Information Quality: Principles and Foundations

Date: March 21-25, 2005

Course: ESD.IQ1 **Tuition:** $2,995

The course is designed to give participants the capability and fundamental skills to understand and overcome data quality challenges, and to deliver the benefits of improved information quality. Participants will learn how to develop and utilize data element maps and apply the principles of managing information as a product.

Day 1

Part I: Information Quality Knowledge (IQK)

Session 1
8:30 – 9:00
9:00 – 10:15

Introduction: Data and Information Quality Landscape
Welcome remark, administrative logistics, and get to know you. What is Information Quality? Why is Information Quality important? We will look at these two questions, discuss IQ concepts, and study principles for managing information as a product. We will assess and evaluate current and emerging trends in the data and information quality landscape

Session 2
10:30 – 12:00

Managing info as a Product
There are considerable advantages to treating data and information as a product. We will focus on the concepts and methods for managing information and data quality. We will examine the four principles of the concept of information as a product; 1. understanding consumers' information needs, 2. considering information as the product of a well-defined production process, 3. managing information as a product with a lifecycle, 4. appointing an information product manager to manage the information processes and resulting product. We will also examine a number of case studies that illustrate the benefits to be gained from managing information as a product.

Session 3
1:00 – 2:15

IQ Definition
Conventional wisdom often equates information quality with accuracy. MIT Research has shown that the definition of Information Quality is actually complex and multi-dimensional. We will explore these complexities, alternate definitions, and IQ characteristics and dimensions. We will provide an understanding of how to define and describe data and information quality beyond mere accuracy.

Session 4
2:30 – 3:45

IQ in Context
How do organizations define information quality? The definition of information quality and the types of IQ problems encountered vary according to context. We will discuss how different types of stakeholders in different contexts define, identify, and resolve IQ issues. We will explore common information quality patterns, look at several case studies, and discover how to map these typical 'potholes' and relate them to specific situations within your organization on the way to IQ improvement.

Session 5

4:00 – 5:15

What IQ Skills Matter?

This session is in the form of an interactive discussion. Participants will get the opportunity to share their insights, experiences and industry specific challenges. They will also be able to share and discuss potential solutions to information quality problems that they are facing.

Day 2

Session 6
9:00 – 10:15

IQ Measurement

How good is the quality of information in my company? How does it compare to others in my industry? How can I measure the quality of information? This session addresses these questions. We will look at types of IQ measurement such as subjective/perceptual and objective/quantitative. Furthermore, we will present categories of *IQ metrics that you can apply to develop for your own work.* Participants will gain an understanding in how to implement information quality metrics in practice.

Session 7
10:30 – 12:00

IQ Tools

We will explore the place of IQ tools in TDQM efforts (Define, Measure, Analyze, and continuously improve IQ). This session includes hands-on exploration of two IQ software tools:
1. Integrity Analyzer (IA)™ is used to analyze database integrity
2. Information Quality Assessment (IQA)™ is used to assess organizational information quality levels, organizational readiness for information quality initiatives, and organizational knowledge of information quality

We will examine and evaluate the usefulness of existing and emerging IQ tools.

Part 2 Data Management

Session 8

Overview and Motivation

1:00 – 2:15

To place process modeling in context, we will discuss the four phases of the systems development life-cycle (planning, analysis, design, and implementation) and examine process-modeling techniques.

Session 9
2:30 – 3:45

Process Modeling

In this session, we will conduct exercises applying the techniques we covered in TDM 1.

Session 10
1:45 – 3:00

Project Presentation

Participants will present their ideas of how the will utilize the principles and theories learned here to their organization.

Day 3

Session 11

Data Modeling

9:00 – 10:15

This session builds skills in translating business data requirements into conceptual designs using ER modeling. Additionally, we will discuss how data quality characteristics can be captured during the conceptual ER modeling phase, using a method developed at MIT TDQM Program. A step-by-step procedure for a course database will be exemplified. Participants will learn to create an ER data model using the modeling constructs of entity, attribute, and relationships.

Session 12
10:30 – 12:00

Normalization

We will examine placing tables in the 1st, 2nd, and 3rd normal forms. The data quality implications of normalization will be discussed. Participants will be able to convert a

relational table (a relation) from the 1st normal form to 2nd Normal form, and then to the 3rd normal form.

Session 13
1:00 – 2:15

SQL for Data Analysis

This session is designed to build skills in the construction of basic SQL statements for the **purpose** of data analysis. Participants will be able to construct basic SQL statements to query a database and understand the data you retrieve.

