DW2.0 and Data Quality

ABSTRACT----------

Does your organization need to deliver BI @ the Speed of Business? This presentation is for Business leaders to Architects who lead their organizations into the future by taking advantage of a sound architectural framework that delivers a high quality data resource. This data resource is the foundation of the data warehouse and is essential to making accurate and quick business decisions.

This paper will describe the optimal data quality process with the aid of the DW2.0 Architecture, DW2.0TM is the architecture of the next generation of data warehousing. It is a statement of what a data warehouse should be and the vision that Bill Inmon has for the future of data warehousing. This architecture gives your organization a sustained quality improvement of the corporation's data warehousing investment. Several features of DW 2.0 include the recognition of the life cycle of data within the data warehouse; inclusion of unstructured data along with structured data inside the data warehouse. Every atomic data element in the warehouse must be of a high quality. This presentation will outline:

- How to achieve DQ for the second generation of data warehouses.
- How to access DQ tool categories to implement your data quality process.
- How to develop the DQ deliverables that promise high ROI.

BIOGRAPHY----------

Linda Kresl
Business Intelligence Manager
Mentor Graphics

Ms. Linda Kresl has held a variety of professional and management positions with world- class companies such as The Boeing Company, Hewlett Packard, PriceWaterhouseCoopers, and Nike. Her professional experience of more than 20 years includes development of enterprise Business Intelligence, Enterprise Information Management and Data Quality Improvement. From 2001-2007 Ms. Kresl established her own consultancy specializing in BI & Enterprise Data Architecture. Ms. Kresl is currently the Business Intelligence Manager at Mentor Graphics.

Ms. Kresl has been a speaker on Information Quality Management at the MIT IQ Industry Forum and the Information & Data Quality Conference. She is a member of the Data Warehouse Institute (TDWI), & is a certified DW 2.0 Architect. She sits on the boards of IAIDQ & DAMA Global Chapter. She has published in DMReview; her articles have also appeared in Oracle Toolbox.

Agenda

- *How to achieve DQ for the second generation of data warehouses.*
- *How to access DQ tool categories to implement your data quality process.*
- *How to develop the DQ deliverables that promise high ROI.*

Who are we?

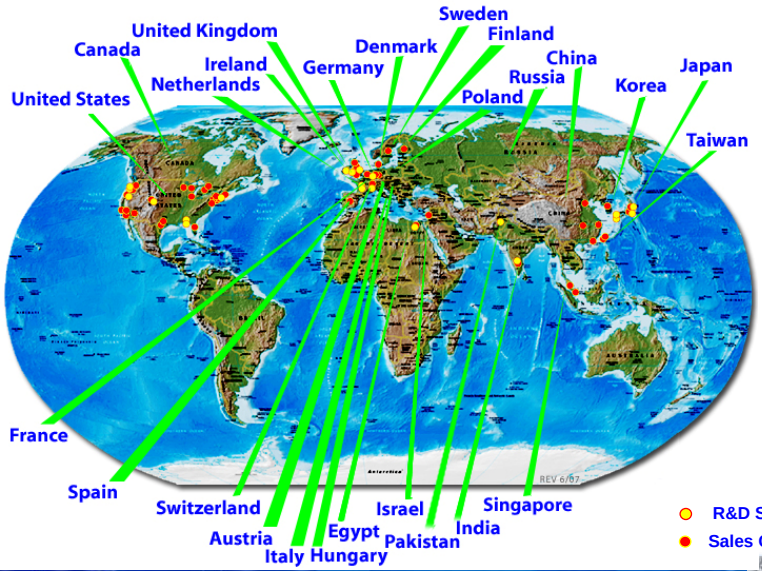
- **Mentor Graphics® is a technology leader in electronic design automation (EDA), providing software and hardware design solutions that enable companies to develop better electronic products faster and more cost-effectively.**
  - **Publicly held (NASDAQ: MENT)**
  - **Founded 1981, headquartered in Wilsonville, Oregon**
  - **approximately 4,350 employees**
  - **Revenue in last reported 12 months: about $789 million**
  - **World-class research and development - 28 engineering sites worldwide**
  - **High-touch, global distribution channel - 48 sales offices around the world**
  - **Strategic partnerships with leading electronics manufacturers, semiconductor and electronic design suppliers for development of new design solutions and methodologies**

Head-Quarters

Wilsonville, Oregon U.S.A.

