

Improving Data Warehouse and Business Information Quality

Methods for Reducing Costs and Increasing Profits

Larry P. English



Improving Data Warehouse and Business **Information Quality**

Methods for Reducing Costs and Increasing Profits



Larry P. English



Wiley Computer Publishing



John Wiley & Sons, Inc.

WEINHEIM · BRISBANE · SINGAPORE · TORONTO

Publisher: Robert Ipsen Editor: Robert M. Elliott Assistant Editor: Pam Sobotka Managing Editor: Marnie Wielage

Text Design & Composition: Benchmark Productions, Inc.

Designations used by companies to distinguish their products are often claimed as trademarks. In all instances where John Wiley & Sons, Inc., is aware of a claim, the product names appear in initial capital or ALL CAPITAL LETTERS. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

TQdM® is a registered trademark of INFORMATION IMPACT International, Inc.

This book is printed on acid-free paper. \otimes Copyright \odot 1999 by Larry P. English. All rights reserved. Published by John Wiley & Sons, Inc. Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in professional services. If professional advice or other expert assistance is required, the services of a competent professional person should be sought.

Library of Congress Cataloging-in-Publication Data:

English, Larry, 1947-

Improving data warehouse and business information quality: methods for reducing costs and increasing profits / Larry English.

p. cm.

Includes bibliographical references and index.

ISBN 0-471-25383-9 (pbk. : alk. paper)

1. Industrial management–Data processing. 2. Business–Data processing. 3. Data warehousing. I. Title.

HD30.2.E54 1999

658.4'038'0285574-dc21

98-33167 CIP

Printed in the United States of America. 10 9 8 7 6 5 4 3 2

Early Reviews for Larry P. English's Improving Data Warehouse and Business Information Quality

"The Information Quality Bible for the Information Age!

"Practical and useful. . . this book has it all in one package: 'concept book, textbook, reference book, practitioner's guide.'

"English's sense of humor is reflected throughout. The rewards from the implementation of his methods should be as enjoyable as the reading"

- Masaaki Imai Founder, Kaizen Institute
- Bud H. Cox
 Managing Director, Kaizen Institute of Japan (Kaizen is a Japanese word that connotes "continuous improvement involving everyone")

"This book is a must for every business bookshelf. Larry English has been on the forefront of the Data Quality issue from the outset. . . [and] has some real wisdom on this vital issue."

John Zachman
 Zachman International,
 Creator of the Framework for Enterprise
 Architecture

"This book is long overdue. As a leading expert on Quality in the world today, Larry English shows the impact that data and information quality directly have on costs and on profitability not just for data warehouses but also for business information. His examples are clear, and vital for management to read.

"This book will maximize your chances for success. No Data Warehousing project and no IT Department should be without it. I predict that it will become the 'Bible' of Quality success."

> — Clive Finkelstein Information Engineering Services Pty Ltd, "Father of Information Engineering"

"Everywhere we go. . . we see the results of data quality problems. In this book, Larry English not only turns up the heat by discussing the sources and nature of data quality problems, he also sheds real light through a practical approach to addressing data quality improvement. Time spent understanding and applying the principles and tips Larry offers will be well worth the investment."

— Vaughan Merlyn The Concours Group

"Very lively reading. The book belongs on the bookshelf of every manager and technician."

— Bill InmonPine Cone Systems,"Father of Data Warehousing"

Acknowledgments

First and foremost to Diane, my partner in business and in life, without whose encouragement, support, and patience, this book would never have become a reality.

To Beth and Bill, my Mom and Dad, who taught me about life, faith, and personal integrity. Isn't this the basis for any kind of improvement?

To Ashley and Chancellor, my children, whom I have seen mature, taking both backwards and forward steps, and from whom I have learned about the challenges of change.

To Clay Steakley, our administrative assistant, who diligently read every word of this book and provided valuable insights.

To Frank Dennis, an associate, who has provided insights on statistical methods and for proofing the section on random sampling.

And especially to the dozens of information professions whose organizations have allowed me the privilege of providing guidance in their journey toward information quality maturity, and from whom I have learned much.



"The best effect of any book is that it excites the reader to self activity."

