

SNIA

STORAGE NETWORKING INDUSTRY ASSOCIATION

EDUCATION

The Many Faces of Classification

Edgar St.Pierre, EMC²

The Many Faces of Classification

If you think that "data classification" is the greatest thing since sliced bread, then come meet the other slices in the loaf. Data classification is an often-referenced practice in Information Lifecycle Management (ILM) that is known to reduce costs and improve service levels in a data center. But before embarking on the effort to "classify your data", you may want to look at the many-faceted aspects of this practice. This presentation will explore the different types of classification that can be undertaken by organizations, and the many benefits to be derived from them. The discussion will range from the many different flavors of information classification, to data classification, to the practice of resource classification, and will help you decide where your organization needs to get started.

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.

Storage Network Industry Association

- SNIA is the trade group for storage networks
 - “ensuring that storage networks become complete and trusted solutions across the IT community”
 - <http://www.snia.org>
- SNIA’s “Dictionary of Storage Networking Terminology” online resource
 - <http://www.snia.org/dictionary>
- SNIA’s Data Management Forum, and its ILM Initiative, is an excellent information resource for data and information lifecycle management
 - <http://www.snia.org/dmf>

Agenda Outline

- At the end of this presentation, you will understand:
 - How classification is a cornerstone of ILM
 - A taxonomy of “classification” related activities
 - How each type of classification is used
 - Benefits associated with the different types of classification

ILM Implementation Roadmap

- Instrument & manage services to ILM practices across sites

Deploy ILM practices across the enterprise

- Automate with ILM Management tools

Automate ILM-based policies & services

- Begin operating services to policies

Pilot ILM-based Solution Stacks

- Set data & info polices across domains
- Tier storage and protection into standard Service Levels

Standardize Information, Data, & Storage Services

- Begin collaborating on requirements
- Identify value, lifecycle, & classification of information for each business unit

Identify information assets & infrastructure resources/ services

Classification

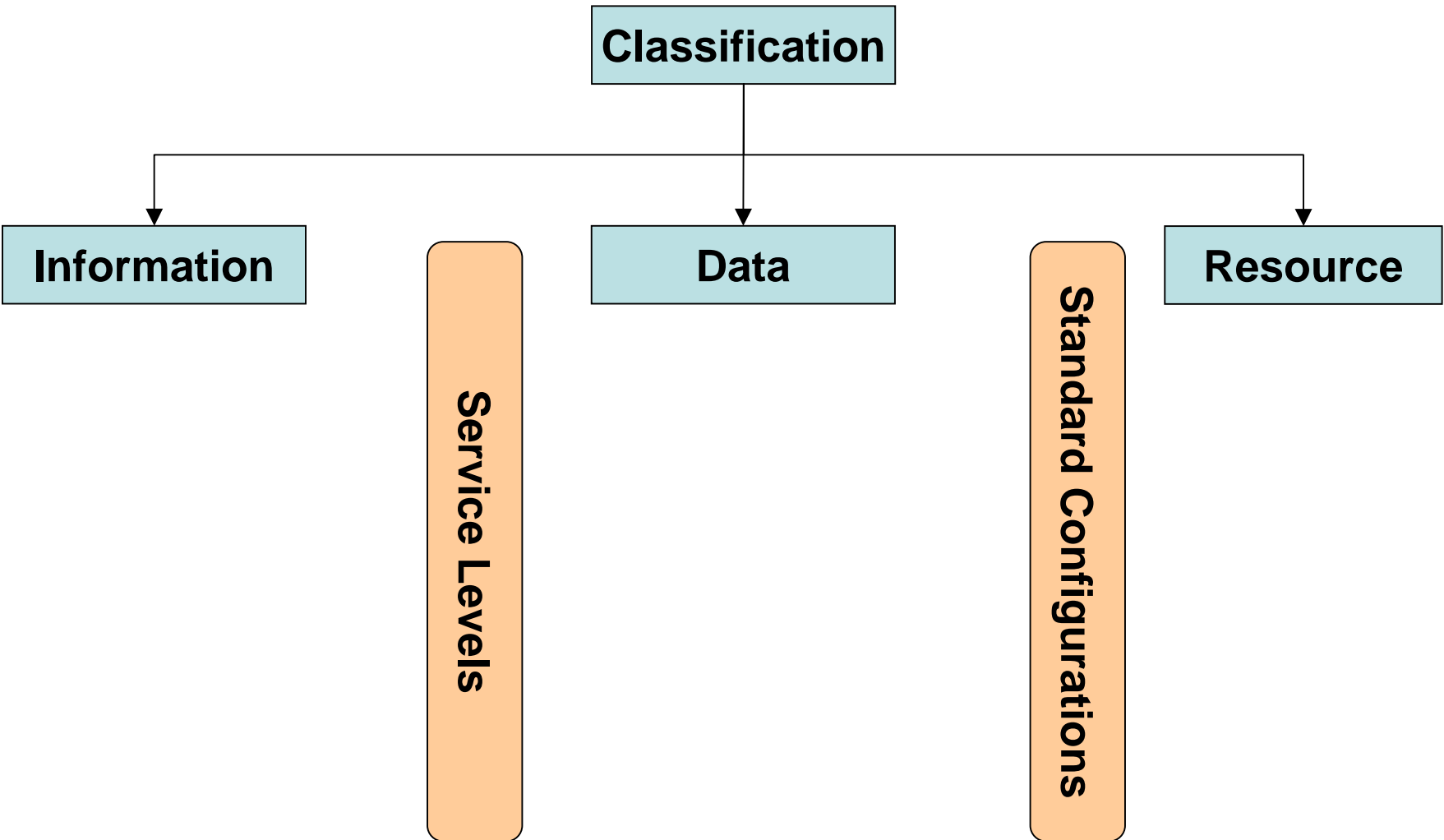
American Heritage® Dictionary:

1. The act, process, or result of classifying.
2. A category or class.
3. *Biology* The systematic grouping of organisms into categories on the basis of evolutionary or structural relationships between them; taxonomy.

