Education

Information-centric Policy Management

Edgar StPierre, EMC

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced without modification
  - The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.

Abstract

Information-centric Policy Management

As enterprises deal with accelerating information growth rates and growing complexity in the data center, one strategy that has proven effective in addressing scalability and effectiveness is the use of IT Service Management – applied from any one of many different disciplines including ITIL, itSMF, CobiT, and others. This has been especially true as it relates to storage service management and its relation to tiers of storage.

This tutorial will look at how this same framework can be used effectively for *information management*-related policies. Starting from the distinctions between storage, data, and information management, attendees will leave this tutorial with an understanding of:

- How multiple perspectives of data classification from multiple information stakeholders across an enterprise can be achieved
- How the use of Service Level Management can bridge the gap between IT, the lines of business, records information managers, legal, and security interests in the enterprise
- How this relates to security, eDiscovery, records retention, information rights management, business continuity, information lifecycles, and archiving policies

About SNIA and the DMF

About the Storage Networking Industry Association (SNIA)

- SNIA's primary goal is to ensure that storage networks become complete and trusted solutions across the IT community
- For additional information about SNIA see www.snia.org
- SNIA's "Dictionary of Storage Networking Terminology" is online at www.snia.org/dictionary

About the Data Management Forum (DMF)

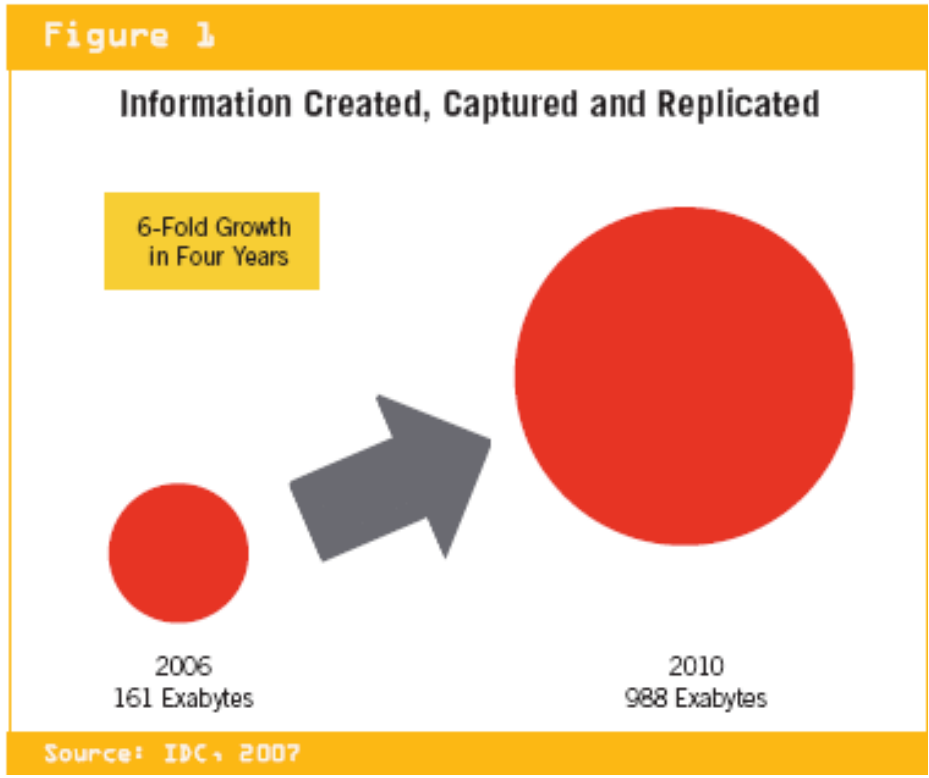
- Founded in 2004, the Data Management Forum is a sub-group of SNIA specializing in data management and protection throughout the lifecycle of information.
- More information about the DMF including resources on data and information lifecycle management can be found at www.snia-dmf.org