-THOMAS CARLYLE

The state of information quality today is worse than it was five years ago, and it is getting worse day by day. In fact, the quality of information in many organizations is enterprise-threatening. Consider some of my recent experiences:

Having just returned from chairing my fourth Data Quality conference, the first in London in November 1998, my renewed excitement about information quality improvement as a trend was brought back to reality. The next week I keynoted conferences in Phoenix, Arizona, and Orlando, Florida. When I checked into the Sheraton Mesa Hotel in Phoenix, the registration clerk asked if I was checking in. I replied, "Yes, my name is English." She entered that into the computer, and asked if my first name was Ron. "No," I replied, "it is Larry." "I'm sorry, but we have no reservation for a Larry English," was her reply. "But that is okay, we have rooms." My confirmation number validated that the reservation was indeed for me, but under the name "Ron." The clerk replied, "that's no big deal, we can change it." "It is a big deal," I contended. "No it isn't," she insisted, probably thinking about the ease of making the change in the database, but not about the customer service aspect of the event. Seeing the history of the record, she asked if the last time I had stayed there was the 22nd of last month. No, I had not. When she gave me the printed copy of the registration form to sign, I discovered Arizona State University had just been "relocated" to my address in Brentwood, Tennessee!

I flew from there to Orlando, where registration went smoothly. I had a letter and fax waiting for me at the mail desk at the Omni Rosen hotel. I was pleased with the information that allowed me to pick them up then, rather than having to return after getting to my room and finding I had to go back to the mail desk. After I got to the room, I opened the fax, noting that only 8 of the 10 sent pages were stapled together and sealed in an envelope to

protect its privacy. The sealing was good; the quality control was not. There is an important reason why you document page counts on faxes: to assure all pages go through. Customer service is more than taking the pages and stuffing them in an envelope. Customer service is what The Hyatt in Phoenix did a month later when they received only 6 pages of an 11 page fax. They called my office, allowing them to re-send the missing pages immediately. This is information quality, comparing the information to reality and taking action to assure information customers have what they need when they are supposed to without having to chase it down. In Orlando, I had checked in after my office had closed, causing me to have to wait for the last 2 (most important) pages until the next day.

This book is not about information quality from an esoteric or theoretical standpoint. It is a practical book about using information quality as a business management tool for reducing the business and systems costs resulting from poor information quality. More importantly, this book is about increasing business profits and business effectiveness as a result of having higher-quality information and the customer satisfaction it generates.

The world is now experiencing a phenomenon that future historians may classify as The Golden Age of Information. The Oxford Dictionary defines *golden age* as a period during which commerce, the arts, and so forth, flourish. "Flourish" means that something is successful, very active, or widespread. With this litmus test, we are indeed living in the Golden Age of Information. Data warehouses that now contain multiple terabytes of data have forever altered the information processing landscape. The explosion of information on the Internet over the past few years confirms the importance and value of accessible information.\(^1\)

With this proliferation of information, the challenge of managing data and providing quality information has never been more important or more complex.

The premise of this book is that:

- Information and data are strategic enterprise resources.
- Quality information enables competitive advantage and business effectiveness.
- Information quality is not an isolated "function"; it is an inherent and integral part of business management.
- Everyone in the organization has a stewardship role for information quality.
- Without information the enterprise will fail, even process-driven organizations. For example, Coca-Cola has its formulas (information) required to produce its products. For the manufacturing firm Optical Fibres, "the

¹L. P. English, "The Golden Age of Information," DM Review, January 1997, p. 20.

key is in the chemicals and the formula for turning sand (silicon di-oxide) into optical fiber capable of transmitting laser signals for thousands of miles without losing information."

- Without *quality* information the enterprise will be suboptimized. In fact some will—and already have—fail.
- The costs of poor-quality information are high. Poor-quality information causes process failure and information scrap and rework that wastes people, money, materials, and facilities resources. The most significant problems caused by poor-quality information, however, is that it frustrates the most important resources of the enterprise—its people resources—keeping them from effectively performing their jobs, and it alienates its customer resources through wrong information about them and to them. Because there is a direct correlation between customer complaints and customer defection, the real cost of poor-quality information is in lost customer lifetime value, profits, and shareholder value.
- Information quality is free. When people ask, "what is the business case for making an investment in information quality improvement?" the answer is, "what is the business case for all the information scrap and rework caused by *not* having quality information?" It is the poor-quality information that costs money. The investment in improving both information product and information process quality is recouped multiple times in decreased costs and increased value of information to accomplish strategic business objectives.

QUALITY IS FREE

"Quality is free. It's not a gift, but it is free. What costs money are the unquality things—all the actions that involve not doing jobs right the first time.

