

Data Quality: A Success Factor

Data Management Association

National Capital Region

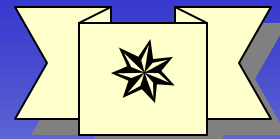
March 10, 2009

Edith Purdie

Strategic Business Consultant
EDP Enterprises

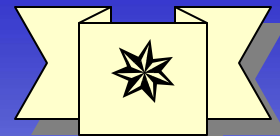
Alan Harmon

U.S. Naval Academy, Technical Director
Institutional Research, Planning,
and Assessment Office



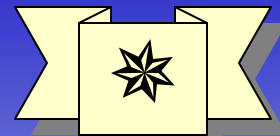
Data Quality — The Abstract

Data quality can mean the difference between success and failure. Quality, although a key component in Data Governance and Master Data Management, is often included as a token effort or as an afterthought. This presentation will focus on an “inside-out,” holistic approach with emphasis on best practices and lessons learned.



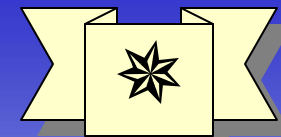
Data Quality Overview

- Data Quality and customer satisfaction are inextricably linked.
- Data quality is holistic as it is linked to every component in your business. DQ transcends organizational lines and crosses over enterprise architecture arenas.
- DQ is a process, not a one time occurrence or task. You can't check the box and say you're done.
- DQ needs to be embedded organizationally and procedurally.

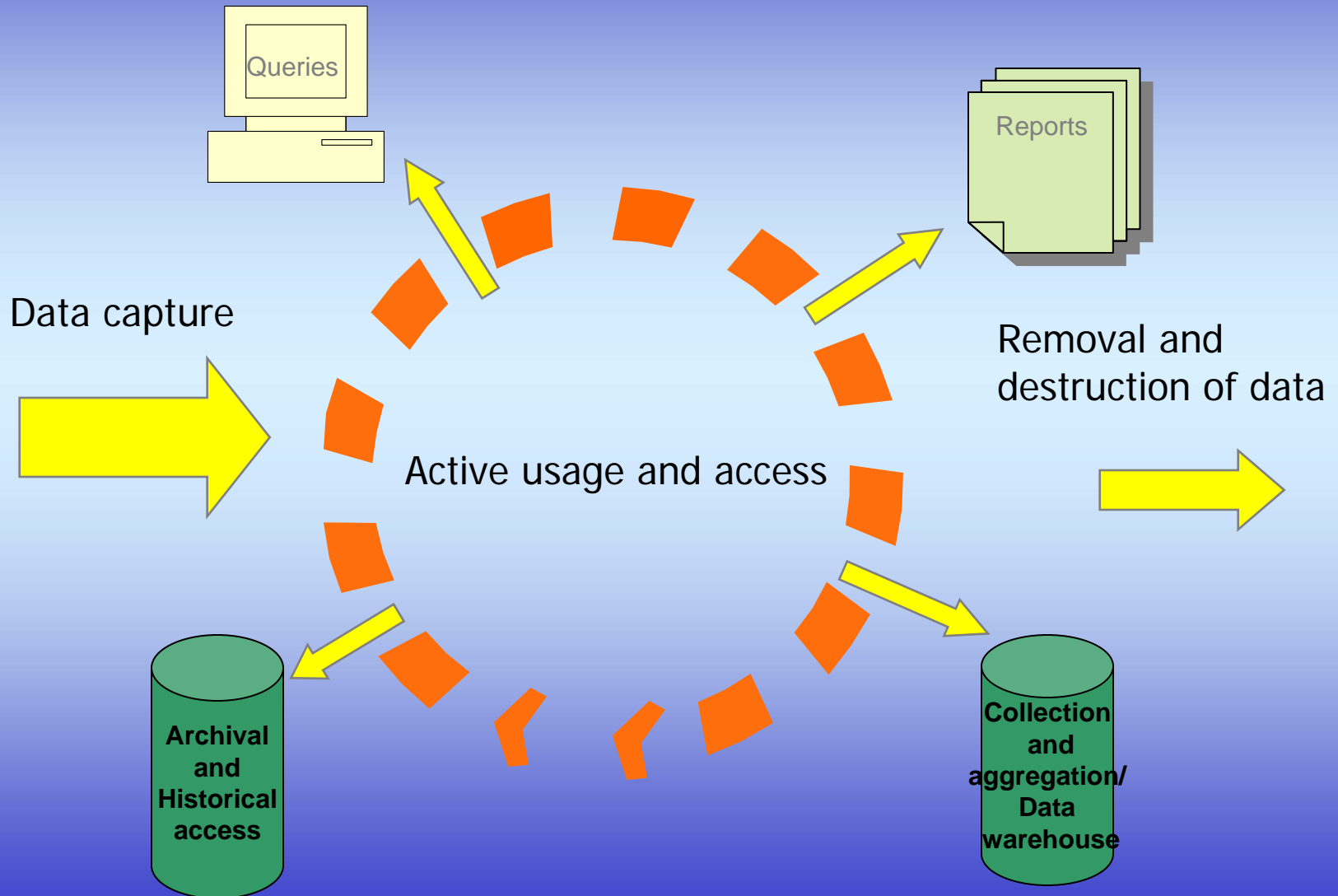


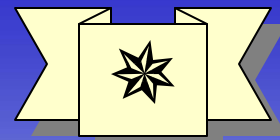
The Data Tsunami

- Increasing amounts of data to be stored and archived
- Types of data, especially unstructured, need capture
- Compliance and other regulations
- Need to evolve from data reporting to business intelligence for agility
- Provide performance assessment metrics as validation
- Cautionary note about hackers, worms, and data corrupters.



