Understanding the potential financial implications of data loss and downtime can help justify the need for comprehensive backup and recovery strategies.

The remainder of this section is focused on backup details along with considerations for data recovery.

7.2.3.2 Types of Backups

There are two common types of backups: file backup and image backup. Each type is described in this section and accompanied with relevant considerations for both.

A “file” backup or a file/folder-based backup is designed to do exactly what the name implies; the smallest unit that could be restored is a file or folder. The typical use for such a system is to restore a file or folder that has been lost on an otherwise healthy system. This is a “selective” backup, where the business chooses what data should be backed up, and only those files and folders are backed up. User data is typically the focus rather than system-generated files, executables/applications, elements of the operating system, or similar content.

An “image” backup consists of the block-by-block contents of a data set, virtual machine (VM), or storage device. Image backups copy an entire system at a point in time – the operating systems, application data, system settings, patches, and user files. All this data is backed up as a single file, called an “image”.

From a recovery perspective, file/folder backups can offer fast and efficient recovery of specific data for systems that are functional. When the system itself or core applications have also been impacted, these elements need to be recovered by other means before the backups can be used to recover the data.

With “image” backups, it is possible to restore the system and its data in the event of a necessary rebuild or to perform a restore to a completely different system (physical or virtual). Such backups can be particularly important for disaster recoveries.

Image or file backups can be performed in many different ways. These backups can be done in conjunction with snapshots, and/or often with copies on various types of storage media, stored at various locations. For individual users or small work groups, file/folder backups are often employed. Larger organizations can use both types of backups, depending on the criticality of data and the RPO requirements. Also, compliance obligations can require different DP levels be deployed (e.g., a data set that includes financial trading logs for a financial trading firm can have a stringent DP policy).

7.2.3.3 Centralized Backup Strategies

To meet resilience and recovery expectations, organizations develop and implement specific backup strategies. One such strategy is the use of automation, which is core to ensuring that backups are

performed in a way that data is protected in a predictable manner. Organizations often perform backups daily and retain these backup copies for a fixed period of time (e.g., every 30 or 90 days).

Another backup strategy involves the way backup copies are managed. For data worth protecting, especially mission critical and high value data, keeping one instance of a backup is often not adequate. Many organizations, including the Cybersecurity and Infrastructure Security Agency (CISA), advocate following the 3-2-1 backup rule[10] that can be summarized as:

- 3 – *Three copies:* Maintain three copies of protected data – the original backup and two additional backup sets.
- 2 – *Two media types*: Store these backups on at least two different media types to protect against various hazards. This could include combinations of hard disk drive, solid state drive, tape, and cloud storage.
- 1 – One Offsite and Immutable Copy: Ensure one backup copy is stored offsite on immutable media, ideally behind an airgap, to protect against data corruption and ransomware attacks.

It is important to note that to create multiple copies of a backup set, there is no need to rerun the backup job since most likely the data has changed, and backup copies will not be consistent. Additionally, running multiple backup jobs for the same system will needlessly waste system resources. For that, backup vendors include the ability to clone the backup without impacting protected system/application. Some operating systems provide this capability at no extra cost.

Producing backup copies can be time consuming as well as requiring a significant amount of storage for the copies. A successful backup strategy should consider the following elements:

Backup Types and Considerations:

- Full backups: where 100% of the data is backed to the backup medium; these backups are typically performed over weekends. Consider also:
  - Any files that need to be restored are a part of a single backup, and therefore the restore is usually faster.
  - No need for multiple “mountings” for a restore.
  - The backup time is usually much longer and can interfere with other production operations.
  - The space required to store the full backup is greater than the other backup types.
- Incremental backups: where only the data that has changed since the last full backup and the last incremental backup (if any) is backed to the backup medium; these backups are typically performed daily. Consider also:
  - The backup time is shorter for incremental backups versus full or differential backups.
  - The restore will need to involve the most recent full backup, as well as all the incremental backups up to the most recent daily incremental that is available, which will be longer than restoring from a combination of full and differential backups.

[10] Peter Krogh, a photographer, writer, and consultant introduced the 3-2-1 backup rule when he published his book, *The DAM Book: Digital Asset Management for Photographers*, in 2005.