"Quality is not only free, it is an honest-to-everything profit maker. Every penny you don't spend on doing things wrong, over, or instead, becomes half a penny right on the bottom line. If you concentrate on making quality certain you can probably increase your profit by an amount equal to 5 to 10 percent of your sales. That is a lot of money for free."

Who Should Read This Book

This book is not for everyone. It is for people who care about their customers and their information customers. This book is for people who do not like to see people and money resources wasted on information scrap and rework when they could be doing things that add value. This book is for people who seriously

²Philip B. Crosby, *Quality Is Free*, New York: Penguin Group, 1979, p. 1.

want to see shareholder value increase on a long-term basis, not merely from quarterly statement to quarterly statement.

If you do not want to rock the boat, make waves, or change your own behavior, please do not buy this book. This book is for people who are discontent with the status quo of their organization's practices in information management. If you are *reactive*, not *proactive*, this book is not for you. This book is for people ready to be change agents. If you are looking for a silver bullet or a magic panacea to solve your information quality problems, skip this book. It is for people ready to roll up their sleeves and make information quality happen.

This book is not just for companies in the private sector that face competitive pressures just to survive. It is for government and other not-for-profit organizations that desire to truly serve their constituents—or customers—and accomplish their mission. This book is for those who desire to provide quality services at the lowest cost.

You are a candidate to receive value from this book if:

- You recognize that information is an important business resource and you want to maximize its value.
- You care about your customers, both internal and external, and desire to maintain accurate information about them and for them.
- You are fed up with the high costs of low-quality information and the resulting problems, and are asking, "is there a better way?"
- You are a business person who requires quality information or who creates information and you don't just want to do your job, you want to do the *right* job, efficiently and effectively.
- You are an information systems professional and you don't just want to build applications or databases, you want to build applications and databases that *add value* to the business and to the end customers.

Who should read this book:

- Information quality managers and staff responsible for information quality processes.
- Data resource managers and staff, and those responsible for developing data models and databases that represent and house the enterprise knowledge resources.
- Data warehouse managers and staff responsible for data architecture, data acquisition, and cleansing, transforming and loading data into the enterprise's strategic knowledge base.
- CIOs and information systems managers responsible for application development processes who are responsible for creating and managing

the information infrastructure—not just the information technology infrastructure—for the enterprise.

- Systems analysts and designers who desire to add value to the business not just create technology "solutions."
- Business information stewards who are responsible for care taking of parts of the information resources for the enterprise.
- Business managers who are owners of processes that create information used by others outside of their business area. This book will be of special value to business managers of information-intensive business areas such as customer relationship management, marketing, sales, order entry, claims processing, customer service, accounting, accounts receivable, human resources, account management.
- Business personnel, such as business analysts, actuaries, and other knowledge workers who are intensive information customers.
- Senior management who are concerned about the high costs and low success of IT and who desire to deliver shareholder value. It is senior management who must understand the absolutes of quality as a management tool and who must establish a management environment that enables information quality to increase business performance.

Why You Should Read This Book

In the *Harvard Business Review*, Schaffe and Thompson cite a survey showing 63 percent of companies that had embarked on TQM-based programs had failed to improve quality defects in products by even as little as 10 percent.³ This book aims to help you understand how to avoid the pitfalls when conducting information quality improvements and when implementing an effective information quality environment.

The Gartner Group states that most reengineering initiatives will fail because of lack of attention to information quality. Experience is revealing that more than half of data warehouses built fail to meet expectations because of poor information quality. This book seeks to help you be successful in all information-related projects by addressing and solving the real problems and causes of poor quality information.

Making any kind of change to the status quo requires effort and work. Information quality is neither automatic nor easy. If it was, there would be minimal information quality problems today. Most of us need guidance in applying new skills. This book seeks to provide that guidance to minimize your risk in making information quality happen.

³Harvard Business Review, 1/92 volume.

Organization of This Book

This book is organized into four sections:

Part One, "Principles of Information Quality Improvement," deals with the fundamental principles of quality and of improving information quality.

Part Two, "Processes for Improving Information Quality," describes how to measure and improve information quality.

Part Three, "Establishing the Information Quality Environment," outlines how to implement an information quality environment.

Part Four, "Appendixes," provides an extensive glossary, recommended reading, and bibliography.

Part One: Principles of Information Quality Improvement

Part One describes the fundamental principles of information quality. They are not theory—they are very real and practical principles, even though they are foreign to many organizations. They provide the basis for understanding the background to information quality improvement as a management tool. Without understanding the principles of quality improvement, implementing the processes may be a hollow and empty exercise that performs the actions but lacks the soul. This may result in loss of motivation for any information improvement initiative, no matter how well intentioned.