Typical Data Life Cycle





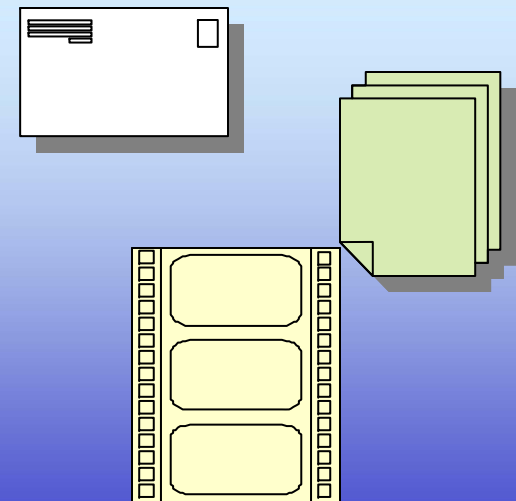
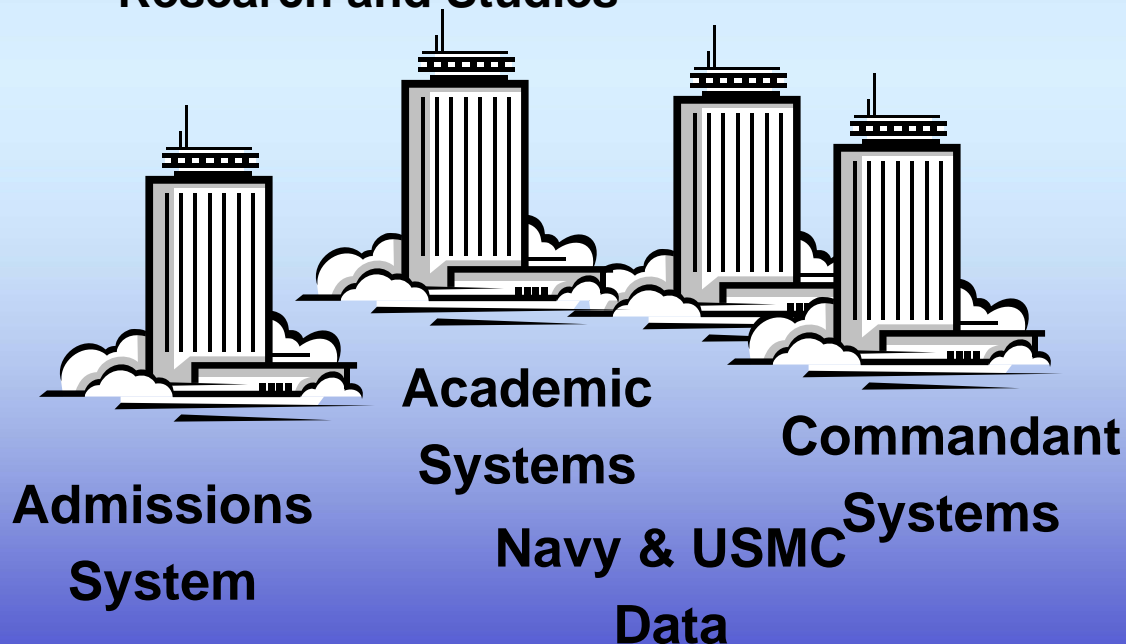
Types of Data: Structured and Unstructured

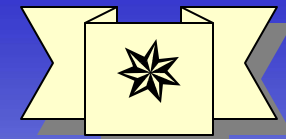
- **Structured Data**

- Production Systems
- Administrative Systems
- Historical Data
- Research and Studies

- **Unstructured Data**

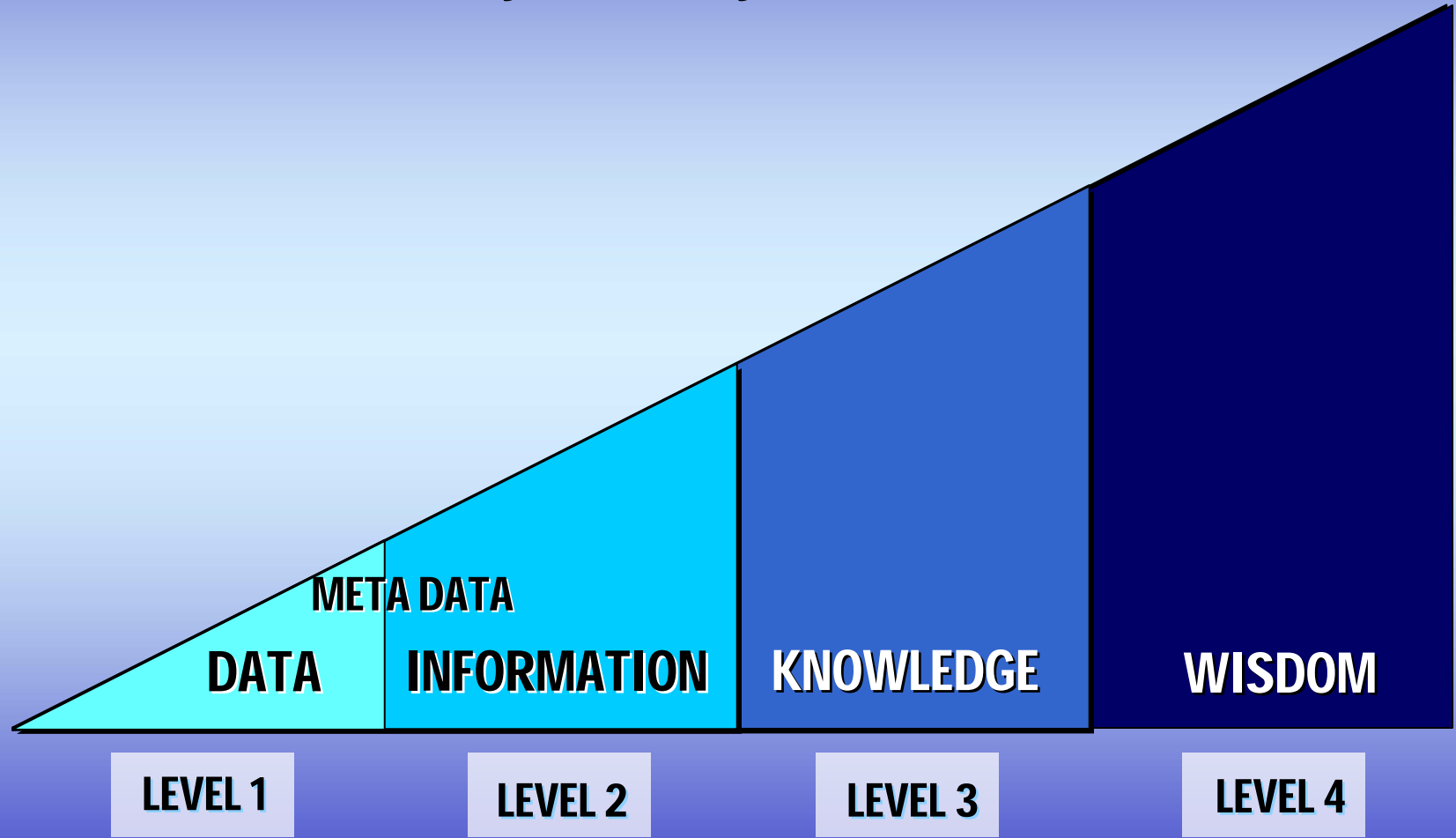
- Email
- Documentation
- Meeting Notes
- Video Clips/Movies
- Web pages & services
- Instructions/ Notices

