PRACTICAL DATA MIGRATION
Second edition

Johny Morris
PRACTICAL DATA MIGRATION
Our mission as BCS, The Chartered Institute for IT, is to enable the information society. We promote wider social and economic progress through the advancement of information technology science and practice. We bring together industry, academics, practitioners and government to share knowledge, promote new thinking, inform the design of new curricula, shape public policy and inform the public.

Our vision is to be a world-class organisation for IT. Our 70,000 strong membership includes practitioners, businesses, academics and students in the UK and internationally. We deliver a range of professional development tools for practitioners and employees. A leading IT qualification body, we offer a range of widely recognised qualifications.

Further Information
BCS The Chartered Institute for IT,
First Floor, Block D,
North Star House, North Star Avenue,
Swindon, SN2 1FA, United Kingdom.
T +44 (0) 1793 417 424
F +44 (0) 1793 417 444
www.bcs.org/contact
To Josephine, with my love, for being there.
Figures and tables | xi
Author | xii
Acknowledgements | xiii
Abbreviations | xiv
Glossary | xv
Useful websites | xviii

**INTRODUCTION**
What is the purpose of this book? | 1
What types of data migration are covered? | 1
What is not covered in this book? | 1
Who is this book aimed at? | 2
What is new to version two of *Practical Data Migration*? | 3
Paragraph styles | 3

**SECTION 1: EXECUTIVE OVERVIEW**

1 **DATA MIGRATION: WHAT’S ALL THE FUSS?**
   What is data migration? | 7
   First the bad news | 8
   So what usually goes wrong? | 9
   The responsibility gap | 10
   How to avoid the responsibility gap | 15

2 **GOLDEN RULES AND SUPER SMART TASKS**
   Golden Rules: the most important lessons | 16
   Golden Rule 1 | 17
   Golden Rule 2 | 23
   Golden Rule 3 | 24
   Golden Rule 4 | 26
   Super SMART Tasks and the mid-change cycle slump | 28

3 **PDMv2 OVERVIEW**
   Introducing *PDMv2* | 35
   Introducing the technology | 40
   Technology overview | 41
   The case for specialist technology | 43
   *PDMv2* provides all the answers | 44
## CONTENTS

### 4 CREATING A MIGRATION STRATEGY

- Starting out on the right foot
- *PDMv2* data migration strategy
- What is in a *PDMv2* data migration strategy document?
- Data migration strategy checklist
- Getting the data migration strategy implemented

### SECTION 2: TOOLS AND TECHNIQUES

### 5 METADATA AND KEY BUSINESS DATA AREA DECOMPOSITION

- Metadata and *PDMv2*
- Uses of data models in *PDMv2*
- Instantiating data models
- Developing a conceptual entity model
- Project decomposition and key business data areas

### 6 DEMILITARISED ZONES AND KEY DATA STAKEHOLDER MANAGEMENT

- The impact of the demilitarised zone
- Introduction to key data stakeholders

### 7 LANDSCAPE ANALYSIS

- What is landscape analysis and why do you do it?
- Legacy data stores
- Landscape analysis as a Super SMART Task
- When to perform landscape analysis
- Documenting legacy data stores
- Profiling, data gaps and master data management
- Software tools to support landscape analysis
- Impact of the DMZ

### 8 SYSTEM RETIREMENT PLAN

- Why have a system retirement plan?
- What are system retirement plans?
- How and when to implement SRPs
- SRP minimum requirements
- SRPs and Super SMART Tasks
- Impact of the DMZ on SRPs

### 9 DATA QUALITY RULES

- Introducing data quality rules
- How do you generate DQRs?
- The DQR process
- How a DQR board works
- Software support
- Role of the DMZ

