

SNIA

STORAGE NETWORKING INDUSTRY ASSOCIATION

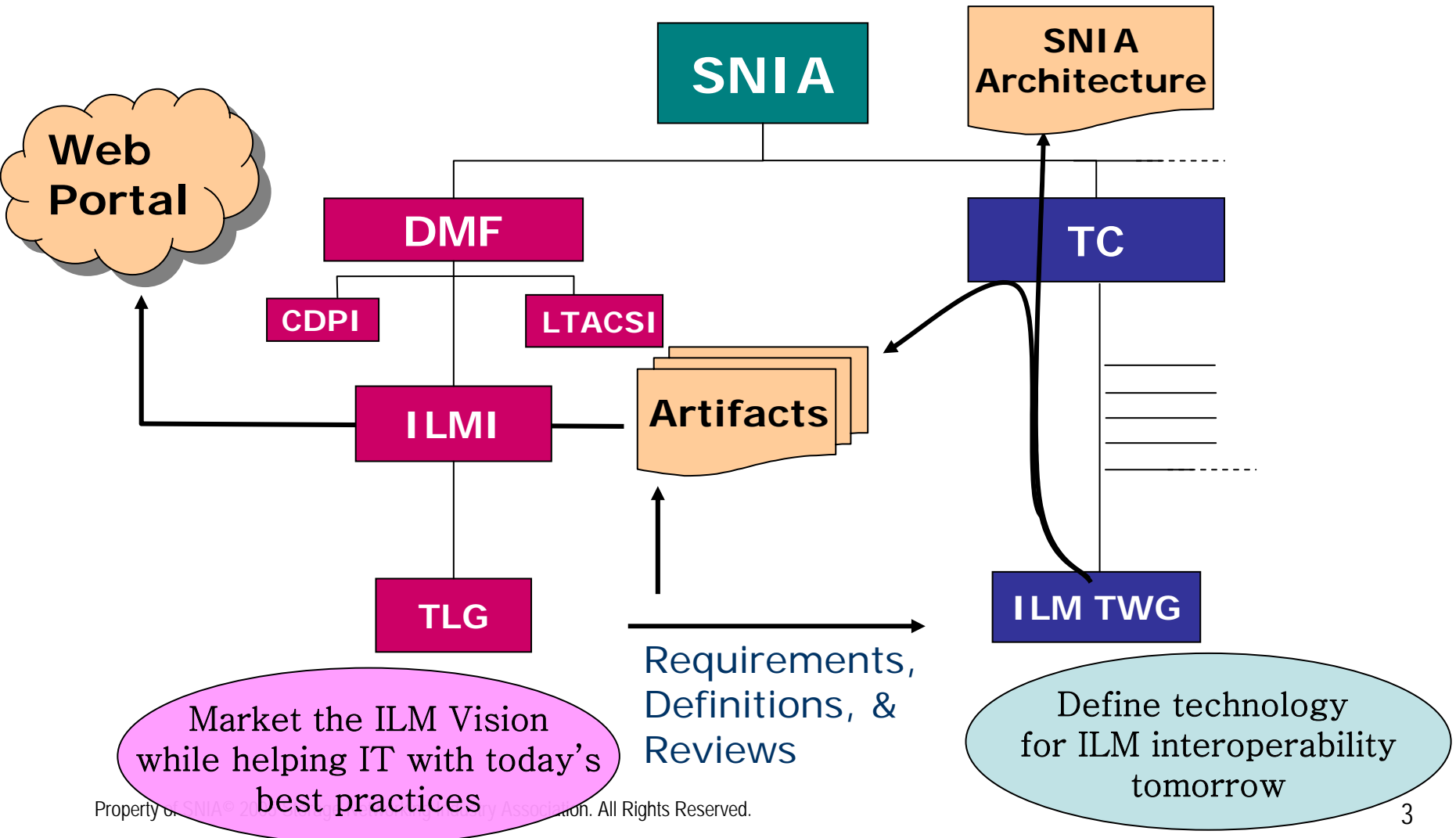
ILM Foundations:
Data Classification –
who, what, and why

Edgar St.Pierre, EMC²

Agenda

- Setting the table
- Information vs. Data
- Bringing knowledge domains together
- Business Process Analysis v.
- Information Classification v.
- Data Classification

SNIA ILM Groups



Assumptions and Value

- DMF assumptions about EUC participant understanding:
 - ILM value proposition
 - Optimize TCO of storage & data mgmt
 - Optimize service level delivery
 - Ensure regulatory compliance
 - Reduce complexity through automation
 - Data Classification is cornerstone for optimization and compliance
 - Automation requires standardization
- Value to EUC members:
 - Knowledge is power in discussions with vendors
 - Scope of issues, terminology, options
- Value to DMF members:
 - Prioritize requirements & activities
 - Reality check in data center processes
 - Provide focus for best practices development
 - Prioritize roll out of SMI-S definitions