Chapter 1, "The High Costs of Low-Quality Data," outlines the business case for information quality improvement. It describes why data that appears to be of satisfactory quality is, in fact, not. Examples highlight the high costs of low-quality data. Failure to solve information quality problems can be fatal to organizations.

Chapter 2, "Defining Information Quality," defines information quality, what it is and is not. Information quality is not a soft measure. It in fact can be quantified in bottom-line terms.

Chapter 3, "Applying Quality Management Principles to Information," describes the principles of quality in general: customer focus, continuous process improvement, and the use of scientific methods. It describes the concept of information as a product, and knowledge workers as information customers. We outline who has accountability for information quality. The answer may surprise you.

Part Two: Processes for Improving Information Quality

Part Two is the guide and road map of the processes to assess and improve information quality. It defines the processes of information quality improvement as a management tool for business performance excellence.

Chapter 4, "An Overview of Total Quality data Management (TQdM)," provides an overview of the TQdM (Total Quality data Management) methodology. It provides a thumbnail sketch of information measurement, assessment, and improvement processes. It further outlines a methodology for guidance in the data warehouse context.

Chapter 5, "Assessing Data Definition and Information Architecture Quality," outlines how to measure and assess the quality of data definition and information architecture. This represents the product "specification" of the information product. Without quality of information architectures that store the enterprise's knowledge resources, information quality will be much more difficult to achieve.

Chapter 6, "Information Quality Assessment," describes how to measure, analyze, and report information quality in databases, data warehouses, or produced by the business processes.

Chapter 7, "Measuring Nonquality Information Costs," describes the process of analyzing and quantifying the costs of poor information quality in business terms. It provides a road map for measuring the devastating impact poor-quality information has on business operations, mission accomplishment, customer satisfaction, and profits.

Chapter 8, "Information *Product* Improvement: Data Reengineering and Cleansing," describes the process of information product improvement; that is, reengineering and cleansing. It describes how to audit and control the extract, transformation and cleansing, and load processes for data warehousing.

Chapter 9, "Improving Information *Process* Quality: Data Defect Prevention," outlines how to improve the quality of the information product through business process improvement. It describes how to identify root causes of information quality problems, and how to plan and implement permanent information quality improvements.

Chapter 10, "Information Quality Tools and Techniques," describes the various categories of information quality tools and techniques that support the processes described in this section.

Part Three: Establishing the Information Quality Environment

Information quality improvement is not simply "scrubbing" data to put it into the data warehouse. Information quality is not simply auditing data to measure it. Information quality improvement *means* fundamental changes in how the information systems organization defines, develops, and delivers its products and services, and fundamental changes in how the enterprise plans, organizes, manages, and performs its business processes, and measures its business performance.

Sustainable information quality improvement will be accompanied by a change in the way people think about their information products and information "customers." Part Three describes the culture shift required to create a sustainable information quality environment.

Chapter 11, "The 14 Points of Information Quality," outlines Deming's 14 Points of Quality along with their direct ramifications for information quality improvement.

Chapter 12, "Information Stewardship: Accountability for Information Quality," describes Information Stewardship, the people roles and accountabilities for information products.

Chapter 13, "Implementing an Information Quality Improvement Environment," describes the steps to implement an information quality environment. It begins with how to conduct an information quality management maturity gap analysis and describes a set of steps to take from where you are.

Chapter 14, "Epilogue: Reaping the Benefits of Quality Information," concludes with an epilogue rather than a conclusion. The information quality journey will never bring you to a final destination. Rather, it will bring you incredible adventures and joys as you bring your organization into the *realized* Information Age.

Part Four: Appendixes

Appendix A is an extensive glossary that defines terms from the information management information, general quality, and statistical analysis domains.

Appendix B contains an extensive bibliography for further reading, beginning with a recommended starter set.

Internet Resources Available in Conjunction with This Book

Because of currency of information and space limitations, there are resources about information quality products, techniques, and information quality best practice case studies available for book holders at www.infoimpact.com under *Information Quality Resources*.

How to Use This Book

This book is a concept book, a textbook, a reference book, and a practitioner's guide. Depending on your work and interests, your use of this book may vary.

Do not assume that you must implement every step described in this book to implement information quality. There are many steps and activities that could be performed. You must identify your priority needs and concentrate on those processes and process steps (see Table I.1). Be eclectic and pragmatic.

There are many ways to use this book; however, shelfware is not one of them.