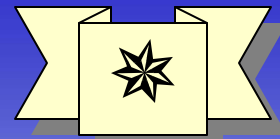




Data Quality vs Information Quality

Information Delivery Maturity Model



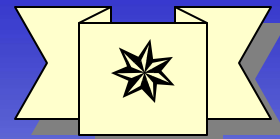


Data Content Quality Dimensions

Accurate (to reality)	Non-Duplication
Accurate (to surrogate source)	Precision
Completeness	Relationship Validity
Concurrency	Timeliness
Consistency	Validity
Derivation Integrity	

from Federal DAS Data Quality Framework, Version 1.0, October 1, 2008

[.http://colab.cim3.net/cgi-bin/wiki.pl?FederalDataArchitectureSubcommittee](http://colab.cim3.net/cgi-bin/wiki.pl?FederalDataArchitectureSubcommittee)

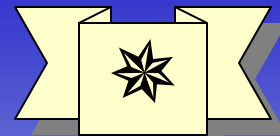


Data Presentation Quality Dimensions

Accessibility
Contextual Clarity
Usability
Rightness

from Federal DAS Data Quality Framework, Version 1.0, October 1, 2008

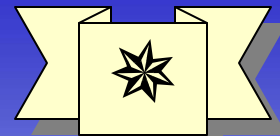
. <http://colab.cim3.net/cgi-bin/wiki.pl?FederalDataArchitectureSubcommittee>



Quality Classes for Each Dimension

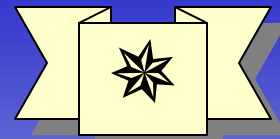
Absolute
Second Tier
Third Tier

from Federal DAS Data Quality Framework, Version 1.0, October 1, 2008
· <http://colab.cim3.net/cgi-bin/wiki.pl?FederalDataArchitectureSubcommittee>



Data Quality Tools and Techniques

- Data Profiling
- Data Standardization
- Data Defect Prevention
- Metadata Management
- Matching and Linking
(aka Data reengineering and correction)
- Business Rule Tools

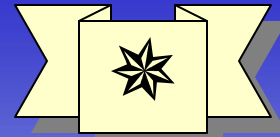


Benefits of a Data Quality Program

- Allow ready access to and sharing of data
- Facilitate effective data analysis
- Enable better, faster decision-making
- To meet regulatory compliance
- Reduce data maintenance costs
- Maintain data integrity

