dandelon.com
© 2008 AGI-Information Management Consultants
May be used for personal purposes only or by
libraries associated to dandelon.com network.

Tourney to Data

Yang W. Lee, Leo L. Pipino, James D. Funk, and
Richard Y. Wang

The MIT Press
Cambridge, Massachusetts
London, England

t

Contents

Preface ix
Acknowledgments xi

1 Introduction 1
Can Information Be Shared? 2
A New System Is Not the Answer 3
Starting the Data Quality Journey 5
Stories of Successful Beginnings 6
The CEO Leads the Journey 7
The Challenges of the Journey 8
Data Quality: Why Is It Important? 9
Overview of the Book 10

2 Cost/Benefit Analysis 13
The Challenge 13
The CostyBenefit Trade-off 15
A Case Example 18
Further Cost/Benefit Analysis Techniques 21
Concluding Remarks 25

3 **Assessing Data Quality, Part I 27**
Assessment Techniques and Associated Approaches 28
Assessment Approaches in Practice *i23*
Gap Analysis Techniques 34
Data Integrity Assessment 38
Concluding Remarks 39
Appendix: Information Quality Assessment (IQA) Survey 40

I

4 Assessing Data Quality, Part II 53
Codd Integrity Constraints 53
Metrics for Data Quality 55
Automating the Metrics 60
Process-Embedded Data Integrity Approach 61
Concluding Remarks *66*

5 Sampling for Data Quality Assurance 67
Fundamental Concepts 67
Selecting a Sampling Procedure 69
Determining the Sample Size 71
Performing Sampling in a Transaction Database Environment 72
Expanding the Environment: Distributed Databases and Data Warehouses 77
Concluding Remarks 78

6 Understanding the Anatomy of Data Quality Problems and Patterns 79
Ten Root Conditions of Data Quality Problems 80
Manifestation of Data Quality Problems 91
Transformations of Data Quality Problems 106
Concluding Remarks 108

7 **Identifying Root Causes of Data Quality Problems: A Case of a Health Care Organization 109**
Case Study: Feelwell Health Systems 109
Identifying Problems 110
Building a Cross-Functional Team 114
Adopting a Framework: Building and Testing Hypotheses 114
Key Information 115
Uncovering the Causes of Data Quality Problems 116
Concluding Remarks 123

8 Managing Information as Product ^125
Information Product 125
Four Cases 126
Four Principles 128
Managing Information as By-product Will Not Work 131
Concluding Remarks 135

I

9 Developing Information Product Maps 137
Basic Definitions, Concepts, and Symbols 137
Procedure to Generate an IP-Map 140
Building the IP-Map: A Case Example 143
Concluding Remarks 149

10 Data Quality Initiatives: A Case of a Large Teaching Hospital 151
Case Study: LTH Health System 151
Data Submission Improvement Project 155
Information Product Maps 159
Improvements Initiated: Current Process and Future Plan 165
Concluding Remarks 166

11 Data Quality Policy 171
Ten Policy Guidelines 172
Concluding Remarks 182
Appendix 1: Job Descriptions of Data Quality Positions 183
Appendix 2: Data Architecture Policy Example from a Global Manufacturing Company 189
Appendix 3: Data Quality Practice & Product Assessment Instrument 191

12 **End of the** Journey? **201**
Recapitulation 201
Future Challenges and General Concerns 203
Formally Capturing the Quality Characteristics of Data 205
Corporate Householding 208
Data Mining / 210
Data Integration 212
Security 213
The Wired and Wireless World 213
A Final Observation 214

References 215
Index 223