Agenda

1. Introduction
2. Classification
3. Service Level Management
4. Policy Management

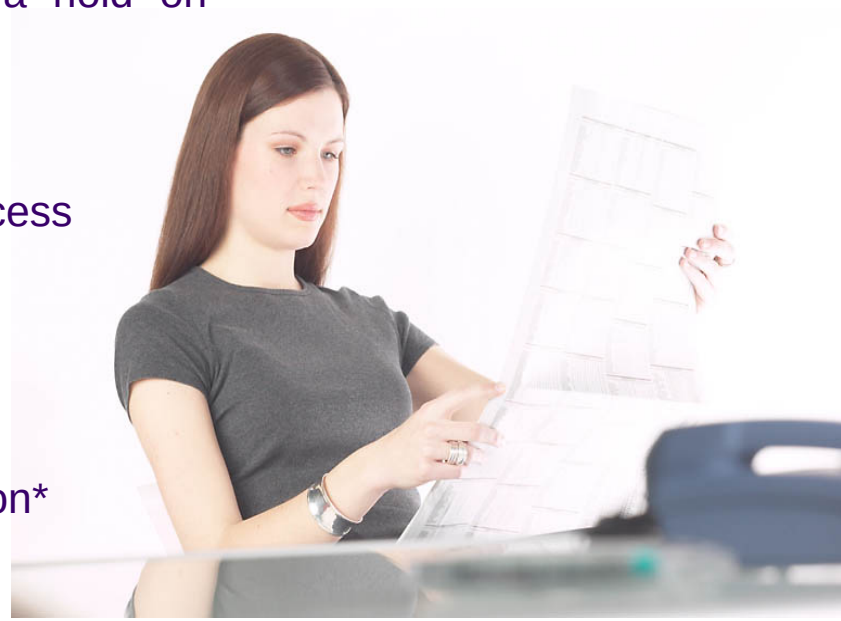
Information Growth

- Save everything!
  - 161 Exabytes in 2006
  - 988 Exabytes in 2010
    - *Source: IDC "The Expanding Digital Universe"*
- Growth for both size and units
- Major new drivers
  - Digital images, voice and TV
- Beyond "size":
  - 70% created by *individuals*
  - *Organizations* responsible for managing 85% of:
    - Security
    - Privacy
    - Reliability
    - Compliance
- *…and with tremendous scaling issues!*



Competing Requirements for Service Delivery

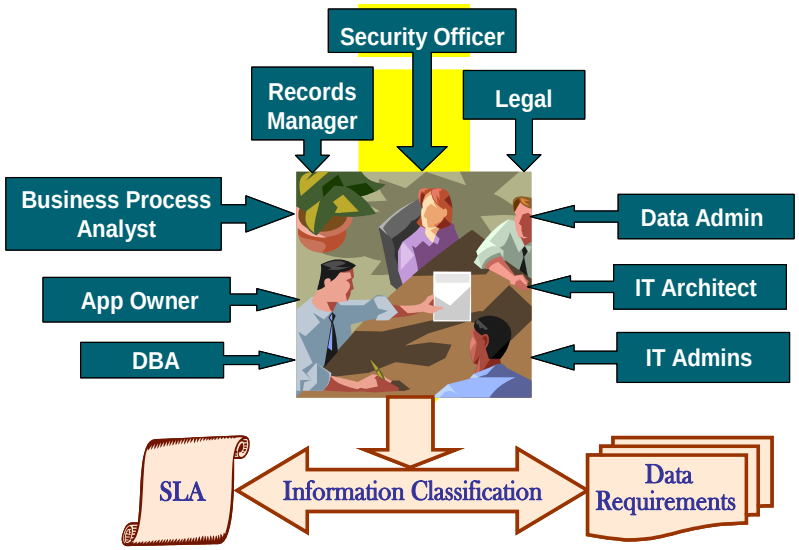
- Senior IT Management:
  - Optimize storage costs with a flat budget
- Security Officer:
  - Manage enterprise-wide information security and risk
- Legal Counsel:
  - Need to find the right information at the right time
  - Multiple simultaneous injunctions to place a "hold" on Electronically Stored Information (ESI)
- Compliance Officer & Records Information Manager (RIM):
  - Comply with regulations for retention & access
- Chief Risk Officer:
  - Addressing corporate risk management
- Business users:
  - The average knowledge worker spends *six hours per week* searching for information*



*Sources: IDC; Kahn/Blair

Gathering requirements from stakeholders

- Information is simply *data* to the data center
  - I.T. manages data: software, files, volumes, bits and bytes
  - Information is data with context: business decisions are based on *information*
  - *Use a collaborative process to identify information service requirements*



	Risk Management			TCO
Perspective / *Stakeholder*	Security & Risk Mitigation	Litigation Support	Records Mgmt	Cost & Performance Mgmt
Chief Security Officer	✓			
Chief Legal Officer	✓	✓	✓	
Corporation Counsel		✓	✓	
Records Information Manager		✓	✓	
Chief Compliance Officer	✓	✓	✓	
Chief Risk Officer	✓			
Chief Financial Officer	✓			✓
Line of business	✓	✓	✓	✓
Chief Information Officer	✓	✓	✓	✓

- Collaboration enables I.T. to:
  - Identify and mitigate competing stakeholder requirements
  - Create data management policies

- ➲ Classification must address multiple *perspectives*
- ➲ Support both overlapping and non-overlapping requirements

