



Information Quality Management I: *Principles and Foundations*

May 2003

Course Description & Objectives

Organizations are increasingly aware of the overwhelming advantages of high-quality information. They are also painfully aware of the significant costs of low-quality information—costs that translate into hard dollars, reduced productivity, waste, and myriads of other consequences—even affecting quality-of-life.

As organizations' awareness of the importance of the quality of their information increases, and as information becomes more and more complex, there is a growing demand for Information Quality management (IQM) practitioners. Based on research findings and experiences with leading organizations conducted by the MIT Information Quality Program and Total Data Quality Management (TDQM) Program, this intensive course prepares participants with the fundamental knowledge and skills needed to recognize, solve, and prevent, IQ problems and risks.

IQM I readies participants for real-world IQ problem-solving. Participants are presented with foundational IQ knowledge, seen from a continuous improvement approach, IQ-related technical knowledge, and essential concepts, methods, and techniques for statistical and quality improvement processes as they relate to IQM. Participants are also prepared to deal with the special challenges of managing IQ projects.

This course is designed for, but not limited to, IQ team leaders, business analysts, data analysts, corporate planning and policy analysts, data warehousing specialists, data administrators, and project managers. It is also well suited for those who wish to gain in-depth experience of the leading research and practice in this field. Enrollment is limited in order to permit small group interaction while ensuring diversity.

➤ **Equipment**

Participants are required to bring a laptop with Excel® installed

➤ **Required Textbook**

For Part III: Quality Processes & Measurements (QPM):

Measuring Quality Improvement in Healthcare: A Guide to Statistical Process Control. Raymond G. Garey and Robert Lloyd. ASQ. 2001 (Available for purchase on-site)

➤ **Recommended Textbooks**

For Part II: Technical Data Management (TDM):

Any recent Systems Analysis and Design textbook can be used. (Chapters on the System Development Life Cycle and Process Modeling.)

Examples of recent ones are:

Modern Systems Analysis & Design by Hoffer, George and Valacich, 3rd Ed., Addison Wesley, 2002,
Systems Analysis and Design by Alan Dennis and Barbara Haley Wixom, 2nd Ed., John Wiley & Sons, 2003
Database Systems: Design, Implementation, & Management by Rob and Coronel, 5th Ed., 2002, Course Technology

NOTE: All other required reading will be supplied.



DAY 1 (May 19)

Session 1

8:00 – 8:30

8:30 – 10:00

RW

Part I: Information Quality Knowledge (IQK)

Continental Breakfast

Introduction, Managing Information as a Product

Welcome to the MIT Information Quality Management program, 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 also explore the role of the 'Information Product Manager' (IPM).

Session Objectives: We help you feel comfortable and go over the objectives and logistics of the weeklong, intensive course; you will be able to explain the concepts of 'information as a product', and be able to apply principles for managing information as a product.

Supplied Reading:

Chapters 1-2, *Quality Information and Knowledge* by Kuan-Tsae Huang, Yang Lee, and Richard Wang, Prentice Hall, 1999.

"Manage Your Information as a Product," *Sloan Management Review*, Summer 1998, Volume 39, No. 4. pp. 95-105. Richard Wang, Yang Lee, Leo Pipino, and Diane Strong.

Questions To Think About:

1. Who are some of the information consumers in your organization? How can you determine their specific information needs?
2. How do the 'product' and 'by-product' views of information differ? What are some of the consequences of each view?

Session 2

10:15 – 11:45

RW

Defining Information Quality (IQ)

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.

Session Objectives: You will be able to explain the relative merits of IQ definitions, list IQ characteristics and dimensions, and begin to apply the multi-dimensional IQ concepts to understand and frame problems you encounter in your work.

Supplied Reading:

Chapter 3, *Quality Information and Knowledge* by Kuan-Tsae Huang, Yang Lee, and Richard Wang, Prentice Hall, 1999.

"Beyond Accuracy: What Data Quality Means to Data Consumers." Wand. *Journal of Management Information Systems*, Spring 1996, 12:4. pp. 5-33. Wang, Richard and Yair

"Anchoring Data Quality Dimensions in Ontological Foundations." *Communications of the ACM*, November 1996. pp. 86-95, Volume 39, No. 4. pp. 86-95. Richard Wang, Yang Lee, Leo Pipino, and Diane Strong.

