PRACTICAL DATA MIGRATION

Second edition

Johny Morris

bcs
The Chartered Institute for IT

PRACTICAL DATA MIGRATION

BCS, THE CHARTERED INSTITUTE FOR IT

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

PRACTICAL DATA MIGRATION
Second edition

Johny Morris

bcs
The Chartered Institute for IT

First published in 2006

Published by BCS Learning & Development Ltd, a wholly owned subsidiary of BCS The Chartered Institute for IT, First Floor, Block D, North Star House, North Star Avenue, Swindon, SN2 1FA, UK.
www.bcs.org

ISBN: 978-1-906124-84-7
PDF ISBN: 978-1-780170-05-3
ePUB ISBN: 978-1-780170-41-1
Kindle ISBN: 978-1-780170-32-9

British Cataloguing in Publication Data.
A CIP catalogue record for this book is available at the British Library.

Typeset by Lapiz Digital Services, Chennai, India.
Printed at CPI Antony Rowe Ltd, Chippenham, UK.

To Josephine, with my love, for being there.

CONTENTS

FIGURES AND TABLES

AUTHOR

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.

ACKNOWLEDGEMENTS

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

ABBREVIATIONS

API	Application program interface
COTS	Commercial off-the-shelf
CRM	Customer relationship management
DBA	Database administrator
DMZ	Demilitarised zone
DQR	Data quality rule
ERD	Entity relationship diagram
ERP	Enterprise resource planning
ETL	Extract, transform and load
GAM	Gap analysis and mapping
KDSH	Key data stakeholder
KDSM	Key data stakeholder management
LA	Landscape analysis
LD	Legacy decommissioning
LDS	Legacy data store
MDE	Migration design and execution
MDM	Master data management
MIS	Management information system
MSG	Migration strategy and governance
PDM	Practical data migration
PID	Programme initiation document
PMO	Project management office
SI	Systems integrator
SME	System matter expert
SRP	System retirement plan
UAT	User acceptance testing

GLOSSARY

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.iergo.com The author's personal website.

SECTION 1:
EXECUTIVE OVERVIEW

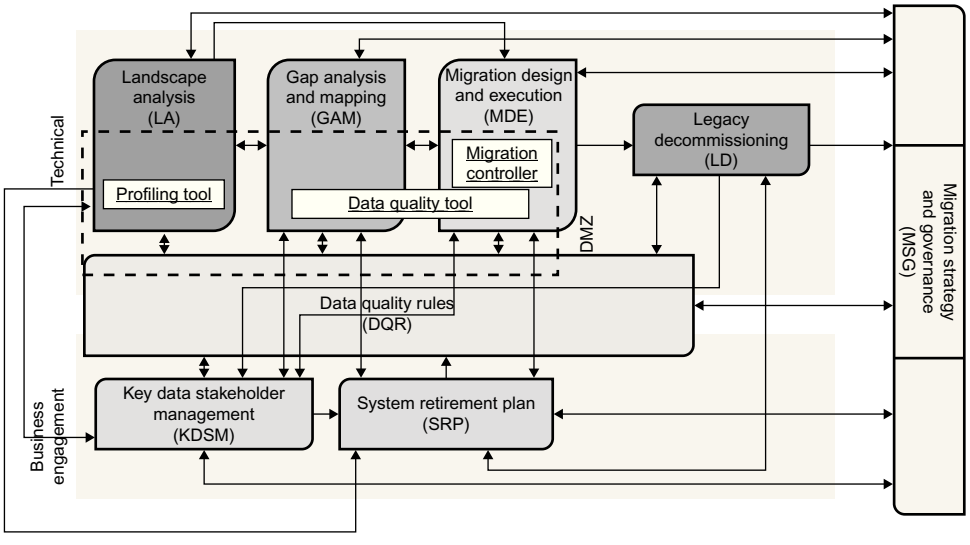
3 *PDMv2* OVERVIEW

In this chapter I outline PDMv2 *and show how it overcomes all the issues identified in Chapter 1 by using a set of integrated modules that cover the whole scope of a data migration from project start-up to legacy decommissioning and beyond. I give a brief overview of the types of software technology available to support data migration.*

INTRODUCING *PDMv2*

PDMv2 is modularised with seven functional modules and one overarching governance and project control module. This modularisation helps, as you will see, in tailoring *PDMv2* to other project delivery methods. Figure 3.1 illustrates *PDMv2*. The arrowed lines illustrate product flows between modules. It is clear from the diagram that it anticipates a degree of recursion in a data migration. For instance, at any point a new requirement for data might emerge from the larger programme of which a data migration forms a part. You might be concentrating at this point on testing your data migration solution, but you still have to handle the requirement. This is something over which you have no control within the data migration project, but it is something you can manage.