Definition of ILM

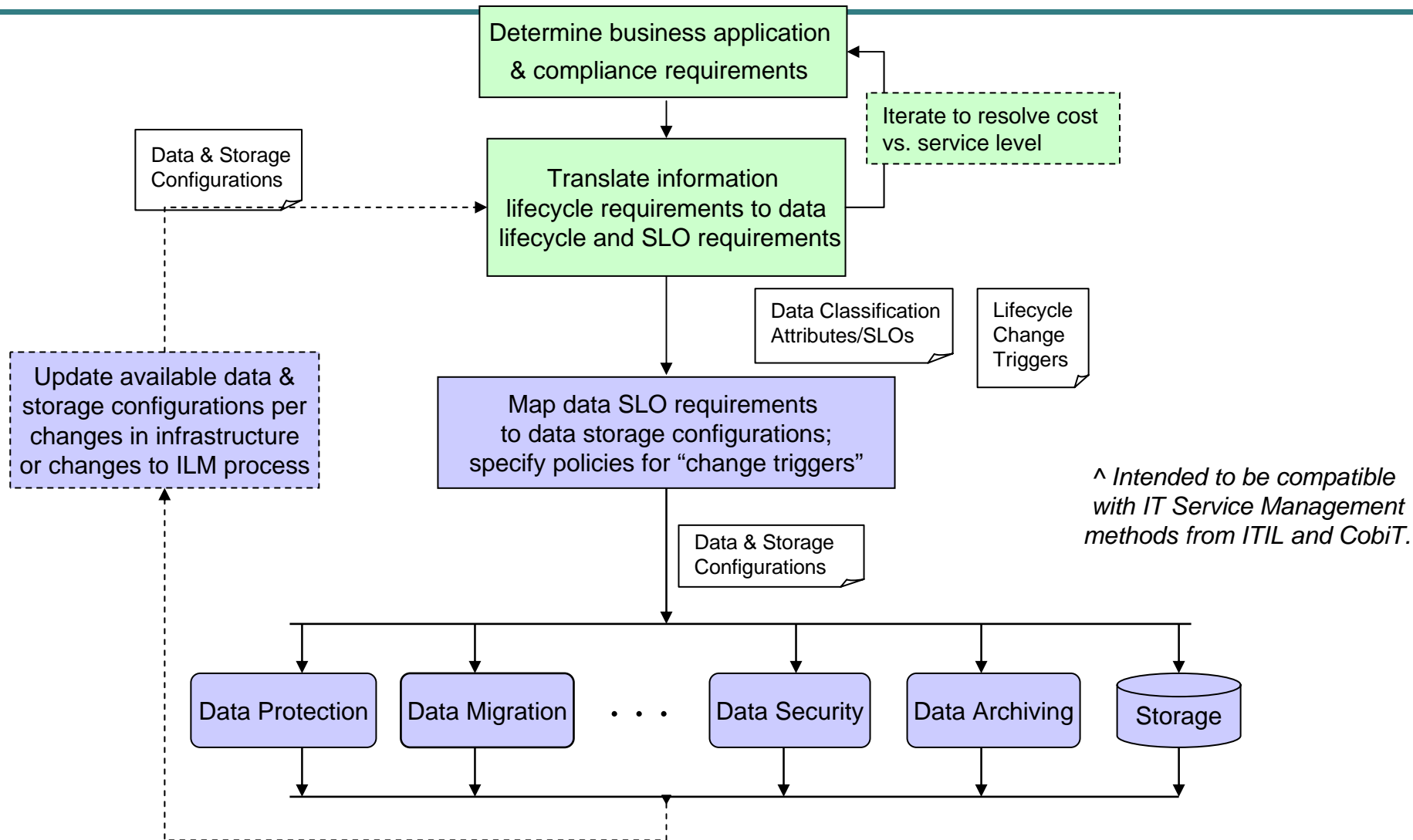
The policies, processes, practices, services and tools used to align the business value of information with the most appropriate and cost-effective infrastructure from the time information is created through its final disposition.

Information is aligned with business requirements through management policies and service levels associated with applications, metadata and data.