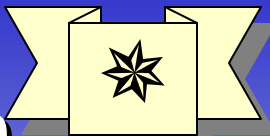
And, last but not least,

- Minimize the cost of poor data quality

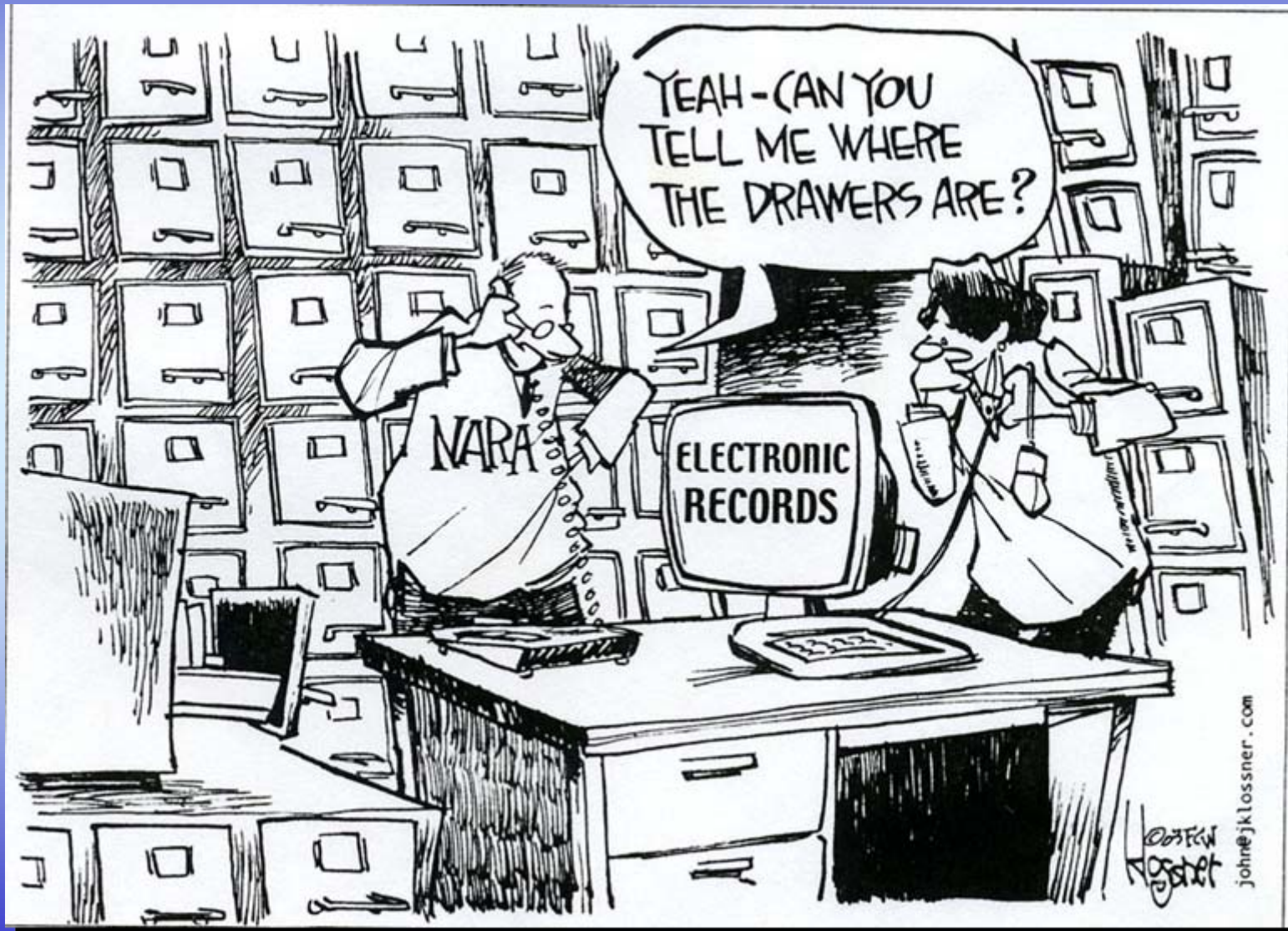


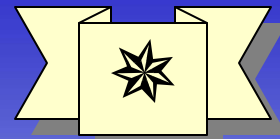
Recent Experiences

- **USNA Data Warehouse and Metadata Repository**
- **Proactive: Top Down Approach**
- **Reactive: Bottom Up Approach**



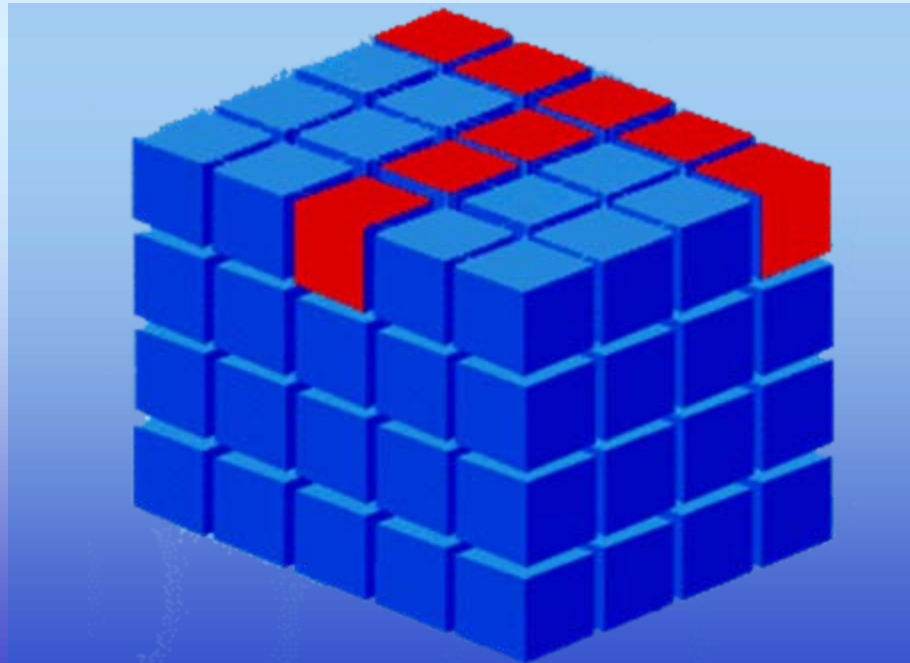
The Reality: Data, Data and More Data...

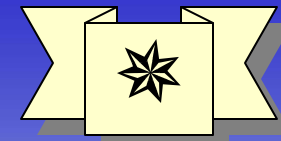




Data Warehouse

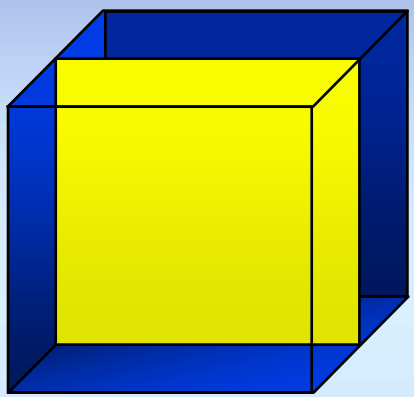
A data warehouse is a consolidated view of the enterprise data, optimized for reporting and analysis



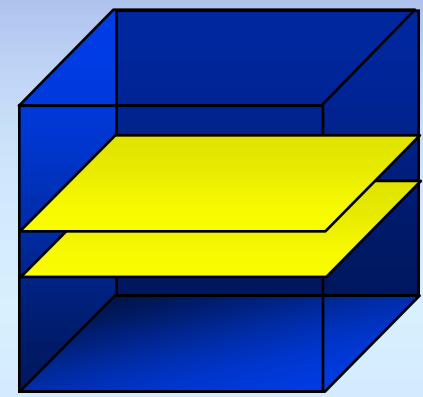
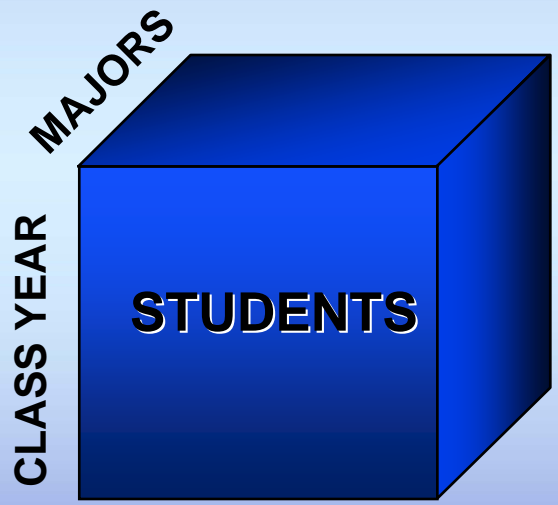


Data Warehouse--Continued

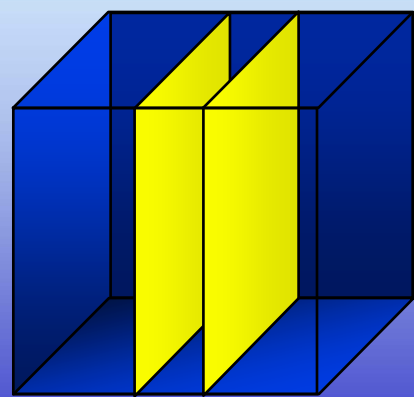
Online Analytical Processing—A Multi-Dimensional View