Information-oriented policies

For each information object:

1. "What is this information?"
2. "What are its requirements?"
3. "What services satisfy those requirements?"
4. Place the corresponding data accordingly
5. Notify appropriate Policy Administration Services

Information-oriented policies

For each information object:

1. "What is this information?"
2. "What are its requirements?"
3. "What services satisfy those requirements?"
4. Place the corresponding data accordingly
5. Notify appropriate Policy Administration Services

- Automated Data Classification
- Service Level Management
- Service Level Alignment
- HSM, Archiving, or integrated data mgmt
- Some homogeneous integrations (so far…)

Education

Classification

Defining “classification”

Aka: “categorization”

Online Dictionary (Answers.com):

1. A way or condition of being arranged
2. A subdivision of a larger group

Records Management perspective (Indiana University*):

1. Classification is the systematic identification and arrangement of records into categories according to logically structured conventions, methods, and procedural rules represented in a classification scheme. …

SNIA Dictionary perspective (data classification):

1. 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.

Tutorial use of “classification”:

1. An organization of data, information, or resources into groups for management purposes. A purpose of a classification scheme is to associate requirements or policies for the handling of that data or information

“Classification” must address multiple *perspectives*

* http://www.indiana.edu/~libarch/ER/mlaslides.ppt

Organizing your data into logical groups

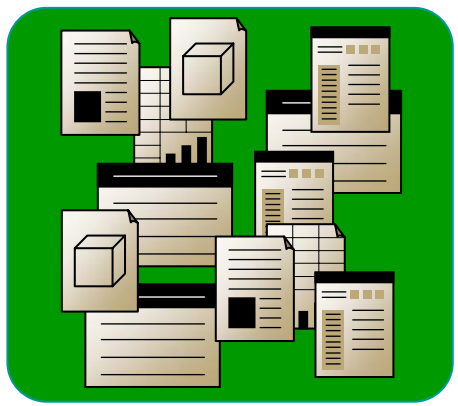
Define Information Requirements per Information Category

Data may be classified by:
- Application or business process
- Metadata (e.g., time last accessed)
- Content

Different performance requirements per classification category

- High Throughput
- Five 9's Availability
- Fast Recoverability

- High Throughput
- Four 9's Availability
- Fast Recoverability



- Medium Throughput
- Four 9's Availability
- 24 hour recovery

Organize same data into multiple logical groups

Multiple classifications allow you to define requirements per aspect

Classification perspectives:
- Application performance
- Records Retention
- Archive / Lifecycle Mgmt
- Security
- Privacy
- Information Rights Mgmt
- Legal Discovery
- and more

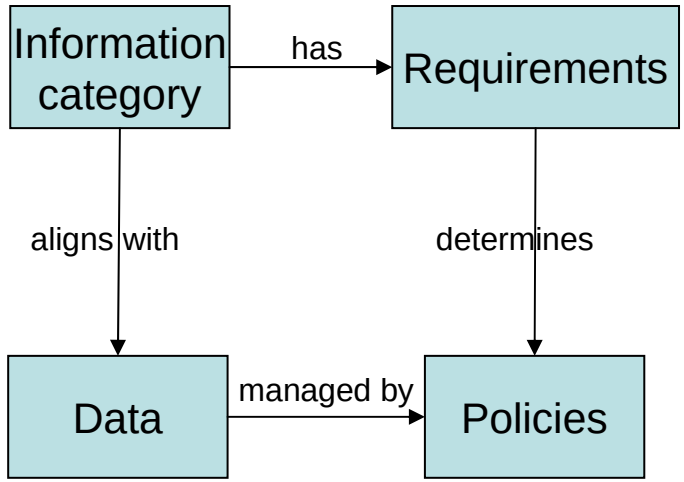
•Retain for 3 years

- Immutable
- Retain for 7 years
- DoD Shred

• Retain Forever

From Information Requirements to Data Classification

- Now that we've gathered Information Requirements – w*hat do we do with it?*
  - How do context-specific "Information categories" align with non-contextual data files or email?
  - Does this work with Service Level Management?



<u>Other issues to consider:</u>

- What data resources are applicable?
  - Production data
    - Start here – it's also the focus of this tutorial
  - Backup and archive data
    - These data copies have their own issues
    - E.g., how many recovery points are needed to meet legal discovery obligations?
- Where can automation be introduced?

How is Data Classified?