### 10 GAP ANALYSIS AND MAPPING

- Gap analysis
- Gap analysis as a service
- Data quality monitoring as a service
### CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping</td>
<td>163</td>
</tr>
<tr>
<td>Data lineage, data audit and the 'one-way street' problem</td>
<td>167</td>
</tr>
<tr>
<td>Impact of the DMZ</td>
<td>170</td>
</tr>
<tr>
<td>Software support</td>
<td>173</td>
</tr>
<tr>
<td><strong>11 MIGRATION DESIGN AND EXECUTION</strong></td>
<td>175</td>
</tr>
<tr>
<td>Migration end-to-end design</td>
<td>175</td>
</tr>
<tr>
<td>Extract, transform and load designs</td>
<td>181</td>
</tr>
<tr>
<td>Detailed decommissioning design</td>
<td>195</td>
</tr>
<tr>
<td>Migration build, test and execute</td>
<td>198</td>
</tr>
<tr>
<td><strong>12 LEGACY DECOMMISSIONING</strong></td>
<td>205</td>
</tr>
<tr>
<td>Executing the decommissioning design</td>
<td>205</td>
</tr>
<tr>
<td><strong>SECTION 3: FAILING DATA MIGRATION PROJECTS</strong></td>
<td>209</td>
</tr>
<tr>
<td><strong>13 RESCUING FAILING DATA MIGRATION PROJECTS</strong></td>
<td>211</td>
</tr>
<tr>
<td>Introduction</td>
<td>211</td>
</tr>
<tr>
<td>Stabilisation</td>
<td>212</td>
</tr>
<tr>
<td>Planned activity</td>
<td>214</td>
</tr>
<tr>
<td>Post-implementation mop-up</td>
<td>215</td>
</tr>
<tr>
<td>Impact of the DMZ</td>
<td>216</td>
</tr>
<tr>
<td>Conclusion</td>
<td>217</td>
</tr>
<tr>
<td><strong>APPENDICES</strong></td>
<td>219</td>
</tr>
<tr>
<td><strong>A1 CONFIGURABLE ITEMS</strong></td>
<td>221</td>
</tr>
<tr>
<td><strong>A2 IMPACT ANALYSIS FRAME</strong></td>
<td>223</td>
</tr>
<tr>
<td><strong>A3 DATA MIGRATION STRATEGY CHECKLIST</strong></td>
<td>226</td>
</tr>
<tr>
<td><strong>A4 SPECIMEN DQR FORM</strong></td>
<td>229</td>
</tr>
<tr>
<td><strong>A5 DATA MAPPING EXAMPLE</strong></td>
<td>231</td>
</tr>
<tr>
<td>Index</td>
<td>245</td>
</tr>
</tbody>
</table>
FIGURES AND TABLES

Figure 2.1  The three circles of activity  
Figure 2.2  Standard reaction to change  
Figure 3.1  A diagrammatic representation of PDMv2  
Figure 4.1  Quality versus time versus budget  
Figure 5.1  An example ERD  
Figure 5.2  ERD relationship symbols  
Figure 5.3  The involuted data relationship  
Figure 5.4  Example conceptual entities  
Figure 9.1  The DQR process  

Table A1.1  Some PDMv2 configurable items  
Table A2.1  PDMv2 impact analysis frame  
Table A3.1  Data migration strategy checklist  
Table A4.1  Specimen DQR form  
Table A5.1  Target system equipment table  
Table A5.2  Legacy Equipment Item Type table  
Table A5.3  Target Equipment Type table  
Table A5.4  Equipment type cross-reference table  
Table A5.5  Equipment cross-reference table  
Table A5.6  Location cross-reference table  
Table A5.7  Partially completed example data mapping table
Johny Morris has over 25 years’ experience in IT working as a programmer, analyst, project manager and data architect. He has worked as an independent contractor for the last 20 years and in that time has worked for some of the biggest names in IT consultancy including CSC, Logica CMG and Price Waterhouse Coopers (PwC). He specialises in data migration and integration and for the last 14 years has been involved with large-scale migrations at blue chip clients like Barclays Bank, National Grid Transco and British Telecom.
I would like to thank my publisher, Matthew Flynn, who showed great patience as I struggled with the text; Jutta Mackwell for her patient chiding that finally got me over the line; and Nina Turner with whom I started this data migration journey and whose insistence on only providing the best we can is a rule I have stuck with. Dylan Jones of Data Migration Pro deserves a mention for the many conversations we have had over the years where ideas have been batted around – he has been one of the most consistent voices demanding that this under-developed corner of IT deserves better recognition. Finally, my thanks go to my wife, Jo, without whose support this book would not have been completed.