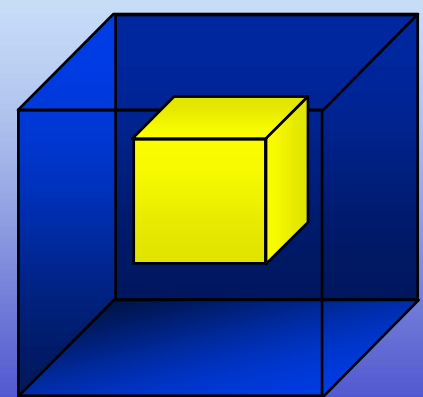
By MAJOR



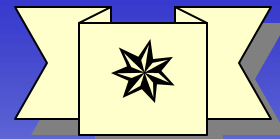
By GENDER



By CLASS YEAR



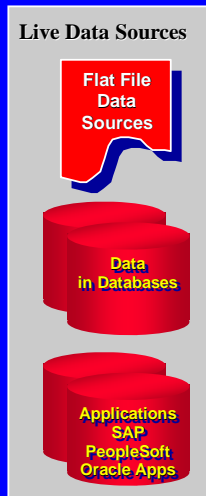
By SAT SCORE RANGE



Data Cleansing

Loading and Cleaning Data

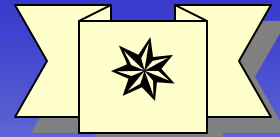
Opportunity to Integrate, Correct, and Validate Data



Data Extraction and Cleaning
(can be very complex)



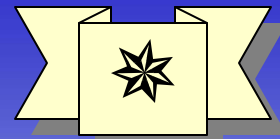
- ◆ Integrate multiple data sources
- ◆ Correct data problems (cleanse)
- ◆ Validate Data
- ◆ Summarize and roll-up data
- ◆ Update Metadata



DW Benefits to the Agency

- Allow ready access to, and sharing of, data
- Facilitate effective data analysis
- Enable better, faster decision-making
- To meet regulatory compliance
- Reduce data maintenance costs
- Maintain data integrity after new loads



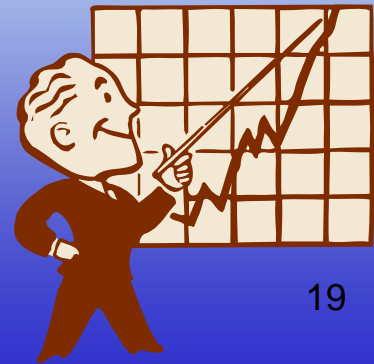


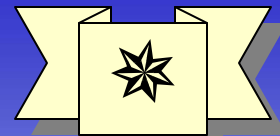
First Practical Steps to get started...

1. Develop a DQ ASSESSMENT

- Can be elusive; not just 100's of error reports
- We took samples & verified e-records
- Extract DQ Rules from Data Dictionary (from 1st stage)
- Exposes YOUR data and your ability to deal with it

2. Build an aggregated DQ SCORECARD

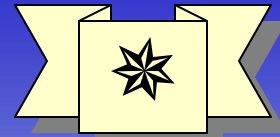




USNA DW & Metadata DQ Lessons

- What are your DQ rules? Do you break every rule?
- Identify DQ rules through modeling & profiling
- Must develop rules for time-dependent data
- Must develop rules for state-dependent data objects
- Must develop rules for complex data relationships