Information vs. Data

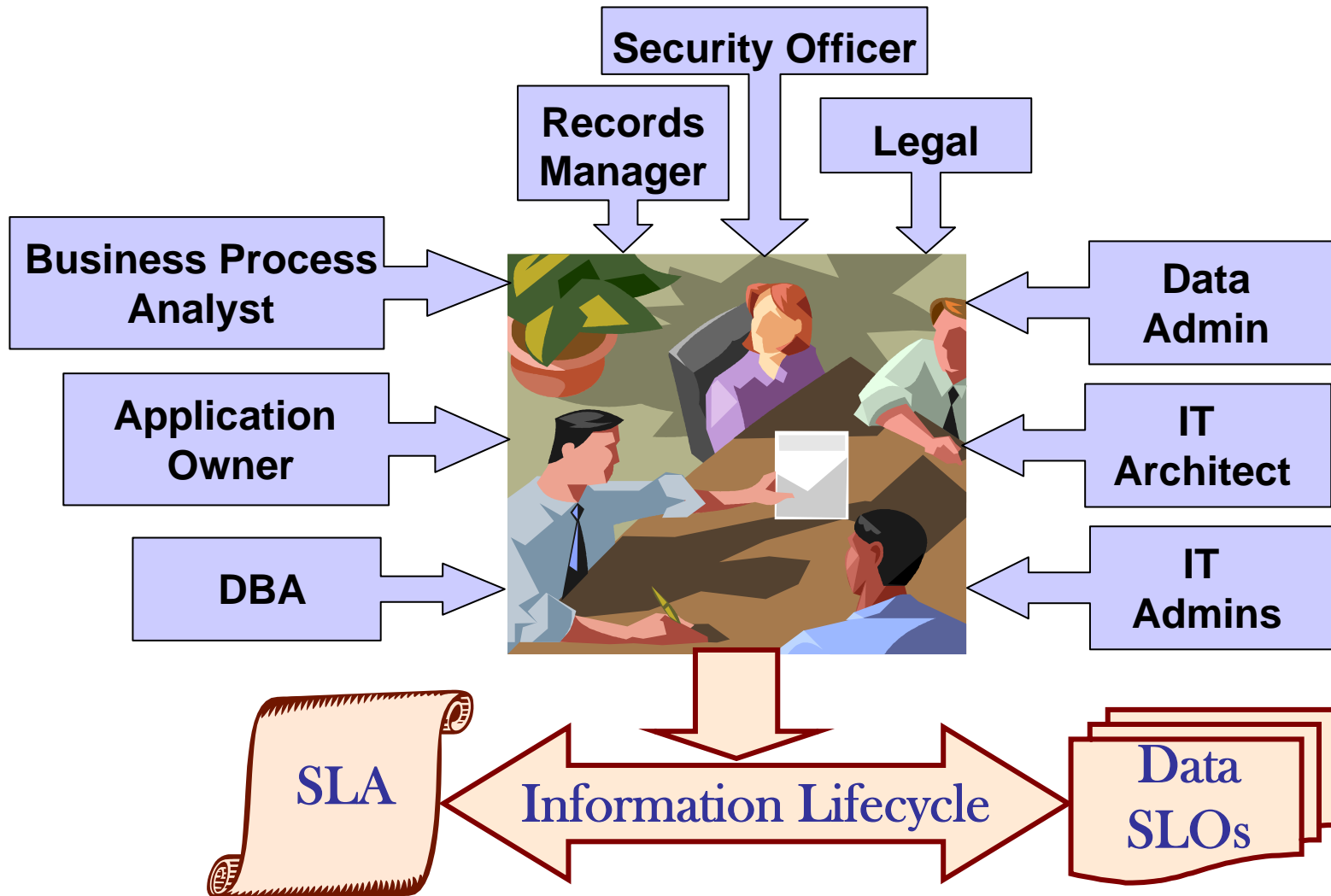
- ❖ Information is data with context
 - ❖ Information runs the business
 - ❖ In some cases, information *is* the business
 - ❖ Information is your competitive advantage
 - ❖ The business unit understands information
 - ❖ Based on business processes and requirements
 - ❖ Defines its value – or cost if it is unavailable
 - ❖ Defines its lifecycle
- ❖ Data is what IT manages
 - ❖ Files, volumes, bits and bytes
 - ❖ Databases may become middleware for applications
 - ❖ To some extent, applications simply group data for consistency
- ❖ Data lifecycle supports the information lifecycle

Classification Process Flowchart *



^ Intended to be compatible with IT Service Management methods from ITIL and CobiT.