Johny Morris
July 2012
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application program interface</td>
</tr>
<tr>
<td>COTS</td>
<td>Commercial off-the-shelf</td>
</tr>
<tr>
<td>CRM</td>
<td>Customer relationship management</td>
</tr>
<tr>
<td>DBA</td>
<td>Database administrator</td>
</tr>
<tr>
<td>DMZ</td>
<td>Demilitarised zone</td>
</tr>
<tr>
<td>DQR</td>
<td>Data quality rule</td>
</tr>
<tr>
<td>ERD</td>
<td>Entity relationship diagram</td>
</tr>
<tr>
<td>ERP</td>
<td>Enterprise resource planning</td>
</tr>
<tr>
<td>ETL</td>
<td>Extract, transform and load</td>
</tr>
<tr>
<td>GAM</td>
<td>Gap analysis and mapping</td>
</tr>
<tr>
<td>KDSH</td>
<td>Key data stakeholder</td>
</tr>
<tr>
<td>KDSM</td>
<td>Key data stakeholder management</td>
</tr>
<tr>
<td>LA</td>
<td>Landscape analysis</td>
</tr>
<tr>
<td>LD</td>
<td>Legacy decommissioning</td>
</tr>
<tr>
<td>LDS</td>
<td>Legacy data store</td>
</tr>
<tr>
<td>MDE</td>
<td>Migration design and execution</td>
</tr>
<tr>
<td>MDM</td>
<td>Master data management</td>
</tr>
<tr>
<td>MIS</td>
<td>Management information system</td>
</tr>
<tr>
<td>MSG</td>
<td>Migration strategy and governance</td>
</tr>
<tr>
<td>PDM</td>
<td>Practical data migration</td>
</tr>
<tr>
<td>PID</td>
<td>Programme initiation document</td>
</tr>
<tr>
<td>PMO</td>
<td>Project management office</td>
</tr>
<tr>
<td>SI</td>
<td>Systems integrator</td>
</tr>
<tr>
<td>SME</td>
<td>System matter expert</td>
</tr>
<tr>
<td>SRP</td>
<td>System retirement plan</td>
</tr>
<tr>
<td>UAT</td>
<td>User acceptance testing</td>
</tr>
</tbody>
</table>
Check point  A decision point at which it is agreed a new system is stable enough to go forward with or from which fallback occurs. (Also sometimes known as a ‘go/no-go point’.)

Churn  The relative frequency with which records of different types are added, amended or deleted from a data store.

Conceptual entity model  A form of data model where atomic entities are grouped together to form higher level entities that are meaningful to the enterprise.

Control total  Either the sum of some meaningful value within the data being transferred or a count of the number of units of migration being transferred.

Data architect  The person responsible for the design of how the data required for an organisation, possibly held over multiple applications, is held.

Data audit  The verifiable proof that all the units of migration in the legacy data stores are accounted for in the migration.

Data freeze  The prevention of updates to records after they have been extracted for data migration and before they have loaded into the new system.

Data lineage  The history of transformation that shows how an individual data item is transformed from one system to another.

Data mapping  The rule(s) by which one or more items in the Legacy Data Store will have their values moved to one or more items in the new system.

Data migration  The selection, preparation, extraction, transformation and permanent movement of appropriate data that is of the right quality to the right place at the right time and the decommissioning of legacy data stores.

Data owners  All the people within or outside an organisation who have the legitimate power to stop a migration from happening.

Data quality rules  A set of processes and deliverables that are used to measure the quality of the data within a data migration project and to resolve or mitigate data quality issues.
Data size The amount of data to be loaded.

Data stakeholder Any person within or outside an organisation who has a legitimate interest in the data migration outcomes.

Data transitional rules The temporary business operating procedures put in place to cope with the disturbance caused by data migration itself.

Demilitarised zone The interface between the technology provider and the wider programme.

Entity type The generic description of an entity.

Fallback The steps that will be taken to get an enterprise back into the position it was in prior to a data migration.

Fallback window The length of time between starting up a new system and taking the final check point that allows for the full decommissioning of legacy data stores according to the system retirement plans.

Homonym Two words that are spelled the same way, but have different meanings.

Instance A particular example of an entity type.

Key business data areas The segments into which a large data migration project is broken down for management and planning purposes.

Landscape analysis The systematic discovery, review and documenting of the legacy data stores, including their linkages, data quality and key data stakeholders.

Legacy data store A data repository of any type that holds data of interest to the new system.

Metadata Data about data. The data that technologists hold about the data in the business.

Migration form The technical style of a migration.

Navigation The links in the data that allow software to move from one data item to another. An example would be a foreign key that allows a program to get from a holding company record to all the operational company records beneath it.

‘One-way street’ problem Occurs when an algorithm transforms data in such a way that the original values cannot be identified.

Policies The explicit or tacit underlying drivers for a project and for the surrounding environment in which a project operates.

Project A one-off enterprise event with a beginning, middle and an end.
Semantic issue  A disagreement about the definition of a business term or the use of fields in corporate systems.

Sequencing  The ordering of update processes into a tenable progression.

Source data store  Synonym of legacy data store.

Synchronisation  Enabling changes to data items in legacy data stores to be reflected in the target (forward synchronisation) or changes in the target to be reflected in the legacy data stores (reverse synchronisation) or in both directions (bidirectional synchronisation).

Synonym  Two words that are spelled differently, but mean the same thing.

System retirement plan  The user-side requirements of a data migration that will allow a legacy data store to be decommissioned.

Target  The final destination system or systems.

Topography  The map of data store linkages.

Training lag  The length of time it takes to train all the staff who need to be trained in the target system.

Transitional data store  A temporary database created during the process of data migration.