- Differential backups: where only the data that has changed since the last full backup is backed to the backup medium; these backups are typically performed daily. Some organizations perform only one full backup and continue with differential backups indefinitely. Consider also:
    - The restore will be shorter than using incremental backups, since the restore will only require a maximum of two “mountings”, the full and the differential backup in either order.
    - Differential backups save backup time versus a daily “full”, since the daily differential backup will only include the data that has been modified since the last “full” backup.
    - Differential backups can take longer than incremental backups.
    - The amount of daily backup storage will be more with differential backups than performing daily incremental backups, since the daily backup will include the data that has been modified since the last “full” backup.
- Synthetic full backups: where the incremental or differential backups are combined with the last full backup to create a new full backup. This is done offline to avoid impacting data availability. Consider also:
    - Faster backup times, since only need to perform incremental backups after the initial “full backup”.
    - Less load on the source system, since only incremental backups are done after the initial “full backup”.
    - “Synthetic full backups” or “incremental forever” backups can save backup space versus a daily “full backup”.
    - Additional load on the backup server due to the processing required to create a synthetic full backup.

Backup Scheduling:

To minimize RPO, backups should be performed daily at a minimum. For critical business data, more frequent incremental backups are often necessary. Backup scheduling also depends on the capabilities of the backup software. For instance, if the software does not support differential forever, a combination of weekly full and daily incremental backups (typically Monday through Thursday) can be the best approach.

Backup Storage Options:

Traditionally, tape was the primary backup medium due to its cost-effectiveness. However, the declining cost of disk storage and the rise of cloud storage have made disk-based backups more mainstream. Deduplication technology has further enhanced the appeal of disk storage by significantly reducing the volume of backup data. Organizations should analyze their data to determine the suitability of deduplication, as not all data types (e.g., compressed or encrypted data) benefit from this technology.

Off-Site Backup Considerations:

- Operational Recovery – Ensuring a backup set is available within the same data center to minimize recovery time for routine data restores.
- Disaster Recovery – Maintaining a backup set at a designated disaster recovery site to enable recovery if the primary site becomes inaccessible.

Backup Security:

Given the mobile nature of backup sets (whether on tape, replicated to a disaster recovery location, or stored in the cloud), security is an essential element of a backup strategy. Securing backups involves a multi-faceted approach that includes stringent access controls, secure storage practices, and compliance with data privacy regulations. Additional security measures such as regular audits, immutable backups, air-gapped storage, multi-factor authentication (MFA), continuous monitoring, employee training, and a robust incident response plan can enhance backup security. By implementing such backup security strategies, organizations can significantly enhance the security of their backup data and ensure its integrity and availability in the face of evolving threats.

Data Compression of Backup Images:

Backup operations can involve, for example, a combination of weekly full and daily incremental backups, with retention periods ranging from two to thirteen weeks. This can lead to a significant increase in storage requirements. For example, a 100TB data set could require up to 1500TB of backup storage. Data compression of backup images can reduce storage volume by 50% to 70%, making it a critical component of an effective backup strategy.

Note that some of the above strategies are not necessarily appropriate for individual or small office/home office (SOHO) scenarios.

7.2.3.4 Backups for Operational Recovery

In most environments, individual backup data sets are retained for a limited time period, usually between 30 to 90 days, since backups are not typically considered a good method for creating archives. If an organization elects to do full backups of all data daily, the most recent day's successful backup is the one that will be used for any necessary restores. This makes all previous days' backups obsolete except as future references for particular points in time, such as the state of financial files at the end of a month or quarter before accounting procedures are run. In some jurisdictions, backup sets can be subject to electronic discovery[11]; therefore, consider that the backups only be kept for a minimal amount of time, and they comply with the organization's retention policies for backups.

Performing lengthy full backups daily, however, is seldom considered now, given large data volumes and the need to run most organizations on a 24x7 basis. Although snapshots can be used as the source for a backup rather than the actual files, backing up from snapshots that are stored on the original devices can cause significant performance degradation, depending upon the specific vendor implementation.

A common practice is to use incremental backups, whereby a full backup is done periodically, such as once per week, followed by backing up just the changed files during each succeeding day. A variation on this approach is to use "differential" backups, whereby the files that are backed up are all files that have changed since the last "full" backup. A key difference between an incremental and a differential backup is the time needed to restore data. When using a combination of full and differential backups, only two backups need to be restored – the most recent full and the most recent differential.