Classification *requires* Stakeholder Communication

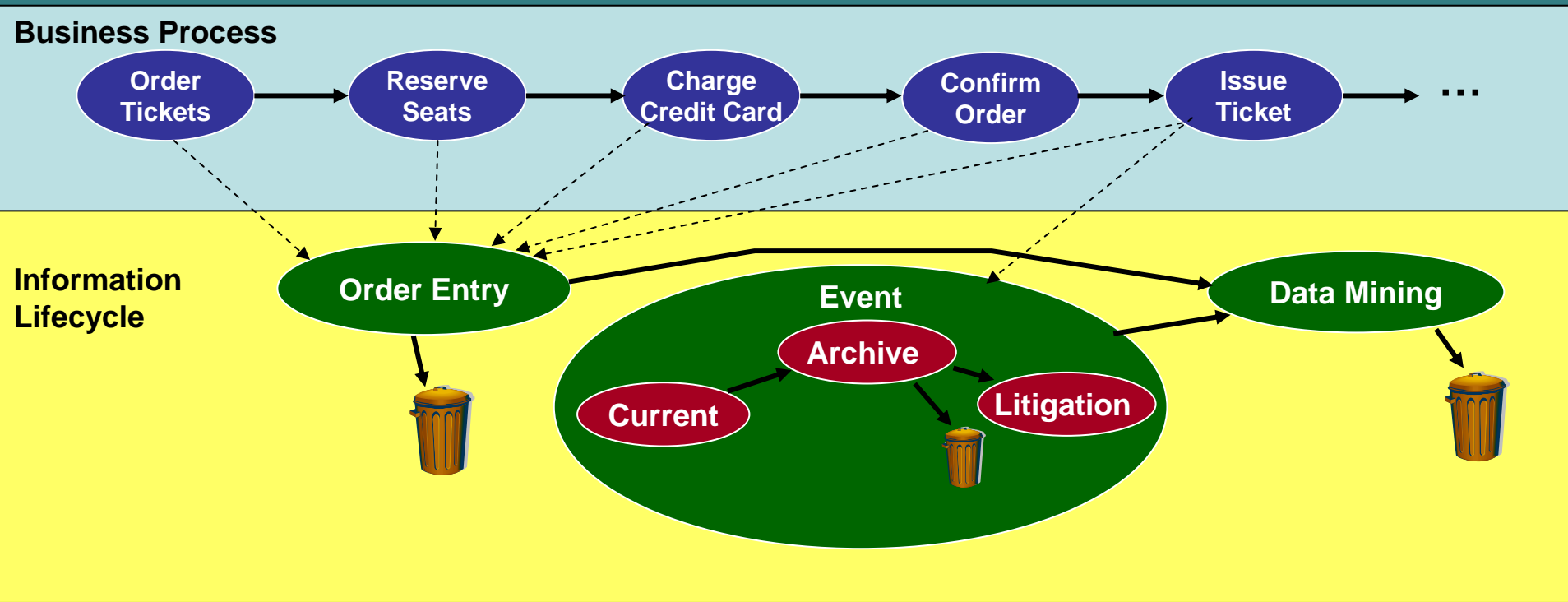


BP, Info, & Data Lifecycles

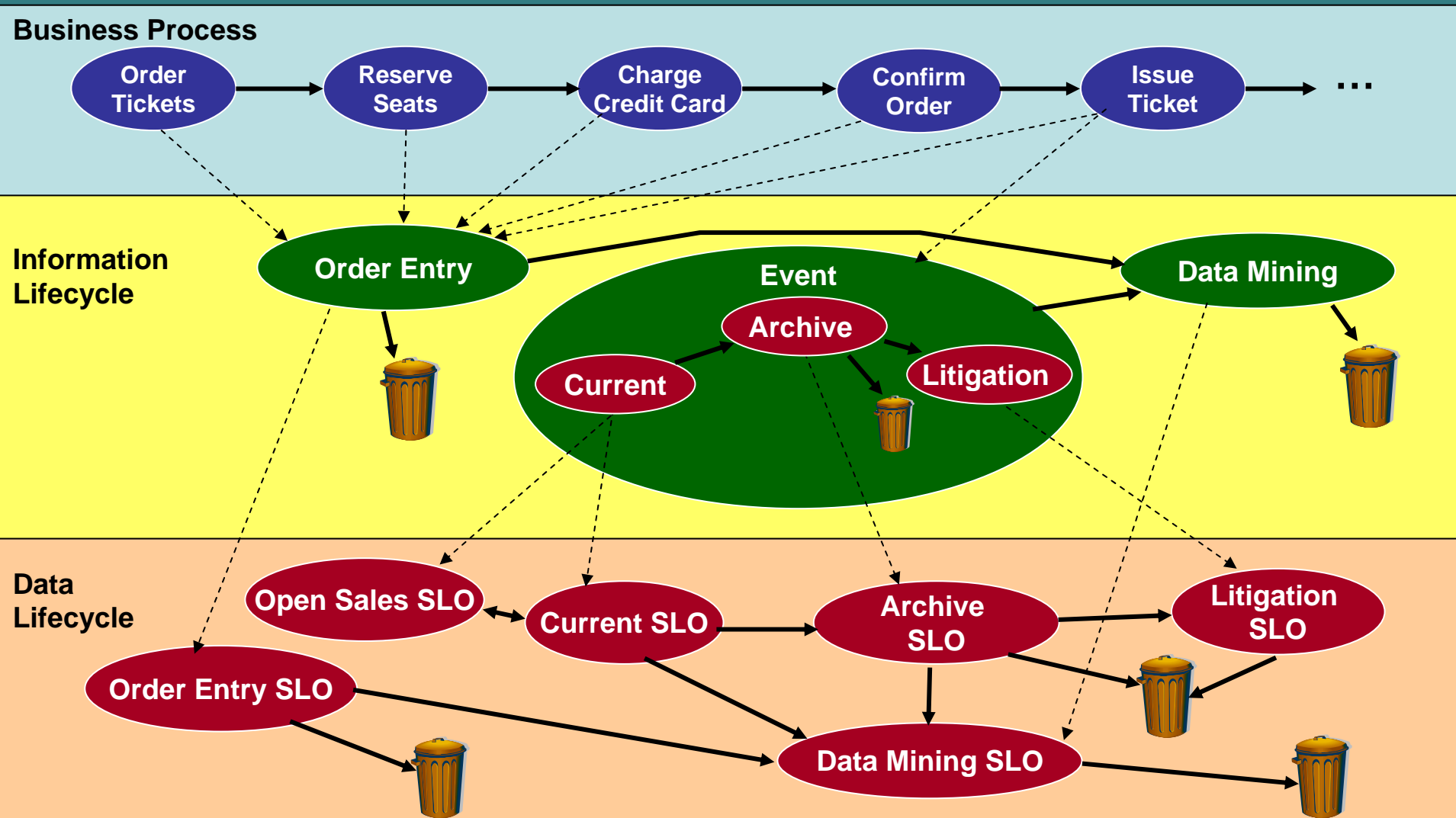
Business Process



BP, Info, & Data Lifecycles



BP, Info, & Data Lifecycles



Feedback:



- Is this, or something similar to this, a process used in your company today?
 - If not, why not?
- Is this a process that you think SNIA could help you with through tutorials and/or best practice white papers?
 - SNW presentations and tutorials
 - Enterprise Information World presentations and tutorials
 - Best Practice white papers on DMF web site
 - Which is most important to you?
 - What else could we do?