Figure 3.1 A diagrammatic representation of *PDMv2*



The functional modules within *PDMv2* are split over two work streams: business engagement and technical. Data quality rules (DQRs) span the two. It is important to note that even at this high level, *PDMv2* is designed from the ground up to integrate the business as well as the technical sides of the project. For *PDMv2*, business engagement is not a separate task, but is built into the way you go about your job.

Each of the modules is briefly explained below. More detailed explanations are given in Section 2, which provides sub-module, workflow-level descriptions.

Landscape analysis
The landscape analysis (LA) module uses various techniques to discover and catalogue legacy data stores (LDSs) and their relationship to one another. It is here that you look inside data stores to see how they work, what data they have and what data challenges they might contain. This is data profiling. It is performed by using both available software tools and manually. It is necessary to seek out consciously all the available LDSs, not just the official enterprise ones, welcoming all contributions. That way *PDMv2* turns a necessary technical activity into a set of Super SMART Tasks. Landscape analysis can commence prior to the design or even the selection of the target system.

HINT

If the decision on the target is still some way off, you can start on a small scale in an area that is likely to be rich in unofficial data stores. Start building the virtual team you will need to succeed. This also provides you with the opportunity to tailor the *PDMv2* deliverables to your programme management standards and to learn *PDMv2* by using it. Finally, data migration is a risky business. *PDMv2* helps de-risking by moving as much activity up the timeline as possible. Your mantra should be 'start small, start early'.

Do not expect that your selected systems integrator (SI) or software supplier can or will do much to help you in this area. Many of these LDSs are hidden away on departmental desktops, inaccessible to the reach of the SI. It is important to analyse and document how these data stores are linked (known as 'system topography') and include links across the scope boundary that receive or give data from and to stores that will be replaced. It is also a growing aspect of best practice that a fast iteration of LA is performed to quantify the scale of the data migration task you have in front of you prior to setting the budget for the remainder of the project.

Gap analysis and mapping
The gap analysis and mapping (GAM) module is where the data mapping takes place once the target system is available. Data mapping is the linking of fields in the LDSs to fields in the target, plus defining the transformation logic that is needed to split data up and merge fields. A classic example of this is reformatting name and addresses where perhaps a source database has the name in a single field, but the target holds the forename and surname separately and only holds the first line of the address, the rest being derived from a national postal file based on a postal or zip code.

HINT

It is perfectly normal for the target system to be delivered later than expected and in phases. Waiting for the target to be completely ready prior to starting data migration activities is a recipe for disaster. *PDMv2* gets around this by moving all discovery and profiling activity up the timeline to LA and the use of migration data models, which are explained in Section 2.

GAM is also where the mapping and gap analysis for legacy decommissioning design is performed so that an integrated target and archive design is developed. Again, it is absolutely standard for most SI or software suppliers not to be involved in designing a solution for those data items that the business might need (except rarely, for instance, where old transactions are preserved in case there is a tax inspection), but for which there is no place in the target. Therefore the design, build, test and execution of the archive solution is normally outside the DMZ, but driven by and dependent on the target migration design. After all, any item that is not moved to the target, but is needed by the business, has by definition to go into the archive, so these are activities that will need to be performed. *PDMv2* has them covered.

HINT

When reviewing proposals from suppliers, no matter how comprehensive they seem, comparing them against *PDMv2* will show you what is missing: in other words, what you will have to do yourself and therefore budget for. There are perfectly legitimate reasons, as you have seen from the discussion of the DMZ, for a supplier to limit their offering. Being aware of these missing elements, however, will allow you to plug the gaps.

Migration design and execution

The migration design and execution (MDE) module is where the physical design, test and execution of migration and archiving are carried out. Data migration is about more than just moving bits and bytes around. You have to be aware of business limitations, timings, audit requirements, data lineage, fallout, fallback, archiving requirements, reporting, management and control etc. MDE integrates all these elements within a single module solidly based on the business requirements expressed in the system retirement plans.

Legacy decommissioning

The legacy decommissioning (LD) module covers the physical or logical removal of legacy databases, hardware and software. It also covers the delivery of archived data storage for data items that have to be retained but which are not to be migrated to the target. There are also project close-down processes, including the handover of data quality issues (which it was not possible to fix within the project's time and budget constraints) to the in-life data quality teams (where they exist). Again it is absolutely standard for SI and software vendors to ignore the legacy decommissioning aspect of a data migration. You need to work with them through the *PDMv2* mechanisms to develop a single coherent design.

