ILM Technical Workgroup

Managing Data and Storage Resources in Support of Information Lifecycle Management

This document provides an overview of the concepts associated with aggregated data service management based on data Service Level Objectives (SLOs). This includes the ability to identify and manage data and storage resources in the delivery of dynamic, distributed computing environments, and the ability to manage and optimally apply these resources using the concepts of Information Lifecycle Management (ILM).

This document offers a snapshot of a work in progress within the SNIA ILM TWG

July 19, 2006

SNIA
STORAGE NETWORKING INDUSTRY ASSOCIATION

Authors:

Dr. Jack Gelb, Sr. Software Engineer, IBM

Co-chair SNIA ILM Technical Workgroup

Edgar St.Pierre, Sr. Staff Software Engineer, EMC^2

Co-chair SNIA ILM Technical Workgroup,

Co-chair SNIA DMF ILM Initiative

Dr. Alan Yoder, Sr. Member of Technical Staff, Network Appliance, Inc

SNIA Technical Council Member

Table of Contents

1 Introduction

This is a forward looking paper intended to spur discussions among various industry groups regarding the management of resources associated with storage and data management in a distributed, dynamic, computing environment. There are many related efforts underway in different industry forums, all of which deliver some, or all, of these concepts in the same or similar fashion. This paper is a stake in the ground intended to engender discussions among these various forums.

The intended audience includes groups defining standards using web services related to business process management, enterprise content management, grid computing and records information management. To a slightly lesser extent, the intended audience may also be anyone interested in the activities of SNIA with regard to data management standards definition in support of Information Lifecycle Management .

This paper is derived from the effort to standardize the management of storage resources as they relate to Information Lifecycle Management (ILM). ILM is a business-driven management practice that uses the value of information and the processing requirements for that information, to set policies and service level objectives (SLOs) for data placement, data protection, and data security services.

IT resources need to respond to changing business requirements. The transformation from business requirements to data processing and data management requirements necessitates the collaboration between line of business users and data center personnel. How data is managed becomes the focal point of that transformation.

The management of data services is best delivered as an aggregation of several underlying services. The underlying services of storage, data protection, and data security are so tightly related that the delivery of one can influence the behavior of another in ways that can either benefit or detract from the overall service delivery. Hence the delivery of services via a single aggregation allows the data center to plan and test configuration templates for service delivery with predictable behavior, reliability and performance.

There is no doubt that this is only one piece of the puzzle needed to deliver truly distributed, dynamic, and heterogeneous solutions to customers. The intent of this paper is to spur discussion among industry groups in an effort to align, and eliminate overlap of, standardization efforts for at least part of the overall solution.

1.1 SNIA Perspective of ILM

ILM is the policies, processes, practices, services and tools used to align the business value of information with the most appropriate and cost-effective infrastructure from the time information is created through its final disposition. Information is aligned with business requirements through management policies and service levels associated with applications, metadata and data.[1]

Since this seems to cover all the bases, let’s reduce the definition to those parts relevant to SNIA and to this paper. ILM provides the means to align business requirements to infrastructure. From a SNIA perspective, this means that there must be a defined, standard, set of data-centric services, Service Level Objectives (SLOs) and lifecycle management functions that support the business process-centric information lifecycle. Likewise, there must be a standard set of attributes that describe the capabilities of heterogeneous storage configurations and associated data management functions that service data.

The SNIA ILM effort *supports* the definition of business process workflows that might include information states for documents to transition through, but SNIA does not anticipate defining any standards in this area as that is more appropriate in other industry groups. Likewise, the use of SLOs to specify compute and network requirements are certainly part of ILM, but SNIA does not anticipate defining standards that are best defined by other industry experts.

This paper focuses on areas that SNIA does expect to contribute to the standards effort:

[1] SNIA Dictionary: http://www.snia.org/education/dictionary

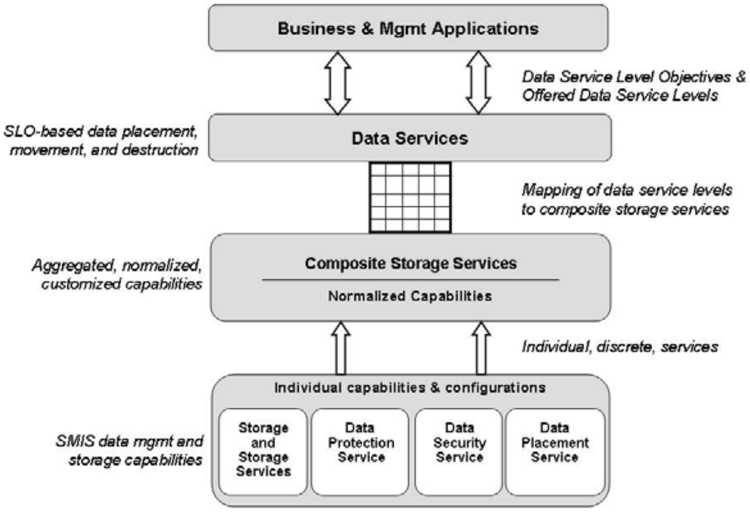
- Definition of standard data SLOs that can be serviced by storage and data management products such as backup, replication and networked storage. These SLOs are a type of Key Performance Indicator[2] (KPI) used to describe service levels with respect to data rather than service levels with respect to application performance. As such, they can be used across any type of data or application.
- Service aggregation for storage and data management products that address key customer management issues associated with compliance, reference data and resource classification.
- Policy-based data lifecycle management to allow for proper classification and placement of data over its lifecycle. Such management provides for standard data classification definitions, and it also allows for application and content-based classification methods to drive data classification. The definition of such application and content-based classification methods are beyond the scope of this work. For example, the classification of data to determine which governmental regulations to which it may be subject is a knowledge-driven activity. When this activity is easily categorized based on file system metadata, then this work will provide for those policies; when this activity requires content-based classification, then this standards work will allow for it through external classification services.

1.2 Abstracting Data Services

This specification introduces levels of abstraction, as depicted in Figure 1 .This is not a formal architecture, but rather, a representation of a hierarchy of capabilities.

It is organized into three layers of abstraction, plus a fourth consumer layer of Business and Management Applications. At the bottom are the raw capabilities available from storage and data management products. In the case of storage, many of these services have existed since the earliest versions of SMI-S. In the case of data management, many of these are to be defined.

Figure 1: Levels of abstraction


Business & Mgmt Applications
Data Service Level Objectives & Offered Data Service Levels
SLO-based data placement, movement, and destruction
Data Services
Mapping of data service levels to composite storage services
Aggregated, normalized, customized capabilities
Composite Storage Services
Normalized Capabilities
Individual, discrete, services
Individual capabilities & configurations
SMIS data mgmt and storage capabilities
Storage and Storage Services
Data Protection Service
Data Security Service
Data Placement Service

Above the individual resources is the combination of these resources based on specific configurations planned for a data center. The Composite Storage Set (CSS) is an aggregation of storage and services representing capabilities of different configurations to produce equivalent function. This layer also allows for a customization to site-specific requirements and preferences. It also allows for both manual and automated creation of configurations.

Data Services is comprised of Data Classification, Lifecycle Management, and Service Level Management. It provides a translation between *storage* behavior and offered *data* service levels. This is the data service management interface through which applications may specify data requirements as service level objectives (SLOs), and through which the data management layer offers

[2] KPIs have a fairly broad definition in ITIL: http://www.ogc.gov.uk/embedded_object.asp?docid=1000379 whereas Data SLOs are intended to narrow that definition to that which is common across any application’s use of data. A Data SLO may be part of a Service Level Agreement, though application-specific KPIs seem better suited for use in an SLA. Data SLOs are best used in describing the capabilities of the service catalog or in specifying requirements for data..

services that are described by data SLOs. This management interface is used to define and manage the data lifecycle as a set of policies. Data center administrators (IT architects, storage admins, etc) define the relationship between Offered Data Service Levels and Composite Storage Sets. In fact, all of the relationships within this hierarchy are based on mappings performed by administrators either manually, or – as technology evolves, using management software with varying degrees of automation.

Overall, the key elements for this architecture include:

- **Standard definitions for Service Level Objectives (SLOs) for data:** This concept provides a standard framework for communicating SLO attributes either as requirements or as a defined service level offering. This enables the specification of data service levels to exist independent from the capabilities that can meet those requirements. As capabilities change due to advances in technology or environmental modification, existing SLO attributes can be simply – and non-disruptively - mapped to the new capabilities. Data SLOs are to be used by client applications to specify data requirements, and by a Data Service Resource Manager to define data service level offerings.

- **Standard properties to describe configurations of data and storage services**: This definition will focus on *expectations* of performance and behavior for specific product configurations. The data center determines what these expectations are rather than vendor-generated tools or documentation. In particular, these capabilities represent not just the performance and behavior of a single product, but of a combination of products, whose combined configuration may be necessary in order to deliver predictable levels of service. The "capabilities" provided by this configuration template is referred to as the **Composite Storage Set**.

- **Management of services in the data layer**: A fundamental aspect of providing these services includes the ability to configure, monitor, and control data management services such as data backup, data replication, data security, and data movement. This includes the ability to perform both ad hoc and policy-based operations.

- **Abstract provisioning and management of storage and services in the storage layer**: This will be provided via current and future versions SMI-S.