Information vs. Data Classification

Business Process

Information Classification

- Industry-specific and company-specific considerations may produce unique information and/or requirements.
- The mapping of business requirements to data SLOs *and* policies for applying those SLOs to data during the information's lifecycle.
- Best practices defined by SNIA DMF ILMI

Information Lifecycle

Data Lifecycle

Data Classification

- Common data center-wide definitions of expected behavior.
- The assignment of service levels to data based on its current state in its lifecycle.
- Standards defined by SNIA ILM TWG

Structured, semi-structured & application-specific data

Business
Process

Information
Lifecycle

Banking

Medical
Imaging

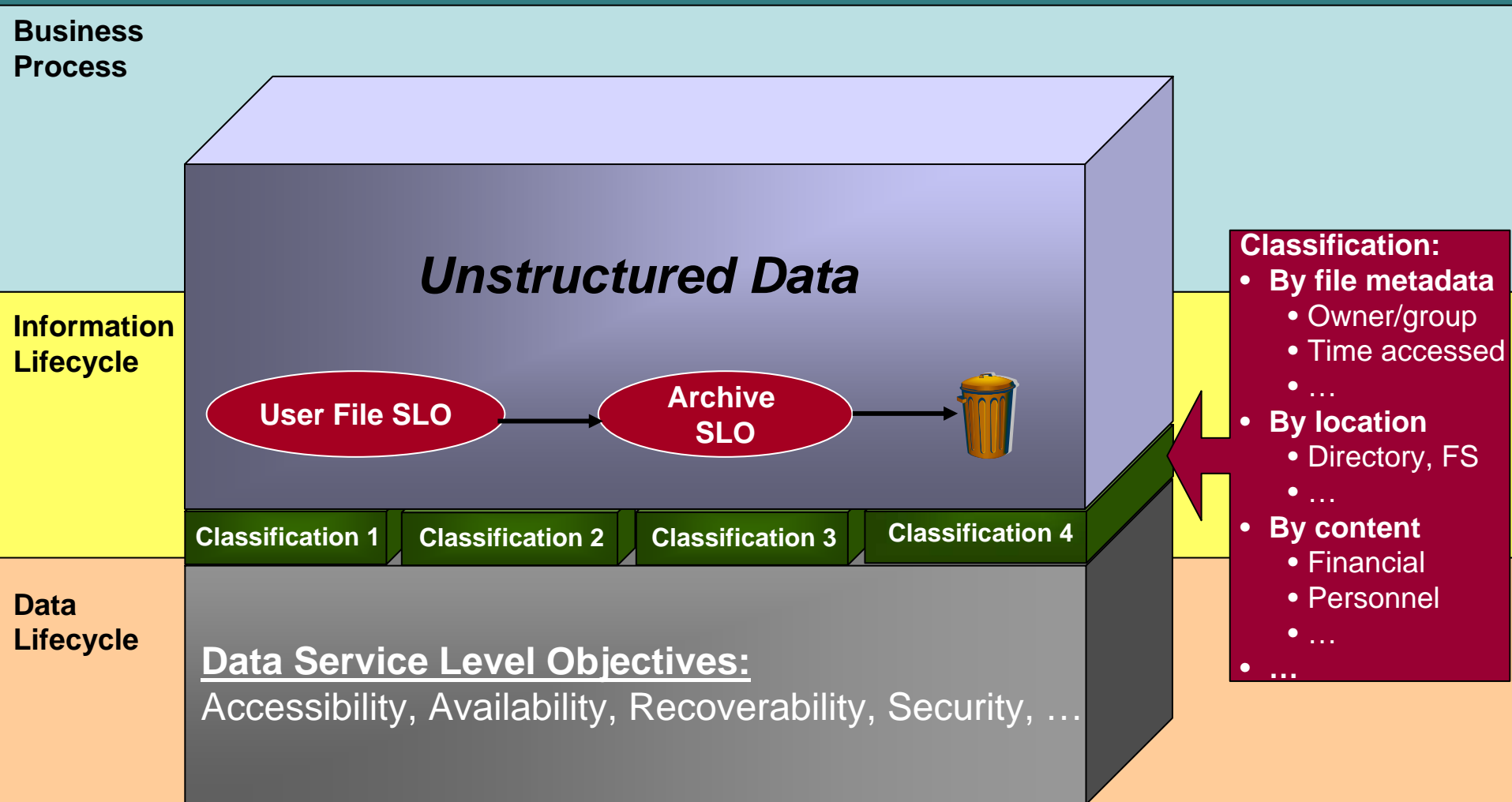
CRM

Email

Data
Lifecycle

Data Service Level Objectives:
Accessibility, Availability, Recoverability, Security, ...