Automated Data Classification methods:

Classify by business process or application policies

- All data assigned same classification
- Simple; good start; a first approximation
- Net effect: ranking of applications to service tiers
- Somewhat effective for TCO Management
- *Possibly* effective for Risk Management (very coarse grained solution)

Classify by metadata-based policies

- Time last accessed, owner, file name, path, etc
- Useful for aligning data to tiers of service
  - E.g., the CEO's email receives different service than yours
- Or for placement of data to appropriate stage within a service tier
  - E.g., Hierarchical Storage Management (HSM) for a file server
- Effective for TCO Management
- *Possibly* effective for Risk Management (coarse grained solution)

Classify by content-based policies

- Content-driven alignment of data to service level requirements
- Added value for Business Intelligence, Compliance & eDiscovery
- Mostly not appicable for TCO Management
- Most effective for Risk Management

SNIA

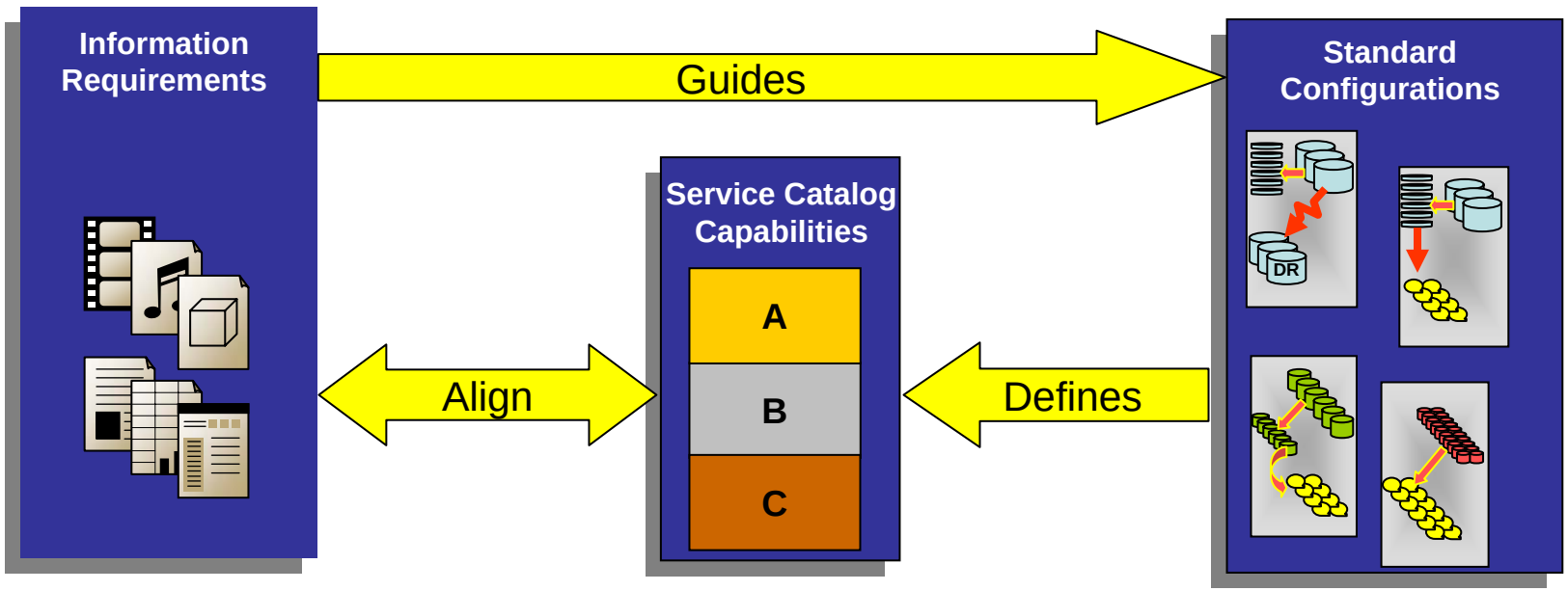
Education

Service Level Management

Service Level Alignment

To address TCO, Service Delivery, and/or Risk Management

- Information Classification to gather stakeholder requirements
- Create standard configurations
- Data classification to organize data based on information requirements
- Align data using Service Level Management