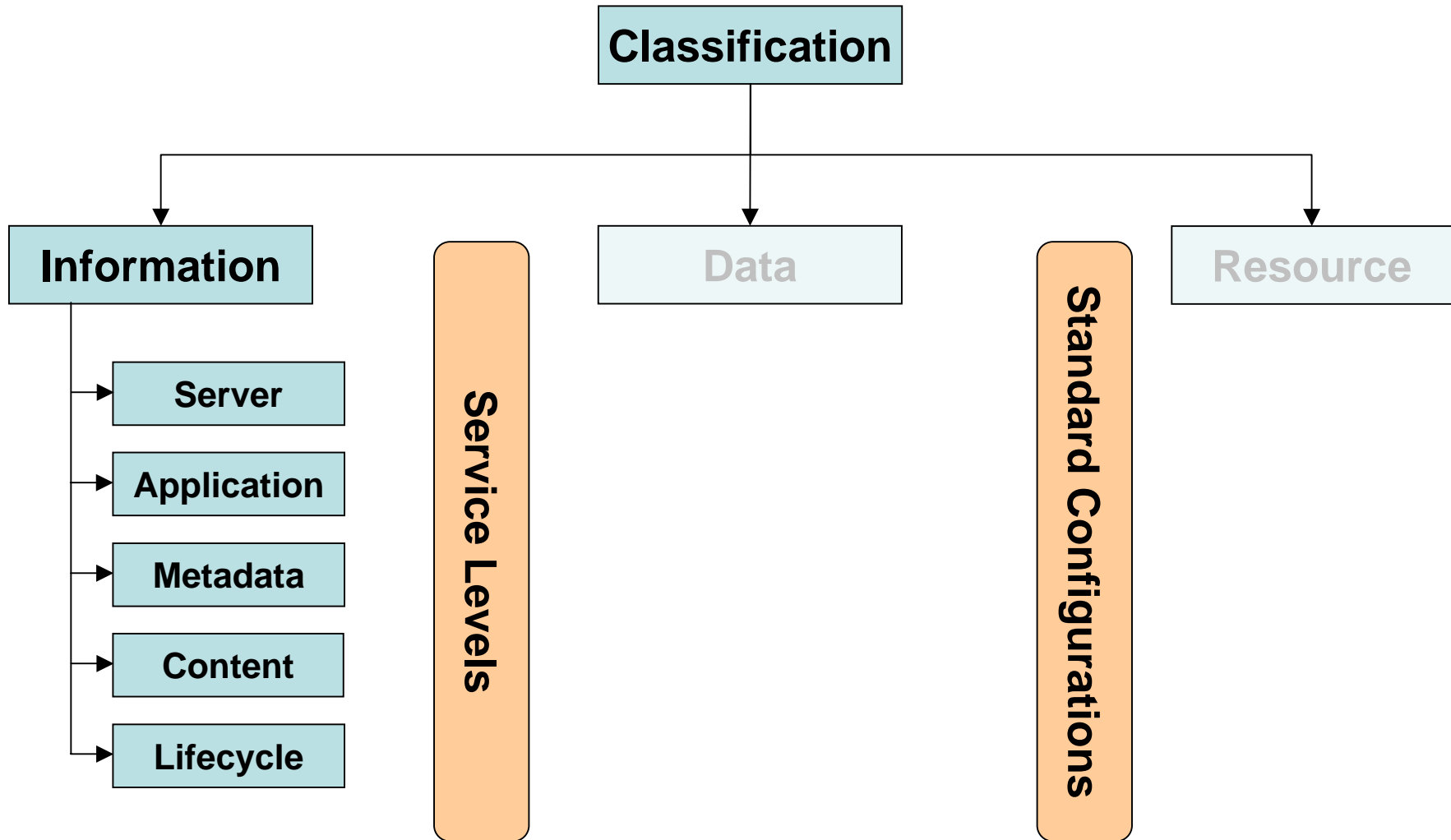
Roget's II New Thesaurus:

1. A way or condition of being arranged: arrangement, categorization, deployment, disposal, disposition, distribution, formation, grouping, layout, lineup, order, organization, placement, sequence. See ORDER .
2. A subdivision of a larger group: category, class, order, set

Taxonomy of “classification”



Taxonomy of “classification”



Information classification

Classification

Information

Server

Application

Metadata

Content

Lifecycle

Server: *(Old school)*

- All apps on the same server receive same services
- May be DAS-driven
- *Simple to define and implement*
- *Define same requirements for all information on server*

Application: *(First wave)*

- Establishes a ranking of applications
- Determine requirements per application
- All information in the application treated the same
- *Good first approximation to align services to requirements*
- *Define same requirements for all information in application*

Information classification

Classification

Information

Server

Application

Metadata

Content

Lifecycle

Metadata: *(First wave and early adopters)*

- Tags: "Creating Application", "Author", ISBN, ...
- Location: all user data on share partition X:\
- *Method to organize data*
- *Define requirements for each information type*

Content: *(Enterprise Content Mgmt-related apps)*

- Determine information *type* by content/key word taxonomy
 - Distinguish Weather→Hurricane→Katrina from Shopping→Gifts→Katrina's Crafts and Gifts
- *Categorization and organization of data*
- *Define requirements for each information type*

