The MIT Information Quality Industry Symposium, 2008

Information Quality for Business Intelligence

Projects

Earl Hadden
Intelligent Commerce Network LLC

The MIT Information Quality Industry Symposium, 2008

Objectives of this presentation

- Understand Information Quality Problems on BI/DW Projects
- Define Strategic and Tactical Approaches to addressing Information Quality Problems
- Demonstrate how TDQM methods can augment BI/DW methodologies

The MIT Information Quality Industry Symposium, 2008

What's the Problem

- "Data quality is the most significant problem in our efforts to integrate information." Al Alborn, consultant to the Chief Architect, Department of Homeland Security
- The cost of non-quality is 5% of US GDP
- In service companies, information non-quality costs can cost up to 20% of gross revenue

The MIT Information Quality Industry Symposium, 2008

Defining Information Quality

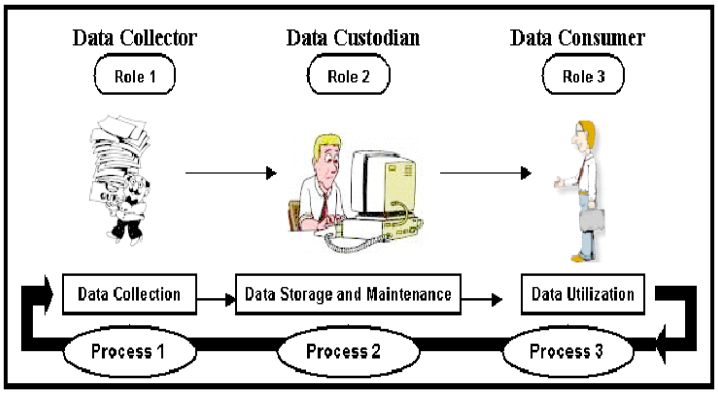
DQ Category	DQ Dimensions
Intrinsic **DQ**	Accuracy, Objectivity, Believability, Reputation
Accessibility **DQ**	Access, Security
Contextual **DQ**	Relevancy, Value-Added, Timeliness, Completeness, Amount of data
Representational **DQ**	Interpretability, Ease of understanding, Concise representation, Consistent representation

Top 3 BI/DW Information Quality Problems

1. Believability – international steel manufacturer with multiple production schedules
2. Completeness – health insurance provider with over 50% of claims records incomplete
3. Timeliness – multinational bank spent US$15 million on DW, warehouse is available on the 15th day after the close of the month, business information required by the 5th

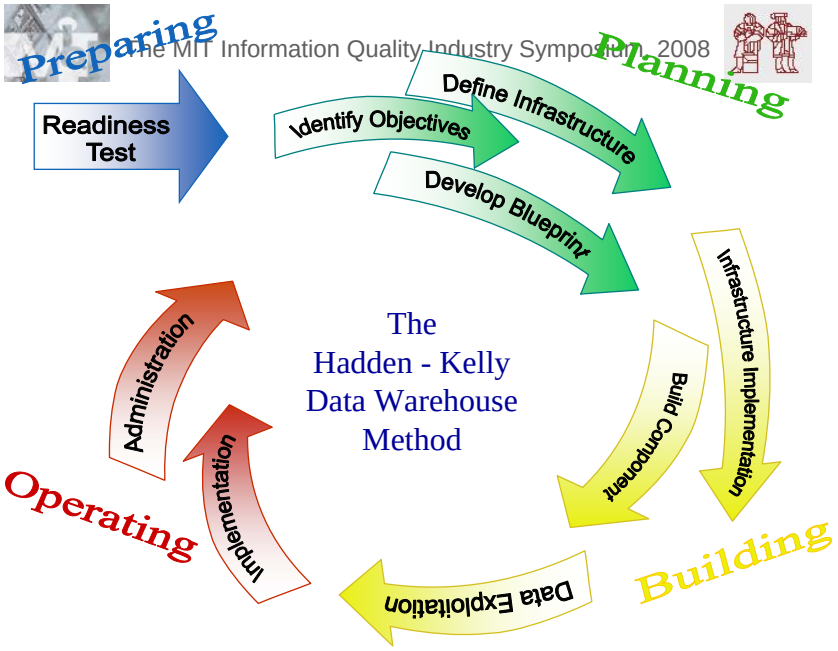
Data Quality Problems vs. Information Quality Problems

Where's the Problem?