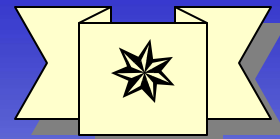




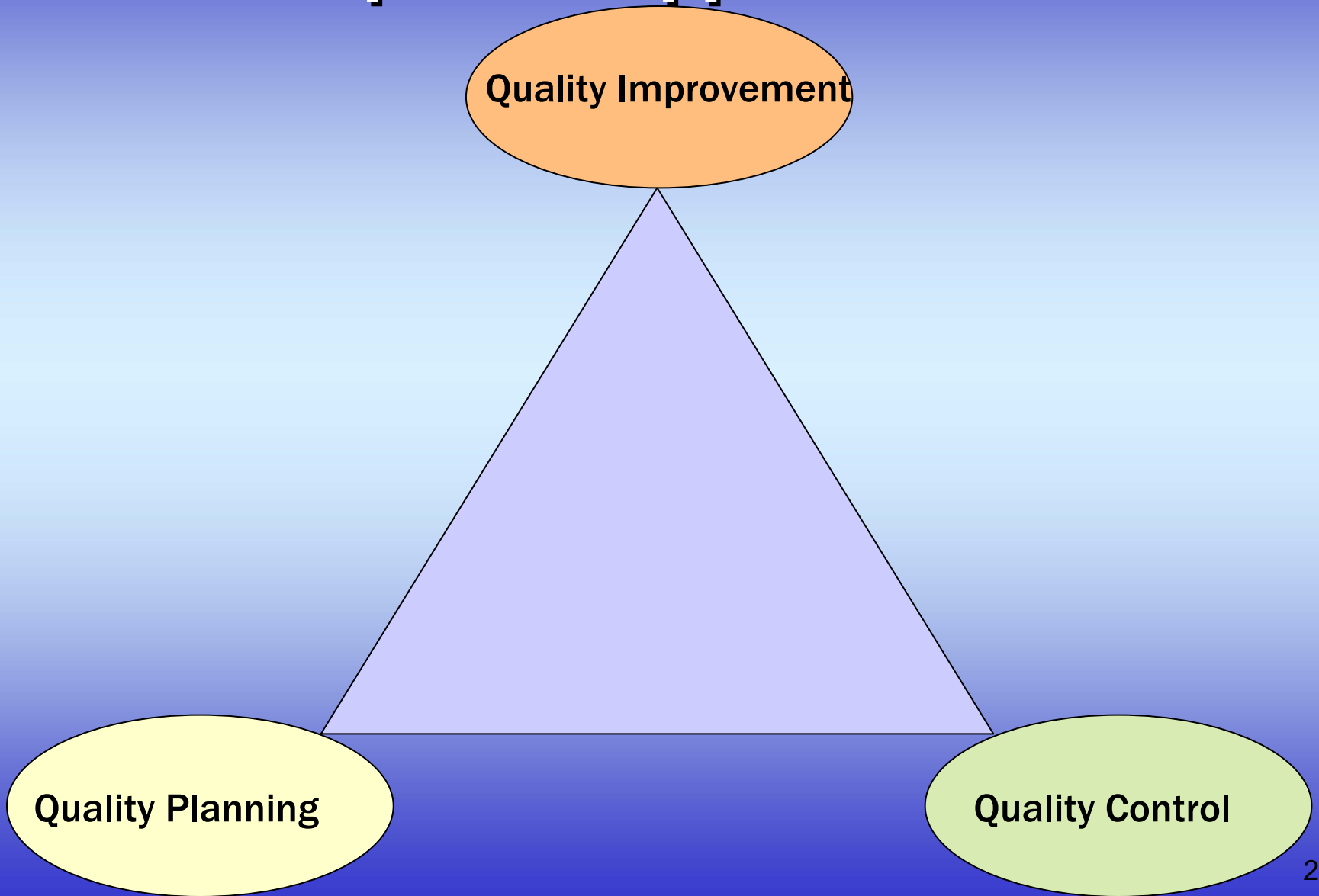
Metadata Management

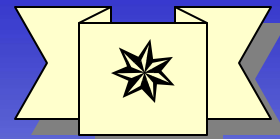
Archive and Retention Management System (ARMS)

- Provides a centralized directory of USNA historical data and meta data
- Ensures data is retained and usable for analysis
- Establishes a process for managing institutional information
- Allows us to understand, manage and evolve our data, information and knowledge architectures