- **300,000 Square Feet of Office & Laboratory Space**
- **4,350 Employees Worldwide**
  - **1,000 at Wilsonville, Oregon Headquarters**

Locations



Siloed Data

Challenge

- Streamline DW & business processes for implementation of Financial Data Warehouse
- Improve the data quality of the existing BI environment
- Standardizing product, finance and customer data across global locations

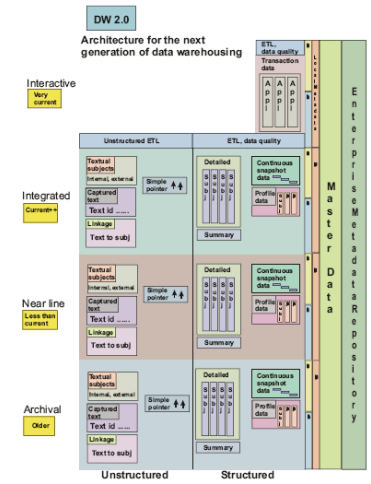
Solution

- Defining internal data standards
- Validating & standardizing information from Mentor offices around the world
- Providing centralized control while enabling local data analysts to ensure DQ to local & global standards
- Tracking DQ via dashboard process

Results

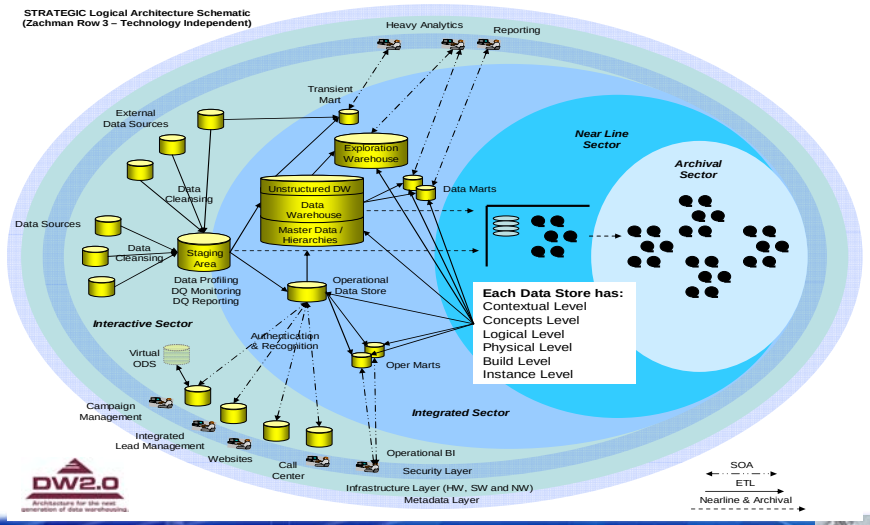
- Mentor utilized DW2.0 architecure & Data Quality processes to improve the data migrating into the BI environment
- Significantly mitigated the risk associated with production defects as a result of poor quality data
- Increased operational efficiency due to single reliable view of corporate data