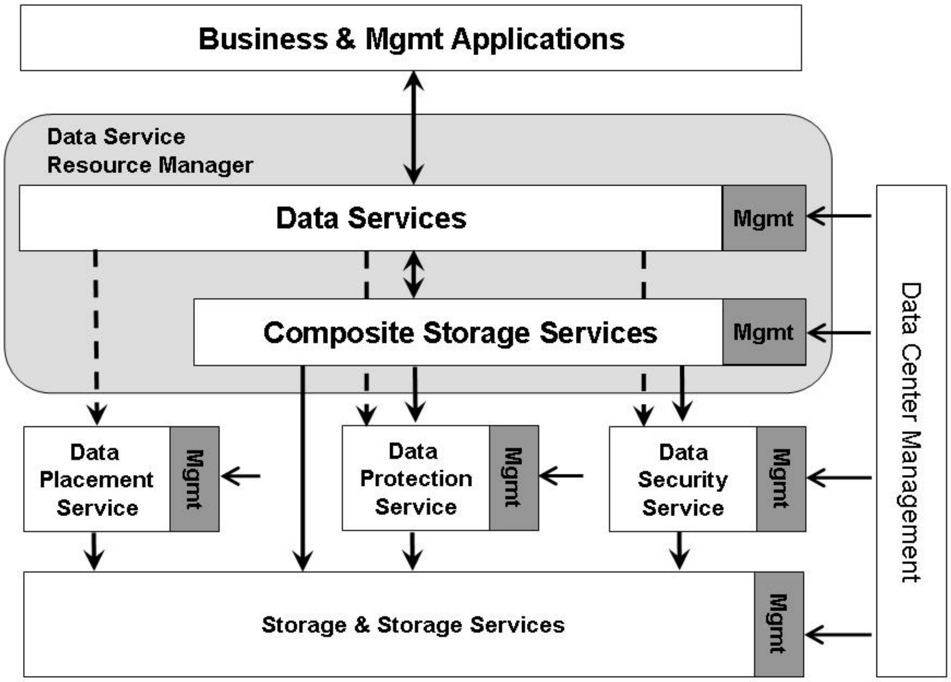
- **Data Lifecycle Management Policies**: Policies that allow for automating the decision-making in relation to events that cause a change of SLOs to be applied to data. These policies are derived from the information lifecycle requirements and provide the specification of that lifecycle within the data management layer.

- **Data Classification**: Structures that allow data to be organized into groups for management purposes such as service level management, lifecycle management, etc. This may include policies for automating the classification of data.

2 Conceptual Model

Data services are embodied in a Data Service Resource Manager, whose components are illustrated at an abstract level in Figure 2. This section defines the basic functions associated with each of its components and its underlying services. The “standardization” of ILM is concerned with how these components and services interface, and with the definition of each of their management interfaces.

Figure 2: Data Service Resource Manager Components



Data Placement, Data Protection, Data Security, and Storage & Storage Services: These are each distinct services to the Data Service Resource Manager. Each must provide for discovery, configuration and management of resources that are part of a data center and their capabilities so they may be used as part of a Composite Storage Set (CSS). These raw services may be managed as distinct management domains, or integrated into an overall management environment. Distinct services are aggregated into a single service offering via the CSS. Data Services may also offer pass-through management functions (e.g., copy, delete, make immutable) directly from these underlying services.

Composite Storage Set (CSS): The CSS provides for the configuration and management of the combination of disparate resources that are known to work together to provide a predictable level of service. The CSS provides for the definition of expected behavior and performance from those resources as part of a CSS configuration. This implies that there is expert knowledge used to combine the resources into working configurations, and to define the expected behavior. The CSS model allows for implementations to vary from discovery and documentation of existing configurations within a data center, to manual configuration of CSSes by data center architects, to a fully automated combination of resources selected dynamically to satisfy data SLOs on a per-object or per-transaction basis.

Data Services: Data Services provides for classification of data, data service level management, and data lifecycle management. The Data Services component provides a service oriented interface to **Business and Management Applications**, who use data Service Level Objectives (SLOs) as attributes to specify service requirements. A collection of SLOs to specify data requirements is referred to as a SLOG, or SLO Group. Such requirements are generally the result of a classification process performed by consumers to determine what service is needed by the data over its lifecycle. Data Services may also use SLO properties to describe the Offered Data Service Levels

(ODSLs) that are supported by the Data Service Resource Manager. Data Services will provide management functions to manage the mapping between ODSLs and CSSes. Finally, Data Services also provides for configuration and management of data lifecycle policies used to transition data through its information lifecycle.

The following sections provide further definition of the role of the Data Service Resource Manager.

2.1 Composite Storage Set

In the implementation of a data center solution for an application, there is often a strong dependency between the data management functions, the storage and the storage services to be used in delivery of that solution. A Composite Storage Set is a specific collection of data management functions, storage, and storage services used by a data center architect (aka IT Architect) to define a solution which can deliver a predictable level of service for data.

The conundrum that has faced both vendor and customer alike is that various combinations of different product configurations may produce equivalent levels of service for data. The measurement of throughput or availability at any one point in the product stack did not provide an adequate picture of overall throughput or availability for the various combinations of sometimes very different product configurations.

In addition, the vagaries of data center management are such that every data center has its own particular nuances associated with vendors, configurations, and measurements. To address these issues, the CSS provides for the ability to normalize and customize the raw data center capabilities into a set of specific configurations managed by the data center that are known to deliver a particular level of service in that environment. Minimally, CSS is a documentation scheme for the IT Architect's selection of available technology configurations so they may be mapped to and from Offered Data Service Levels.

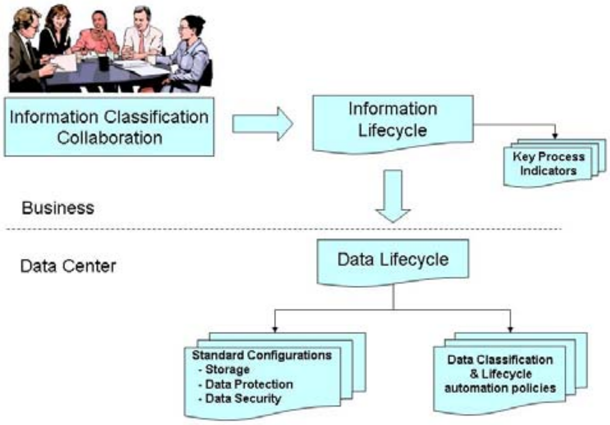
2.2 Data Services

Data Services is comprised of Data Classification, Lifecycle Management, and Service Level Management.

2.2.1 Data Classification

Data classification is the organization of data into groups for management purposes. A purpose of a classification scheme is to associate service level objectives with groups of data based on their value to the business and their processing and storage requirements.

Figure 3: Information & Data Classification



A precursor to data classification is information classification, which is used to associate groups of data with a particular lifecycle. An information classification scheme is generally built through collaboration between the data center, lines of business, records information managers, and other corporate stakeholders in the availability and use of information. These concepts are illustrated in Figure 3 and in the top of Figure 4.

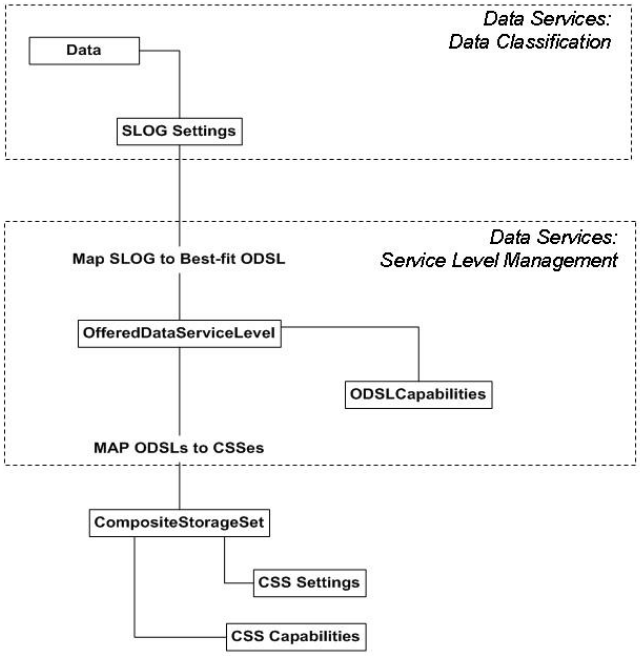
Information classification is not part of the SNIA standards effort; the management of data classification functions is part of that effort.

2.2.2 Service Level Management

Offered Data service levels (ODSLs) are advertised to consumers, such as email or enterprise content management applications in the Business and Management Application Layer, using the "capabilities" model as currently implemented in CIM[3].

The properties of ODSL capabilities are based on data SLO attributes, which are the same set of attributes used as "settings" when assigning data SLO requirements. The settings for data requirements are referred to as its SLO Group, or SLOG. Data SLO attributes are defined later in Appendix B.

Figure 4: Data Classification and Service Level Management



Service Level Management provides the functions to create and manage ODSLs and their capabilities, as illustrated in Figure 4. Service Level Management also provides the ability to perform "best fit" mappings from Data's SLOG Settings to an ODSL, and from an ODSL to CSSes.

These mappings may be manually created by an IT Architect, policy-driven based on a static rule set defined by the IT Architect, or even a dynamic set of mappings derived by an adaptive policy engine.