ANECDOTE

It is with some regret that I have to confess to not always finding willing recipients of those data quality issues that could not be fixed in the time and budget of a data migration project. However, at a corporate level, more organisations are making a conscious effort to manage data quality issues. As good corporate citizens we should endeavour to pass on what we can.

Data quality rules

The data quality rules (DQR) module is the centrepiece of what makes *PDMv2* unique. Sticking to the principle of Super SMART Tasks, this module manages all data quality and preparation-related activity on the programme. It integrates the legacy technical system experts, the target system experts and the business domain experts to prioritise, manage and complete all data issues, including the selection and exclusion of data sources. It is Super SMART because it builds the team by linking into the resources of the rest of the enterprise, creating a single virtual team; it builds the individual by empowering business colleagues and giving them the skills and opportunity to make a positive contribution to the project; and it completes the task by bringing much needed enterprise knowledge into a collaborative framework. DQRs are so important that Chapter 9 is dedicated to them alone.

HINT

To test how essential (and often missing) this element is in most data migration approaches, try the following. When analysing the bids of various suppliers, ask how they handle data quality issues. The common response is a long technical description of how they have sophisticated tools to trap errors. Your next question should be: 'And then what?' If you are lucky they will explain how they have an issues log that records the issues with appropriate dashboarding etc. Try another: 'And then what?' and you will be really pushing the boundaries of their method. The honest answer is: 'And then we wait while you, the project, come up with a fix.' They have no mechanism for getting that fix, only ones for uncovering errors and logging them. This illustrates the boundary of the DMZ. If you employ *PDMv2* and its DQR processes you will have one.

Key data stakeholder management

PDMv2 has its own specific role definitions for each key data stakeholder (KDSH). Key data stakeholder management (KDSM) manages the discovery, briefing and management of these individuals. *PDMv2* is very business-focused, so there are as many business as technical roles. *PDMv2* is very prescriptive when it comes to KDSHs. A full description of KDSHs is provided in Chapter 6, but from the business side the two most important stakeholders are data owners and business domain experts.

PDMv2 is quite clear about the definition of data owners.

DEFINITION

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

This definition is not based on organisation charts of who has titular responsibility for a database. If a person can legitimately stop the migration occurring because their information is not adequately managed, then they are the data owner of that piece of information. This means that the often forgotten, but powerful, individuals like financial controllers, who can stop migrations if the results compromise their needs, are equal data owners with the people who work directly with a data store (but only for the data items that impact them). Each data owner is expected to take part in the system retirement planning process, all of which is explained in more detail in Chapter 8.

The system owners, however, are often senior executives who cannot be expected to answer every query coming out of the project personally and who, in any case, often have no direct, day-to-day, hands-on experience of using the systems in question. They will generally defer to nominated business domain experts who they empower to input to the various meetings and provide the detailed knowledge to the project.

This is all very well, you might say, but how do you get the commitment of these powerful few to your project? We all know how difficult it is to get real commitment from colleagues working in silos different from our own. In this case, you leverage the compelling event of the forthcoming system retirement to grab attention and make your needs real to people.

System retirement plans

PDMv2 does not begin its conversation with the business by asking about data mappings, data quality, data gaps, data lineage or any other esoteric technical feature. *PDMv2* begins with the ultimate goal of a data migration: turning off legacy data stores.

HINT

It might seem crudely brutal, but I have found in practice that the bold unvarnished question 'You know system X that you depend on to do your day job? Well, we are turning it off on (enter date). How will you be certain that you will be able to continue with your day job once that has happened?' works better than anything. I leave it to you to dress it up in language appropriate to your situation, but I find the unchallengeable simplicity of this statement works best.

Going back from this you uncover all the things that must be done so that data owners will be comfortable with signing off the decommissioning certificate. You seek and encourage objections to going forward. Getting these objections shows

you have moved the subject over the hump of denial, down the slopes beyond rejection, to at least the negotiation of reluctant acceptance, if not to positive acceptance. Starting the conversation this way makes real the compelling event of a data migration.

You provide reassurance by explaining that you will be proceeding through a controlled sequence of iterations (the data owners will have to sign off the system retirement plan (SRP) at least three times before you get to the decommissioning certificate) that allows them to be ready and confident in the migration and that they have made you aware of all the things you must do to satisfy them.