Service catalog capabilities may also include information policies

Service Catalog Client Sample

	Scheme	Specification	Tier 1	Tier 2	Tier 3	Tier 4
Storage	Guaranteed Performance	Performance throughput per port (IOPS)	5,000+	Up to 5,000	Up to 3,500	Up to 1,500
		Response time (ms)	< 8ms	7-14ms	12-30ms	12-30ms
	Availability	Maximum unplanned downtime per year (mins)	< 26.5	< 26.5	< 52.5	< 263
Archive	Performance	Response time	< 1 second	< 1 second	< 24 hours	
		Throughput	<= 300 Mbps	<= 700 Mbps	<= 280 Mbps	
	Availability	Maximum downtime (yr)	<5.25 mins	<52.56 mins	< 175.2 hours	
	Retention & Disposition	Retention period	< 30 years	< 10 years	< 3 years	
		Data shredding compliance	Yes	No	No	
	Data Integrity	Guarantee of authenticity	Yes	No	No	
	Accessibility	Read / annual access frequency	< Hourly	< Hourly	Daily	
	Offsite	Recovery point objective	< 1 minute	< 28 hours	< 38 hours	
Operational Recovery (OR)	Recovery Classification	Recovery classification	Complete app. restore	Complete app. restore	File or file sys. restore	File or file sys. restore
	Recovery Point Objective (RPO)	Amount of data loss	1 hour	24 hours	24 hours	30 days
	Recovery Time Objective (RTO)	Time required for recovery	< 30 minutes	< 30 minutes	7 GB/minute	.5 GB/minute
	Recoverability	Ability to recover backed up data	100%	100%	98%	95%
	Retention period	Time data is retained	2 hours	24 hours	3 Weeks	15 months
Disaster Recovery (DR)	Recovery Point Objective (RPO)	Amount of data loss	0 minutes	< 4 hours	24-48 hours	24-48 hours
	Recovery Time Objective (RTO)	Time to restore data	< 2 hours	<12 hours	< 48 hours	<72 hours

Service Catalog – Info-centric Policy Mgmt

Education SNIA

	Scheme	Attributes	Tier 1	Tier 2	Tier 3	Tier 4
Primary Storage	Performance	Throughput (IOPS)	Up to 5000	Up to 3,500	Up to 1,500	Up to 500
		Response time (ms)	< 8ms	7-14ms	12-30ms	12-30ms
	…	…	…	…	…	…
Indexed	Data search capability	Index method	Full Text	Metadata	None	
Archive	Performance	Response time	< 1 second	< 1 second	< 24 hours	
		Throughput	<= 300 Mbps	<= 700 Mbps	<= 280 Mbps	
	…	…	…	…	…	
	Retention	Retention period	< 30 years	< 10 years	< 3 years	
	Data Integrity & Authenticity	Guarantee of immutability	Yes	No	No	
User ID	Authentication	ID challenge method	Physical	Logical	Logical	None
	User Login	Scope of login	Unique	Unique	Federated	Federated
Audit	3rd party log gather	Audit action level	Intervene	Alert	Monitor	None
	Audit Log Storage	Guarantee of immutability	Yes	Yes	No	No
Data Security	Encryption	Data at rest	Tamper-Proof	Centralized	None	
		Secure data movement	Enterprise	Application	None	
		Data shredding at deletion	Destroy	DoD Shred	Simple Delete	
	Information Rights Mgmt	Customer Data Access	PII-Private	PCI-CC	Confidential	Basic
Operational Recovery (OR)	Recovery Point Objective (RPO)	Amount of data loss	1 hour	24 hours	24 hours	30 days
	Recovery Time Objective (RTO)	Time to restore data	< 30 minutes	< 30 minutes	7 GB/minute	.5 GB/minute
	Retention	Retention period	2 hours	24 hours	3 Weeks	15 months
Disaster Recovery (DR)	Recovery Point Objective (RPO)	Amount of data loss	0 minutes	< 4 hours	24-48 hours	24-48 hours
	Recovery Time Objective (RTO)	Time to restore data	< 2 hours	<12 hours	< 48 hours	<72 hours

Corresponding Solutions

Education SNIA

	Scheme	Attributes	Tier 1	Tier 2	Tier 3	Tier 4
Primary Storage	Performance	Throughput (IOPS)	Hi-end FC RAID-1	Hi-end FC RAID-5	Mid-tier FC RAID-5	Mid-tier ATA RAID-5
		Response time (ms)				
	…	…				
Indexed	Data search capability	Index method	Text Indexing policy			
Archive	Performance	Response time	CAS	ATA	Tape	
		Throughput				
	…	…				
	Retention	Retention period				
	Data Integrity & Authenticity	Guarantee of immutability				
User ID	Authentication	ID challenge method	HW Token, Local Login	SW Token, Local Login	SW Token, LDAP Login	LDAP Login
	User Login	Scope of login				
Audit	3rd party log gather	Audit action level	SIEM Intervention, on CAS	SIEM Alerts, on CAS	SIEM Monitoring, on disk	No SIEM Monitoring
	Audit Log Storage	Guarantee of immutability				
Data Security	Encryption	Data at rest	Secure/HW key vs. central key manager			
		Secure data movement	IPSEC, SSL, None			
		Data shredding at deletion	HW destroy, DoD “shred”, standard delete			
	Information Rights Mgmt	Customer Data Access	IRM Key Manager policies			
Operational Recovery (OR)	Recovery Point Objective (RPO)	Amount of data loss	Many Snaps	Fewer Snaps	B2D / VTL	B2T
	Recovery Time Objective (RTO)	Time to restore data				
	Retention	Retention period				
Disaster Recovery (DR)	Recovery Point Objective (RPO)	Amount of data loss	Sync Mirror	Async Mirror	Remote DiskCopy	B2T
	Recovery Time Objective (RTO)	Time to restore data				