Unstructured data



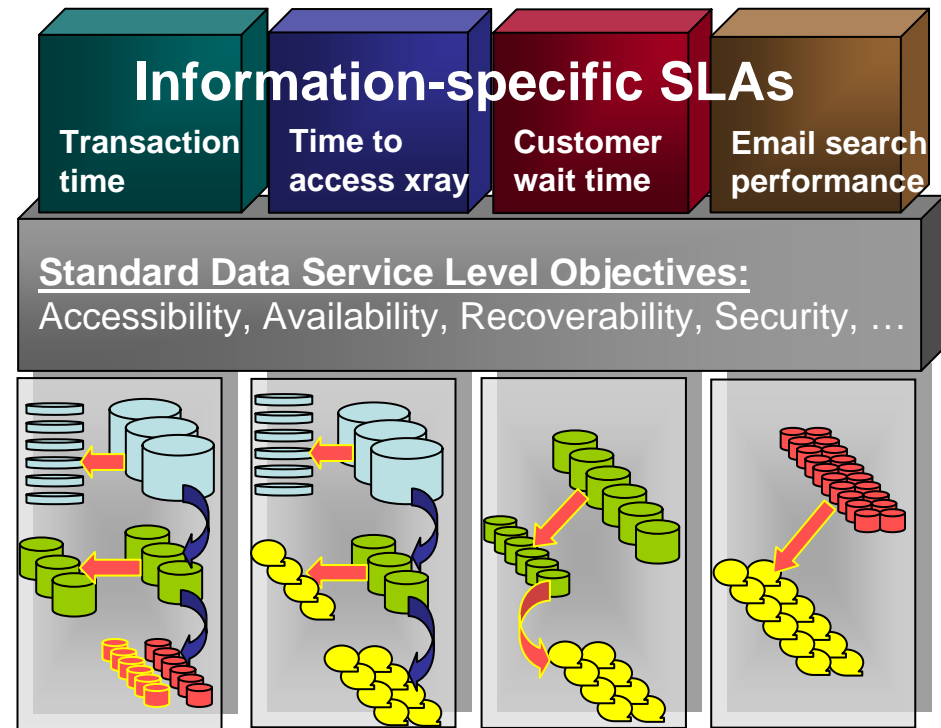
Feedback:



- What are you doing with respect to lifecycle management today with respect to:
 - Structured data?
 - Unstructured data?
- Are “best practice information lifecycles” something that you think SNIA could help you with through tutorials and/or white papers?
 - Is it something you would use today?
 - Which area should SNIA address?

Standard data SLOs: Defining data requirements

- **Different technology combinations of:**
 - Storage hardware
 - Storage software
 - Data Protection
 - Data Security
 - Etc
- **May produce equivalent service levels for:**
 - Performance
 - Availability
 - Data recoverability
 - Data integrity & security
 - Etc
- **Defined in terms similar to a Tower of Babel**



Benefits:

- Consistent data center metrics
- Consistent terminology across solutions
- Gap analysis
- *Enables next step for automation*

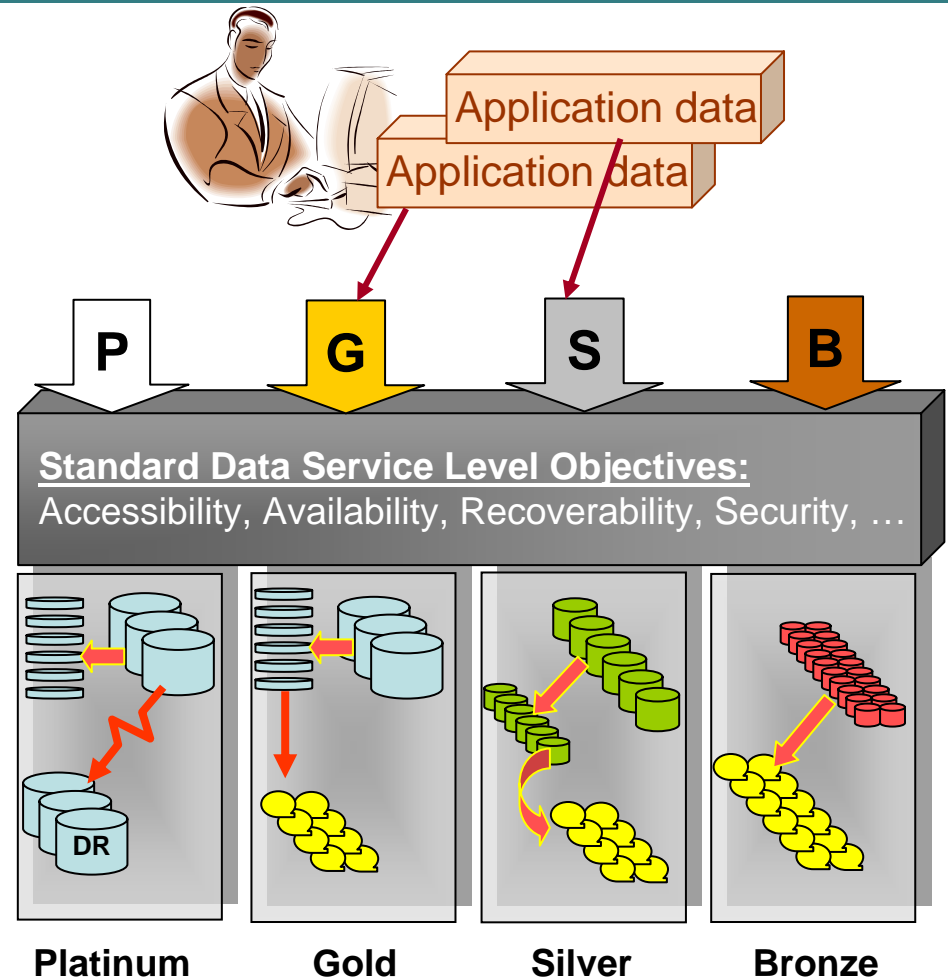
Immediate benefit: standard SLO metrics for SLAs