Under the guidance of the *PDMv2* model you ask a series of structured questions that elicit the business view of the migration, looking at necessary items like business migration audit requirements, data lineage requirements, data retention requirements, migration restrictions, user acceptance testing requirements, go-live restrictions, fallback requirements, units of migration definitions, migration resource requirements and data transitional processes (all of which are explained in detail in Section 2). All responses are expressed in business terms.

The SRP is your business colleagues' main view of the migration and is a key input to migration design and execution.

Migration strategy and governance

The migration strategy and governance (MSG) module covers all the standard programme management functions that are expected on a well-managed project, plus some unique activities that are mandated by *PDMv2*. Section 2 gives a full description of MSG; however, from an overview perspective there is one task that must be completed with the involvement of the senior management of the whole programme, that is the creation of a data migration strategy and that follows in Chapter 4.

Demilitarised zone

As you have seen, the demilitarised zone (DMZ) is the interface between the work of the technology supplier and the responsibilities of its clients. The DMZ is a key component of *PDMv2* that will be, to an extent, formally defined in the contract with the supplier. However, the DMZ is wider than the contract and its formal definition will help both sides understand and manage their reciprocal dependencies. Throughout the rest of this book, I will constantly refer to the DMZ and its impact on each of the *PDMv2* modules.

INTRODUCING THE TECHNOLOGY

So far I have looked at the softer issues around a data migration because these are typically where the project is likely to go wrong. I will now introduce the standard technology that underpins data migration projects. A more detailed description can be found in appropriate places in Section 2 where the use of the technology is covered.

TECHNOLOGY OVERVIEW

There has been an explosion in the technology available to assist with data migration over the last few years and it seems that each newcomer to the market has a different take on how to perform the necessary tasks. However, broadly speaking, there are three phases of activity where specific technology support is available: data profiling; data quality; and migration control. I will also discuss generic project-supporting software (hubs and workflow tools).

Data profiling tools

Data profiling tools allow you to analyse the legacy databases and discover unforeseen features at column and row level. Some tools on the market also allow cross data source analysis. This means that you can compare, say, customer names in a sales ledger with customer names in a customer relationship management (CRM) database, even though they are from different vendors. The majority of tools work at the database level, analysing and comparing fields, but some tools will even interrogate source code to find relationships that are in the code but not in the database management system. (This only works, of course, when you have access to the source code, but is very helpful in old, locally built legacy systems where validation rules have been lost in the mists of time.)

The use of profiling tools is a vital first step in producing data of the quality that will load into your new system. As you will now be aware, you must anticipate that there will be surprises in the LDSs. It is a common misconception that the use of tools can only start once the target system has been fully defined. This is not the case. As you have seen, semantic issues, the most difficult and time-consuming to resolve, will be issues whatever the new system is. Be prepared to profile data prior to the definition of the target system.

However, data profiling tools are almost too good these days and will generate large amounts of information. You will need a process to winnow the wheat from the chaff. Fortunately *PDMv2* comes complete with just such a process: DQRs. By just using data profiling tools without the DQR process you risk drowning in a sea of possible issues. Data profiling tools are also limited in what they can access. Although some can compare spreadsheet data, most are restricted to looking at corporate standard databases and none can look easily into hard copy sources like rolodex, notebooks etc.

Similarly, data profiling tools will not uncover, on their own, the hidden data sources that are not linked by database level exchanges of data. The clerk with the data stick, moving data from one machine to another before creating figures that are re-entered manually into an old application might be essential to regulatory processes, but will be invisible to any tools. To discover these you need to rely on the more pedestrian methods within the LA module and the creation of your virtual team.

However, having dwelt on the limitations of profiling tools, best practice now recommends the use of profiling tools for projects of any size prior to setting budgets and plans for data migration. For smaller migrations there are plenty of free-to-use, studio editions of some of the leading toolsets. They tend to have enough

functionality for small datasets, but are limited in their integration, and therefore application, to larger programmes. On the other hand, the studio editions are also a good way to get a feel for what is on the market even if eventually you need to scale up to enterprise level software.

Data quality tool

Once you have an idea of the target, on the one hand, and the constraints of the LDSs, on the other, data quality software allows the speedy (most are 'point and click') implementation of validation and cleaning rules. There is obviously some overlap in functionality between profiling and data quality tools. Some software can span both spheres, but most are stronger in one than the other. The difference is that profiling tools discover relationships and possible data quality issues, whereas data quality tools check for and enforce known data quality rules. It is for this reason that the leading vendors tend to have both in their software sets. Ideally the rules discovered in profiling should be passed seamlessly for implementation in the data quality software, which in turn is fully integrated with the migration controller.