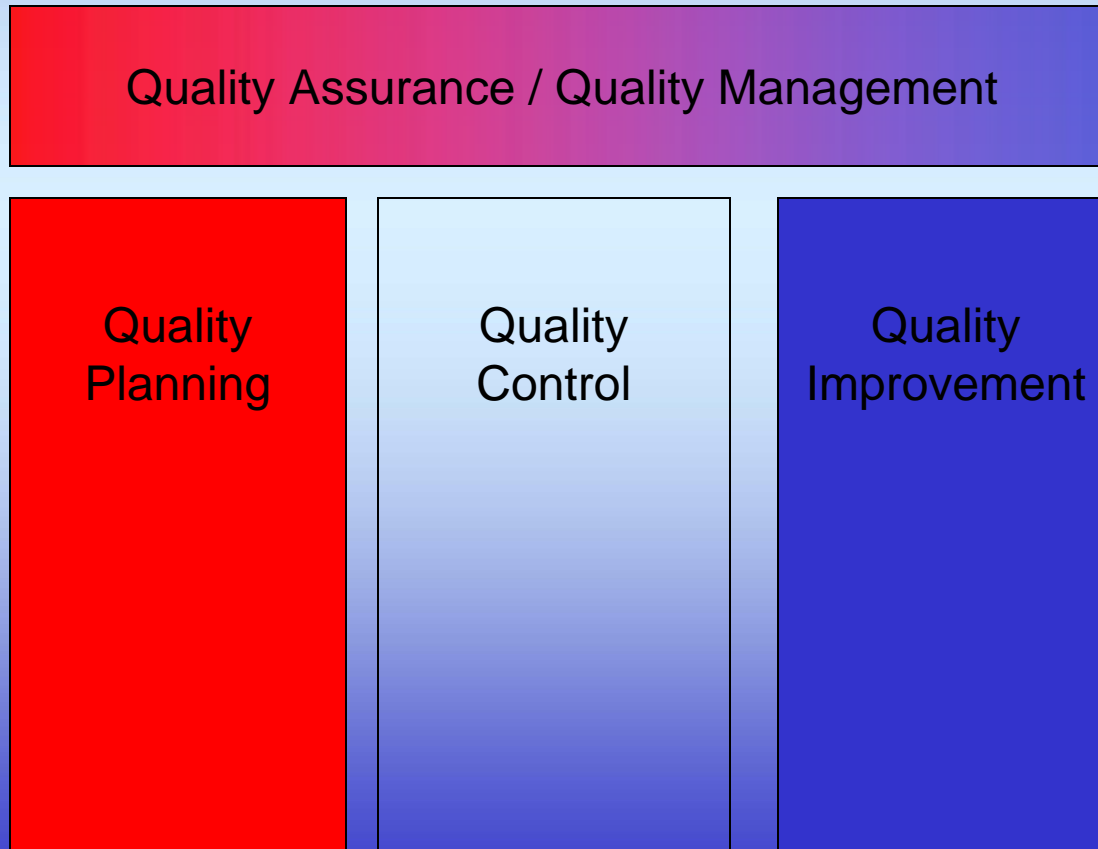


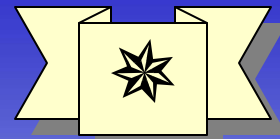
Proactive: Top Down Approach



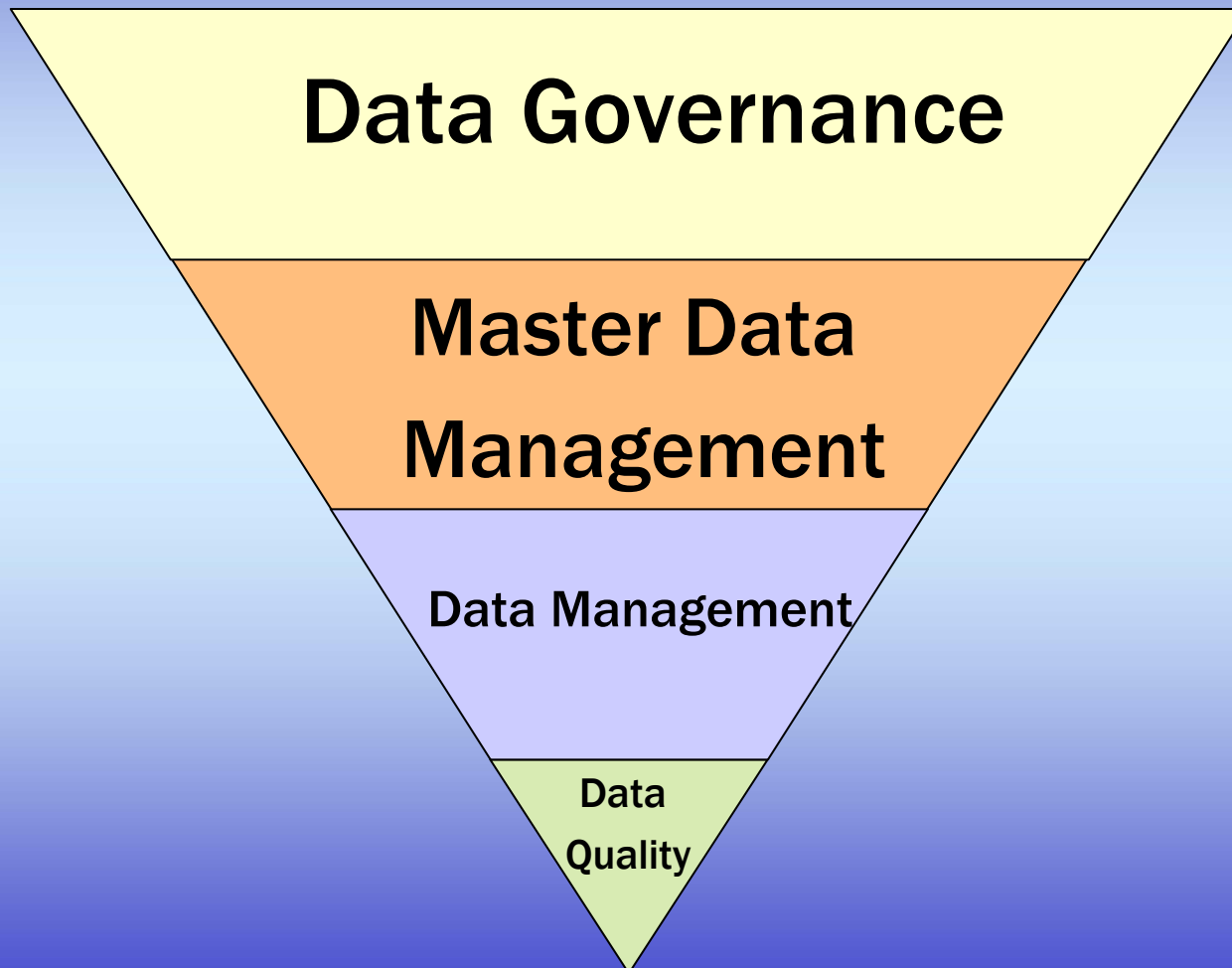


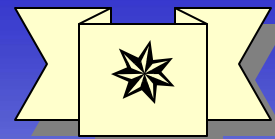
Quality Management / Quality Assurance



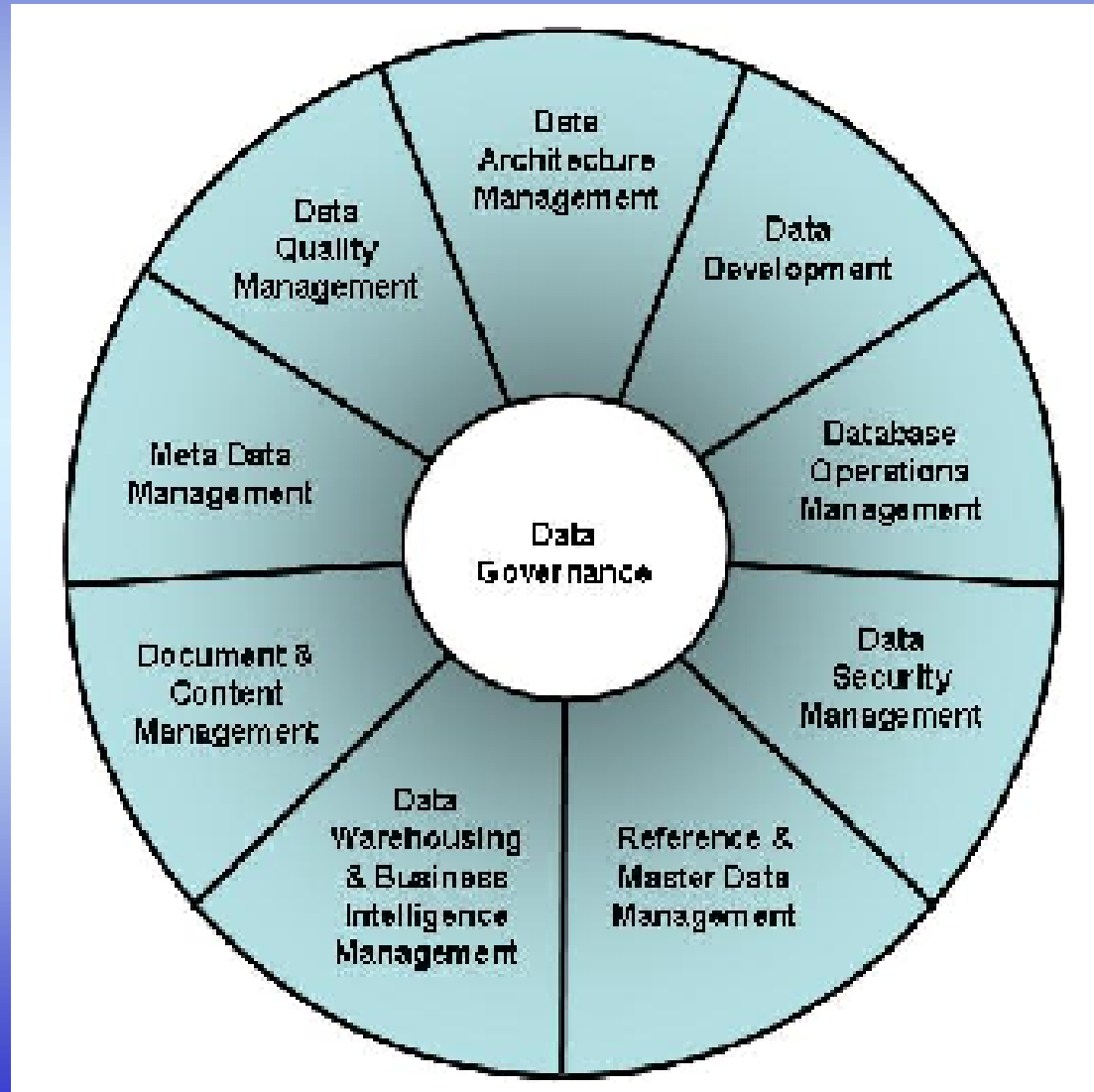


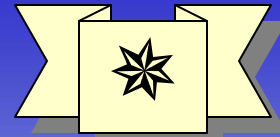
Top Down Governance Approach





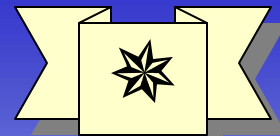
DAMA DMBOK Functional Framework v.3





Formulate Standards

- Service Level Agreements (SLAs)
- Obtain Measurements and Performance Samples
- Review Work in Progress
- Quality Assurance Evaluator (QAE) Reports
- Quality Audits, Evaluations, and Inspections
- Testing and Inspection Procedures
- Verification and Validation of Test Results
- Independent Audits and Research
- Project-specific standards