- Some “fuzzy” metrics
 - Coarse enumerations
 - Mapped to site-defined specifics
- Some quantifiable metrics
 - Usage descriptors to aid provisioning
 - Operations metrics for service level monitoring



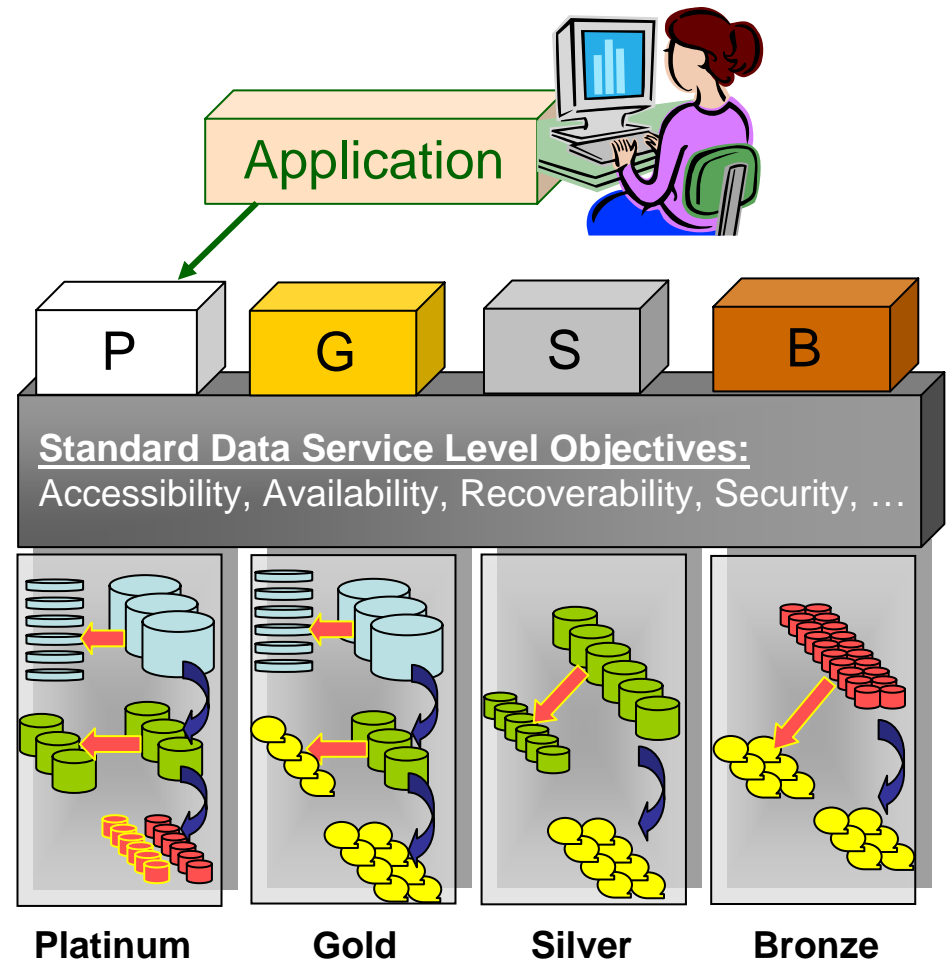
Future benefits

- Advertisement of data services
- Application developer chooses an available data service level for his data