Unit of migration  The lowest level of data granularity of meaning to the business.
USEFUL WEBSITES

www.bcs.org  Publisher of this book and home of ‘Johny’s Data Migration Blog’ – the author’s regular commentary and insights into data migration.

www.dama.org  Data Management International – some good data quality and modelling information, but light on data migration.

www.datamigrationmatters.org  The Data Migration Matters event series with forthcoming events and past presentations.

www.datamigrationpro.com  Probably the best community website devoted to data migration.

www.dataqualitypro.com  Sister website to datamigrationpro with lots of product reviews, comment etc.

www.iaidq.org  The International Association for Information and Data Quality – good for local events and discussions, especially via the LinkedIn® website.

www.iерго.com  The author’s personal website.
INTRODUCTION

WHAT IS THE PURPOSE OF THIS BOOK?

This book is aimed at practitioners, project managers and purchasers of consultant resources who have a data migration project to deal with. It is designed as a teach yourself guide to data migration. I have written it as a consultant with many years’ experience in data migration to give you a series of steps developed in real-life situations that will get you from an empty new system to one that is populated, working and backed by the user population.

WHAT TYPES OF DATA MIGRATION ARE COVERED?

Data migration projects take many forms. The classic form is where a new system is to be implemented and needs to be primed with data from the legacy systems. There are also system consolidation programmes, either spawned by businesses merging or by a drive for standardisation. There are system upgrades, and these also require data migration. However, whatever the spur for data migration, the same problems will have to be faced, and this book will guide you past the pitfalls. For the sake of simplicity and consistency, unless there is a specific reason to indicate differing approaches for different types of data migration, I am going to refer to the old/existing data as the ‘legacy’ and the destination as the ‘target’.

WHAT IS NOT COVERED IN THIS BOOK?

This book is system-neutral. It is aimed at large-scale data migration projects visible to the end-user population. It does not cover the detail of migrating, for example, from one version of Oracle® or SAP® to another. It does not cover changes to operating systems or hardware. There are courses available for changes like these, and, if there is a sufficiently large market, books will emerge. This book is aimed at the gap in the methodologies that allow you to develop the perfect system but then say nothing about how you get the best legacy data out of the flaky old systems you are trying to leave behind. This book also does not cover the regular movement of data that supports business information type applications, be they data warehouses, data marts, master data management (MDM) or management information systems (MIS). This book is aimed at a project environment where there is a clear need to move data as a ‘one off’ to populate a new database.
HINT

A project is a one-off enterprise event with a beginning, middle and end. A business process is continuous: individual items will move through from the start event to the final transformation, but the process itself never stops. Learn the difference. Projects require different management skills from processes, have different deliverables and different timelines, but it is surprising how easy it is to confuse them. Superficial similarities hide the essential differences. Having said that, there are techniques in this book to perform data analysis and data cleansing, and they can be used to bring the enterprise into the project, so it might still be useful reading for anyone interested in cyclic data cleansing or data quality issues.

WHO IS THIS BOOK AIMED AT?

There are two types of reader who will find this book essential reading: the executive and the practitioner. The first is the person with management responsibility for seeing the project to a successful conclusion. You might be a practising or a lapsed technologist, or you might have no technological experience at all, but you want to know how to control a data migration project. The practitioner is the technologist with a data migration project looming in front of them and who is sensibly reaching for assistance.

To the executive

You might be surprised to learn that there were no non-proprietary (i.e. not tied to one particular technology or consultancy supplier) methodologies for data migration until the first edition of this book was published in April 2006. In other words, no one had created a widely accepted series of steps that would guarantee to get the dirty old data out of old systems and transform it into clean new data to be placed into a new system in which a company has invested so much of its money. The first edition of Practical Data Migration quickly established itself as the primary text, with thousands of copies sold worldwide. This is the second edition of that book (PDMv2), updated to take account of technological changes and the maturing of the supply of services in this area.

PDMv2 will demystify the plethora of terms with which technologists love to surround their activities. It will illustrate the sort of controls you should expect to see from a well-managed data migration project, and the sorts of contracts you should write with suppliers and the amount of work you must be prepared for, even in the best managed projects. It will illustrate the steps you should expect an experienced data migration consultant to execute.

So if you are responsible for hiring consultancy resources or are overseeing an in-house project, this book will arm you with the ammunition you will need to stay on top of the project.

Section 1 is for you. It explains why data migration projects are intrinsically difficult and why there is such a high failure rate. It explains why you should insist that all
parties involved in the migration work to PDMv2 standards if you are to succeed. It gives you an overview of PDMv2 so that you can converse with the practitioners with confidence.

The practitioner
If you are a practitioner about to embark on a data migration project for the first time (or even second or third time) you are right to feel daunted by the scale of the task. Bad start-up data is the curse of many good projects. It is not a subject that is well covered in most computing courses. It might not even seem that glamorous. Well, do not worry. Follow the methods and principles in this book and you will be guided to success. You will even make lasting allies. As explained, this book is also intended for the executive, so you might occasionally find yourself being told things that are the common currency of your daily working life. I would advise you to stick with it. Data migration uses many commonplace concepts in subtly different ways.