Source: Prof. Richard Wang

The MIT Information Quality Industry Symposium, 2008



Choosing your IQ Path

- Enterprise-wide, executive sponsorship

  Advantages
  - Broad sponsorship
  - Organizations tend to “stay the course”

  Disadvantages
  - Hard sell
  - Expensive
  - Takes a long time to get measurable results

Choosing your IQ Path

- Subject based – “middle out”

  Advantages
  - Can be tied to a specific project with business goals, benefits
  - Eliminates a lot of time wasted on data of lesser importance

  Disadvantages
  - Hard to get business units not directly receiving value to participate (therefore limits value)
  - Adds time (and costs) to integration projects
  - Takes a long time to get measurable results

Choosing your IQ Path

- "Bottom up" – data cleansing

  Advantages
  - Limited to organization units directly involved in the project
  - Can be done in "stealth" mode

  Disadvantages
  - May create as many problems as it solves – multiple versions of the truth, conflicting rules…
  - Tends to get lost when the deadline approaches

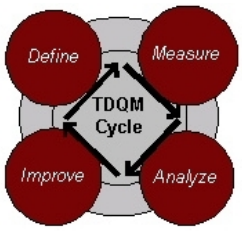
Choosing your IQ Path

- Do nothing

  Advantages
  - BI results match production system reports/queries

  Disadvantages
  - Lack of "believability" compromises use of the BI solution

Enterprise-wide – executive sponsor



- *Define* and establish data quality to be
  - A multi-dimensional concept beyond accuracy
  - Both objective and subjective
- Measure DQ with software tools such as
  - *Integrity Analyzer* and *Information Quality Assessment*
- Analyze DQ with models, methods & principles
  - Modeling Information Manufacturing Systems to deliver high-quality information products
- Improve

The MIT Information Quality Industry Symposium, 2008

Subject based – "middle out"

- Identify business objectives for the BI/DW project
- Identify organization units involved
- Identify other stakeholders interested in the outcome
- Identify information needed by the organization units and stakeholders to ensure the objectives are met

Subject based – "middle out"

- Establish IQ Environment (Policies, Roles & Responsibilities, etc.)
- Conduct preliminary information quality assessment
- Determine where the information is needed
- Identify technology that will be used to deliver the information
- Develop the project plan for the BI/DW implementation project

Subject based – "middle out"

"The only way to achieve integration is to work from a common data model."

-- John Zachman

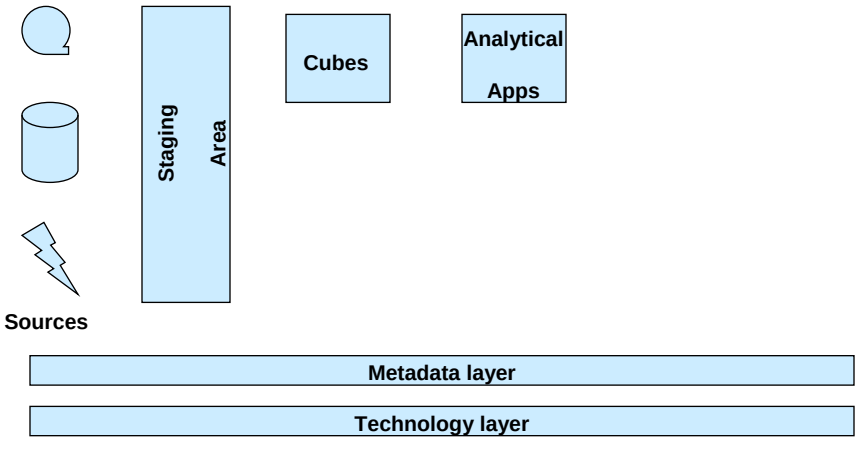
The MIT Information Quality Industry Symposium, 2008

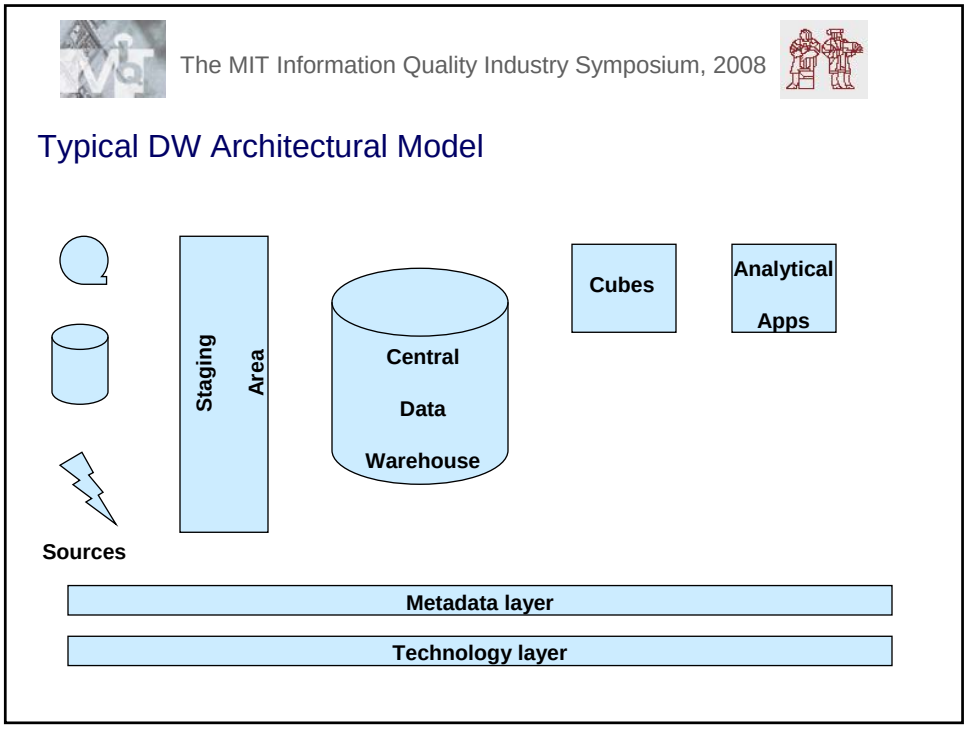
Subject based – “middle out”