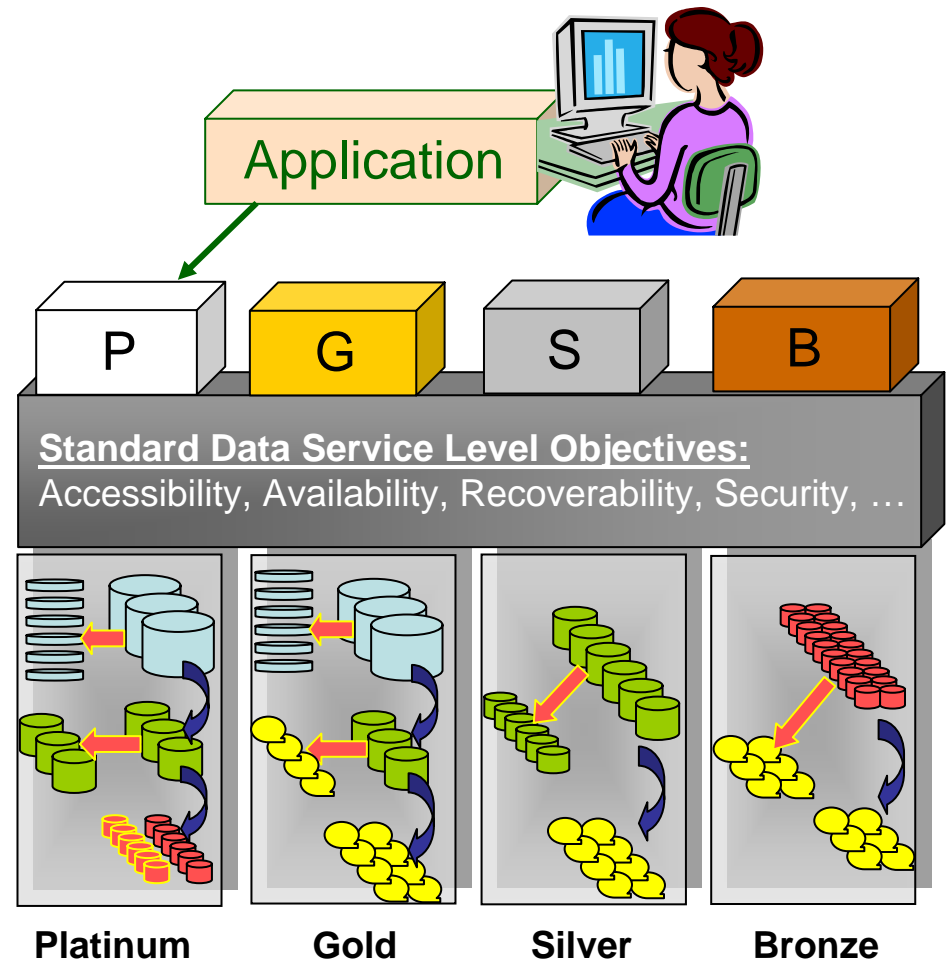
Future benefits

- Advertisement of data services
- Application developer chooses an available data service level for his data
- Advertisement of pre-defined corporate data lifecycles
- Application developer chooses a data lifecycle for her app



Future benefits

- Advertisement of data services
- Application developer chooses an available data service level for his data
- Advertisement of pre-defined corporate data lifecycles
- Application developer chooses a data lifecycle for her app
- Application developers design software that specify data SLOs and data policies
- Integration with grid computing service level management



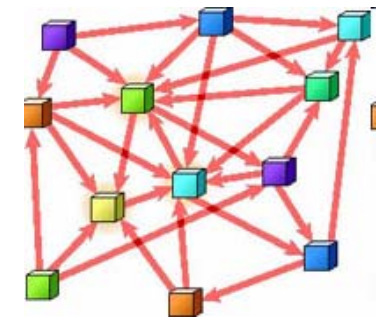
Data SLOs: Accessibility

Name	Data Type	Value / Quantifier	Notes
AccessType	Enum	<i>See next slide</i>	Fuzzy, extensible enumerator of storage access requirements.
AccessPattern	Enum	Random, Sequential	Distinguish access patterns.
SpaceUsage	Enum	Static, dynamic, sparse, fluctuate	How this would influence allocation is TBD
InitialAccessTime	Enum	Immediate, Sub-second, Seconds, Minutes, Hours	To distinguish media access MTFB characteristics. How long is application willing to wait?
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	Describe growth characteristics for data for capacity planning purposes
SizeGrowth	Integer	Bytes	Will grow this much during the GrowthPeriod
MaxSize	Integer	Bytes	Up to a maximum size
...			More SLOs are TBD – mostly focused on usage descriptors that optimize allocation

Draft – this is still under development by the SNIA ILM TWG

Data Accessibility SLOs: “AccessType” Enumerator

1. High throughput, Low R:W Ratio
2. High throughput, High R:W Ratio
3. Medium throughput, Low R:W Ratio
4. Medium throughput, High R:W Ratio
5. Low throughput, Low R:W Ratio
6. Low throughput, High R:W Ratio
7. Low planned usage



On target or oversimplified?

Draft – this is still under development by the SNIA ILM TWG

Data SLOs: Availability

Name	Data Type	Value / Quantifier	Notes
AvailabilityPeriod	Enum	Day, Week, Month, Quarter, Year	Availability measurement period for planned and unplanned downtime
PlannedDowntime	Integer	Seconds	Seconds of acceptable planned downtime for AvailabilityPeriod. Timing for such downtime would part of the SLA
MaxUnplannedDowntimeAggregate	Integer	Seconds	Seconds of unplanned downtime for the AvailabilityPeriod
MaxUnplannedDowntimePerOutage	Integer	Seconds	Seconds of unplanned downtime for any one occurrence of downtime

Draft – this is still under development by the SNIA ILM TWG

Data SLOs: Data Restore

Name	Data Type	Value / Quantifier	Notes
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.
DataConsistency	Boolean	True = consistency required (default) False = fuzzy acceptable	E.g., can a file system be backed up over time rather than having a point-in-time consistent copy?

Draft – this is still under development by the SNIA ILM TWG

More Data SLOs

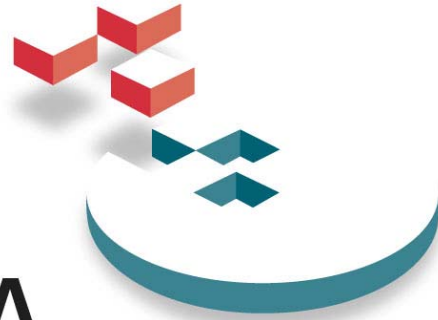
- **Budget & Tariff:**
 - Budget: Defined by business as the most they'll pay
 - Tariff: Assigned by IT to an advertised set of data and storage capabilities
- **Security:** TBD in conjunction with Security TWG
 - Regulatory compliance driven
 - Privacy and confidentiality-aware

Draft – this is still under development by the SNIA ILM TWG

Feedback:



- Are standard data SLO definitions something that would be of value to your data center today?
- Is the use of “fuzzy” specification for data SLO requirements useful?
- Any general comments on our choice of SLOs?
- Any specific comments on particular SLOs?
 - E.g., Are there any SLOs that are way off base or missing?



SNIA

STORAGE NETWORKING INDUSTRY ASSOCIATION

Thank you!