You should read Section 1 for an overview, then Section 2, where the PDMv2 modules are covered in detail. If your project is in deep trouble and you are buying this book in the hope that it will get you out of it, there is also Section 3, but browse the rest of the book first because it explains key concepts.

WHAT IS NEW TO VERSION TWO OF PRACTICAL DATA MIGRATION?

PDMv2 builds on the success of the original. Anyone who has mastered the underlying principles of practical data migration (PDM) will find that they are unchanged here. The most significant changes reflect the introduction of new technology over the years since the first edition was published, the introduction of the demilitarised zone (DMZ) concept that reflects the maturing of the market for data migration services and the redefining of PDM in a modular fashion. This final point is in response to the many comments that asked for a more precise definition of how the products and methods of PDM interact. Modularisation also makes it easier to see how PDM can be tailored to wrap around different technologies and implementation practices. Of course there are also the many small and subtle changes that over five years of additional practice have suggested.

PARAGRAPH STYLES

I have placed the following items in boxes so that you can find them easily:

- **Anecdotes** – These record my experiences and, hopefully, amplify the point I am making in the main body of the text.
- **Hints** – These are tricks and tips that I have found to work. These should, of course, be applied with circumspection based on your knowledge of the culture and structure of the environment in which you are working.
- **Golden rules** – You will be introduced to four Golden Rules that underlie and govern the approach used. They are the most significant things to take away
from this book. Learn them by heart and, whatever else you find expedient to change, stick with them and you will have increased your chances of success many times over.

- **Definitions** – As well as the Golden Rules, there are also other key ideas, unique to this approach, that need to be carefully defined.

Additionally, each chapter starts with a brief overview of what it includes, and ends with a summary of what you should take away from it.
INDEX