Other backup methods, such as "incremental forever" and "synthetic full backups" allow the backup administrator to do one full backup and then only do incremental backups thereafter. The name and implementations of these vary by vendor, each having their pros and cons. The common characteristic

[11] Discovery that includes the identification, preservation, collection, processing, review, analysis, or production of Electronically Stored Information. [ISO/IEC 27050-1]

of these methods is that they can eliminate the need for periodic full backups, however, building a synthetic full backup consumes computational and storage resources and takes time.

The consolidation processing can be done as part of the backup or alternately by post-processing. When it is done as part of the backup, the backup window can be increased. Methods which consolidate in post processing can delay the availability of the backup for restore.

Another consideration for backups is the data sensitivity of the data set(s). Based on the sensitivity and/or criticality of the data set(s), the backups need to be handled differently. For example, if there is a concern of backup media being stolen, the data should be encrypted. Another example is if the backups of specific data set(s) need to be restored within specific time frames (see **Error! Reference source not found.**), then the media selected should be appropriate to allow for "fast" recovery of those data set(s) after a disaster/disruption has taken place.

Another important best practice regarding backups is to ensure that the backups that are executed can be used for successful data recovery. This can be accomplished with a regular routine of data recovery testing, to ensure data recoverability. The worst time to find out that the backups are not recoverable is when a restore needs to be executed just after a disaster took place. Keep in mind also that the restoration time for a given data set will highly depend on the backup type (e.g., incremental versus full), the backup media to be restored from (e.g., tape versus disk), network topology (e.g., LAN versus WAN), and the overall size of the data set.

There are a few additional considerations for backing up virtualized environments. For virtualized environments, it is best to use a backup application designed to work with virtual machines, so that the backup will take place directly at the hypervisor layer (without involving the guest operating system layer or the VM host) where all the appropriate resources are made available for the backup session workload. If there is external data (outside of the VM image), other methods will be required to make sure that all external data is backed up appropriately. Also ensure that the VM application programming interfaces are used whenever possible for backups, enabling direct access to the VM disk files, and that the "Changed block" feature is used, which allows for quicker incremental backups. Always remember to back up the individual host servers, along with the appropriate VM server(s) configuration files.

Additional backup requirements can be driven by specific business requirements, including the criticality of each data set, regulations, contractual commitments, audit logging, etc.

7.2.3.5 Backups for Cyber-attack Recovery

When production data is damaged or lost, organizations should be able to recover it using replicated or backed up data copies. If the damage is the result of a malicious attack, and the attackers were also able to compromise all recent backup data copies, the attack on the production environment can have a devastating effect since the organization will not have the ability to recover. To improve resilience of backup copies, sufficient isolation should be guaranteed between data assets and their recovery copies.

NIST Special Publication 800-209 suggests that organizations have a separate DP approach for cyber-attack recovery, wherein data copies are hardened, locked, and kept in isolation. The design of such an approach strives to achieve a state where these copies could not be impacted by *anything*, including scenarios where production volumes or other types of copies they are linked to have been compromised.

While not mandatory, cyber-attack recovery is recommended for critical and sensitive systems. To be effective, the cyber-attack recovery scheme needs to be sufficiently compatible with the operational recovery scheme.

Cyber-attack recovery schemes have many considerations that are above and beyond those associated with operational recovery schemes. NIST Special Publication 800-209, sections 4.7 and 4.8, provides relevant guidance.

7.2.4 Snapshots

A snapshot is a point-in-time copy of a defined collection of data. A “delta snapshot” is a point in time copy that preserves the state of the data at an instant in time, by storing only those blocks that are different from an already existing full copy of the data.

Snapshots are a way to create distinct “point-in-time” views of a data set, when performing DP actions such as backups and/or replication; the “view” of the data set is “frozen” in a known state and usually alleviates issues such as open files. Snapshots can offer a near instantaneous, alternative recovery mechanism. The exact implementation of snapshot execution will vary by supplier or provider (whether proprietary or open source).

Snapshots are usually taken regularly, as part of a backup strategy (see 7.2.3). The interval of the snapshots is usually based on the granularity requirements for restoring from a specific point in time. For example, taking snapshots every minute versus every hour will provide greater granularity in restore point capability. The criticality of the data will help in deciding how often snapshots should be executed. So, if the business requires data to be restored to a point in time with a granularity of one-minute, then snapshots should be executed every minute.