Education

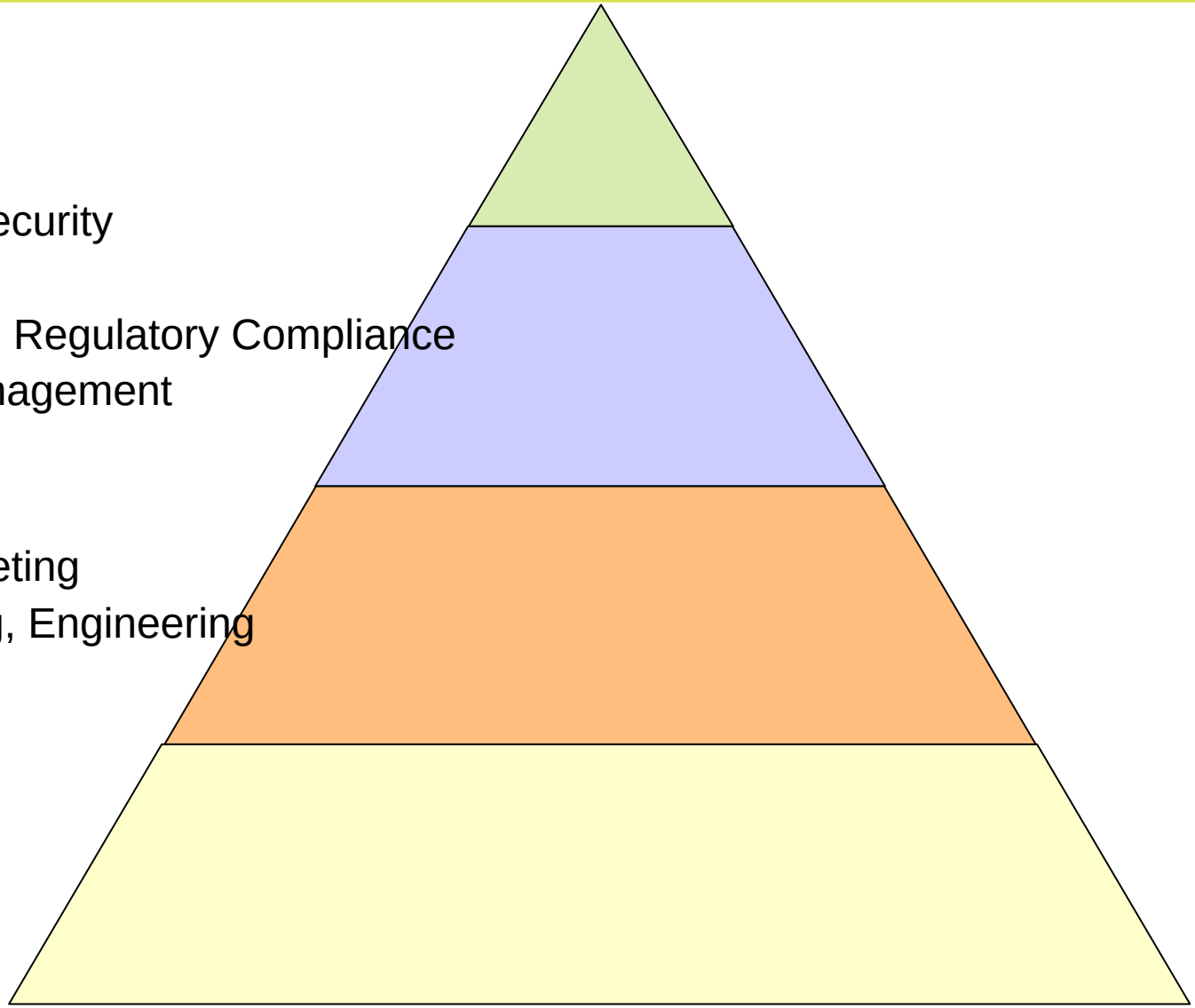
Policy Management

One Policy Management Framework…

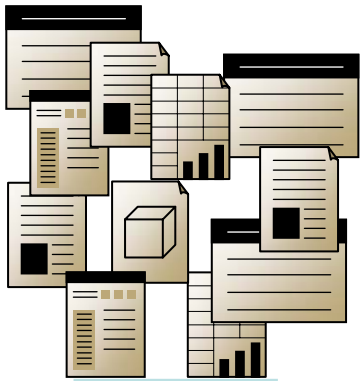
- Policy:
  - An official or prescribed plan or course of action
- Policy Administration
  - Define, manage, and assign policies
  - Using rules defined by conditions and actions
- Policy Decision
  - Evaluation of a policy rule's conditions
  - Leads to a policy decision and possibly an action
- Policy Enforcement
  - Execution of a policy decision