The intent of SLOGs/ODSLs and CSSes is to provide two, "semi-independent" abstractions: one to characterize *data* requirements and services, the other to organize *storage* system capabilities. By relating data to SLOGs and ODSLs, we allow requirements to change over time (or business necessity) through re-assigning the data-to-SLOG relationship. The SLOGs themselves remain unchanged (unless, of course, there are new objectives introduced). Similarly, changes in topology, technology, or characteristics in the storage layer or the data management layer may not change the CSSes themselves, just the mapping of how those CSSes are physically realized. Or they may create new CSSes to which existing ODSLs may be mapped, allowing data to migrate to more efficient resources over time with no change in requirements..

2.2.3 Lifecycle Management

Data lifecycle is an abstract concept that is implemented through the use of SLOGs and SLOG assignment policies. A data lifecycle is derived from, and supports, the information lifecycle. This includes ad hoc lifecycle events which are not pre-planned such as "apply this SLOG now". The distinctions between the information and data lifecycle concepts include:

- "Information lifecycle" is the definition of business requirements and business value that apply to data, and the events that may change the value and requirements over the data's lifetime. These are expressed in terms of Key Performance Indicators (KPIs) that are specific to the information and/or the application. The information lifecycle is not modeled in SMI-S.

[3] Common Information Model defined by DMTF http://www.dmtf.org and used by SNIA in SMI-S.

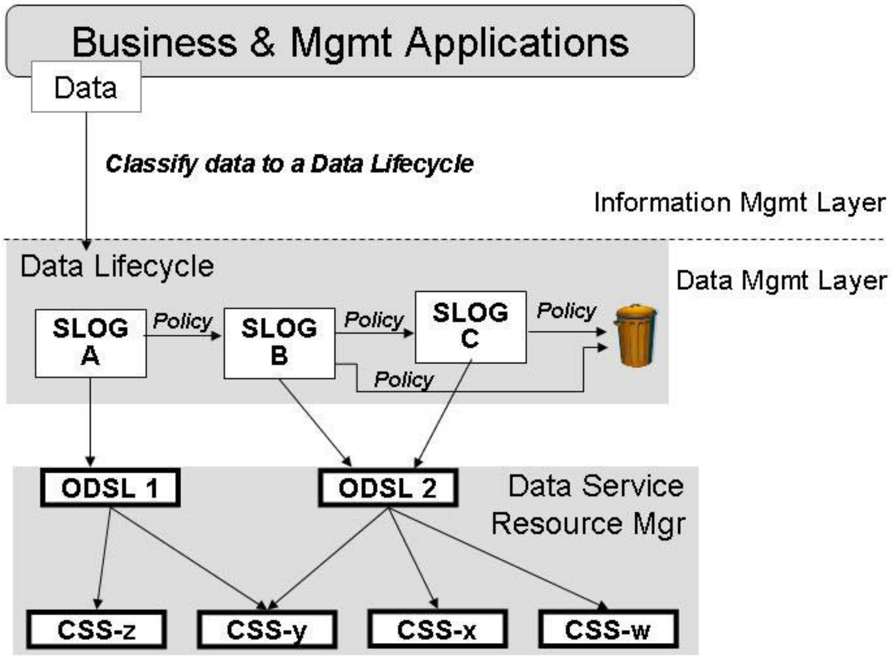
- "Data Lifecycle" is the definition of data SLOGs that apply during each phase of the information lifecycle, and the SLOG assignment policies used to transition data to either the initial or a new SLOG. The data lifecycle supports the information lifecycle, but is expressed in data-specific terms. The data lifecycle is an abstract concept that is modeled in SMI-S through the use of data SLOGs and SLOG assignment policies.

See Figure 3 on page 8 for a conceptual illustration of this relationship.

The concepts of data classification, service level management and data lifecycle management are combined in the illustration of **Figure 5**. Business applications at the Information Management Layer produce data. ILM-aware Business Applications may classify data when it is created. Management Applications may be used to classify data after it's been created by non-ILM-aware Business Applications.

There is a "best fit" service level mapping performed from the desired SLOG at each stage of the data lifecycle to an Offered Data Service Level. Today, this is typically performed as a discussion or negotiation between a Line of Business Application Administrator and an IT Architect or Storage Administrator. Also, the IT Architect has defined a set of valid mappings from each ODSL to one or more CSSes that are capable of meeting the service levels defined in the ODSL.

Figure 5: Classification, Service Level Mgmt, and Data Lifecycle



SLOG assignment policies, shown as "***Policy***" in **Figure 5**, are used to apply a new set of data requirements for each stage of the data lifecycle. Sometimes, the application of a new SLOG will result in the mapping to a new ODSL, and sometimes it will not. Likewise, the mapping to a new ODSL may or may not result in the mapping to a new CSS (see the case of CSS-y which is capable of supporting both ODSLs). The mapping to a new CSS may have consequential actions such as the movement of data to a new storage location. There are also SLOG assignment policies that determine that it is time to delete the data.

3 Use Cases

These use case narratives describe a small sample of business and system level interactions related to Data Service Resource Management. They are not intended to be a complete representation. As a further simplification, these use cases utilize three archetype actors:

- **LOB**: The Line of Business representative that has a profit and loss interest in information. This role may be performed by one or more distinct actors such as an Application Owner, a Product Manager, a DBA that reports into the Line of Business, or Business Process Analyst.
- **ITA**: The data center's Information Technology Architect. This role may be performed by one or more actors such as the CIO, the IT Director, Data Architect, Solutions Architect, System or Storage Administrators.
- **RIM**: The Records Information Manager role is used as a front end to corporate governance responsibilities associated with legal and security concerns. Alternatively, this role may be performed by one or more actors from corporate such as the CSO, CCO, corporation counsel, or their representatives.

The "system under discussion" for the majority of these use cases is a "Data Service Resource Manager" which provides an interface between the management of data service level requirements and the resources that deliver those services. This includes the management of data Service Level Objectives (SLOs), service level alignment using Offered Data Service Levels (ODSLs) over the course of the data's lifecycle, and aggregation of storage, data protection, and data security into Composite Storage Sets (CSSes).

3.1 Business Process Workflow Integration

3.1.1 Classify a New Application's Information [UC-2005-010]

A business is introducing a new line of products that requires new production, sales, and back office processes. The business processes for this have been defined and the application to create and manage the information for the business processes identified by the LOB. This use case describes the process associated with classifying the new application's information.

The use case begins when the LOB representative briefs the IT Architect on the requirements for the new application. The IT Architect then arranges a meeting for all Classification Team participants, which includes all corporate stakeholders such as represented by the RIM and ITA, plus other LOBs that may have overlapping data with the new application. At the classification meeting, the LOB representative describes the applicable business process, its use of the new application, the information associated with the new application, and how the use of that information might change over time or based on business events. Each participant is expected to contribute their expertise to define the lifecycle for the information, and the information requirements at each stage of its lifecycle. At each stage, the LOB expands on information requirements in terms specific to the business use of the information, such as "20 seconds for a web page to update or display while in the shopping cart". The RIM and other LOBs identify requirements on the information that may be either related to, or independent of, its intended business use, such as required retention and access restrictions. The team also defines the conditions that may cause a change in the information requirements. The IT Architect leads the team in matching the information requirements for each stage to an ODSL provided by the data center. The IT Architect reports on the total projected cost for the new application and its information lifecycle, which includes the costs specific to data services over the information's lifecycle. If the data services cost is acceptable to the LOB, then the ITA is free to configure the data service using any of the CSSes which satisfy the requirements without exceeding the specified LOB budget, and the use case is complete. If the data service cost is not acceptable to the LOB, then each participant must consider reducing requirements in exchange for additional risk. This includes modification of the lifecycle and/or modification of data service requirements over the lifecycle. This repeats until the cost does not exceed the LOB's maximum budget, which then ends the use case. For follow-up by the IT Architect, see Section 3.2.3 Define Data Policies.

3.1.2 Request Grid Data Service Resources for New Application Instance

A web based retailer has organized its IT resources into grid-based solutions for many of its applications, including those that require additional resources during peak retail periods around the holidays. A grid computing automation program operating on behalf of a LOB monitors application activity to add/delete application instances as required.

This use case begins when the grid computing automation program determines that a new application instance is required due to a spike in internet order activity. The automation program requests server, network, and data service resources for the new application instance (the details of server and network are out of scope for this use case[4]). The automation program has been preconfigured to request the ODSL named "Business Critical" from the Data Service Resource Manager. The grid automation program sends the request to the Data Service Resource Manager along with instance-specific criteria such as requested capacity and the name of the allocated server. The Data Service Resource Manager determines which CSSes were previously identified that could meet the service level objectives associated with "Business Critical" and that have the resources available to meet the requested capacity requirements. The Data Service Resource Manager selects the CSS that meets the requirements based on selection policies previously specified by the IT Architect. The Data Service Resource Manager then initiates provisioning activities for the storage hardware, storage software stack, data protection (backup, replication), and applicable security characteristics. The Data Service Resource Manager returns the access information for the data storage to the grid automation program. The grid automation program then sends a request to the Data Service Resource Manager to replicate data for the application to the new location. The Data Service Resource Manager identifies the best Data Placement service to copy the data and initiates the copy. Once the copy is complete, the Data Service Resource Manager informs the grid automation program and the use case is complete.

3.1.3 Automatically Add Resources to Existing Application

A newspaper has launched a for-fee online news delivery service with access to its archived news stories. It has defined the ODSLs that are part of the news story archive's lifecycle as being "Immediate Access" first, then move the news story to "Get It Tomorrow Access" after one year. It knows that growth of the news archive is inevitable due to time and due to a continued increase in the amount of news being delivered to its readers. The business has budgeted for Immediate Access growth to sustain an increasing amount of news delivery. The filesystem used by the LOB for the news archive is capable of automatically requesting additional data storage resources when %utilization thresholds are reached. This use case assumes that growth within predefined capacity planning limits may be fully automated, whereas capacity increases beyond predefined limits require human workflow intervention.