Part 3: Quality Processes & Measurements

Session 14
2:30 – 3:45

Improving IQ

We will discuss how Quality programs typically get started in organizations; we will cover principles common to Quality programs, selection of first IQ projects, identifying customer IQ specifications, selecting a quality metric, and collecting measures.

Session 15
4:00 – 5:15

Sampling to Measure

Sampling large data stores (e.g., millions or billions of rows) adequately poses some challenges. We will explore methods, as stated in the steps below, of overcoming these challenges.
1. Using statistical methods we will identify methods to use in Simple Random Sampling (SRS) with Proportional Allocation and understand the purpose of this type of sampling for IQ analysis
2. We discuss baseline assessments of data stores

Day 4

Session 16
9:00 – 10:15

Understanding Variation

We will cover frequency distributions, numerical &graphical techniques, the Empirical Rule and standard deviation, and common probability distribution patterns.

Session 17
10:30 – 12:00

Using Control Charts

We will discuss the Voice of the Process (VOP) vs. the Voice of the Customer (VOC), chance vs. assignable causes of variation, and meeting goals. We will also cover Control Charts: types, development, analysis, and use in IQ improvement.

Session 18
1:00 – 2:15

Tools & Techniques

We will explore tools such as check sheets, Pareto charts, Ishikawa (fishbone) diagrams, scatter plots, and defect concentration diagrams. We will also touch on some advanced techniques.

Session 19
2:30 – 3:45

After the Analysis—Where Do You Go from Here?

We will discuss changes as it applies to implementing, tracking, and documenting changes and cover techniques for measuring the cost vs. benefit of changes. We will also provide you with direction and tips on where to go for additional help.

Session 20

4:00 – 5:15

Change Management

" Changing a culture requires simultaneous, well-coordinated interventions in a system of complex and evolving patterns that is reinforced through its members' mutual adjustment and shared environment. In order to introduce an innovative change, make sure you understand the problem you are trying to solve." In this session we will discuss techniques that will help participants initiate change in their organization.

Day 5

Session 21

9:00 – 10:15

Lessons Learned and Directions

This program is designed to give participants the capability and fundamentals skills needed to understand and overcome data quality challenges. We will provide an opportunity to consolidate and distill all the lessons learnt during the week and clarify any concerns that participants have.

Session 22

10:30 – 12:00

Project Presentation

Participants will present their ideas of how the will utilize the principles and theories learned here to their organization.

Session 22

12:00 – 12:30

Ceremony, Feedback, & Discussion

Instructors

Richard Wang, Director of MIT Information Quality (MITIQ) Program at the Center for Technology, Policy, and Industrial Development (CTPID) and Co-Director for the Total Data Quality Management (TDQM) Program at MIT Sloan School of Management. Dr. Wang has served as a professor at MIT for a decade, a professor at the University of Arizona, Boston University, and a Visiting Professor at the University of California, Berkeley.

At MITIQ Program, Dr. Wang has developed a well-received information quality curriculum to certify practitioners for positions such as corporate IQ analyst, manager, and trainer. His outreach programs such as the MITIQ consortium extend Information quality principles and theories to intelligence quality in homeland security, information architecture in both public and private sectors, and intelligent commerce. At TDQM Program, Dr. Wang co-heads the innovative Corporate Household research that investigates relationships of business units within the firm and across organizational boundaries such as those in supply chains.

Dr. Wang has put the term Information Quality on the intellectual map with myriad journal and conference publications. His books on information quality include Quality Information and Knowledge (Prentice Hall, 1999), Data Quality (Kluwer Academic, 2001) and Journey to Data Quality (MIT Press, forthcoming). In 1996, Prof. Wang organized the premier International Conference on Information Quality at MIT, which he has served as the general conference chair, and currently Chairman of the Board. Dr. Wang received his Ph.D. degree from MIT. He can be reached at rwang@mit.edu, or (617) 739-7234, http://mitiq.mit.edu, http://www.iqconference.org

FOR FURTHER INFORMATION OR TO REGISTER, CONTACT:

MIT Professional Institute
Room 9-435
Cambridge, MA 02139
TEL: 617 253 2101 * FAX: 617 253 8042
e-mail: professional-institute@mit.edu
http://professional.mit.edu/

MIT PEP MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROFESSIONAL EDUCATION PROGRAMS