abstraction 79
accreditation in PDM v2 61
always up migration 54
amnesties, legacy data stores 66, 100, 107
application program interfaces (API) 182, 191
architectural diagram 178
archived data storage delivery of 37
design 37, 195–197
implementation 205
testing 131
assessments, data quality 110, 154–155
audit see data audit
audit experts on DQR board 149
identifying 101–102
automated fixes of data 145–146
awareness training 130
benefits, post-project 206–207
big bang migration 54
‘black boxing’ 90
budgeting 68–69
budgets for project 49
for software 57–58
see also quality versus time versus budget
build, migration 198–199
business colleagues gaining support of 32–34
losing trust of 31–32
see also key data stakeholders (KDSH)
business domain experts on DQR board 148
identifying 95–96, 137
and testing 131
business engagement communications strategies 65–67
need for 9, 17–23
overcoming resistance 28–34
ownership 91–92
see also key data stakeholders (KDSH)
business engagement work stream Fig. 3.1, 36
business knowledge, importance of 17, 23–24
business readiness 74
business readiness teams 67, 110, 129
business-class plane flight metaphor 13–14
change, emotional reactions to Fig. 2.2, 29–31
change control 64–65
in failing projects 212–213
of legacy data store list 110, 212
change management 70–71
check points 186–187
churn 183–184
close down, project 206
commercial off-the-shelf (COTS) packages 43–44, 118, 182
communications strategies 65–67
communications teams 67, 110, 129
complex datasets 111
concatenating data 168–169
costs of systems integrators 60
see also budgets
customer experience, in system retirement plans 135
cutover plans 194, 195, 204
cutover timelines 178
dashboards 74, 142
data architects definition 99
on DQR board 149
identifying 99–101
data audit 167, 170
definition 133
and demilitarised zone 172
design 194, 195
in end-to-end design 177, 179
fallback 189
manual 159
testing 200, 201
data customers 102
data dictionaries 161
data gaps 114–120, 158–159
internal 115–116, 160
migration model 116–120, 121–122, 160
target model 114, 145, 158, 160–161
topographical 159
see also reality gaps and checks
data governance 101
data lineage 133–134, 167–168, 170
in end-to-end design 177, 179
fallback 189
software feature 43
testing 200
data mapping 36–37, 163–170
and demilitarised zone 171–172
‘one-way street’ problem 169–170
parsing and concatenation 168–169
post-project benefits 207
software for 173–174
worked example 231–244
data migration common problems 9–10
definition 7–8
where to start 75
data migration analysts chairs DQR board 149
identifying 97–98
data migration strategy 46–75
checklist 226–228
communications strategies 65–67
contents of 47–48
implementation 75
initial migration plan 55–56
migration form 54–55
policies 50–53
project office functions 64–72
project scope 48–49
software selection 56–60
systems integrator selection 60–63
data models
data store models 83
entity relationship diagrams Fig. 5.1, Fig. 5.2, Fig. 5.3, 80–81
instantiating 84–85, 161–162
target data models 83, 117
types of 80
uses in PDMv2 81–83
see also conceptual entity model; migration models
data owners
definition 38–39
identifying 94–95
and system retirement plans 128–130
data profiling
in landscape analysis 36, 114–120
services 125–126
tools 41–42, 125–126
data quality
assessments 110, 154–155
compromise 24–26
enhancing 24, 26
monitoring 163, 171
removing duplicates 120–121, 124
responsibility of the business 10–12, 17–23
see also quality versus time versus budget
data quality rules (DQR) 38, 140–157
control of in failing projects 213
definition 140–141
and demilitarised zone 157
documentation 153–156, 229–230
escalation path 68
fix methods 143–144
fix types 144–146, 159
initial list of 72
monitoring 163, 171
post-project benefits 206–207
priorities 142, 154
process Fig. 9.1, 142–147, 151–152, 207
reporting on 73, 146–147
software for 156–157
specimen form 229–230
see also DQR board
data quality tools 42, 125, 146
and migration models 161–162
data retention 131–132, 179
data sizing 181–182
data stakeholders see key data stakeholders (KDSH)
data standards in legacy data stores 100
data store models 83
data transitional rules 143, 183–184, 235
demilitarisation see legacy demilitarisation (LD)
demilitarised zone (DMZ) 20–23, 40, 89–91
and data quality rules 157
definition 22
and end-to-end design 176–177
establishing 61–62
and extract, transform and load 190, 191, 192, 193, 195
and failing projects 216–217
and gap analysis and mapping 170–173
and in-house resources 23
and landscape analysis 125–126
and legacy decommissioning 177
logical 90
physical 90
and programme experts 98
and system retirement plans 139
and transitional business processes 193
design see extract, transform and load (ETL) design; migration design and execution (MDE)
DMZ see demilitarised zone (DMZ)
documentation 69
data quality rules 153–156, 229–230
legacy data stores 109–114
downsizing policies 53
DQR see data quality rules (DQR)
DQR board
composition of 148–149
frequency of meetings 152
running meetings 150–151
size 150
tasks 142–143
emotional reactions to change
Fig. 2.2, 29–31
end-to-end (E2E) design 175–181
engagement see business engagement
enterprise application migration see data migration
entity model see conceptual entity model
entity relationship diagrams (ERD) Fig. 5.1, Fig. 5.2, Fig. 5.3, 80–81
escalation path, data quality rules 68
estimating see underestimating
exclusion rules 145, 166, 171
execution
legacy decommissioning 205–208
migration 203–204
extract, transform and load (ETL)
design 181–195
and demilitarised zone 190, 191, 192, 193, 195
extract design 190
fallback design 185–189
load design 191–192
non-functional requirements 181–184
outputs 193–195
transformation design 190–191
transitional business processes design 192–193
transitional interface design 192
extract, transform and load (ETL) tools 42–43
extract design 190
extraction rules 165, 166, 171
falling projects, rescuing 211–218
fall forward plan 136, 188
fallback
contents of plan 186–189
definition 185
design of 185–189
forms of 187–188
reasons for 185
in strategy 66
in system retirement plans 135–136
testing 199
fallback window 189
fix methods 143–144
fix types 144–146, 159
framework purchasing agreements 62
full fallback 187
gap analysis and mapping (GAM) 36–37, 158–174
data quality monitoring 163, 171
demilitarised zone 170–173
gap analysis 159–162, 171
software for 173–174
see also data mapping
gaps see data gaps
Golden Rules 16–27
and data quality rules 140, 152
importance of 16
rule 1 17–23
rule 2 23–24
rule 3 24–26
rule 4 26–27
go-live restrictions 134–135, 179
governance
data 101
migration strategy and governance module 40
project 49–50
hardware considerations 184
hubs 43
impact analysis 70, 71
frame 223–225
in-flight transactions, handling 138, 178, 192–193
initial lists 72
instance, definition 120
integration testing 199
internal gaps 115–116, 160
issue management 67–68
key business data areas 86–87, 177, 216
key data stakeholder management (KDSM) 38–39, 102–103
key data stakeholders (KDSH) 91–103
definition 93
documenting 110
in failing projects 213–214
identifying 92–102
initial list of 72
managing 38–39, 102–103
and ownership 91–92
landscape analysis (LA) 36, 105–126
data profiling 36, 114–120
and data quality tools 161–162
definition 106
and demilitarised zone 125–126
documenting legacy data stores 109–114
identifying legacy data stores 36, 107–108, 110
master data management 120–124
software for 125
as Super SMART Task 108–109
migration form
migration controllers
migration build
mid-change cycle slump Fig. 2.2,
metrics and data quality rules
methods, fix
methodologies
metaphors for data migration
meetings, DQR board
master data management (MDM)