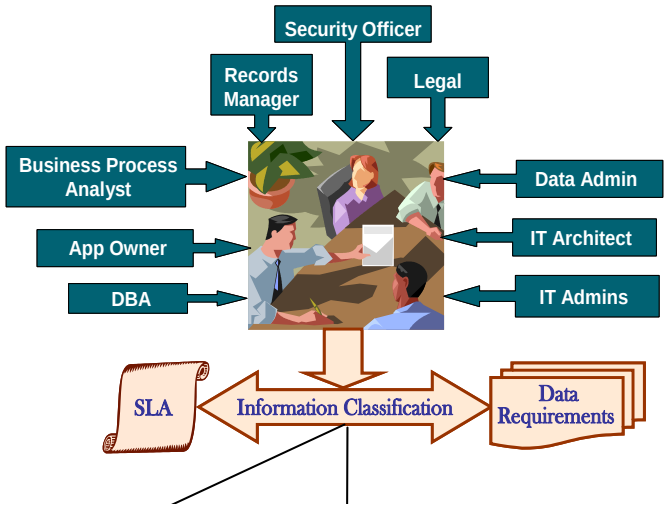
…for Many Policy Domains

- Corporate
  - Information Security
  - Legal
  - Governmental Regulatory Compliance
  - Email, IM Management
- Business Unit
  - Finance
  - Sales & Marketing
  - Manufacturing, Engineering
- User
  - Security
  - Distribution
  - Retention
- …



Assigning Policies via Automated Data Classification





2. Classify "Data"

1. Assign Data Policies to Service Levels

Automated Data Classification

3. Assign Service Level To Data

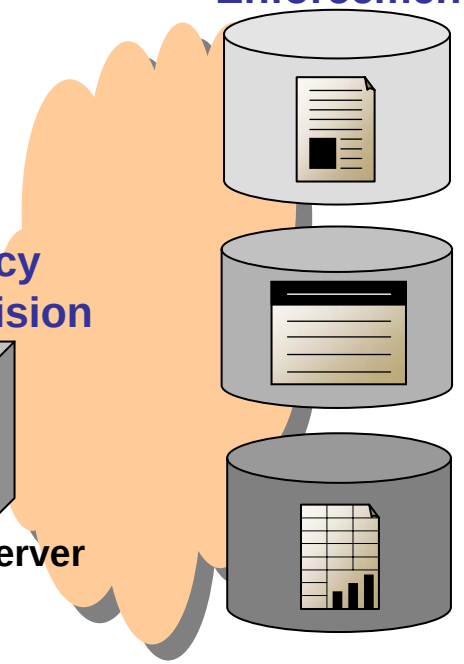
Policy Admin

4. Enforce Policies For Data

Policy Decision

Key Server

Policy Enforcement



Information-centric Policy Management

- Assign policies to data based on classification
  - Application, metadata, and/or content
- Policies derived from Information Requirements
  - Security
  - Information Rights Management
  - Data Leakage Protection
  - Data retention
  - Search & Index
  - Authentication & Authorization
  - And others…
- Automated data classification provides scalability

Automated Data Classification

Features of this emerging product area

- Automated data classification setup/execution
  - Create classification schemes
  - Based on metadata, content or both
- Policy engines
  - Classification
  - Service Level Management
  - Data lifecycle management
- Catalog for metadata & indexing
  - Independently or as extensions of metadata
- Leverage data placement capabilities
  - OS, HSM, etc.
  - Interfaces to archive devices to manage retention
- Interact with other policy administration services
  - Security, Information Rights Management
  - Records Management
  - Data Leakage Prevention