The use case begins when the filesystem detects that the %utilization for the Immediate Access data has exceeded a predefined threshold. The filesystem sends a request to the Data Service Resource Manager to allocate additional resources. The file system always grows its resources by a predetermined amount consistent with its predefined capacity growth plan. The Data Service Resource Manager determines if the request meets predefined capacity growth rate criteria. If not, then an email is sent to the ITA to begin a manual workflow process to approve additional resource allocations or other remediation actions such as ad hoc movement of data to "Get It Tomorrow Access". If this is within the predefined growth rate, then the Data Service Resource Manager determines if the current CSS has sufficient resources to add to the existing allocation. If not, then the Data Service Resource Manager initiates a data movement of the file system's data to a CSS that will support the required capacity and growth. (See 3.3.2: Implement ad hoc SLO Change Request from LOB) This data movement is transparent to the file system. Otherwise, the Data Service Resource Manager uses the storage provisioning service to add the additional resources to the existing file system's resources. The use case ends when the file system is informed that additional resources are now available.

[4] We assume for the sake of this paper that server provisioning includes virtual machine allocation that is comprised of a virtual processor, application and file system for processing, but no storage allocation for database or web hosting, which is provisioned through data services as a second step.

3.2 Data Center Management

3.2.1 Discover & Classify Storage Resources [UC-2005-023]

The ITA wishes to create standard data storage configuration templates to be used throughout the enterprise as part of a storage consolidation initiative that is intended to reduce capital and operating expenses. These configuration templates, also known as Composite Storage Sets, or CSSes, are comprised of existing storage and data management resources as well as planned purchases of the same.

Using a Data Service Resource Managment facility, the ITA first discovers all of the existing storage (block, file, and other), data protection (backup, replication, and other), and data security resources in his environment using both automated and manual discovery tools.[5] For each CSS, the ITA identifies the storage resources that may be used for that CSS and the configuration in which to use that storage for the template. The ITA may have organized those storage resources into smaller specific configurations or into larger pools of multi-use resources beforehand using storage resource management tools. The ITA then selects zero, one, or more data protection services to use in conjunction with those storage resources.[6] The ITA had previously configured and organized the data protection service to provide specific data protection services such as backup to tape, backup to disk, rotation of snapshots, etc. The ITA then identifies any specific security features associated with these configurations, such as the use of encryption devices or WORM functionality.

Having previously tested this combination of product features and functions, the ITA then identifies expected behavior with respect to storage performance, data restoration, and security capabilities for this specific configuration template. These attributes describe such capabilities as known throughput for defined configurations, data restoration capabilities for specific data types and sizes, as well as security attributes associated with accountability, confidentiality, and more. This process repeats for each CSS to be maintained by the data center.

3.2.2 Create ODSLs from CSSes

The business has decided to assign IT costs to each LOB budget in order to optimize spending on IT resources. IT has been given the responsibility for service delivery and chargeback to the LOB, and the responsibility to reduce service delivery costs over time. The LOB, ITA, and RIM collaboratively determine classifications for information over its lifecycle (see 3.1.1 Classify a New Application's Information [UC-2005-010]). The ITA has created a number of CSSes for the data center using a Data Service Resource Manager, and is ready to create Offered Data Service Levels as part of the service catalog for use by the LOBs. Whereas the CSSes are comprised of specific technologies and solutions, the ODSLs are abstractions of the data service levels that will be delivered to the LOB with those configurations. *[It is assumed that the creation of initial ODSLs is the result of consultation with each LOB about data requirements and the creation of CSSes with solutions that can meet those requirements. The definition of each ODSL is the last step in what is likely an iterative process.]*

The use case begins when the ITA uses the Data Service Resource Manager to create a new ODSL based on a CSS. Whereas the CSS template describes the technologies, solutions, and configurations used to deliver data storage and data management services, the ODSL describes the data service levels that are to be provided as a result. The ITA uses previously conducted test results[7] of that template to define the data SLOs provided by this configuration in terms of:[8]

- Cost
- Avg I/O Characteristics (Rate, Throughput, Latency)
- Min/Max Allocation size and growth rate

[5] For detailed analysis, this must be split into TWO use cases.

[6] Operational Recovery and Disaster Recovery data protection solutions may be quite distinct and separate (e.g., rotation of snapshots and B2T), related (B2D and copy to tape), or one and the same (just B2T).

[7] While not preferred, it's also possible for the ITA to use vendor-supplied values for service level characteristics.

[8] See Appendix B for a detailed discussion of Data Service Level Objective attributes.

- Availability
- Data restore characteristics (RTO, RPO)
- Security and compliance capabilities

The ITA then assigns a name to the ODSL, such as “Business Important”. Names are generally assigned based on a classification naming scheme such as “Mission Critical, Business Critical, Productivity, etc.”, or “Gold, Silver, Bronze, etc”. The ITA may then map one or more additional CSSes to the same ODSL. This mapping will allow the introduction of newer, more efficient, and less expensive solutions over time in order to meet the goal of reducing service delivery costs over time. The use case ends when the ITA has finished mapping all of the CSSes that can and should be mapped to the ODSL.

3.2.3 Define Data Policies

A business is introducing a new application, as described in Section 3.1.1. Based on input from the Information Classification meeting, the ITA must define new data policies to implement the correct placement of data at each stage of the information lifecycle.

The use case begins when the ITA uses the Data Service Resource Manager to identify data that either exists, or will be created, as part of the information lifecycle. All data that shares the same information lifecycle is organized as an Application Data Group. This data is identified using metadata such as file system mount points, file owners, or keyword content of the data in order to identify its Application Data Group. The ITA then defines the initial placement policy for data in this Application Data Group by assigning it an ODSL (e.g., “Business Critical”).

The ITA then uses the defined information lifecycle to determine what conditions will cause a new ODSL to be assigned to the data. These conditions are derived from the information lifecycle, but are specified in data management terms available to the ITA. The ITA uses the Data Service Resource Manager to specify the appropriate condition and the new ODSL to be applied to the data when the condition is met. These conditions might be based on file age, owner, or other metadata. The possible conditions are limited by the lifecycle management tools available in the Data Service Resource Manager.

The use case ends when all pre-defined changes in the information lifecycle have been defined, including the destruction of data, if applicable.

3.2.4 Move Data to Non-disruptively Replace Storage Equipment

The data center has leased a new storage array to replace an older array whose lease is about to expire. The new array provides greater capacity, performance, and reliability than the array it is replacing. The IT Architect wants to transparently move all the data from old to new without disrupting active users.

The use case begins when the IT Architect uses the Data Service Resource Manager to create one or more new CSSes specifying the new storage array resources with its associated configurations of data protection and data security. Each CSS with resources from the new array may provide increased levels of performance and reliability. The IT Architect maps the new CSSes to new and existing ODSLs, as desired, until all CSSes to be retired have new CSSes mapped to the same ODSLs. The IT Architect then initiates the movement of data from the old to new storage array using the Data Service Resource Manager to determine which Data Placement Service has access to both old and new CSSes. From this list, the IT Architect selects a Data Placement Service that provides for movement of data that will be transparent to applications. The IT Architect reviews then schedules that specified data movement. The IT Architect is notified when all data movement from old to new storage arrays has completed. The applications continue to receive the desired service level (or better) before, during, and after the data movement. The use case ends when the old storage array can be removed.

3.3 Data Management Scenarios

3.3.1 Classify New Data When Created [system level]

In order to reduce costs through consolidation, a business keeps all of its user data in a single file system that is capable of delivering different levels of service. The data ranges from users’ personal music files to spreadsheets that capture current sales rollup data that must be protected from unauthorized access per SEC regulations regarding insider trading. The file system supports the creation of data by any desktop application. Data policies have been defined in the Data Service Resource Manager in support of each Application Data Group's information lifecycles. This use case describes system level operations that occur when a new data file is created by a desktop application.

The use case begins when a desktop application creates a file in the file system. The file system's data placement service will pass all of the metadata associated with the file to a Data Service Resource Manager for classification, lifecycle, and service level assignment. (See use case in Section 3.2.3.) The metadata includes information such as filename, owner, etc. The Data Service Resource Manager applies a predefined set of classification rules to the data in order to determine its data lifecycle and in particular, its initial ODSL. The initial ODSL is then mapped to one or more Composite Storage Services that are available to the data placement service that originally requested the classification. This list of CSSes is sent to the data placement service for it to choose from based on local optimizations. The use case ends when the data placement service stores the data according to this initial data placement policy.

3.3.2 Implement ad hoc SLO Change Request from LOB

The newspaper in the use case 3.1.3 Automatically Add Resources to Existing Application has decided to temporarily change its policy regarding the age of news articles moved from "Immediate Access" to "Get It Tomorrow Access" because the growth rate of the Immediate Access resources has surpassed budget planning estimates.