Information classification

Classification

Information

Server

Application

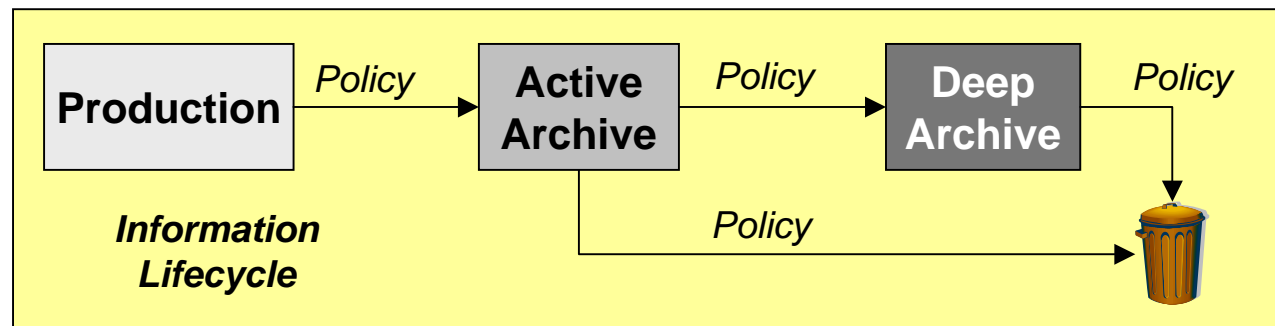
Metadata

Content

Lifecycle

Lifecycle: *(Both old school & early adopters)*

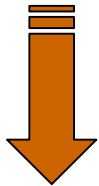
- Define stages & policies for the data
 - Production, archive, compliance, ...
 - Last access time, access frequency, ...
- Used in conjunction with information type classification
- *Derived from business use of information*
- *Define requirements at each stage of lifecycle*
- *Align services with the value of the information as it changes*



Information classification output

Classification

Information



Criteria for Service Level Agreement (SLA)

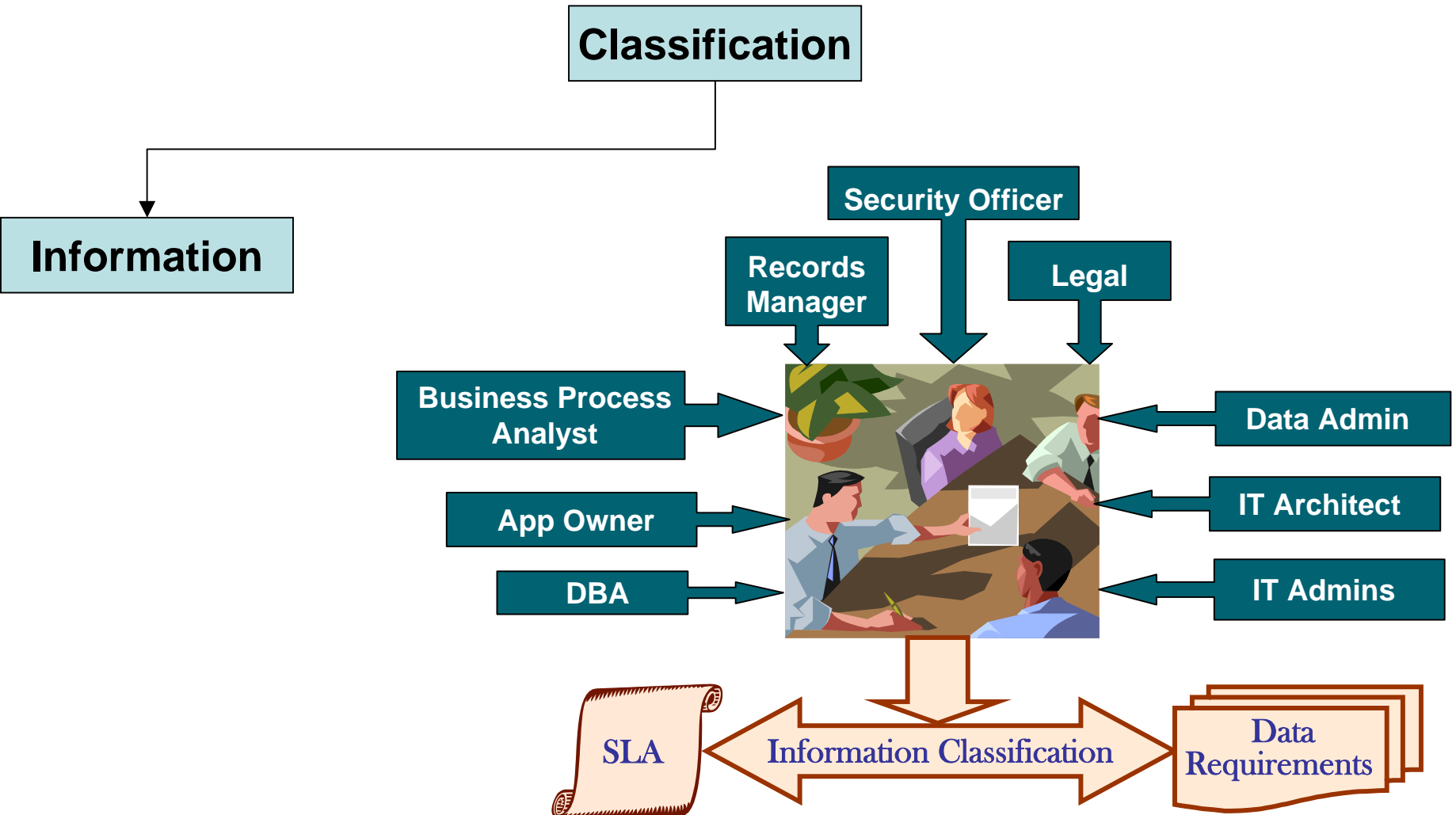
- End user performance requirements
 - Screen refresh, availability, throughput, transactions, ...
- Monitoring & Reporting
 - Report content & frequency
- Chargeback
- Capacity Planning
- ...

The Many Faces of Classification - Property of SNIA© 2005 Storage Networking Industry Association. All Rights Reserved.