109

LD see legacy decommissioning (LD)
legacy data stores (LDS)
amnesties 66, 100, 107
control of in failing projects 212
definition 106
documenting 109–114
identifying 36, 107–108, 110
initial list of 72
modelling 83
technical details of 111–112
technical system experts 148
topography of 112–113

legacy decommissioning (LD)
and demilitarised zone 177
design 195–198
execution 205–208
introduction 37
reporting on 74
signing off 130, 138, 203
testing 199
lessons learned 207
library services 69, 73
load design 191–192
see also extract, transform and load (ETL) design
load testing 199
see also mock load testing
loading rules 145, 167, 172
logical migration models 161
‘long tail problem’ 206
manual fixes of data 145
mapping see data mapping
master data management (MDM)
101, 120–124
objective of 122
policies 52–53
post-project benefits 207
MDE see migration design and execution (MDE)
meetings, DQR board 150–151, 152
metadata
definition 79
see also data models
metaphors for data migration 13–14
method statements 155–156
methodologies
project 49, 51
supplier-side 23, 61, 90
see also PDMc2 methodology
methods, fix 143–144
metrics and data quality rules 156
mid-change cycle slump Fig. 2.2,
29–31
migration build 198–199
migration controllers 42–43
migration design and execution (MDE)
decommissioning design 195–198
to-end to end design 175–181
introduction 37
migration build 198–199
migration execution 203–204
migration testing 199–203
see also extract, transform and load (ETL) design
migration form 54–55
definition 54
impact of training lag on 130

and key business data areas 87
and ‘one-way street’ problem 170
and software selection 58
and systems integrator
selection 60
see also phased migration
migration model gaps 116–120,
121–122, 160
migration models 83, 116–120
instantiating 119, 161–162
logical 161
physical 162
migration plans, initial 55–56
migration readiness 66, 74, 147
migration strategy and governance (MSG)
introduction 40
see also data migration strategy;
governance
mock load testing 201–202
see also trial migrations
models see data models; migration models
MSG see migration strategy and governance (MSG)
navigation, definition 159
navigation rules 166
network considerations 184
non-functional requirements of design 181–184

off-the-shelf software 43–44, 118,
182
‘one-way street’ problem 169–170,
189, 196
open source software 41–42,
57–58, 60
open transactions, handling 138,
178, 192–193
orchestration 177
ownership 91–92
see also data owners
parallel migration 54
and ‘one-way street’ problem 170
parallel processing 58
parsing data 168–169
partial fallback 187
PDMc2 methodology
accreditation 61
overview Fig. 3.1,
35–40
solutions to common problems
44–45
people management 28–29
phased migration 54, 87
due to training lag 130
and ‘one-way street’ problem 170
switching to 188
and transitional interfaces 192
physical decommissioning design 197
physical migration models 162
plane flight metaphor 13–14
planning, project 67
initial migration plan 55–56
updating after end-to-end design
180–181
plans
cutoff 194, 195, 204
initial migration 55–56
see also system retirement plans (SRP)

policies
in data migration strategy 50–53
definition 51
framework purchasing
agreements 62
risk aversion 51–52, 57, 147
software 58
post-implementation mop-up in failing projects 215–216
priorities, data quality rules 142,
154
procurement processes 21, 89, 97
profiling see data profiling
programme experts, identifying 98–99
programme initiation document (PID) 48
project close down 206
project decomposition 54, 86–88
in failing projects 216
project governance 49–50
project office functions 64–72