Questions To Think About:

1. What happens to data when they are not used? Why?
2. Which IQ dimensions does your organization struggle with the most? How do problems with these dimensions affect their work? (Please give specific examples.)

11:45 – 12:45

Lunch break



Day 1 con'd

Session 3

12:45 – 2:15

YL

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 identify typical 'potholes' on the way to IQ improvement.

Session Objectives: You will be able to explain varying IQ contexts, patterns, and potholes and identify what these map to in your organization.

Supplied Reading:

Chapter 3, *Quality Information and Knowledge* by Kuan-Tsae Huang, Yang Lee, and Richard Wang, Prentice Hall, 1999.

"Data Quality in Context," *Communications of the ACM*, May 1997. pp. 103-110. Diane Strong, Yang Lee, and Richard Wang.

"10 Potholes in the Road to Information Quality," *IEEE Computer*, August 1997, Volume 30, No. 8. pp. 38-46. Diane Strong, Yang Lee, and Richard Wang.

Questions To Think About:

1. What are three of the "Ten Potholes"? What are the consequences of 'falling into' them? Through examples, describe how problems evolved through different information production processes.
2. What are three of the 'Lessons Learned'? Give an example of each of these three.

Session 4

2:30 – 3:30

RW

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*.

Session Objectives: You will be able to explain different types of measurements and when to use each. In addition, you will be able to develop your own IQ metrics based on three general categories of measures.

Supplied Reading:

Chapters 4-5, *Quality Information and Knowledge*

"Data Quality Assessment," *Communications of the ACM*, April 2002. pp. 211-218. Leo Pipino, Yang Lee, and Richard Wang.

"Information Quality Benchmarks: Product and Service Performance," *Communications of the ACM*, April 2002. pp. 184-192. Beverly Kahn, Diane Strong, and Richard Wang.

Questions To Think About:

1. Define Codd's Integrity Rules. How is each important to IQ?
2. Give 2 examples each of subjective and objective measures. Why and when would you use each of these measures?

Session 5

3:45 – 5:00

RW

IQ Tools for Integrity Analysis & Information Quality Assessment

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



Day 1 con'd

Session Objectives: You will be able to explain the place of IQ tools and be able to use IA and IQA. A demonstration copy of IA will be handed out for the participants to use for hands-on experience; access to the IQA instrument will also be made available.

Supplied Reading:

Chapters 4-5, *Quality Information and Knowledge*

“AIMQ: A Methodology for Information Quality Assessment,” forthcoming in *Information & Management*, published by Elsevier Science (North Holland). Yang Lee, Diane Strong, Beverly Kahn, and Richard Wang.

Questions To Think About:

1. Define ‘column integrity’. What are the IQ implications of column integrity?
2. Define ‘user-defined integrity’. Give four examples of user-defined integrity rules as they apply to your organization.

6:00 – 9:00
RW, EP

DQ Project Workshop (optional)
Hotel@MIT

DAY 2 (May 20)

Session 6
8:30 – 9:45
RW, All

What Information Quality Management Skills Matter?

Participants will be asked to read job descriptions and working papers on information quality skills and curriculum prior to this lecture so they can engage in a lively discussion based on their own working experience and explore answers to questions, such as: What skills do IQM professionals need? How to classify different types of capabilities? What does research have to say? Practitioners? You?

Session Objectives: You will know what you need to further develop your IQM capabilities. You will be able to write up a job description for Information Quality Analyst, Information Quality Manager, Director, and Information Quality Trainer.

Part II: Technical Data Management (TDM)

Session 7
10:00 – 11:45
LP

Overview of Technical Foundations & Process Modeling Using DFDs

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 Objectives: You will be able to explain how/where process modeling fits in the systems development life cycle and explain how to model systems processes using DFDs and process modeling.

Supplied Reading:

TBD

Recommended Reading (not supplied):

Modern Systems Analysis & Design by Hoffer, George and Valacich, 3rd Ed., Addison Wesley, 2002, **Chapters 1 and 8.**

Systems Analysis and Design by Alan Dennis and Barbara Haley Wixom, 2nd Ed., John Wiley & Sons, 2003, **Chapters 1 and 6.**



Day 2 con'd

Questions To Think About:

1. Describe the classical System Development Life Cycle approach.
2. Given the following DFD, where are the errors in representation, that is, where are the symbols incorrectly applied?