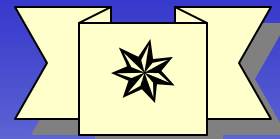
Reactive: Bottom Up Approach

"Would you tell me, please,
which way I ought to go from here?"

"That depends a good deal
on where you want to get to," said the Cat

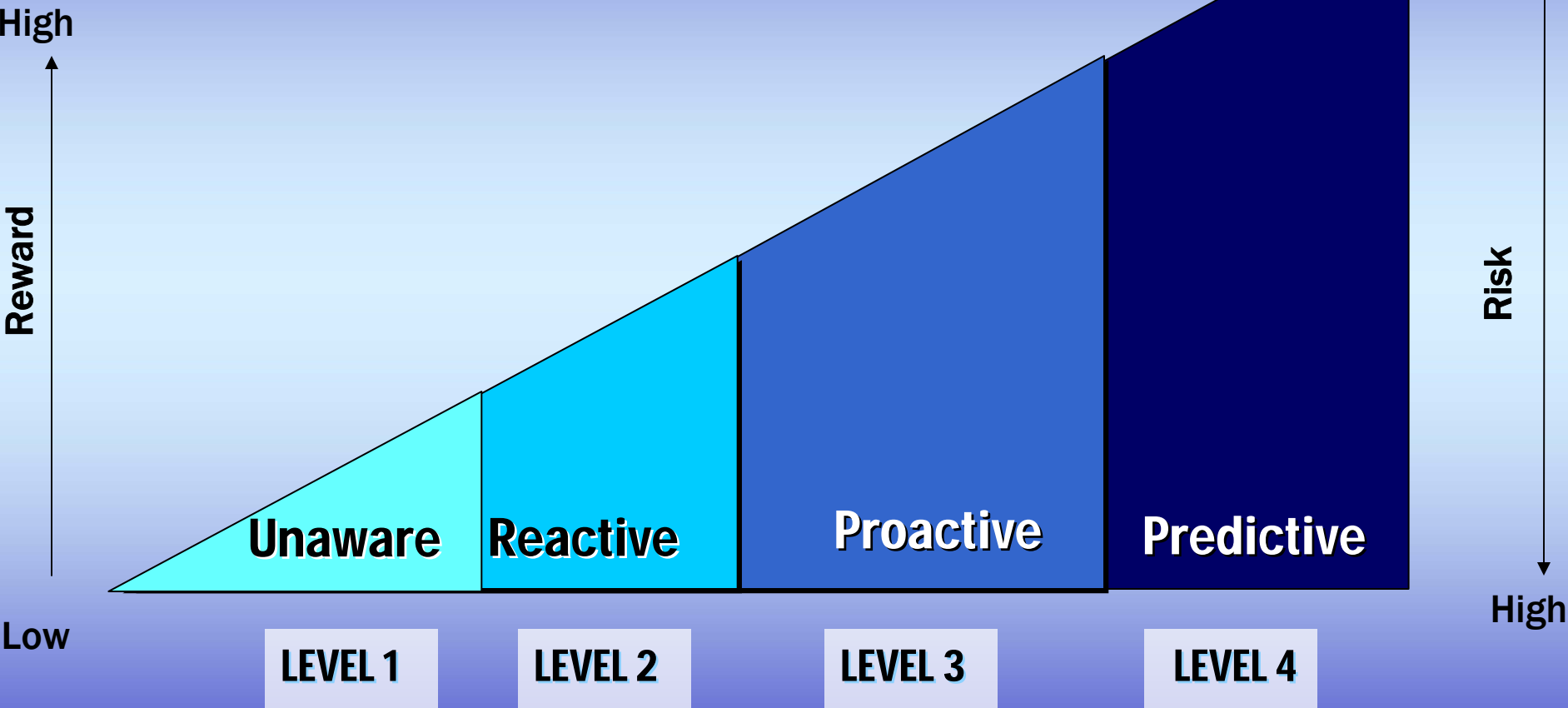
From Lewis Carroll's 1865 novel, Alice's Adventures in Wonderland

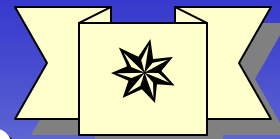




Maturity Model

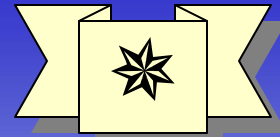
Data Quality Management Maturity Model





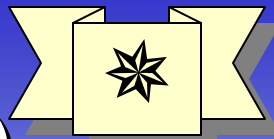
Challenges to a Data Quality Program

- Weak data governance
- No data management infrastructure
- Organizational Turf Wars
- Siloed organization preventing a holistic vista
- No perceived value
in Earned Value Management (EVM)



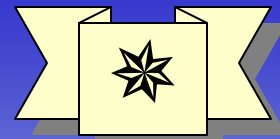
Tips For Success

- Line up management commitment first
- Understand the “pieces”, the dimensions
- Make a plan
- Begin with a measurable, prioritized goals
- Show results in the first six months



Some References (Non Inclusive List)

- **Data Management Association -- www.dama.org**
- **The Data Administration Newsletter -- www.tdan.com**
- **Data Governance Institute --
<http://www.datagovernance.com/>**
- **DM Review -- www.dmreview.org**
- **International Association for Information and Data Quality --
www.iaidq.org**
- **MIT, Total Data Quality Management Program --
<http://web.mit.edu/tdqm>**



Questions?



Go **NAVY!**
Beat Army!