



Data Quality for Data Stewards

by Arkady Maydanchik and Dave Wells



Module 0. About the Course (7 min)

Module 1. Quality Basics (43 min)

- *Quality Defined*
 - What is Quality?
 - What isn't Quality
 - Quality vs. Qualities
 - Some Examples
- *Quality and Defects*
 - Defect Defined
 - Sources of Defects
 - Responding to Defects
 - Process Improvements
- *Quality Economics*
 - Cost of Quality
 - Cost of Defects
 - Cost of Quality Management
- *Quality Management Practices*
 - Quality Control
 - Quality Assurance
 - Quality Improvement Cycles
- *Quality Management Methodologies*
 - Statistical Process Control
 - SPC Control Charts
 - Total Quality Management
 - Six Sigma
 - Six Sigma Performance

Module 2. Data Quality Basics (52 min)

- *Data Quality Defined*
 - Defect Free
 - Conforming to Specifications
 - Suited to Purpose
 - Meet Customer Expectations
 - Defining Data Quality
 - Qualities of Data
- *Dimensions of Data Quality*
 - Data Quality is Multidimensional
 - Content and Quality
 - Structure and Quality
 - Time and Quality
 - Business and Quality
 - Usage and Quality
 - Presentation and Quality
- *Data Quality Projects*
 - Common Data Quality Projects
 - Assessment Projects
 - Data Cleansing Projects
 - Process Improvement Projects
- *Building Data Quality*



DS-02: Data Quality for Data Stewards

- Data Quality and System Architecture
- Data Quality and IT Projects
- Application Development Projects
- Data Conversion and Migration Projects
- Data Warehousing Projects
- Business Analytics Projects
- *Data Quality Programs*
- *Data Quality Profession*

Module 3. Common Causes of Data Quality Problems (29 min)

- *Introduction*
- *Manual Data Entry*
- *Data Integration*
 - Initial Data Conversion
 - System Consolidations
 - Batch Feeds
 - Real-Time Interfaces
- *Data Manipulation*
 - Data Processing
 - Data Cleansing
 - Data Purging
- *Data Decay*
 - Changes Not Captured
 - System Upgrades
 - New Data Uses
 - Loss of Expertise
 - Process Automation

Module 4. Introduction to Data Quality Assessment (59 min)

- *Why Assess Data Quality*
 - Role of Assessment in Data Quality Programs
 - Assessment Objectives
- *Business Value of Data Quality Assessment*
 - Planning Data Cleansing
 - Improving Data Collection Processes
 - Improving Data-Driven Business Processes
 - Supporting Data Migration and Data Integration
- *Data Quality Assessment Approaches*
 - *Data Validation against Trusted Sources*
 - *Sample Data Validation*
 - *Rule-Driven Approach*
- *Project Team*
- *Project Steps*
 - Defining the Project Scope
 - Loading the Data Staging Area
 - Gathering General Metadata
 - Data Profiling
 - Designing Data Quality Rules
 - Implementing Data Quality Rules
 - Fine-Tuning Data Quality Rules



DS-02: Data Quality for Data Stewards

- Building Data Quality Scorecard
- *Data Quality Rules Overview*
 - Attribute Level View of Data
 - Attribute Domain Constraints
 - Entity Level View of Data
 - Relational Integrity Constraints
 - Subject Level View of Data
 - Attribute Dependency Constraints
 - Advanced Business Rules
 - Time Dependent View of Data
- *Building Data Quality Scorecard*
 - School Report Card Example
 - Data Quality Scorecard
- *Recurrent Data Quality Assessment*

Module 5. Introduction to Root Cause Analysis (53min)

- *The Nature of Cause and Effect*
 - Correlation
 - Coincidence
 - Confounding Variables
 - Contributing Variables
 - Influence
 - Complexity and Complications
- *Cause and Effect Misconceptions*
 - Simple vs. Complex
 - Linear vs. Circular
 - Close vs. Distant
 - Immediate vs. Delayed
 - Ordered vs. Non-ordered
- *The Purpose of RCA*
 - Root Cause Analysis Defined
 - Root Cause Defined
 - Root Cause Criteria
 - Root Cause Criteria Examples
- *The Process of RCA*
 - Five Steps of RCA
 - The Problem Description
 - Data Gathering
 - Casual Modeling
 - Root Cause Identification
 - Recommendations
- *A First Look at Cause-Effect Models*
 - Five-Why's
 - Fishbone Diagrams
 - Causal Loop Models
- *Verifying Cause-Effect Conclusions*
 - A Short Reasoning Exercise
 - Exercise Review
 - Nonsensical Thinking
 - Nonsense and Distortion



DS-02: Data Quality for Data Stewards

- Seeing What We Want To See
- Logical Fallacies
- Rethinking Cause and Effect
- Systems Thinking
- Bias of Facts and Logic
- Evidence and Assumptions
- Lateral Thinking
- *Practical Application*
 - Beyond RCA
 - Implement Solutions
 - Measure and Monitor
 - Watch for Side Effects
 - Correct Existing Issues

Module 6. Ensuring Data Quality in Data Integration (41 min)

- *Data Integration Basics*
 - Data Integration Defined
 - Common Applications
 - Data Integration Approaches
 - Comparison of Integration Approaches
- *Data Quality Perspective*
 - Pros and Cons of Data Federation
 - Data Federation Quality Solution
 - Pros and Cons of Data Propagation
 - Data Propagation Quality Solution
- *Data Propagation Overview*
 - Propagation Frequency and Data Latency
 - Real-Time Interfaces
 - Batch Interfaces
 - Using Information Bus
 - Using Staging Area
 - Interface Spider Web Challenge
 - Information Integration Hub
- *Batch Interfaces*
 - Causes of Data Quality Problems
 - Data Quality Management
 - Batch Monitors
 - Change Monitors
 - Error Monitors