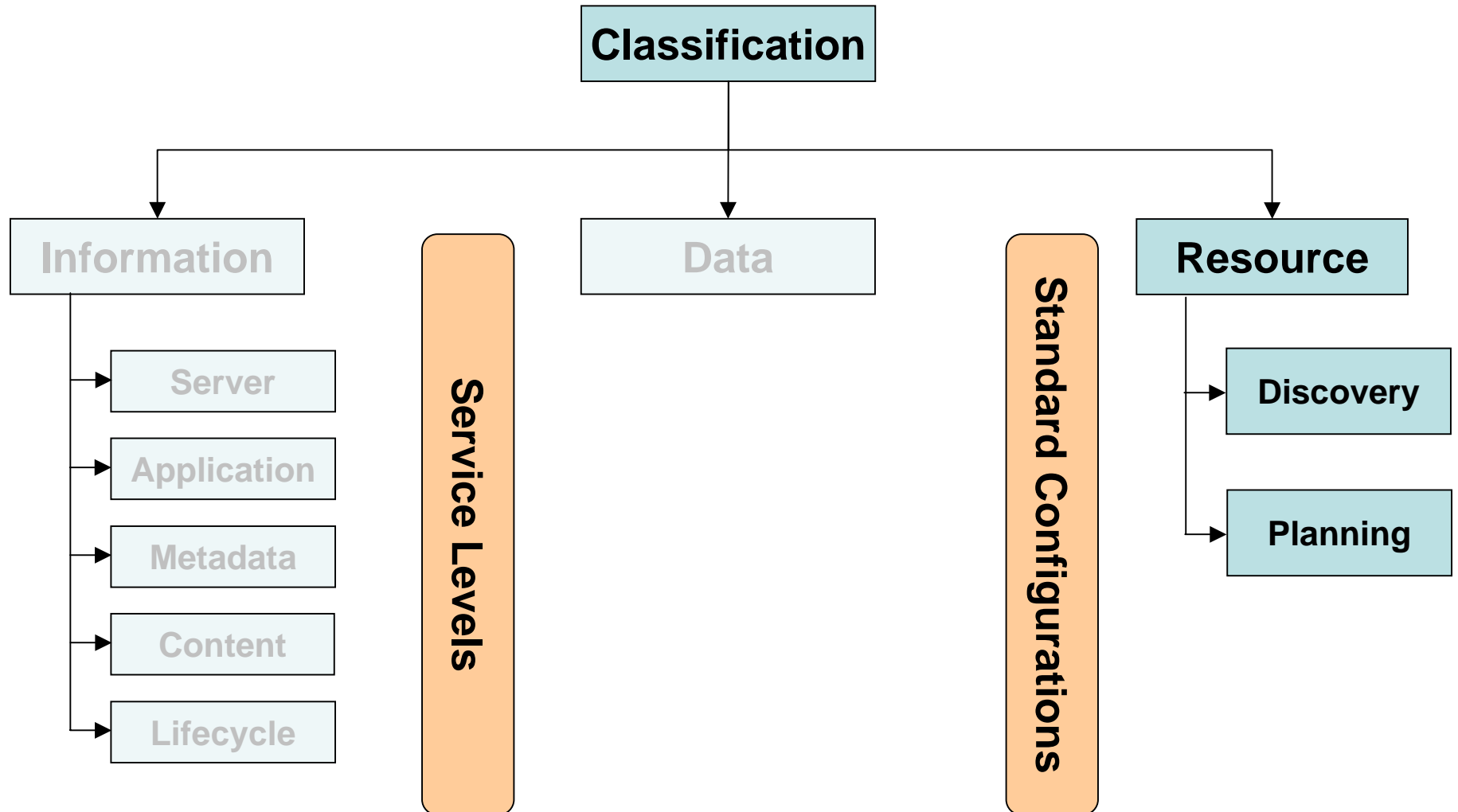
Requirements for Data Mgmt

- Business Value:
 - Budget for IT
- Application & usage requirements:
 - Uptime, bandwidth, IOPS, size, growth, RPO, RTO, ...
- Security requirements:
 - Public, Internal, Confidential, ...
- Compliance requirements:
 - SOX, HIPAA, ...
 - Retention time, immutability, audit requirements, ...

Information Classification *requires collaboration*



Taxonomy of “classification”



Resource classification

Classification

Resource Classification:

Organize IT resources into well-defined configurations with predictable behavior.

- **Discovery:** capture reality
 - Via Configuration Mgmt tools
 - Via SRM tools
 - Manual mapping of applications to resources, ...
 - *Map environment to improve service delivery*
- **Planning:** define desired
 - Examine requirements
 - Create well-defined configurations to meet them
 - *Resource & configuration consolidation to reduce TCO*

Resource

Discovery

Planning

Resource classification

Classification

Account for all Data Center resources:

- Storage, Compute, Network
- Data Management SW
 - Operational Recovery software
 - Backup, replication management
 - Disaster Recovery infrastructure & software
 - Backup/RM, WAN, pickup truck
 - Data Movement SW
 - Archive, HSM, compliance
- Security HW/SW resources
 - Accountability, Integrity, Authenticity, Trustworthiness, Confidentiality, Immutability, Destruction, Physical Security
- Compliance resources
 - Immutable storage, litigation audit & search engines