Here are some of the considerations associated with the use of snapshots:

- Allows for the recovery of files from a specific point in time (based on snapshot schedule).
- Backup applications can use the snapshot as a “quiescent” view of the data set to be backed up, so that there will be no issues with open files, ongoing modifications, etc.
- When doing application backups, using a snapshot as the source of a backup can greatly reduce the time that the application must be in a quiesce state (backup mode).
- The snapshot-based backup can be performed transparently to ongoing processing, but can introduce severe performance degradation since backup processing will compete with user access for system storage resources.
- Space is consumed for each respective snapshot taken (though each snapshot can require only a very limited amount of storage space).
- There could be performance degradation during the execution of the snapshot, and afterwards while the snapshot is maintained (i.e., a single snapshot might have limited impact on storage performance, but having frequent snapshots at short intervals will have considerable performance impact).

Best practices include using snapshots as part of the backup strategy, so that the source of the backup is the snapshot, such that the backup as well as the restore can be executed from a “quiescent” view of the data set. This assures that there are no “open” files[12] in the snapshot from the data set that was

[12] Open files are an issue since they will usually be skipped when the backup occurs.

backed up. Also, the use of the snapshots for restores allows for a finer granularity (versus regular daily backups), and for the restore of a given data set to a more precise point in time.

Another best practice for snapshots includes executing the snapshot interval in line with the RPO requirements of the data sets that are being protected. So, for example, if the business requires that a specific data set needs to be recovered to within a one-minute point in time, then the snapshots should be taken once per minute.

8 Summary

Depending on the DP technology, the protection takes the form of resilience-oriented and/or recovery-oriented capabilities. This distinction can be important as DP requirements can change as data moves through its lifecycle.

The governance of IT within an organization and the corresponding IT management have dependencies and expectations for the availability and usability of data. Organizations with mature governance practices understand the potential impacts of system outages and establish metrics for what is acceptable (e.g., RTO and RPO). Data protection technologies play an important role in helping organizations meet their RTO and RPO.

Developing KPIs for DP activities can be an effective way of ensuring these capabilities are aligned with strategic organizational objectives. Focusing on organizational, people, physical, and technological KPIs can provide a robust way of measuring success against key DP objectives.

Understanding the capabilities of various DP technologies, as well as their common application, can help ensure the right tool is selected for the job. In addition, leveraging the guidance and considerations in this paper can serve as a basis for more effective DP implementations and operations.

9 Abbreviations

Abbreviations used in this paper:

CNSSI	Committee on National Security Systems Instruction
DMBOK	Data Management Body of Knowledge
DP	data protection
EC	European Commission
ECC	error correction coding
EU	European Union
FIPS	Federal Information Processing Standards
GDPR	General Data Protection Regulation
I/O	input/output
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
IT	information technology
KPI	key performance indicator
LAN	local area network
MTBF	mean time between failures
MTTF	mean time to failures
MTTR	mean time to repair
NIST	National Institute of Standards and Technology
RAID	redundant array of independent disks
RPO	recovery point objective
RTC	recovery time capability
RTO	recovery time objective
VM	virtual machine
WAN	wide area network

10 Acknowledgments

10.1 About the Authors

Eric A. Hibbard, CISSP-ISSAP, ISSMP, ISSEP, FIP, CIPP/US, CIPT, CISA, CDPSE, CCSK

Eric A. Hibbard is the Director, Product Planning – Security at Samsung Semiconductor, Inc. and a cybersecurity and privacy leader with extensive experience in industry (PrivSec Consulting LLC, Hitachi, Raytheon, Hughes, OAO Corp), U.S. Government (NASA, DoE, DoD), and academia (University of California). Mr. Hibbard holds leadership positions in standards development organization and industry associations, including ISO/IEC, INCITS, IEEE, SNIA, ABA, and CSA. He has also served as editor of ISO/IEC 27040, ISO/IEC 27050 series, ISO/IEC 22123 series, IEEE 1619-2018, and IEEE P3454.

Mr. Hibbard possesses a unique set of professional credentials from (ISC)2, IAPP, ISACA, and the Cloud Security Alliance. He has a BS in Computer Science. Learn more at https://www.linkedin.com/in/ericahibbard/.

Mounir Elmously

Mounir is a Senior Manager in EY's Infrastructure and Service Resiliency practice. He brings more than 25 years of relevant resilience experience, including business continuity, disaster recovery, incident/crisis management, cloud and service resiliency. He has served a major technical role for multiple relevant Power and Utility resiliency projects, including both business continuity and disaster recovery work. Additionally, he has led multiple BIA engagements across various industries including education, healthcare, financial services and manufacturing to help management identify critical business processes, recovery requirements (RTO/RPO), and to determine potential financial and operational impact if the recovery times are not met. In addition to evaluation of on-premises, Disaster Recovery as a Service and cloud-based disaster recovery.
Mounir focuses on leading practices for:

- Data storage, protection, archiving and compliance including on premise and cloud implementation.
- Data center modernization (e.g., technology platforms, public and private cloud, software-defined, workload mobility, etc.)