Architecture Landscape



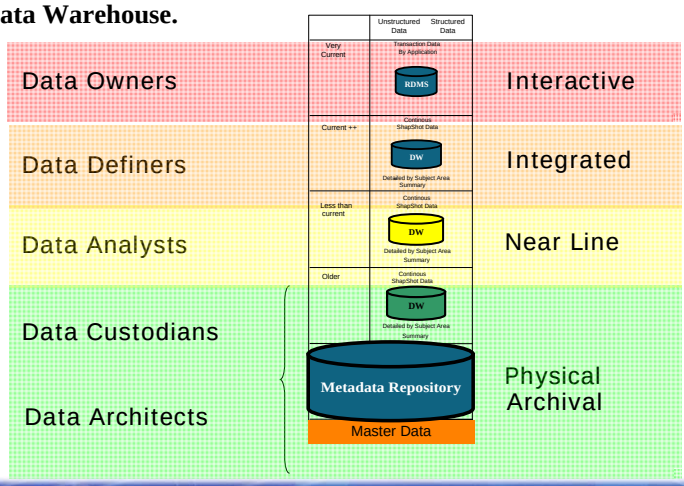
- **4 Sectors**
  - **Interactive**
  - **Integrated**
  - **Near line**
  - **Archival**

Database Landscape



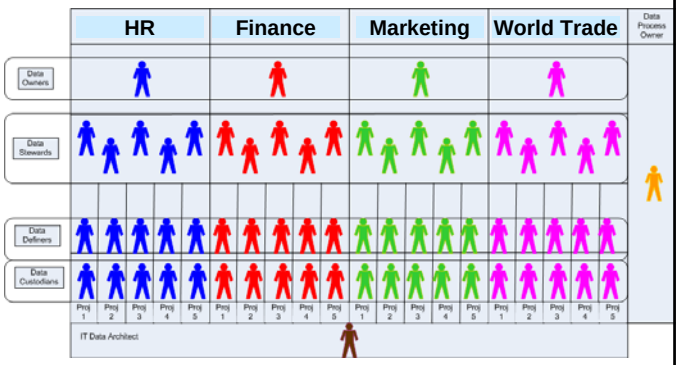
DQ Roles

- **Using the DW2.0 Architecture as a reference, we can define corresponding Roles and Responsibilities for the Data Warehouse.**



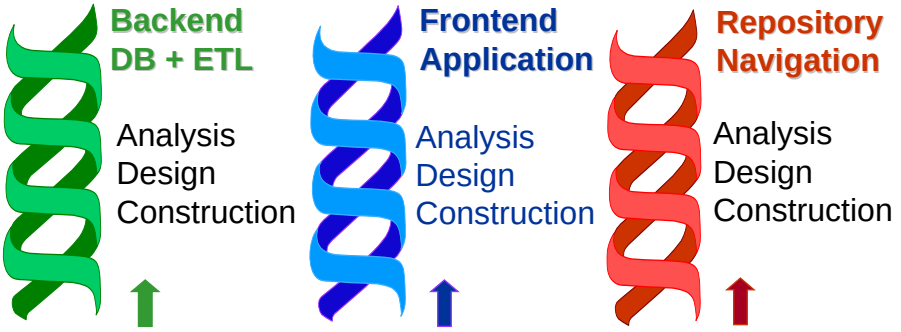
Govern Data

- **Data Governance Roles tied to Business Area**
  - **Data Process Owner**
  - **Data Owners**
  - **Data Stewards**
  - **Data Definers**
  - **Data Custodians**
  - **Data Architect**