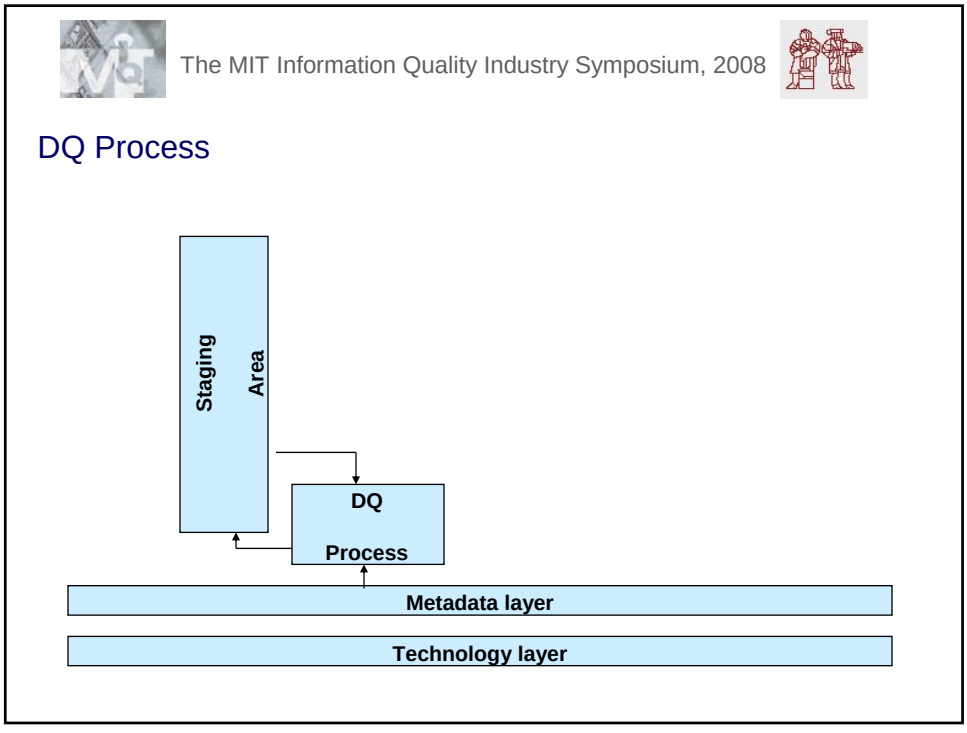
- Identify attributes required to provide desired information
- Define IQ standards for each attribute
- Perform source analysis for each attribute
- Establish sourcing logic (if there are multiple candidate sources)
- Define extract and transform specifications

The MIT Information Quality Industry Symposium, 2008

Typical BI Architectural Model




The MIT Information Quality Industry Symposium, 2008
Typical DW Architectural Model
Staging Area
Central Data Warehouse
Cubes
Analytical Apps
Sources
Metadata layer
Technology layer


The MIT Information Quality Industry Symposium, 2008
DQ Process
Staging Area
DQ Process
Metadata layer
Technology layer

Subject based – "middle out"

- Note quality failure, notify data steward, load non-quality data, indicate non-quality attributes to consumers
- Automated information quality correction, load corrected data
- Level 1 manual intervention – hold data until corrected, then load corrected data
- Level 2 manual intervention -- hold data until corrected, then load corrected data
- For each of these activities, notify Information Product Manager

Subject based – "middle out"

- Provide business metadata to consumers
- Provide training in IQ to consumers
- Audit compliance with IQ standards with information generated

The MIT Information Quality Industry Symposium, 2008

Questions?

Earl Hadden
Earl@Hadden-Kelly.com
(919) 593 1804