The use case begins when the LOB uses the Data Service Resource Manager to query the size of all data currently classified as Immediate Access whose age is older than 10 months. The Data Storage Resource Manager scans for all data matching those criteria, and then provides a report to that effect. The LOB determines that the reported size of that data is sufficient to allow the file system's use of Immediate Access resources to grow at the current rate through the rest of the current fiscal year, at which time the new budget will be adjusted accordingly. The LOB then performs a one-time SLO change request by classifying all data whose age is older than 10 months to the ODSL of "Get It Tomorrow Access". The Data Service Resource Manager scans for all data matching these criteria and determines whether that data can be moved to a more cost-effective configuration. Where applicable, the Data Service Resource Manager initiates the data movement using data placement services available to it to perform data movement. This data movement may take place immediately or over time. The use case is complete when all data has been moved to its most appropriate CSS.

3.3.3 Destroy Data per Defined Policy (workflow)

A business has classified its user data in order to specify the business and regulatory requirements for data, and how those requirements may change over time as expressed in an information lifecycle. The lifecycles of some data require that the data be deleted according to definable policies, and the business wants to automate the process of deleting data while still providing safeguards against deletion of data should be treated as exceptions.

The use case begins when the Data Service Resource Manager scans the files in the user data for conditions which would qualify for deletion according to a particular lifecycle. This may be part of an overall scan for files that require a change of SLOs. When Data Service Resource Manager identifies a file that requires a change of SLOs involving deletion, it determines who the LOB owner is for the file, and which RIM is responsible for the regulatory conformance for that lifecycle. The Data Service Resource Manager notifies the RIM about files that should be deleted according to a particular information lifecycle. The Data Service Resource Manager notifies the LOB file owner about each of his files that need to be deleted according to the information lifecycle. If both LOB and RIM acknowledge and approve the deletion, then Data Service Resource Manager executes the data deletion using the data destruction facility specific to the data's lifecycle, which ends the use case. If one or the other does not approve of the deletion, then the other is notified and the Data Service Resource Manager will place the activity on hold pending manual intervention and this use case ends.

3.3.4 Search Data for Litigation Discovery

A competitor has filed a patent infringement lawsuit. The time frames associated with initial invention from their patent has come into question however, as engineering believes that it had actually produced prior art in this area without filing a patent. The legal department needs to search all engineering archives for proof of when engineering's invention was first identified.

The use case begins when the RIM requests that a copy be made of all engineering documents associated with the "Omega" product from 5 through 8 years prior. In order to reduce the scope of the search, the RIM determines that only documents (of any type) that were created by the CTO, who claims the contested functionality was part of one of the first few releases of "Omega". The search is also broadened to include any member of the project team that the CTO claims may have been involved with development of the invention. The RIM specifies the search criteria and invokes a search to the Data Service Resource Manager. The Data Service Resource Manager scans the file systems associated with the engineering dept for files that match the requested criteria. When a list of files are produced, the RIM requests the Data Service Resource Manager to copy each of the files for further discovery to a new data store for which the RIM has specified a set of Data SLOs that include WORM functionality. The RIM then informs legal of where the target files have been stored so they can conduct further discovery searches.

3.3.5 Replicate Data for Repurposing

A LOB needs to update its current business process to accommodate new features in the order entry process. The LOB determines that the process transition requires a database modification and must be tested using recent data from the application.

The use case begins when the LOB queries the Data Service Resource Manager for a list of Application Data Groups associated with the application. The Data Service Resource Manager returns a list of Application Data Groups, from which the LOB selects all those related to the business process update. The LOB requests the Data Service Resource Manager to replicate all data from the selected Application Data Groups to new storage accessed by a test server using a replication method that provides a consistent copy across all Application Data Groups. The Data Service Resource Manager determines which CSS has sufficient resources to meet the service level requirements for the test data (as defined by the LOB) and provisions the CSS resources to the test server. The Data Service Resource Manager then determines which Data Placement Service provides the most efficient method of copying the data with consistency and asks the LOB for when he would like to schedule the replication. The LOB selects the desired time for replication. At the scheduled time, the Data Service Resource Manager uses a Data Placement Service to create a point-in-time copy of all the Application Data Groups. The Data Service Resource Manager then uses a Data Placement Service to copy all of the data from the point-in-time copy to the newly provisioned CSS storage resources to end the use case.

3.3.6 Scan Data For Service Level Alignment

A data center has several NAS file servers for serving up its user files. One LOB using these file servers has determined that 60% of its user files have not been accessed in over a year, and it could reduce its service chargeback costs significantly by using a less expensive data service classification for such files. In addition, it could save even more by deleting files that have not been accessed in more than 2 years. The LOB has already conducted a classification meeting with the ITA as described in Section 3.1.1Classify a New Application's Information [UC-2005-010], and the IT Architect has entered the appropriate Data Policies into the Data Service Resource Manager as described in Section 3.2.3.

The use case begins when the LOB schedules a "Scan Data for Service Level Alignment" activity with the Data Service Resource Manager for each of the LOB's Application Data Groups. At the scheduled time, the Data Service Resource Manager directs a file level Data Placement Service to discover each file and its metadata in each file server serving up one of the specified Application Data Groups. Each Data Placement Service sends all discovered file metadata to the Data Service Resource Manager. The Data Service Resource Manager classifies each file according to its appropriate lifecycle, and determines to which ODSL the file is entitled. The Data Service Resource Manager then determines on which CSS each file is currently placed and whether there is a mapping between that CSS and the file's current ODSL entitlement. For each file that does not have a valid mapping between the file's current ODSL and current CSS, the Data Service Resource Manager assigns the most appropriate CSS. Where applicable, the Data Service Resource Manager initiates data movement using data movement services available to it. This data movement may take place immediately or over time. The use case is complete when all data has been placed on its most appropriate CSS.

4 A Phased Approach to Deployment

There is obviously too much for vendors to define and implement all at once. Likewise, there would also be too much for customers to adopt all at once – exactly where the knee of the curve lies is up to the market to determine. The following phases describe a very rough, very general, roadmap for definition of data management standards supporting ILM, how that may relate to product implementations, and therefore possibly foretell the phases in which customers will adopt the use of data management applications in support of heterogeneous environments.

4.1 ILM-Unaware Applications and Storage

Data management applications must start as an overlay to existing environments of applications and resources. As retrofitted solutions, the data management applications must support the most basic ILM goal: place data on the most appropriate resources.

To do so requires two basic facilities for heterogeneous solutions:

- Allow an administrator to define the capabilities of existing storage configurations – not just storage devices; but rather, the overall capabilities of a storage configuration including protection and security. In addition, leverage the overall trending towards configuration consolidation and provide templates that describe the similar capabilities of perhaps many related configurations.
- Allow the movement of data across different storage configurations that provide different capabilities. This should be modeled to support file, block, and object data; however, given the proliferation of unstructured data in customer data centers, the primary goal should be to support file-based movement by products that may be internal to, or external from, file system implementations. These methods must also be supported by mechanisms to ensure authenticity of the moved data.

In this first stage, the policy engines available from vendors today that coordinate data lifecycles, perform classification of data, coordinate between service level requirements and storage configuration capabilities remain proprietary, but they may use the standard facilities described above. All data movement remains transparent to the applications using the data, and the resources themselves are simply organized rather than managed.

4.2 ILM Data Services

The second phase of data management applications extends the retrofit nature of these applications to bring more value-added functionality to the data center.

Evolution in the previous phase continues while the new work focuses on true “data services”:

- Standard definitions for data service level objectives that can be used by data centers to manage the service classifications in their Service Catalog.
- A data management application with a service interface to manage the mapping between standard configurations and service classifications in a Service Catalog. Likewise, this service level management application is able to resolve the negotiation between data service level requirements and service classifications in the Service Catalog.
- Data classification service definitions to allow data lifecycle management applications to use the specialized expertise and capabilities of other classification service providers. This expands the scope of these services to incorporate boutique and business vertical providers.
- Data lifecycle management policy engines with a service interface for managing policies that control the lifecycles of file, block, and object based data, including the automated destruction of data, as determined by business and regulatory requirements
- Standards for metadata definition and use with respect to data management begin to evolve.
- Notifications for failures within data management applications are introduced.

The ultimate consumer in this scenario is likely to be vendor-specific data lifecycle management applications that incorporate and/or utilize the other services. But Vendor A's data lifecycle management application could utilize the standard service level management facilities of a Vendor B's systems management product, and Vendor C's data classification service to make it all happen. It may also be able to provide a service interface for managing lifecycle policies to Vendor D's enterprise content management product..

4.3 ILM-Aware Storage

Storage and data management products emerge that incorporate services defined in the first two phases. This begins the transition from retrofit to integrated solutions.

In addition to continued evolution in the areas above:

- Data may be classified by external or internal classification engines when it is created, and is always placed on the correct storage configuration
- Data protection and data security services are modeled and can be discovered by data management applications. These services are managed with respect to their participation in storage configurations.
- Service delivery for data service classifications are monitored and compared to predicted levels of data service.
- Notifications for data management application failures and service level threshold failures
- Management of storage resources begins to leverage, accommodate, and utilize configuration templates for coordinated storage and data management resources
- The data classification services are used by information management workflow products to push some lifecycle automation "down the stack".
- IT administrators are able to manage storage configurations as a group rather than just one configuration instance at a time.

The focus in this phase is in building up the infrastructure for coordinated management of resources. File systems are able to leverage external policy engines to classify data as it is created. The management of the storage infrastructure provides for integrated policy identification and monitoring for data protection and data security services. I.e., a backup product advertises not only its capabilities but also the protection policies associated with data it is protecting. And most importantly, round-trip management of the infrastructure across heterogeneous products is possible – from provisioning, to monitoring, to automated and manual corrective reaction.