Methods

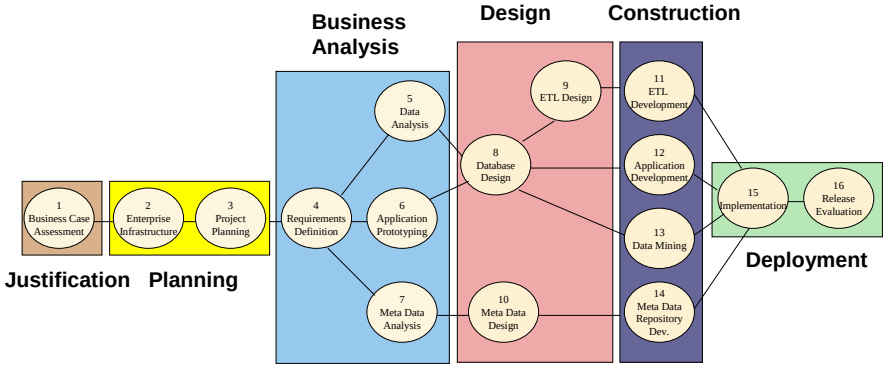
3 Spiral Parallel Development Efforts



After Larissa Moss

Methods

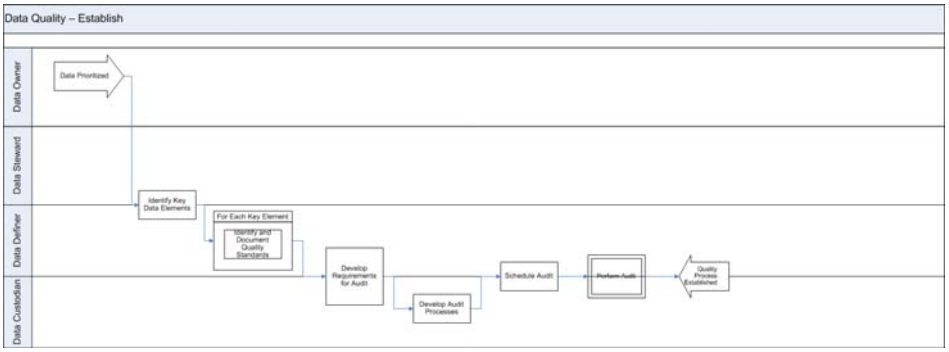
An Example of a Spiral Methodology – Development Steps



After Larissa Moss

Start up

Data Quality Process:



Audit

Data Quality Process:

Audit Scheduled

Run Audit

Assess Results

WHO *Data Definers and Data Custodians*

PROCESS *An audit is scheduled, initiating the audit process. This may be a regularly scheduled audit, an audit using newly defined metrics, of the result of a remediation effort.*

RESULT *Audit will be run*

WHO *Data Definers*

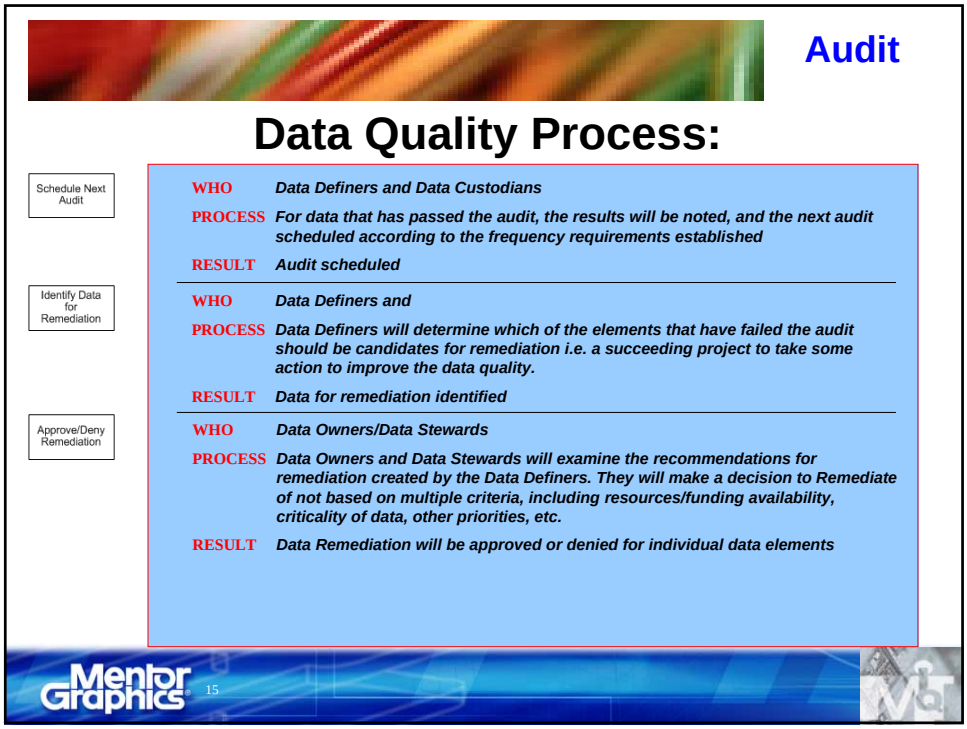
PROCESS *Data Definers will schedule the audit. This will mean running the audit code that was developed during the Establish process*