11:45 – 12:45

Lunch break

Session 8

12:45 – 1:30

LP

Process Modeling Exercises

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

Session Objectives: You will be able to represent process models using DFDs and associated techniques.

Supplied Reading:

TBD

Recommended Reading (not supplied):

Modern Systems Analysis & Design by Hoffer, George and Valacich, 3rd Ed., Addison Wesley, 2002, **Chapters 8 and 9.**

Systems Analysis and Design by Alan Dennis and Barbara Haley Wixom, 2nd Ed., John Wiley & Sons, 2003, **Chapter 6.**

Questions To Think About:

1. Given a specification draw the DFD context diagram, and the level 0 and level 1 diagrams.
2. Represent the given specification using a Decision Table.

Session 9

1:45 – 3:00

LP

ER Data Modeling

This session builds skills in translating business data requirements into conceptual designs using ER modeling.

Session Objectives: You will be able to create an ER data model using the modeling constructs of entity, attribute, and relationships.

Supplied Reading:

TBD

Recommended Reading (not supplied):

Database Systems: Design, Implementation, & Management by Rob and Coronel, 5th Ed., 2002, Course Technology, **Chapter 3.**

Questions To Think About:

Given a specification, draw the ER diagram using the Chen notation.

Session 10

3:15 – 4:30

RW

ER to Relational Conversion and Quality ER Modeling

This session will examine converting ER models to relational models. 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.

Session Objectives: You will be able to convert an ER model. In addition, you will gain an appreciation of how data quality characteristics can be captured in an explicit manner in the database application development process.

Supplied Reading:

“Modeling Quality Requirements in Conceptual Database Design” by Veda Storey and Richard



Day 2 con'd

Wang

Questions To Think About:

1. Distinguish among data quality requirements, application requirements, and application quality requirements
2. Given a specification, draw the QER diagram.
3. Given an ER diagram, transform the diagram into a relation database model.

6:00 – 9:00
RW, LP

DQ Project Workshop (optional)
Hotel@MIT

Day 3 (May 21)

8:00 – 9:00

Continental Breakfast
DQ Project Workshop (optional)

Session 11
9:15 – 10:30
LP

Normalization

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

Session Objectives: You 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.

Supplied Reading:

TBD

Recommended Reading (not supplied)

Database Systems: Design, Implementation, & Management by Rob and Coronel, 5th Ed., 2002, Course Technology, **Chapter 4.**

Questions To Think About:

1. Given the following relations, specify whether or not they violate 1st, 2nd, and 3rd normal forms.
2. Given a relation, place the relation in a specific normal form (1st, 2nd, 3rd)

Session 12
10:45 – 12:00
LP

SQL for Data Analysis

This session is designed to build skills in the construction of basic SQL statements for the purpose of data analysis.

Session Objectives: You will be able to construct basic SQL statements to query a database and understand the data you retrieve.

Supplied Reading:

TBD

Recommended Reading (not supplied)

Database Systems: Design, Implementation, & Management by Rob and Coronel, 5th Ed., 2002, Course Technology, Chapter 5.

Questions To Think About:

1. Given the Premier Products data base, write the SQL query that will retrieve the customer's id, customer's name, and customer's balance and display it in decreasing order of balance.



Day 3 con'd

2. Given the Premier Products database, write the SQL query that will retrieve the customer's id and name, the name and address of the customer's sales representative.

12:00 – 1:30

Lunch Break
DQ Project Workshop (optional)

Part III: Quality Processes & Measurements (QPM)

Session 13
 1:45 – 3:15
 EP

Principles of Quality

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 Objectives: You will be able to explain Quality principles; you will be able to identify possible IQ projects you can start when you return to work; you will be able to begin development of a plan to select one (or more) quality metrics for your IQ project, to learn the quality characteristics most important to your customers, and to collect measures.

Required Reading (not supplied)

Measuring Quality Improvement in Healthcare: A Guide to Statistical Process Control. Raymond G. Garey and Robert Lloyd. ASQ. 2001, **Chapters 1 – 3.**

Questions To Think About:

1. How ready is your organization to undertake an information quality improvement project?
2. Has your organization collected any information quality data? If yes, has this data proven useful?