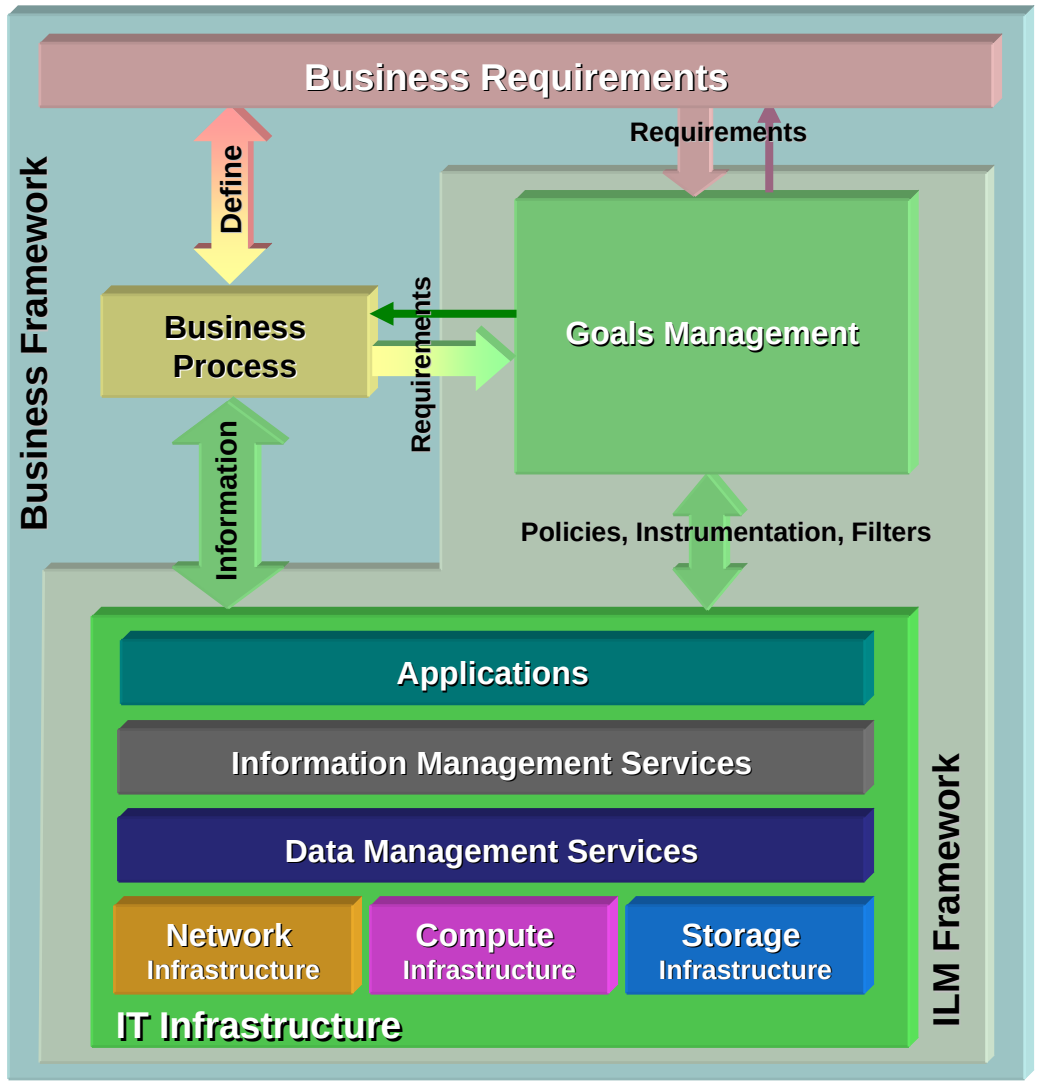
DMF's ILM Framework for the Datacenter

For more information on SNIA's Data Management Forum (DMF) visit the DMF website at

http://www.snia-dmf.org

Related ILM tutorials at SNW:

- Information Classification: The Cornerstone to Information Management
- The Secret Sauce of ILM: The Professional ILM Assessment



Q&A/Feedback

Please send any comments on this tutorial to SNIA: trackdatamgmt@snia.org

Many thanks to the following individuals for their contributions to this tutorial:

Edgar StPierre
Bob Rogers
Sheila Childs
John Field

Abbreviations used in this tutorial

- **Async**: Asynchronous
- **ATA**: Advanced Technology Attachment
- **B2D**: Backup To Disk
- **B2T**: Backup To Tape
- **CAS**: Content-Addressable Storage
- **DoD**: Department of Defense
- **FC**: Fibre Channel
- **IOPS**: Input/Ouput Operations Per Second
- **IPSEC**: Internet Protocol Security
- **IRM**: Information Rights Management
- **LDAP**: Lightweight Directory Access Protocol
- **HSM:** Hierarchical Storage Management
- **HW**: Hardware
- **ms**: milliseconds
- **SIEM**: Security Incident and Event Management
- **SSL**: Secure Sockets Layer
- **SW**: Software
- **Sync**: Synchronous
- **TCO**: Total Cost of Ownership
- **VTL**: Virtual Tape Library