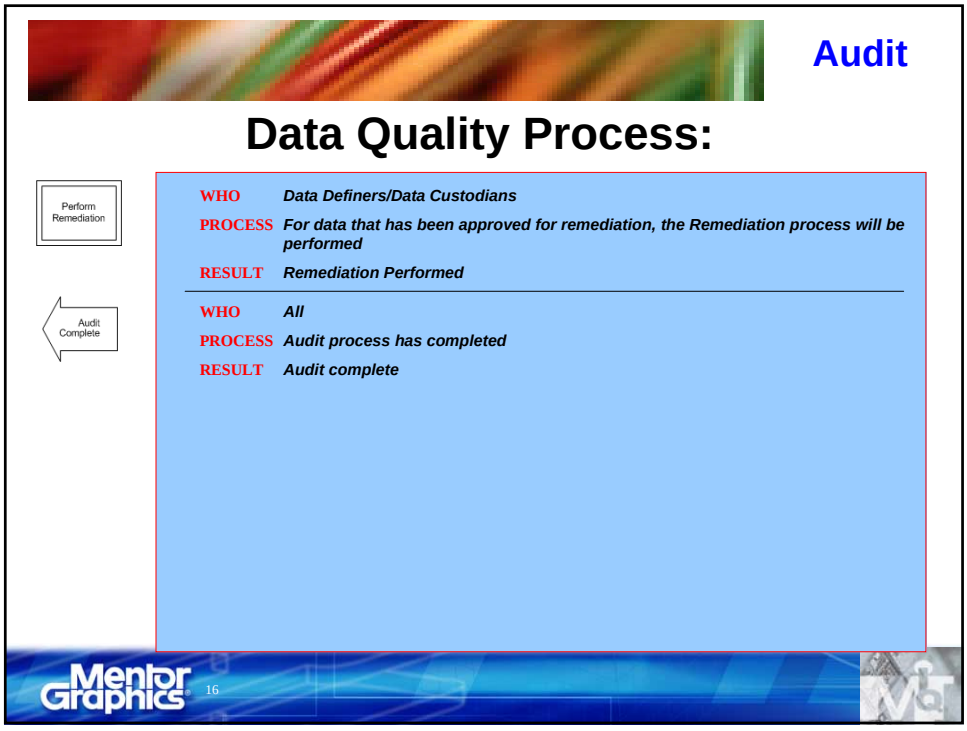
Deliverable *Audit report run*

WHO *Data Definers and Data Stewards*

PROCESS *The Data Definers and Data Stewards that have defined the quality metrics for the specific data elements will examine the quality results.*

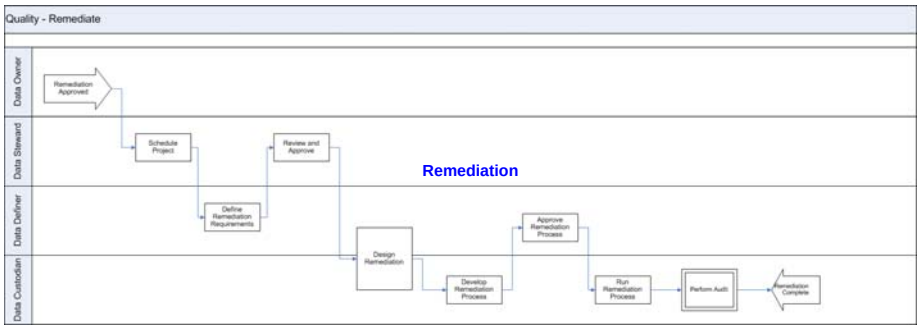
RESULT *Pass/Fail results for each data element*