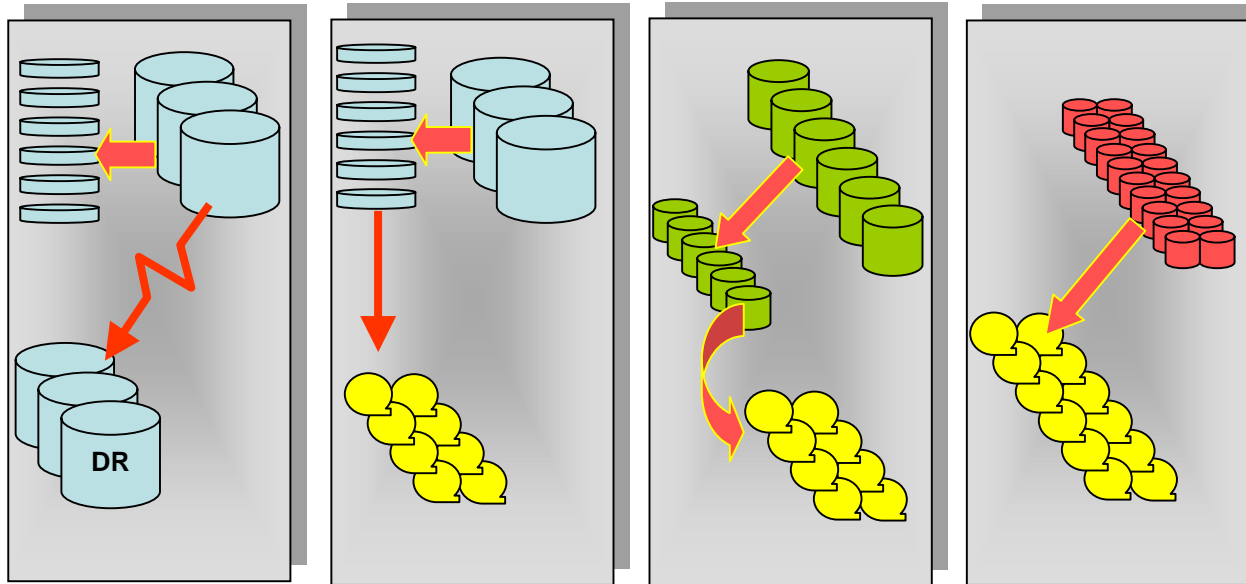
Resource

Discovery

Planning

Resource classification

Classification



**Hi Perf
Hi Avail
Fast OR
Fast DR**

**Hi Perf
Hi Avail
Fast OR
Normal DR**

**Std Perf
Std Avail
Std OR
Normal DR**

**Low Perf
Std Avail
Slow OR
Slow DR
Immutable**

Resource

Discovery

Planning

Resource classification

But how?!

Classification

Reduce 10's or 100's of configurations to a few templates of standard configurations?

- Yes, this is the “magic”
- Satisfy largest and most important apps first
- Map remaining apps as feasible
- Business priority: improve service levels vs. reduce TCO
- Gap analysis provides input to resource planning

How many configuration templates?

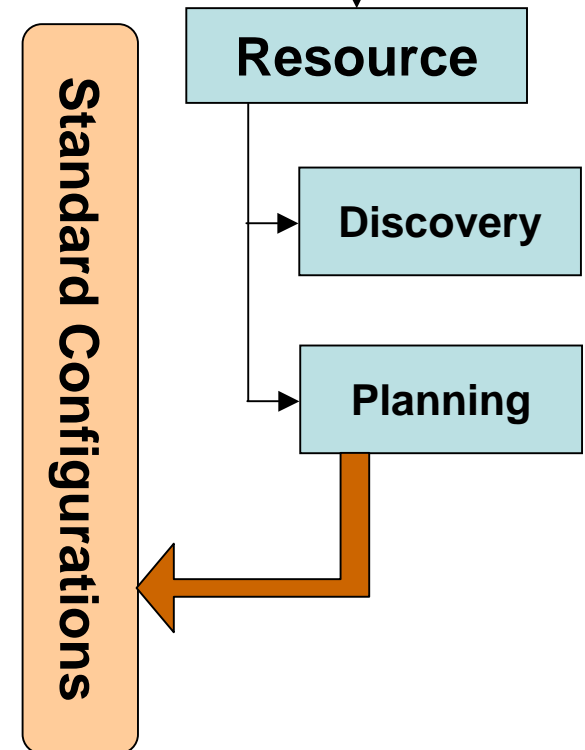
- There is no “right” number
 - Consensus best practice: few – not many
- May include “whatever I have now”

Are configuration templates == service levels?

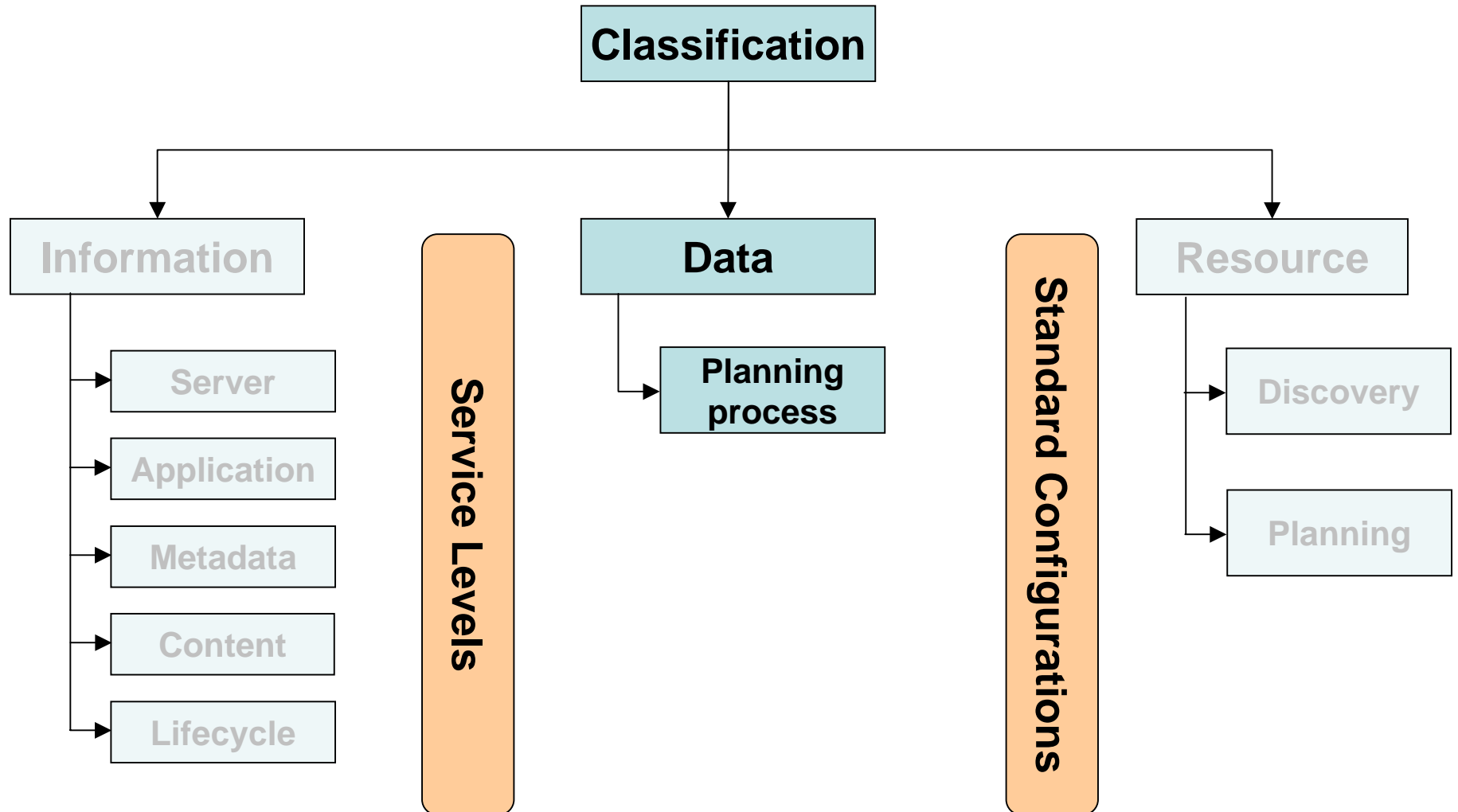
- No - multiple templates may map to one service level