4.4 ILM-Aware Applications and Storage

The final stage incorporates ILM awareness into applications by specification of requirements and classifications for data directly from the application.

- Applications may specify the requirements or the classification of data when it is created. If necessary, the data management application assigns the data to a new storage configuration instance that is provisioned on the fly along with the appropriate data protection and security services.
- Grid management applications replicate data to newly provisioned storage configurations to expand the compute power of a grid-enabled application.

These data management standards evolve to support the visions of utility- and grid-style computing advocated by leading industry experts today. This extends the concept of round-trip management into the application domain so that business applications interact with data management services to extend automation in the data center to its desired level of automation and utilization. Ideally, we eventually reach the level of automation so that storage devices and services do not need to be explicitly configured through management interfaces; but rather, they are automatically configured to meet defined data requirements.

Appendix A. Terms and Acronyms

The SNIA ILM TWG is producing definitions and proposed standards for the management of storage and data management resources. This appendix captures some of the key definitions that are part of this work in progress.

The following terms and acronyms are taken from the SNIA dictionary and/or the work of the SNIA ILM TWG. For more information, see http://www.snia.org/education/dictionary.

Composite Storage Set	An aggregation of storage and services representing capabilities of different configurations to produce equivalent function. The aggregation includes storage, data protection, and data security.
data	The digital representation of anything in any form.
Data classification	An organization of data into groups for management purposes. A purpose of a classification scheme is to associate service level objectives with groups of data based on their value to the business.
Data Lifecycle Management – DLM	The policies, processes, practices, services and tools used to align the business value of data with the most appropriate and cost-effective storage infrastructure from the time data is created through its final disposition. Data is aligned with business requirements through management policies and service levels associated with performance, availability, recoverability, cost, etc. DLM is a subset of ILM.
Data Services	The control of data from the time it is created until it no longer exists. Data Services are not in the data path; rather, they provide control of, or utilize, data in the delivery of their services. This includes services such as data movement, data redundancy, and data deletion.
information	Information is data that is interpreted within a context such as an application or a process.
Information Lifecycle Management – ILM	The policies, processes, practices, services and tools used to align the business value of information with the most appropriate and cost-effective infrastructure from the time information is created through its final disposition. Information is aligned with business requirements through management policies and service levels associated with applications, metadata and data.
information management services	The processes associated with managing information as it progresses through various lifecycle states associated with a Business Process. These services exploit information about data content and relationships in making decisions, Examples include records management and content management applications.
metadata (from ISO 14721)	Data about other data.
Offered Data Service Level	The abstraction/translation of Composite Storage Set Settings into a data service level expressed in terms of Data Service Level Objectives.
Service Level Objective – SLO	Partitions an SLA into individual metrics and operational information to enforce and/or monitor the SLA. "Service Level Objectives" may be defined as part of an SLA, an SLS, or in a separate document. It is a set of parameters and their values. The actions of enforcing and reporting monitored compliance can be implemented as one or more policies.

Appendix B. Data Service Level Objectives (SLOs)

There are two types of data SLOs described:

- Service level metrics define requirements that affect the provisioning of services initially, and then require ongoing measurement and validation as the characteristics of the delivered service may change over time based on external conditions. Examples include performance-related attributes such as throughput and data recovery times. In general, these SLOs may be used as Key Performance Indicators in an SLA, but are not always well suited to the task since they tend to be data-specific as opposed to application-specific.

- Usage properties define requirements that affect provisioning of services initially, but do not require ongoing measurement and validation. Examples include Location and Initial Size for the data.

An SLO Group, or SLOG, is a collection of SLO attributes that specify the desired service characteristics for data. The SLO attributes in Table 1 below are organized into categories such as “Accessibility” and “Availability” for convenience of discussion only. SLOGs must be evaluated as a whole in order to map to an Offered Data Service Level. Likewise, the capabilities of an ODSL must be considered as a whole to map to one or more Composite Storage Services (CSS) to provide services to satisfy the requirements.

When specified by a client, a SLOG may contain one or more of the individual SLO attributes. Some SLO attributes are expressed as a min/max range, an enumeration of fixed values, or a goal, as indicated. Any non-specified SLO attribute in a SLOG is treated as a “Don’t Care” condition.

Attributes in italics are proposed only – pending agreement within ILMTWG.

Table 1: Data SLO Attributes

Group				
	SLO Attribute	**Data Type**	**Value / Quantifier**	**Notes**
Budget				
	MaxBudget	Integer	Cost/GB/month	When specified as input, it is set by the user of the Data Management Layer; when advertised as a Data Management Layer capability, then it is “Cost” and is set by the IT Architect.
	CurrencyUnits	String	User-specified	Description of cost units described by MaxBudget. E.g., “dollars” or “pounds”.
Accessibility				
	ReadWriteRatio	Enum	HighReadRatio, HighWriteRatio, BalancedRWRatio	Used to distinguish between different storage configuration optimizations
	AvgIORate	Enum	HighIO, MediumIO, LowIO	Describes the IO rate requirements for the data
	AvgDataThroughput	Enum	HighThroughput, MediumThroughput, LowThroughput	Describes the data throughput/bandwidth requirements for the data.
	InitialAccessTime	Enum	Immediate, Sub-second, Seconds, Minutes, Hours	To distinguish media access MTFB characteristics. How long is application willing to wait?

Group				
	SLO Attribute	Data Type	Value / Quantifier	Notes
	MaxSize	Integer	Bytes	
	AccessPattern	Enum	Random, Sequential	
	SpaceUsage	Enum	Static, dynamic, sparse, fluctuate	How would this influence allocation
	InitialSize	uint64	Bytes	Provides hint as to initial allocation size for data
	AllocationGuarantee	Boolean	True = must allocate	Must never return out of space errors
	GrowthPeriod	Enum	Seconds, minutes, hours, days, weeks, months, quarters, year	
	SizeGrowth	Integer	Bytes	Prediction for allocation growth rate.
	DataSharing	*Enum*	*SingleHost, CrossPlatform, HardCluster, LooseCluster*	*Requires more definition – probably jointly with the FSM TWG*
	Affinity	*Enum*	*Co-location, separation*	*TBD*
	Location – See SMI-S CIM_Location	String	User-specified	For location-matching if necessary, as in SMI-S
	UserSpecified	String	User-specified	Arbitrary use by data center
Availability				
	AvailabilityPeriod	Enum	Day, Week, Month, Quarter, Year	Availability measurement period for planned and unplanned downtime
	PlannedDowntime	Integer	Seconds	Seconds of acceptable planned downtime for the AvailabilityPeriod. Timing for such downtime would be part of a separate SLA
	MaxUnplannedDowntimeAggregate	Integer	Seconds	Seconds of unplanned downtime for the AvailabilityPeriod
	MaxUnplannedDowntimePerInstance	Integer	Seconds	Seconds of unplanned downtime for any one occurrence of downtime
Data Restore				
	RPO	Integer	Seconds	Recovery Point Objective – how long data may be at risk of loss.
	RTO	Integer	Seconds	Recovery Time Objective – time required to restore data to promised state.

	Consistency	Boolean	True = consistent required (default)
			False = Fuzzy acceptable
				Identify what level of recovery requirements will be imposed on the the restore service provider.

Security				
	TBD			*Working with SNIA Security TWG*

Appendix C. Composite Storage Set Capabilities

A Composite Storage Set (CSS) is a collection of data management and storage capabilities that are known to work together at a predictable level of service.

A CSS is the normalization of the myriad of configurations and capabilities that can be used to deliver equivalent levels of service to an application. At the same time, it provides for customization of these capabilities to account for industry-specific, site-specific, or even administrator-specific preferences.

A CSS is created based on expert knowledge of the relationships and interoperability of specific data management technologies such as data replication and backup, with specific storage management technologies such as storage pools and storage replication, and with the characterisitics of the storage and storage fabric devices themselves. The expert knowledge may be provided by data center administrators, or by commercially-available software modules that guide the same user through the creation of a CSS based on the scope of the vendor's expertise.

The following sections describe capabilities and settings of a CSS. These are organized as:

CSS Storage Capabilities and Settings
CSS Data Protection Capabilities and Settings
CSS Security Capabilities and Settings

A CSS has a set of of capabilities to describe the overall scope of its possible capabilities and one set of Settings that define instance-specific configurations. The definition of a CSS's capabilities may include any one or more of these, with no more than one of each. A specific CSS configuration, or setting, may include any one or more of these, with no more than one Storage, Security, and Placement Settings. There may be multiple Data Protection Settings specified if multiple levels of data restoration are provided by the CSS[9]. The capabilities of a CSS are defined by the combination of its individually-defined member capabilities. The settings of a CSS are defined by the combination of its individually-defined member settings. The same properties are used for capabilities and settings and are defined below.

CSS Storage Capability and Setting Properties

CSS Storage Capabilities defines the intended behavior characteristics (e.g., performance, reliability, etc) that each of its storage resources is capable of satisfying.

The description of the storage capabilities will vary, depending on how the storage resources are configured. E.g., the same pool of storage extents could be used for a RAID 1 configuration which will produce a different set of performance characteristics than a RAID 5 configuration of the same storage extents. Since the Storage Administrator is responsible for configuring the CSS Storage Capabilities, it is his/her option as to whether to configure a very wide band of storage options from a single CSS, or a very narrow band of options. The variety of performance characteristics is specified as an array of values. A wide band of configuration options for a single CSS that has a large number of storage pools and extents associated with it will likely have a very large number of elements defined in this array. A narrow band of configuration options for a CSS that has only one storage pool pre-configured for RAID 5 would likely have fewer elements.

