

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Facilitating Evolution in Relational Database Design:

A procedure to evaluate and refine novice
database designers' schemata

A thesis presented in partial fulfilment of the requirements for the degree of
Master of Business Studies in Information Systems
at Massey University

Michael Robert Ryder

1996

Acknowledgments

I feel the desire to thank several people for the support they have shown me throughout the duration of this research.

Many thanks go to Clare Atkins for continually sparking my enthusiasm for issues surrounding database and data modelling, for acting as a second researcher and for her unyielding support and proofreading of this thesis. Also to Wen van Kersbergen from the Amsterdam School of Business for his expert knowledge and assistance.

Thanks also go to Jon Patrick for his advice, guidance and support throughout the process of conducting this research. To the survey respondents and Angus, I also extend my gratitude.

To the staff of the Information Systems Department, thank you for your interest in the thesis's progress and your patience whilst sitting through the seminars it has given rise to.

Lastly, to my friends and family, your perseverance and patience and encouragement is greatly appreciated.

“A little learning is a dang’rous thing;
Drink deep, or taste not the Pierian spring:
There shallow draughts intoxicate the brain,
And drinking largely sobers us again.”

(Alexander Pope)

Ars longa, vita brevis

Abstract

Relational database management systems (RDBMS) have become widely used by many industries in recent years. Latterly these systems have begun to expand their market by becoming readily available at minimal cost to most users of modern computing technology. The quality of applications developed from RDBMSs however is largely dependent upon the quality of the underlying schema.

This research looks at the area of schema design and in particular schemata designed by people who have a minimal understanding of relational concepts. It uses a survey and case studies to help define some of the issues involved in the area. A procedure to modify existing schemata is described, and the schema from one of the case studies used to apply the schema re-design procedure to a real database design. The results are compared to the original schema as well as a schema designed using a conventional application of the NIAM analysis and design methodology.

The research supports the hypothesis that database applications based on schemata designed by lay-persons are currently being used to support business data management requirements. The utility, reliability and longevity of these applications depend to some extent on the quality of the underlying schema and its ability to store the required data and maintain that data's integrity. The application of the schema re-design procedure presented in this thesis reveals refinements on the original schema and provides a method for lay-persons to evaluate and improve existing database designs.

A number of issues and questions related to the focus of this research are raised and, although outside the scope of the research, are noted as suggestions for further work.

Table of Contents

ACKNOWLEDGMENTS	III
ABSTRACT	V
SECTION ONE	1
1 INTRODUCTION	3
1.1 DBMS Utilisation Environments.....	5
2 THE RESEARCH PROGRAMME.....	11
3 SCOPE OF THE STUDY	15
3.1 Reverse Engineering of Database Schemata.....	18
3.2 Data Modelling	18
SECTION TWO	21
4 CONTEXT.....	23
4.1 The Position of the Relational Model.....	23
4.1.1 Hierarchical Model.....	24
4.1.2 Network Model	25
4.1.3 Relational Model.....	26
4.1.4 Relations.....	28
4.1.5 Normalisation.....	30
4.1.6 Sub-language.....	31
4.1.7 Operations on relations.....	33
4.1.8 RMT.....	33
4.2 Other Data Models.....	34
4.2.1 The entity relationship model (ERD).....	36
4.2.2 Extended entity relationship models (EER).....	36
4.2.3 Relational Model replacements.....	37
4.3 Relevance to novice designers	39
4.3.1 Reverse engineering	40
4.4 Data Modelling	41
4.4.1 Conceptual modelling.....	41
4.4.2 Logical modelling.....	43
4.4.3 Physical modelling.....	43
4.5 The NIAM Methodology.....	43
4.5.1 CSDP Step 1: From examples to elementary facts and quality checks.....	44
4.5.2 CSDP Step 2: First draft of conceptual schema diagram and population check.....	46
4.5.3 CSDP Step 3: Eliminate unnecessary schema and find derived fact types.....	48
4.5.4 CSDP Step 4: Add uniqueness constraints.....	49
4.5.5 CSDP Step 5: Arity and logical derivation checks.....	51
4.5.6 CSDP Step 6: Add additional constraints.....	51
4.5.7 CSDP Step 7: Entity identification.....	54
4.5.8 CSDP Step 8: Add additional constraints.....	55