qualitative assessments of data 154–155
quality see data quality; data
quality rules (DQR)
quality versus time versus budget
Fig. 4.1, 51–52, 57, 147
quantitative assessments of data
154–155
quest metaphor 13–14
readiness
business 74
migration 66, 74, 147
see also business readiness teams
reality gaps and checks 115,
119–120, 159–160
in failing projects 216
in master data management 121
recession, uncontrolled
PDMc2 solution 45
problem with 9
regulatory experts
on DQR board 149
identifying 101–102
regulatory requirements,
policies 53
release management 71
in failing projects 214
reporting
on data quality rules 73, 146–147
on legacy decommissioning 74
during migration 191
requirements 177
rescuing projects 211–218
resources, in system retirement
plans 136–137
responsibility gap 10–15
retention, data 131–132, 179
risk aversion policies 51–52
effect on software selection 57
revisiting 147
risk management 67–68
rollback see fallback
rules
data mapping 145, 165–167,
171–172
data transitional 143, 183–184,
235
see also data quality rules (DQR);
Golden Rules
run times 182–183
scope
data migration project 48–49
management of 64–65
programme 48
semantic issues 11–12
sequencing 183–184
rules 167, 172
SI see systems integrators (SI)
sign-off
check points 187
decommissioning 130, 138, 203
system retirement plans 74, 129–130, 194, 203
sizing, data 181–182
SMART tasks 67
see also Super SMART Tasks
social networking 43, 59
software
budget for 57–58
data profiling tools 41–42, 125–126
data quality tools 42, 125, 146, 161–162
in end-to-end design 175–176, 178
for migration build 198
migration controllers 42–43
off-the-shelf 43–44, 118, 182
open source 41–42, 57–58, 60
researching the market 59–60
selection 56–60
support for data quality rules 156–157
support for gap analysis and mapping 173–174
support for landscape analysis 125
synchronising 42, 54, 58, 184, 188–189
and systems integrator selection 60
specialist skills, lack of
PDMv2 solution 44
problem with 9
spreadsheets, interlinked 105
SRP see system retirement plans (SRP)
stabilisation of failing projects 212–214
staging databases 162, 173
stakeholders see key data
stakeholders (KDSH)
strategy see data migration strategy
studio editions 41–42, 57–58, 60
Super SMART Tasks 28–29, 32–34
data migration strategy as 47
data quality rules as 38
landscape analysis as 108–109
system retirement plans as 127, 138–139
supplier/client responsibilities see
demilitarised zone (DMZ)
suppliers see systems integrators (SI)
switch to phased delivery fallback 188
synchronising software 42, 54, 58, 184, 188–189
syncretistic option 237
system retirement plans (SRP)
contents 130–138
definition 127–128
dedicated zone 139
and end-to-end design 179–180
in failing projects 215
implementing 128–130
introduction 39–40
signing off 74, 129–130, 194, 203
as Super SMART Tasks 127, 138–139
systems administrators 96
systems integrators (SI)
costs of 60
selection 60–63
tasks outside scope of 36, 37
target data models 83, 117
target model gaps 114, 145, 158,
160–161
team building 28–29
technical work stream Fig. 3.1
tech-no-centricity
PDMv2 solution 44
problem with 9
technology
case for using off-the-shelf 43–44
overview 41–43, 178
in PDMv2 45
see also software
technique conferencing 157
testing
data migration 199–203
exception handling 202–203
mock load 201–202
responsibility for 176–177
in system retirement plans 131, 179
user acceptance 130, 131, 200
timing
of landscape analysis 109
of migration 134–135, 178
of mock load testing 201
run times 182–183
windows of opportunity 134–135,
179, 182, 183
see also quality versus time
versus budget
timing rules 166
tools see software
topographical gaps 159
topography of legacy data stores 112–113
tracking see data lineage
training 130, 179
training lag 130, 179
transactions, handling in-flight 138, 178, 192–193
transformation
and demilitarised zone 176
design 190–191
see also extract, transform and load (ETL) design
transformation rules 145, 167, 172
transitional business processes 65–66, 124
closing down 206
design of 192–193
in system retirement plans 138, 194
transitional data stores 160
transitional interface design 192
transitional rules 143, 183–184, 235
trial migrations 55–56
see also mock load testing
type, definition 120
UAT see user acceptance testing (UAT)
uncontrolled recursion see
recursion, uncontrolled
underestimating
PDMv2 solution 44–45
problem with 9
unit testing 199
units of migration
definition 27
in system retirement plans 137–138
use in reporting 74, 146–147
user acceptance testing (UAT) 130, 131, 200
validation rules 145, 166, 172, 190
widgets 59, 60
windows of opportunity 134–135,
179, 182, 183
work flow 43
Nearly 40% of data migration projects are over time, over budget or fail entirely. Studies have shown that using a proven methodology greatly increases the chances of on time, on budget and zero defect migrations. Enhance your career prospects by reading this book and making a success of your migration projects.

The author uses a series of steps developed in real-life situations that will deliver a populated, clean, working system backed by the user population.

The first edition of this book provided the first non-proprietary methodology for data migration and quickly became the primary text on the subject. This new edition is updated to take account of changes in technology and the maturing of the market for data migration services.

- New updated edition of bestselling text
- Get the dirty old data out of your legacy systems and transform it into clean new data
- Enriched with anecdotes, hints, golden rules and definitions

ABOUT THE AUTHOR
Johny Morris has over 25 years’ experience in IT working as a programmer, analyst, project manager and system designer. He has worked as a data migration consultant for some of the biggest names in IT consultancy (CSC, Logica CMG and others) and has been involved in data migrations large and small at blue chip clients like Barclays Bank, BT, Network Rail and Jaguar Land Rover.

You might also be interested in:
PRINCIPLES OF DATA MANAGEMENT:
Facilitating Information Sharing
Keith Gordon

BUSINESS ANALYSIS Second Edition
Debra Paul, Donald Yeates and James Cadle (Editors)