How are these named?

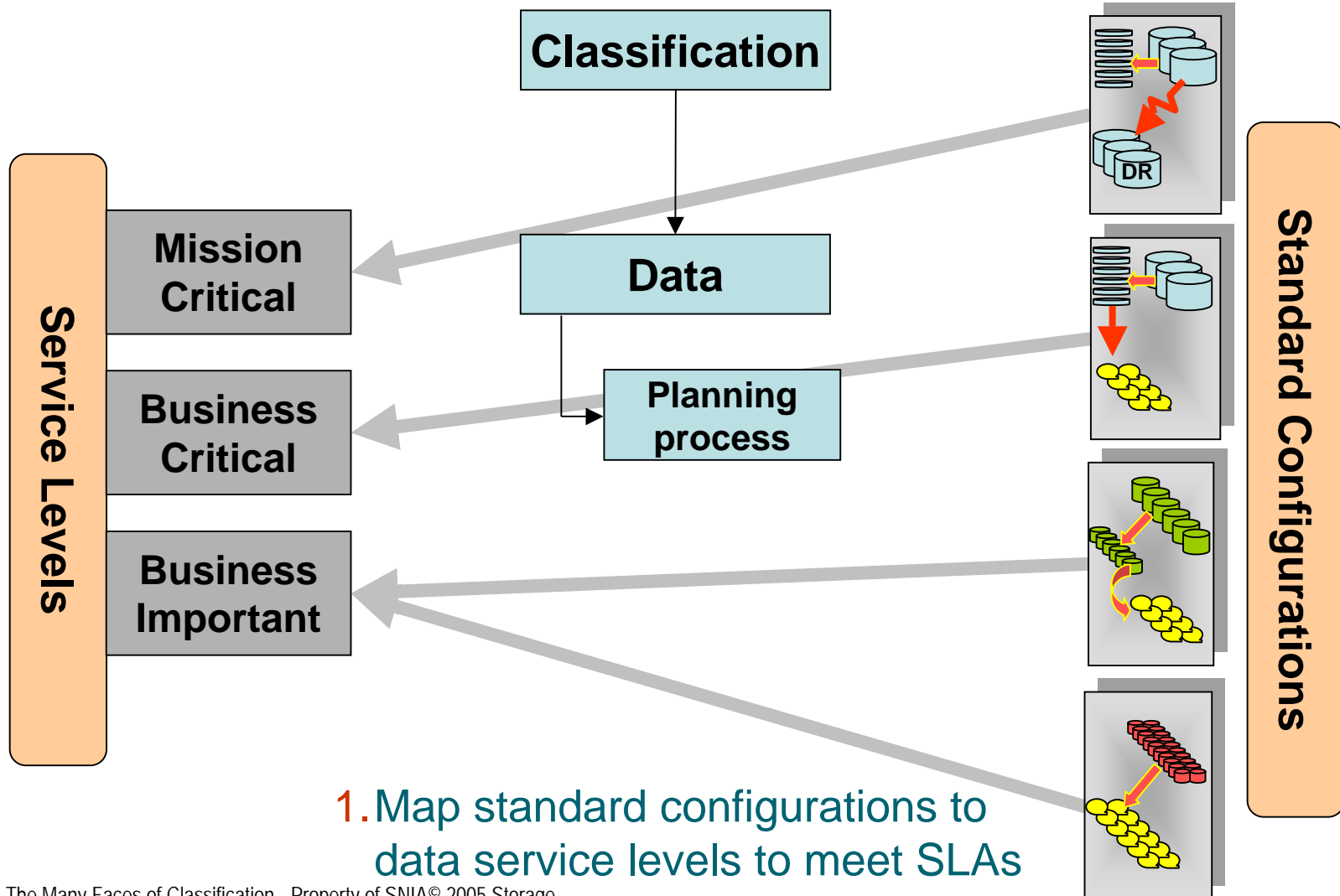
- Any way you want – these are IT-internal



Taxonomy of “classification”