Table I.1 How to Use This Book

AREA OF INTEREST:	SEE BOOK SECTION:
What is information quality?	Part One
Principles of information quality.	Chapter 3 and Chapter 11
Cost justification for information quality.	Chapter 1, states it; Chapter 7 describes how to justify it
How to determine what level of maturity your enterprise has in information quality practices.	Chapter 13, with an example information quality management maturity grid, along with guidance in how to conduct a maturity assessment
How much is poor quality information is costing your enterprise?	Chapter 7
Developing a data warehouse with focus on identifying authoritative source databases, and cleansing and transforming data.	Chapter 8
How to perform an information quality assessment.	Chapter 6
You know you have information quality problems, but need to know how to eliminate them once and for all.	Chapter 9
What tools are available to assist in information quality and how to evaluate them.	Chapter 10, and Internet at www.infoimpact.com: Information Quality Resources
Developing a data model and creating a quality model.	Chapter 5
How to set up an information or data stewardship program.	Chapter 12
Best practices in information quality.	Chapter 9, last section
How to implement an environment for sustainable information quality.	Part Three
Case studies or organizations making information quality happen.	Throughout the book, especially in Chapter 9, Chapter 13, and on the Internet at www.infoimpact.com: Information Quality Resources

isolated applications and islands of information independent of or only loosely interfaced with other applications and databases.

Information quality problems have been masked by layer upon layer of interface programs in which inconsistent data of one application is "transformed" into usable data structure and values required by another application area. Organizations have accepted this as necessary. However, the validity of this approach is seriously challenged as the weight of those layers of interfaces consume the time and resources of information systems organizations to maintain them. One organization discovered that the equivalent of 120 to 160 of its 250 highly skilled application developers spend their careers maintaining programs that simply copy data from one database, transform it, and put it into another database. Is this a valuable use of developer skills? Hardly, especially considering the people are a *consumable* resource and data is a *reusable* source.

Any new process takes time and experimentation to mature. Agricultural processes continue to be improved to provide greater quality of crops. For example, genetic alteration in foods like tomatoes have extended shelf life and increased taste. The Japanese began improving manufacturing process quality in the early 1950s when Dr. W. Edwards Deming introduced Statistical Quality Control techniques. This led to the dramatic turnaround of the postwar Japanese economy and the quality revolution and maturing of the Industrial Age manufacturing processes. It was not until the 1980s and the Japanese quality invasion of America that American business began transforming and improving American industry's manufacturing processes.

Fifty years into the Information Age we are now seeing the same quality improvement techniques being applied to information as the product of business processes. We are now seeing the maturation of the information management processes and the dawn of the *realized* Information Age. In the Realized Information Age, quality information indeed becomes the new economic currency and the competitive differentiator.

Information Quality Improvement: Beyond Data Cleansing

This entire book addresses not just the techniques for cleansing data for the data warehouse. It addresses a complete set of processes to attack information quality *problems*. Before an organization can significantly improve its information quality, it must understand the paradigms of information as a business resource and as a product. Information quality improvement seeks to measure information quality, both data definition (data specification) and data content; analyze and identify root cause of data defects; and improve processes to prevent defective data. The sole reason for improving information quality is to improve business efficiency and effectiveness and end-customer satisfaction

by eliminating the problems caused by nonquality data. This book addresses the components of mature information management processes and organization culture that embody a customer satisfaction mind set to provide quality information. We will also identify the fundamental changes that are required to move an organization from data as a neglected and proprietary resource to information as a strategic and open business resource for competitive advantage and to achieve information maturity.

The information quality movement signals the beginning of the maturing of the Information Age. A word of warning is due the reader: Information quality improvement does not mean "more of the same" way of doing business. After all, we have been successful in the past. It does not mean building the same kinds of systems faster. It does not mean building bigger databases faster. Information quality improvement will force management to rethink the way it builds (or buys) applications and databases. It will force management to rethink the relationship of business processes and information. It will force management to rethink how it performs work. It will finally force management to rethink its performance measures and accountabilities for the resources of the enterprise.

Author's Warranty

If you are not able to apply ideas contained in this book to achieve value to your organization worth multiple times the cost of the book, I will personally refund to you the purchase price you paid for this book, no questions asked. Simply contact me at Larry.English@infoimpact.com for refund instructions as to where to send the book. All I ask is for you to give me a copy of your sales receipt along with a statement of what you tried that did not work, as well as your assessment of why it failed to result in value. No further questions asked.