Each attribute of the CSS Storage Capabilities and CSS Storage Settings is specified as an n-tuple in which each 0th element is aligned, each 3rd element is aligned, etc. Note that storage in this context applies to disk block or file-level storage. The properties that describe the behavior of storage configurations for use in a CSS are:

- **AvgDataThroughput**: In an ideal world, the IT Architect will configure and test the performance of each storage configuration. This attribute captures the average number of I/Os that can be sustained as it relates to the particular protocol being used in this configuration. This attribute is measured in IOPS.

[9] E.g.: Data restoration from a rotation of hourly snapshots plus daily backup to disk for Operational Recovery, plus a copy of that daily backup copied to tape and sent offsite for Disaster Recovery offers three levels of data restoration for a single CSS.

- **AvgBandwidth**: This attribute captures the average amount of data that can be transferred to/from this storage per unit time. This refers to the speed and number of I/O paths from routers to storage arrays, for example; or the speed of the backbone Ethernet network for file servers. This attribute is measured in Mbytes/second.

- **InitialAccessTime**: This is similar to the enumerated attribute specified as a Data SLO. The approximation provided by the enumerations is better suited for matching capabilities to requirements as it applies the correct order-of-magnitude assessment to the characteristic. Enumerations include:

  - Immediate: for access to storage with no known latencies in the IO path.
  - Sub-second: may be used by IT Architect to distinguish from immediate for storage that has significant IO path latencies such as remote access over high speed networks.
  - Seconds: for access to data on storage media that may need a few seconds to respond to the initial request (example: MAID).
  - Minutes: for access to data on storage that requires time to access, such as tape or DVD media in a library unit.
  - Hours: for access to data on storage that may be offline, such as tape that is offsite or on a shelf.

- **Protocol**: The concept of IOPS and bandwidth applies to many different storage protocol stack configurations. The IT Architect specifies protocol used in the measurement of the above characteristics. From a usability perspective, it may come from a “recommended” list provided by the data management product, such as “SCSI, NFS, CIFS, XAM, …”. This attribute allows the use of these capabilities to describe multiple storage formats such as disk, tape, file systems, and fixed content.

- **Workload**: The measurement of these capabilities by the IT Architect will vary based on the applied workload. This attribute is a string to characterize the workload applied during the measurement, if desired by the IT Architect. It allows for workload variations such as Write vs. Read intensive applications and block size variations.

- **MeasuredCapacity**: This is another workload “normalizer” which describes how the data was distributed across the configured storage for each measurement. It may be the same or different for each workload, and is intended to account for test differences such as number of spindles.

Since the specification of storage performance characteristics is fraught with interpretation problems if left to vendors, the most reliable means of performance characterization is to leave such definitions to the experts in the data center that actually use the storage devices. The advantages to this strategy are numerous, not the least of which is the fact that the performance characterization can now be made with configurations and workloads relevant to, and familiar to, each specific data center.

Note that all of these attributes are optional. An IT Architect may choose to create CSSes and define all, some, or none of these values. The power of these attributes in early products will be for embedded documentation of IT expectations. The power of these attributes in eventual products will enable automated mapping of ODSLs to CSSes. These properties are summarized in Table 2.

Table 2: CSS Storage Capability and Settings Properties

Name	Data Type	Value/Quantifier	Notes
AvgDataThroughput	Integer [ ]	I/O messages/second	IT Architect measured, protocol-specific, IOPS for each workload measured.

AvgBandwidth	Integer [ ]	MBytes/sec	IT Architect measured bandwidth, or capacity for data transfer, for each workload measured.
InitialAccessTime	Enum [ ]	Immediate, Sub-second, Seconds, Minutes, Hours	IT Architect measured "mean time to first byte" of data for any initial access to this storage class.
Protocol	String [ ]	User-specified	IT Architect specifies protocol used in measurement. This allows the use of these capabilities to describe multiple storage formats such as disk, tape, file systems, and fixed content.
Workload	String [ ]	User-specified	A Storage Class has unique performance characteristics for different workloads defined by the IT Architect. This accounts for workload variations such as Write vs. Read intensive and block size.
MeasuredCapacity	Integer [ ]	User-specified	This is a "normalizer" – it's the capacity used for each measurement. It may be the same or different for each workload, and is intended to account for test differences such as number of spindles.

CSS Security Capability and Setting Properties

Security within the context of ILM is multi-dimensional and involves the data itself, the storage infrastructure and resources, the physical environments, the characterization of the technology trustworthiness (e.g., security posture, certifications, etc.), and the ILM solution itself. Each of these is important, but in totality, the sheer volume of details, the interdependencies, and the complexities can be so overwhelming that security may not be addressed by the data center. Consequently, a simplified and abstract (or normalized) view of security is offered for ILM. The expectation is that consumers of these normalized security capabilities will have mappings between specific normalized values and the detailed security capabilities.

The normalized security capabilities along with values that each may take on are briefly described below. It is important to recognize that these normalized security capabilities are highly subjective and will vary in significant ways from one organization to another. However, these capabilities should be used consistently within a single organization or ILM solution; in other words, the mapping of the normalized capabilities to the detailed security capabilities need to be the same.

- **SecurityAccountability** – From a security perspective, accountability is frequently achieved through a combination of measures, including logging (the record), authentication (proof of identity), authorization (proactively restrict access and usage), and non-repudiation (indisputable proof that a particular entity performed a specific action).
    - 0 = *No Accountability* (default) – resource has no or negligible support to address any of the required accountability requirements
    - 1 = *Basic Accountability* – resource has the bare minimum support for some aspect of the accountability requirements (typically, audit logging support)
    - 2 = *Moderate Accountability* – resource has Basic Accountability plus support for an additional capability (e.g., audit logging + authentication)
    - 3 = *Accountability* – resource has some level of support for audit logging, authentication, and authorization
    - 4 = *Strong Accountability* – resource meets the minimum requirements for audit logging, authentication, and authorization; nonrepudiation is typically not available
    - 5 = *Full Accountability* – resource has full support for all the required and desired

authentication, authorization, audit logging, and nonrepudiation services

- **SecurityIntegrity** – The integrity security service includes the following methods: prevention of unauthorized modification of data (both stored and communicated), detection and notification of unauthorized modification of data, and recording of all changes to data. Modification of both stored and communicated data may include changes, insertions, deletions, or duplications. Additional potential modifications that may result when data is exposed to communications channels include sequence changes and replay.
  - 0 = *No Integrity* (default) – resource has no or negligible support to address any of the integrity requirements
  - 1 = *Integrity* – resource meets the integrity requirements with support for secure hashing
  - 2 = *Full Integrity* – resource has full support for all the required and desired secure hashing and nonrepudiation services
- **SecurityAuthenticity** – Authenticity guarantees that a record is not changed or manipulated after it has been created or received or migrated over the whole continuum of records creation, maintenance and preservation. In the context of records as legal evidence, authenticity is an absolute concept in that it either exists or does not. There is no relative degree of authenticity, while there may be for reliability. The status of being authentic, however, can change at any moment as a result of residual effects of an action or migration that has been performed on the record over time.
  - False = *No Authenticity* (default) – resource has no or negligible support to address any of the authenticity requirements
  - True = *Authenticity* – resource meets the authenticity requirements
- **SecurityTrustworthiness** – Trustworthiness can be considered to be the level of assurance one has that a system will perform as intended. The trustworthiness of a system is often demonstrated by external certification, accreditation, and/or as a general characterization of its security posture. In short, it is the level of trust a computing environment is willing to place on a system.
  - False = *No Trustworthiness* (default) – resource has no or negligible support to address any of the trustworthiness requirements
  - True = *Trustworthy* – resource meets the trustworthiness requirements
- **SecurityConfidentiality** – The confidentiality security service is defined as preventing unauthorized disclosure of data (both stored and communicated). One subset of confidentiality is “anonymity,” a service that prevents disclosure of information that leads to the identification of the end user. The provision of the confidentiality security service depends on a number of variables, including location(s) of the data that needs protection, type of data that needs protection, amounts or parts of user data that need protection, and value of data that needs protection.
  - 0 = *No Confidentiality* (default) – resource has no or negligible support to address any of the required accountability requirements
  - 1 = *Basic Confidentiality* – resource has the bare minimum support for some aspect of the confidentiality requirements (typically, data in-flight support)
  - 2 = *Moderate Confidentiality* – resource has Basic Confidentiality plus support for an additional capability (e.g., data at-rest + data separation)
  - 3 = *Confidentiality* – resource has some level of support for data in-flight and data at-rest confidentiality
  - 4 = *Strong Confidentiality* – resource meets the minimum requirements for data in-flight and data at-rest confidentiality; data separation may also be available
  - 5 = *Full Confidentiality* – resource has full support for all the required and desired data in-flight and data at-rest confidentiality as well as data separation services
- **SecurityImmutability** – This capability describes WORM functionality. However, compliance with SEC 17a-4 requires audit logging (from accountability), authenticity elements, and integrity elements.