Audit
Data Quality Process:
Schedule Next Audit
Identify Data for Remediation
Approve/Deny Remediation
WHO Data Definers and Data Custodians
PROCESS For data that has passed the audit, the results will be noted, and the next audit scheduled according to the frequency requirements established
RESULT Audit scheduled
WHO Data Definers and
PROCESS Data Definers will determine which of the elements that have failed the audit should be candidates for remediation i.e. a succeeding project to take some action to improve the data quality.
RESULT Data for remediation identified
WHO Data Owners/Data Stewards
PROCESS Data Owners and Data Stewards will examine the recommendations for remediation created by the Data Definers. They will make a decision to Remediate of not based on multiple criteria, including resources/funding availability, criticality of data, other priorities, etc.
RESULT Data Remediation will be approved or denied for individual data elements
Mentor Graphics
15


Audit
Data Quality Process:
Perform Remediation
Audit Complete
WHO Data Definers/Data Custodians
PROCESS For data that has been approved for remediation, the Remediation process will be performed
RESULT Remediation Performed
WHO All
PROCESS Audit process has completed
RESULT Audit complete
Mentor Graphics
16

Remediate

Data Quality Process:



Remediate

Data Quality Process:

Remediation Approved

WHO	***Data Owners***
PROCESS	***Data has been identified as needing quality remediation, and the Data Owners have approved the remediation***
RESULT	***Data Quality remediation project will be executed***

Schedule Project

WHO	***Data Stewards***
PROCESS	***Data Stewards will prioritize the data remediation project. In cases where there are cross-functional ramifications, Data Stewards from multiple functional areas will be involved in the prioritization***
RESULT	***Data remediation project will be scheduled***

Define Remediation Requirements

WHO	***Data Definers***
PROCESS	***Data Definers will examine the results of the Audit. For each data element that is to have quality addressed, they will define the requirements for improvement based on the results of the audit and the quality metrics that are to be applied to that data element. They will also recommend changes to any existing processes that will improve the quality (e.g. if the quality metrics say that the element is mandatory, but it is not a forced entry on the originating process(es), the Definers will recommend a change to the data entry process(es). Note that there may be multiple Definers from different functional areas involved in this step***
RESULT	***Quality remediation definitions will be defined***

Remediate

Data Quality Process:

Review and Approve

Design Remediation

Develop Remediation Process

Approve Remediation Process

WHO	*Data Stewards*
PROCESS	*Data remediation requirements have been defined. The data stewards will review the requirements, make suggested changes, and/or approve the requirements*
RESULT	*Requirements approved*
WHO	*Data Definers/Data Custodians*
PROCESS	*The Data Definers who developed the requirements will work together with the appropriate Data Custodians to design the processes that will be used to carry out the remediation*
RESULT	*Remediation processes defined*
WHO	*Data Custodians*
PROCESS	*The Data Custodians who were involved in the design process will develop the remediation processes based on the agreed requirements*
RESULT	*Remediation processes ready for approval*
WHO	*Data Definers*
Deliverable	*The Data Definers that were involved in defining the requirements will test and approve the developed code*
RESULT	*Remediation processes ready for application*

Mentor Graphics 19

Remediate

Data Quality Process:

Run Remediation Process

Perform Audit

Remediation Complete

WHO	*Data Custodians*
PROCESS	*The developed and approved data remediation processes will be scheduled and run*
RESULT	*Remediation processes applied*
WHO	*All*
PROCESS	*The audit process will be performed in order to assess the impact of the remediation*
RESULT	*Audit scheduled*
WHO	*All*
PROCESS	*The remediation process is complete. If the remediation was not successful, the audit process will identify further remediation*
RESULT	*Remediation complete*

Mentor Graphics 20

Questions