Session 14
 3:30 – 5:00
 EP

Introduction to Statistical Analysis & Variation

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

Session Objectives: You will be able to explain variation & differences in distribution patterns, develop a histogram, graph a time series, and understand in which context to use which technique.

Required Reading (not supplied)

Measuring Quality Improvement in Healthcare: A Guide to Statistical Process Control. Raymond G. Garey and Robert Lloyd. ASQ. 2001. **Chapter 4**

Supplied Reading:

Using Excel and Statistical Analysis to Understand Variation

Questions To Think About:

1. A questionnaire provides 58 yes, 42 no, and 20 no-opinion answers.
 - a. Construct a frequency distribution for the responses.
 - b. Construct a percent frequency distribution for the responses.
 - c. Construct a pie chart, a bar chart.
2. The following table shows the number of orders processed per month for the past year for a small manufacturing firm.

month	1	2	3	4	5	6	7	8	9	10	11	12
orders	19	27	20	16	18	25	22	24	17	25	15	17

- Construct a relative and percent frequency distribution.



Day 3 con'd

- Construct a histogram, a time series (run chart).
- Calculate the mean, median, mode, range, variance, and standard deviation for this data.

6:00 – 9:00
 EP, LP

DQ Project Workshop (optional)
 Hotel@MIT

Day 4 (May 22)

8:00 – 9:00

Continental Breakfast
DQ Project Workshop (optional)

Session 15
 9:15 – 10:30
 EP

Using Control Charts to Analyze Variation

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 Objectives: You will be able to explain the different types of control charts, know when to use which, and analyze control chart patterns. You will also be able to construct a control chart and to interpret it.

Required Reading (not supplied)

Measuring Quality Improvement in Healthcare: A Guide to Statistical Process Control. Raymond G. Garey and Robert Lloyd. ASQ. 2001, **Chapters 5 & 6.**

Supplied Reading:

Using Excel and Control Charts to Analyze Variation

Questions To Think About:

The following table shows the number of orders processed per month for the past year for a small manufacturing firm.

Month	1	2	3	4	5	6	7	8	9	10	11	12
Orders	19	27	20	16	18	25	22	24	17	25	15	17

- a. Construct a control chart for the individual observations (this type of chart is sometimes called an XmR chart).
- b. Analyze the pattern. Does the order process appear to be in control?

Session 16
 10:45 – 12:00
 RKH

Sampling Large Data Stores for Baseline Analysis

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 will cover SQL statements (DB2) used to randomly sample large data stores
3. We will learn how to create tables and load the sample data to these tables
4. We discuss baseline assessments of data stores

Session Objectives: You will be able to determine sample sizes for SRS with Proportional



Day 4 con'd

Allocation. You will be able to use specified SQL to create tables of sampled data in order to analyze the quality of the data therein. You will be able to conduct a baseline analysis of a data store.

Questions To Think About:

Background information:

Your company promotes various items (i.e., the items are 'on sale') in various markets. Items promoted in one market are not always the same as in another market. A given promotion can last from a day to several weeks.

In a data warehouse, you want to analyze the "Promotion" table, which has following columns:

PROMOTION_ID (PK)	PROMOTION_END_DT
PROMOTION_NM	NBR_ITEMS_SOLD
ITEM_ID (FK)	UPDATE_DT
PROMOTION_MARKET	LOAD_DATE
PROMOTION_EFFECTV_DT	

LOAD_DT = the date new data (e.g., new promotions, new markets, etc.) are loaded to this table.

UPDATE DT = the date new data about existing promotions are added to the table (e.g., when a particular promotion ends, a new row is added with the promotion end date.) The table is updated weekly. This table has 7,456,238 rows

1. For the purposes of SRS, by which column would you stratify the table and why that column?
2. Which 3 columns would you check for errors in a baseline assessment, and why?

12:00 – 1:00

Lunch Break

Session 17
1:15 – 2:30
EP

Tools & Techniques for SPC Problem-Solving

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 Objectives: you will be able to construct these charts/diagrams and explain the purpose of each one.

Required Reading (not supplied)

Measuring Quality Improvement in Healthcare: A Guide to Statistical Process Control. Raymond G. Garey and Robert Lloyd. ASQ. 2001. **Chapters 7 & 8**

Supplied Reading:

Using Excel and SPC Tools for Problem Solving

Questions To Think About:

Select a problem that you encounter often. (Ex. Being late for work or school.) How many SPC tools can you use to discover what is the underlying factors behind this problem?