4.5.9 CSDP Step 9: Completeness and succinctness checks.....	55
4.6 Mapping from conceptual to implementation models.....	56
4.7 Optimal normal form algorithm.....	58
SECTION THREE.....	63
5 DESIGNER SURVEY.....	65
5.1 Sample selection.....	65
5.2 Survey response rate.....	67
5.3 Questionnaire structure.....	68
5.4 Discussion of survey results.....	69
5.5 Conclusions drawn from the survey results.....	73
SECTION FOUR.....	75
6 CASE STUDIES.....	77
6.1 Case A. Socks and All.....	78
6.2 Case B. Public Reservations.....	80
6.3 Case C Apples and Pears.....	82
6.4 Case D Communications Breakdown.....	86
6.5 Questions to be answered from the case studies.....	90
6.6 Results obtained.....	95
6.7 Summary.....	97
SECTION FIVE.....	99
7 SCHEMA RE-DESIGN PROCESS (SRDP).....	101
7.1 Stage one.....	102
7.2 Stage two.....	104
7.3 Stages three and four.....	105
7.4 Stage five.....	108
7.5 Stage six.....	108
7.6 Stage seven.....	109
8 THE ADMINISTRATION DATABASE EXAMPLE.....	111
8.1 The original database schema.....	112
8.1.1 Stage 1: Transformation of existing schema to populated diagram with standard predicates.....	112
8.1.2 Stage 2: Addition of meta-data from the database to explicitly show constraints.....	117
8.1.3 Stage 3: Creation of example sentences.....	120
8.1.4 Stage 4: Verification of example sentences by universe of discourse expert.....	122
8.1.5 Stage 5 - Incorporation of additional semantic constraints.....	123
8.1.6 Stage 6 - Generation of new logical model and completeness check.....	124
8.1.7 Stage 7 - Generation of new physical schema (relational).....	128
8.2 Discussion and comparison with associate researchers' results.....	130

SECTION SIX 133

9 MARXIST EXPERIMENT 135

 9.1 *Experiment methodology* 135

 9.2 *Experiment Design* 137

 9.2.1 Stage 1 - Transformation of existing schema to populated diagram with standard predicates.138

 9.2.2 Stage 2 - Addition of meta-data from the database to explicitly show constraints.....143

 9.2.3 Stage 3 - Creation of example sentences.153

 9.2.4 Stage 4 - Verification of example sentences by universe of discourse expert.....153

 9.2.5 Stage 5 - Incorporation of additional semantic constraints.154

 9.2.6 Stage 6 - Generation of new logical model (NIAM type diagram).....154

 9.2.7 Stage 7 - Generation of new physical schema (relational).....160

 9.3 *Associate researcher’s activities*..... 164

 9.3.1 Comparison and discussion.....168

SECTION SEVEN 171

10 CONCLUSIONS 173

 10.1 *Recommendations for further research* 174

SECTION EIGHT 177

APPENDIX ONE *SURVEY RESULTS* 179

APPENDIX TWO *SAMPLE QUESTIONNAIRE*..... 183

APPENDIX THREE *FRAME-WORK FOR CASE STUDY INTERVIEWS* 189

APPENDIX FOUR *EXAMPLE SENTENCES AND OBJECT DEFINITIONS* 191

APPENDIX FIVE *REVISED GENERAL LEDGER (CREDIT) EXAMPLES* 197

APPENDIX SIX *ADMINISTRATION SCHEMA DIAGRAMS AND TABLE STRUCTURES*..... 199

APPENDIX SEVEN *EXAMPLE SENTENCE SETS WITH VERIFICATION*..... 219

APPENDIX EIGHT *PRODUCT ACKNOWLEDGMENTS* 229

SECTION NINE 231

REFERENCES 233