Michael Dexter

Michael is an independent Support provider who specializes in the cross-platform and vendor-neutral OpenZFS file system and volume manager. Michael has spoken at and organized Open Source events around the world for over 20 years, with a focus on virtualization, storage, and BSD Unix.

Thomas Rivera, CISSP

Thomas Rivera has over 25 years of experience in data storage architecture, with specialties in data protection, cybersecurity, and data privacy. Thomas has held a variety of roles in firms such as Broadcom, VMware Carbon Black, and Hitachi Vantara. He also holds the $(ISC)^2$ CISSP certification.

Thomas is actively involved in organizations developing cybersecurity and privacy specifications and standards. He has served as the Co-Chair of SNIA's Data Protection and Privacy Committee (DPPC) and an active member of SNIA's Security Technical Working Group. Thomas also serves as the

Secretary of INCITS/Cybersecurity & Privacy, Secretary of the IEEE Cybersecurity & Privacy Standards Committee (CPSC), and Chair of the IEEE Zero Trust Security Working Group.

Thomas Pearce, BSc, ABCP, CCSP, CCRP, SNIA Certified Storage Architect

Tom has 30 years' experience working in environments of increasing complexity encompassing cloud computing, datacentre build and resiliency, storage, compute and backup/restore platforms as well as virtualization, systems management and network/security architectures and technologies.

10.2 Reviewers and Contributors

The SNIA Data Protection & Privacy Committee (DPPC) wishes to thank the following for their contributions to this technical paper:

Richard Austin, CISSP

James Borden

Glenn Jaquette

Tom Mancuso (SNIA)

SW Worth

Bibliography

[1] ISO/IEC 10116:2017, *Information technology — Security techniques — Modes of operation for an n-bit block cipher*

[2] ISO/IEC 11179-1:2023, *Information technology — Metadata registries (MDR) — Part 1: Framework*

[3] ISO/IEC 15408-1:2022, *Information security, cybersecurity and privacy protection — Evaluation criteria for IT security — Part 1: Introduction and general model*

[4] ISO/IEC 15408-2:2022, *Information security, cybersecurity and privacy protection — Evaluation criteria for IT security — Part 2: Security functional components*

[5] ISO/IEC 15408-3:2022, *Information security, cybersecurity and privacy protection — Evaluation criteria for IT security — Part 3: Security assurance components*

[6] ISO/IEC 2382:2015, Information technology — Vocabulary

[7] ISO/IEC 27001: 2022, Information security, cybersecurity and privacy protection — Information security management systems — Requirements

[8] ISO/IEC 27002:2023 Information technology — Security techniques — Information security controls

[9] ISO/IEC 27040, *Information technology – Security techniques – Storage security*

[10] ISO/IEC 38500:2024, *Information technology — Governance of IT for the organization*

[11] NIST Federal Information Processing Standards (FIPS) Publications 140-3 (FIPS PUBS 140-3), *Security Requirements for Cryptographic Modules*

[12] NIST Federal Information Processing Standards (FIPS) Publications 199, Standards for Security Categorization of Federal Information and Information Systems, February 2004

[13] Committee on National Security Systems (CNSS) Glossary, Committee on National Security Systems Instruction (CNSSI) No. 4009, April 2015

[14] DAMA International, Guide to the Data Management Body of Knowledge (DMBOK), https://www.dama.org/cpages/body-of-knowledge

[15] IEEE 2883, IEEE Standard for Storage Sanitization

[16] ISO 37000:2021, *Governance of organizations — Guidance*

About SNIA

SNIA is a not–for–profit global organization, made up of member companies spanning the global storage market. SNIA's mission is to lead the storage industry worldwide in developing and promoting standards, technologies, and educational services to empower organizations in the management of information. To this end, the SNIA is uniquely committed to delivering standards, education, and services that will propel open storage networking solutions into the broader market. For more information, visit http://www.snia.org.

SNIA
5201 Great America Parkway, Suite 320, Santa Clara, CA, 95054
Phone:719-694-1380 • Fax: 719-694-1385 • www.snia.org