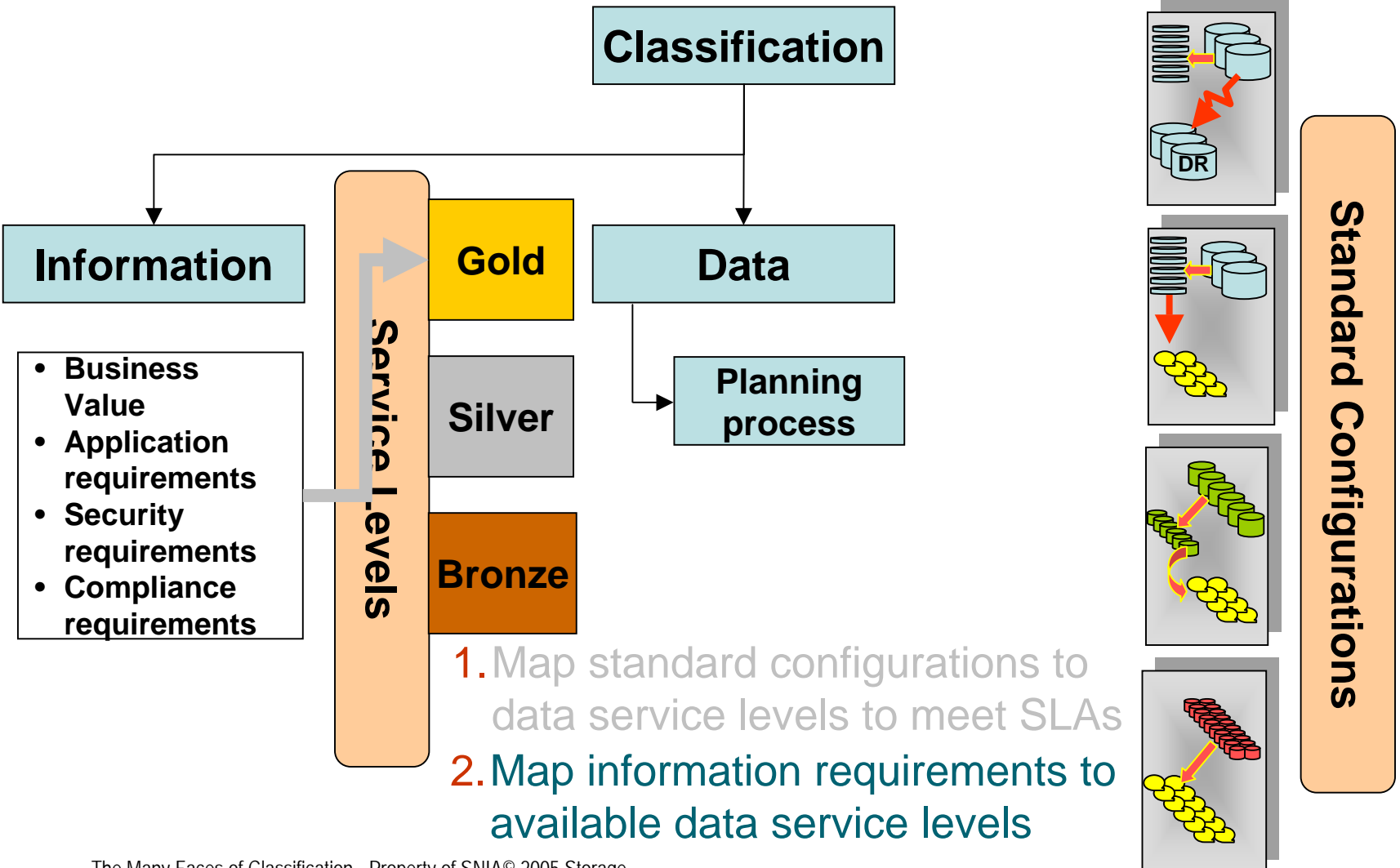
Data classification planning process





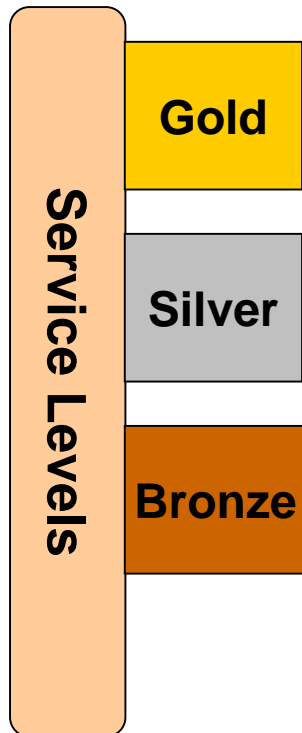
Data classification planning process

ILM aligns information with the most appropriate resources

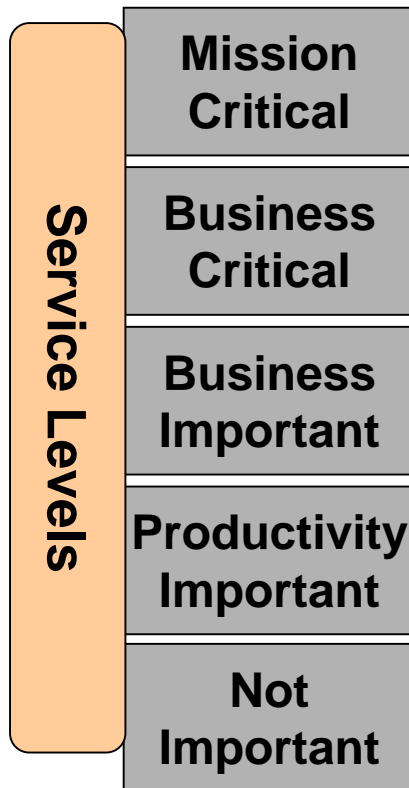


Naming service offerings/levels

- Don't get hung up on service level names

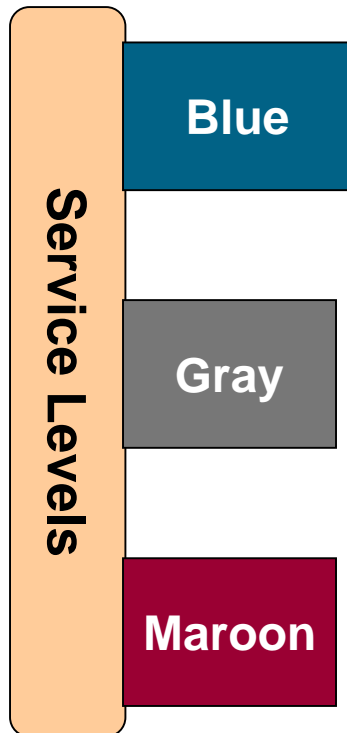


Naming service offerings/levels



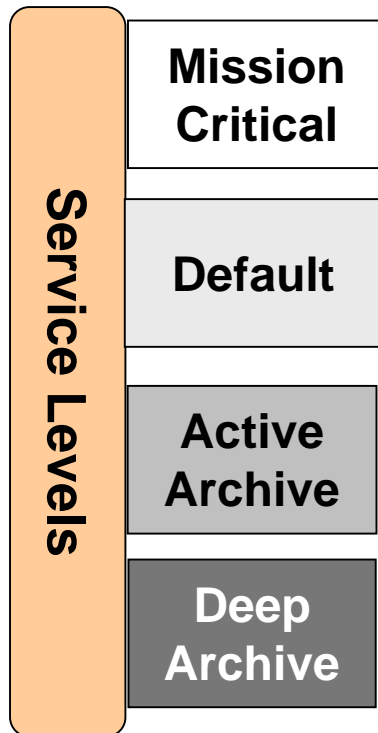
- Don't get hung up on service level names
- Each variation conveys different implications to different people

Naming service offerings/levels



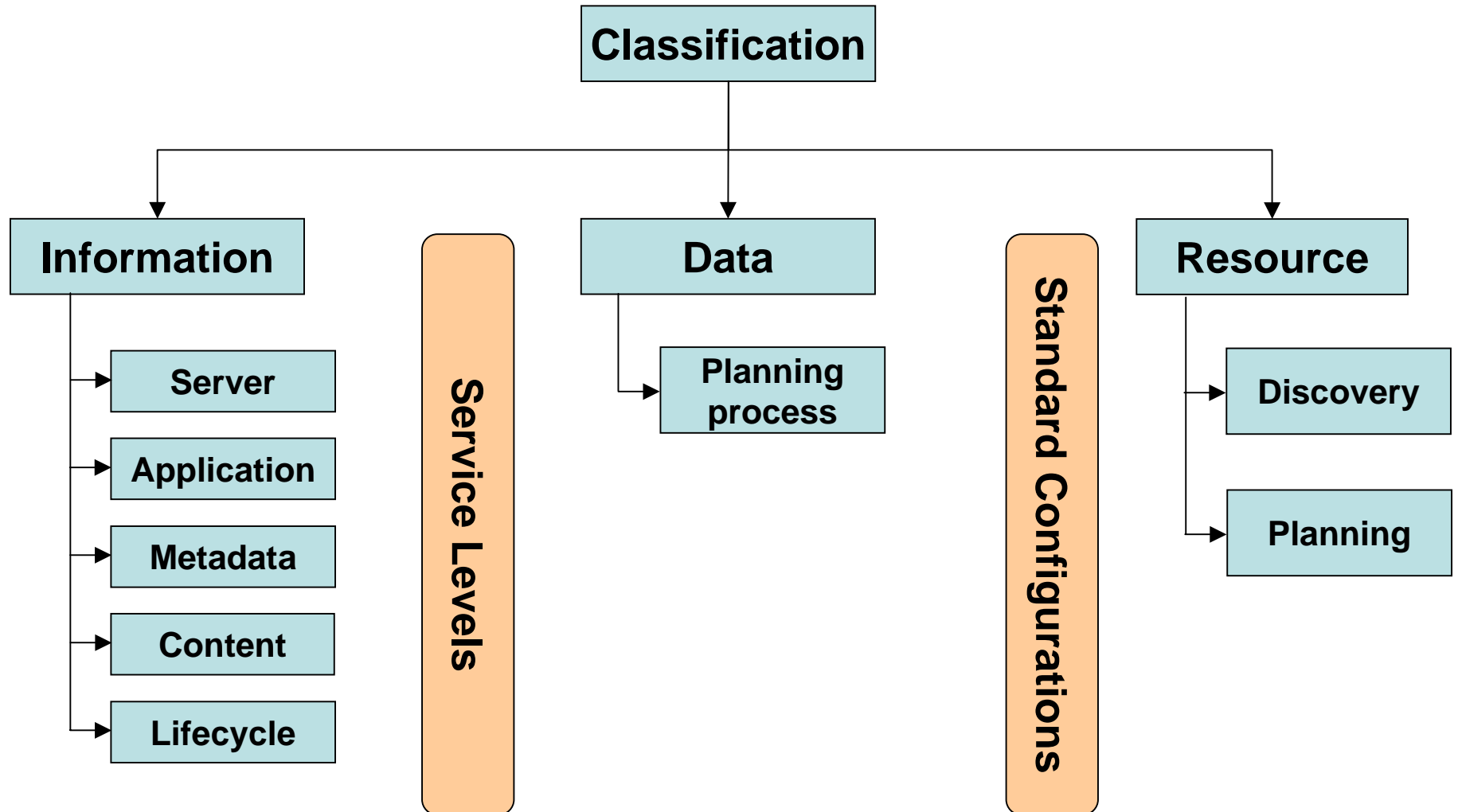
- Don't get hung up on service level names
- Each variation conveys different implications to different people
- Use what's good for you

Naming service offerings/levels



- Don't get hung up on service level names
- Each variation conveys different implications to different people
- Use what's good for you
- Beware of extensibility issues:
 - If “service level #1” is fastest and best available, what happens when you need something faster?
- Descriptive terms work well, too

Taxonomy of “classification”



ILM is a classification-driven management practice

- Many forms of classification
 - Information classification → Define requirements
 - Resource classification → configuration templates
 - Data classification → align requirements to resources
- Value
 - Reduced TCO
 - Improved service levels
 - Simplification of environment

Q&A / Feedback

- Please send any questions or comments on this presentation to SNIA: track-datamgmt@snia.org

**Many thanks to the following individuals
for their contributions to this tutorial.**

SNIA Education Committee

**Edgar St.Pierre, EMC
Terry Yoshi, Intel
Jeff Porter, EMC
DMF ILM Initiative
SNIA ILM TWG**