- Histogram
- Process Flow Diagram
- Cause and Effect Diagram
- Control Chart
- Check sheet
- Pareto Diagram
- Scatter Plot

Session 18
2:45 – 4:15

Root Cause Analysis: Uncovering Hidden Interdependencies

Typically, finding root causes of IQ problems requires uncovering hidden interdependencies between organizational/business, information, and systems processes. We will explore how



Day 4 con'd
RKH

these complex interdependencies come about, issues, frameworks, and methods for uncovering these interdependencies and finding root causes in order to improve/redesign processes that cause poor-quality information. Concepts & methods studied include Systems Theory/Thinking, interviewing & contextual inquiry, mapping business & information processes, control barrier analysis, identifying relationships to uncover root causes, etc.

Session Objectives: You will be able to explain why root cause analysis is critical to IQ improvement & problem prevention and how organizational systems and, root cause analysis, and information processes relate to each other. Given a case study, you will be able to identify root causes of a systemic nature.

Supplied Reading:

Journey to Data Quality (pre-draft), by Lee, Pipino, Funk, and Wang, 2003, **Chapter 8**
System Dynamics and K-12 Teachers, a lecture at the University of Virginia School of Education by Jay w. Forrester. May 30, 1996. p.1, p.5 from (#4) to top of p.8; p.9 (#6) – p.13, 1st two paragraphs

“Hospitals Encourage Staff to Report Medical Errors,” by Laura Landro, The Wall Street Journal, March 25, 2003.

Questions To Think About:

Think of a persistent IQ problem in your organization. Devise a high-level plan to improve the quality of that information, in the context of your organization, using concepts and methods most appropriate to discovering root causes.

Session 19
4:15 –5:30
EP

After the Analysis—Where Do You Go from Here?

We will discuss Change Management 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 Objectives: You will be able to explain how to implement, track, and document changes. You will be able to measure cost vs. benefit and know where to go for additional resources.

Questions To Think About::

1. Can you quantify the costs and benefits of information quality?
2. How difficult is it in your organization to change how things are done?
3. What suggestions do you have for creating an information quality culture within your organization?

6:30 – 9:00
EP, RKH

DQ Project Workshop (optional)
Hotel@MIT

Day 5
(May 23)

8:00 - 9:00

Continental Breakfast
DQ Project Workshop (optional)



Part IV: IQ Project Management

Session 20
9:15 – 10:30
SD

Managing IQ Projects

After covering concepts of traditional project management, we will highlight some of the issues IQ projects face and explore how to modify conventional project management to address the needs of IQ projects.

Session Objectives: You will be able to develop a high to mid-level IQ project plan, taking into account specific issues and modifications needed for IQ projects.

Session 21
10:45 – 12:00
SD

Communications for IQ Professionals

IQM typically involves some type of change and therefore meets with varying degrees of resistance, depending on the organizational culture. We will explore effective means of communicating with stakeholders across the organization in order to eliminate or reduce resistance, to present a framework for understanding IQM, and to encourage participation in IQM processes. We will also discuss listening skills as part of the communication process.

Session Objectives: You will be able to listen to stakeholder concerns, effectively present a framework for understanding IQM (particular frameworks will be different for different people or functions), and communicate in a way that encourages cooperation and trust.

Session 22
12:00-1:30

Ceremony, Feedback and Lunch Break

Session 23
1:30 -4:00

Informal Discussion

- Preview of Introduction to IQM II: *Theory in Action*
- Potential Collaboration with MIT-IQ Program

Lecturers

Richard Wang (*RW*)
MIT-IQ Program
rwang@mit.edu

Elizabeth Pierce (*EP*)
Indiana University of Pennsylvania
empierce@iup.edu

Yang Lee (*YL*)
Northeastern University
y.lee@neu.edu

Leo Pipino (*LP*)
UMass Lowell
leo_pipino@uml.edu

Raïssa Katz-Haas (*RKH*)
Cambridge Research Group
raissa@crg2.com

Stanley Dobbs (*SD*)
Data Integrity Management Center
stanley_w_dobbs@icpphil.navy.mil