  - False = *No Immutability* (default) – resource has no or negligible support to address any of the immutability requirements
  - True = *Immutability* – resource meets the immutability requirements
- **SecurityDestruction** – From a risk management perspective, the only acceptable method of discarding stored records is to destroy them by a method that ensures that the information is obliterated. Documenting the exact date that a record is destroyed is a prudent and recommended legal precaution.
  - False = *No Destruction* (default) – resource has no or negligible support to address any of the destruction requirements
  - True = *Destruction* – resource meets the destruction requirements
- **SecurityPhysical** – The physical facilities can play a role in the selection of data protection and data security measures. A “secure” facility may mitigate the need for encryption, where as an office setting may force its use. In addition, the availability of environmental protections can impact reliability and availability.
  - 0 = *No Physical Security* (default) – resource has no or negligible physical security; resource may also have significant threats from its neighbors, natural disasters, and/or man-made hazards
  - 1 = *Basic Physical Security* – resource has primitive support for site physical security or environmental controls (typically an office setting with environmental controls like battery backup)
  - 2 = *Moderate Physical Security* – resource has basic site physical security (controlled access) and environmental controls; threats from its neighbors, natural disasters, and/or man-made hazards have not be addressed
  - 3 = *Physical Security* – resource has some level of support for site physical security (datacenter) and environmental controls; threats from its neighbors, natural disasters, and/or man-made hazards many not be addressed
  - 4 = *Good Physical Security* – resource meets the minimum requirements for site physical security (controlled-access datacenter) and environmental controls (power, air, fire); threats from its neighbors, natural disasters, and/or man-made hazards partially addressed
  - 5 = *Full Physical Security* – resource has extensive site physical security protections (guarded/secured facility) and environmental controls as well as no or negligible threats from its neighbors, natural disasters, and/or man-made hazards

The following table summarizes the proposed normalized security capabilities for ILM.

Table 3: CSS Security Capability and Settings Properties

Group: Security				
	Capability	**Data Type**	**Value / Quantifier**	**Notes**
	SecurityAccountability	Int	0…5	0 = No Accountability 1 = Basic Accountability 2 = Moderate Accountability 3 = Accountability 4 = Strong Accountability 5 = Full Accountability
	SecurityIntegrity	Int	0…2	0 = No Integrity 1 = Integrity 2 = Full Integrity

	SecurityAuthenticity	Boolean	False, True	False = No Authenticity
				True = Authenticity

Group: Security				
	Capability	Data Type	Value / Quantifier	Notes

	SecurityTrustworthiness	Boolean	False, True	False = No Trustworthiness
				True = Trustworthy

	SecurityConfidentiality	Int	0…5	0 = No Confidentiality 1 = Basic Confidentiality 2 = Moderate Confidentiality 3 = Confidentiality 4 = Strong Confidentiality 5 = Full Confidentiality

	SecurityImmutability	Boolean	False, True	False = No Immutability
				True = Immutability

	SecurityDestruction	Boolean	False, True	False = No Destruction
				True = Destruction

	SecurityPhysical	Int	0…5	0 = No Physical Security 1 = Basic Physical Security 2 = Moderate Physical Security 3 = Physical Security 4 = Good Physical Security 5 = Full Physical Security

CSS Data Protection Capability and Settings Properties

The CSS may also include one or more data protection services, each characterized by a CSS Data Protection Capabilities and Settings. There are a number of products that offer very different technologies to address data protection requirements for Operational Recovery and/or Disaster Recovery (OR/DR). The IT Architect is responsible for combining storage configurations with desired and appropriate data protection services.

The method and means to achieve data protection for OR and DR can be a very site- and product-specific configuration, yet its tight coupling with different storage technologies makes it a necessary service to combine with storage as an integrated solution. Hence, the CSS, is an ideal mechanism by which the IT Architect may combine desired storage configurations with data restoration services that are known (and tested!) to work with those storage configurations.

The need to abstract the capabilities of the underlying services without modeling of some of the more sophisticated features of these products has lead to the definition of the following attributes:

- **RPO**: Recovery Point Objective: The amount of time that active data is left at risk. This translates into the amount of time between copies of changed data. While this may be specified as a specific value in the SLO, the technologies that provide OR and DR solutions will do so in a configurable range specific to that technology for copying changed data. Hence, an enumeration with values ranging from Continuous to Weekly is used to represent the best RPO solution that can be provided by this data restoration service.

- **DataRestoreTime**: A benchmark time expected for this service to restore data from a copy for use by an application, based on the size of data specified in MeasuredCapacity. The DataRestoreTime is combined with the time to detect and repair failed equipment, plus perform application recovery of the data (e.g., replay the redo logs) to satisfy the data's Recovery Time Objective (RTO).

- **MeasuredCapacity**: This attribute is an IT Architect-specified normalizer for varying RTO performance across different data sizes and different restore technologies.
- **Workload**: Each Data Protection Service has unique performance characteristics for different workloads defined by the IT Architect. This accounts for workload variations such as number of files, application type, etc. Data size is accounted for in MeasuredCapacity.
- **MaxDataRetentionCapability**: The maximum amount of time that the IT Architect expects to be feasible as a data retention time for the data copies. E.g., it may not desirable to save snapshot copies on disk for longer than a couple of days. Note, this is intended to reflect the intended capabilities of a managed data restoration service and not necessarily the media. Therefore, if the managed service provides for data refresh and migration to newer technologies forever, then it is possible to support infinitely long data retention. All 1's in this Unit32 value signifies infinitely long.
- **DataCopyLocales**:This is used by the IT Architect to distinguish OR and DR copies, and can also be used to distinguish among multiple DR sites.

The Data Restore attributes are summarized below in Table 4.

Table 4:CSS Data Protection Capability and Settings Properties

Name	Data Type	Value/Quantifier	Notes
RPO	Enum [ ]	Continuous, Minutes, Hours, Daily, Weekly	IT Architect-specified expectations for the frequency with which changed data can be copied.
DataRestoreTime	Uint32 [ ]	Seconds	IT Architect specified data restore time for specified MeasuredCapacity.
MeasuredCapacity	Uint32 [ ]	Bytes	This is a "normalizer" – it's the ITA-specified size of data as measured for RTO.
Workload	String [ ]	ITA-specified	IT Architect-specified workload type for a each RPO/RTO benchmark.
MaxDataRetentionCapability	Uint32 [ ]	Seconds	ITA-specified range for possible retention of restorable copies. "Seconds" covers up to 136 years. All ones signifies infinite.
DataCopyLocale	String [ ]	ITA-specified	The IT Architect defines the target locale of the data copies. (e.g., DR site)

About the Authors:

Dr. Jack Gelb, Senior Software Engineer, IBM

Dr. Gelb is a Senior Software Engineer with IBM's Systems Technology Group, working primarily on strategy, architecture, design, and market support for storage management. He joined IBM in 1970 and his work led, in part, to the realization of the IBM Data Facility Storage Management Subsystem (DFSMS) for MVS and VM, for which he has received several awards and international recognition. His doctorate is in Computer Science from Rensselaer Polytechnic Institute, Troy, NY, and he has authored many papers and presentations on system-managed storage and other topics. Dr. Gelb holds a patent for automated storage management, and is currently focusing on the application and extension of system-managed storage techniques to network storage. He is co-chair of the SNIA Policy-based Storage Management and Information Lifecycle Management technical work groups.

Edgar St.Pierre, Senior Staff Software Engineer, EMC^2

Edgar has over 28 years of experience in software engineering, including the last 10 years in storage software development and the preceding 18 years in the communications industry. At EMC, he has been responsible for requirements analysis and software architecture for several data protection products in EMC's information management product portfolio. He has had 5 storage-related patents issued to date while at EMC. Edgar received his BS in Computer Science from Roger Williams University. He is currently co-chair of the SNIA DMF’s ILM Initiative, and co-chair of the SNIA ILM Technical Workgroup.

Dr. Alan Yoder, Senior Member of Technical Staff, Network Appliance, Inc.

Dr. Yoder has been at Netapp in Sunnyvale, CA since earning his Ph.D. in distributed systems in 1997, working on protocols, management frameworks, management applications, management partnerships, the Manage ONTAP(tm) SDK, and other projects. He also has experience in construction and industrial accounting, CAD design and programming, GUI design and development and project management. He holds Bachelors, MSEE and Ph.D. degrees from Goshen College and the University of Notre Dame. Alan is a member of the SNIA Technical Council, co-chairs the SNIA Disk Resource Management TWG , chairs the Enterprise Grid Alliance Data Provisioning Working Group and participates actively in several other Working Groups in the SNIA and DMTF.

About the SNIA ILM Technical Workgroup

The ILM TWG will develop shared data management and associated storage management services across applications that span networked storage. In particular, the ILM TWG will define processes, control mechanisms, and artifacts to map data management requirements (service level objectives and policies for protection, availability, etc.) into data management services. While individual applications can and do provide some of these services, the ILM TWG will address the need for a common set of data and storage management services and their coordination across multiple applications that share and use networked storage.

About the SNIA Data Management Forum and ILM Initiative

The SNIA Data Management Forum (DMF) is a cooperative initiative of Information Technology Professionals, Vendors, Integrators, and Service Providers formed to define, implement, qualify, and teach improved and reliable methods for the protection, retention, and lifecycle management of electronic data and information. The DMF ILM Initiativeis dedicated to advancing the vision of ILM for tomorrow, and the best practices of ILM for today.

About the SNIA

The Storage Networking Industry Association (SNIA) is a not-for-profit global organization, made up of more than 460 member companies and close to 7,000 active individuals spanning virtually the entire storage industry. SNIA members share the common goal of advancing the adoption of storage networks as complete and trusted solutions. 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 additional information, visit the SNIA web site at http://www.snia.org.