Once again, there are studio editions of many of the leading software offerings. These are worth looking at to get a feel for what is on offer if you have no local expertise.

Migration controllers

Migration controllers are often known as the extract, transform and load (ETL) tools; however, they are expected to do more than merely perform these three functions. They are the essential 'on the night' software that delivers the migration. Migration controllers need to be capable of performing the following functions:

- Reading data from the LDSs (the 'extract' step).
- Validating the extracted data (preferably using the data quality software seamlessly embedded into the migration controller suite).
- Reformatting the data and blending data from multiple sources (the 'transform' step).
- Scheduling, starting and stopping the migration process.
- Writing the data to the target (the 'load' step).
- Managing fallout.
- Managing fallback.
- Reporting on execution.
- Reporting on fallout.
- Providing audits.

More sophisticated products are also capable of many other features including:

- Synchronisation (keeping the changes to source and target data in step after the data has been moved. This means that the source can continue to be used during the migration thus enabling zero downtime migrations).

- Data lineage (tracking individual units of migration to show how they were transformed, combined and written to the target. This is sometimes a necessity for regulatory reasons and is also helpful for other technical reasons covered in Section 2).

Modern, built-for-purpose migration controllers have so many complex features that it is unlikely that you would be successful in replicating them in locally produced software.

Hubs and workflow

It is commonplace these days on large projects to implement some form of hub. This is an area where documents can be shared and different software is used to facilitate collaboration. Good data migration projects are a hive of collaboration, with the disparate groups of workers sharing knowledge. Workflow engines allow the output of, say, records that have fallen out of the migration to be routed from the migration controller to the correct team for analysis. The use of these tools depends on the scale of the migration. Small, co-located teams, have less need for collaboration tools than large project teams spread over a number of time zones and continents, but the ubiquity of this software and its relative low cost compared with the benefits in productivity and knowledge sharing make it all but essential. At the time of writing there has been a surge in interest in using social networking type software for collaboration on projects. This really is an area where you need to discuss what is available for your project with your architectural resource.

THE CASE FOR SPECIALIST TECHNOLOGY

It is possible to complete a data migration using the code writing capabilities of an indigenous IT department, but there are risks. All code writing is inherently risky. Bugs work their way into the code and then need to be weeded out in testing. The more coding you do the more bugs, the more testing, the more time, the more risk. Using software designed for the job reduces the number of bugs dramatically because it is only the logic of the migration that is being tested not the logic as it was instantiated in someone's handcrafted code.

A second compelling reason is that off-the-shelf data migration software these days is incredibly sophisticated. There is no way that any company, other than a rival software vendor, could ever justify the investment to recreate even a tenth of the features available out of the box.

ANECDOTE

I do know of one large technology company that decided in a fit of hubris to build their own migration hub. I have been quietly tracking the project and after five years and considerable expense they have not succeeded in replicating even the features of one of the cheaper off-the-shelf products. I suspect that their business drivers must go beyond merely delivering a solution.

All that being said, it remains a business judgement as to just where the cost–benefit lies. Tools can be purchased separately or as part of a set. Each migration has its own set of issues. For instance, is it a true 24/7 environment (like production control systems or telephony) where you really can never shut down the systems? If so, investment in sophisticated migration controllers that allow zero downtime are called for. Or do you have a complex heterogeneous migration environment to migrate from where the use of leading edge data profiling tools capable of cross database analysis will be useful? Possibly you are in a heavily regulated environment where data lineage is vital. Maybe you have only a short window of opportunity to migrate the data, in which case tools with superior prototyping capability could be key. And so on.

Only you know your own migration challenges and drivers. On the one hand, you do not want to invest in heavyweight technology you do not need. On the other, you do not want to be hamstrung trying to manhandle a migration via an inappropriate vehicle. It really pays to take advice at this point from internal or independent experts who can refine your thinking and define your options.

PDMv2 PROVIDES ALL THE ANSWERS

A review of the issues identified in the preceding text that have historically damaged data migration projects shows that *PDMv2* provides a solution that covers all the bases but is flexible enough to be deployed in partnership with your chosen supplier's own preferred approach.

Techno-centricity

PDMv2 sees data migration as a business-led joint IT–business activity, albeit with a clear view of where technology sits and how to make informed decisions about which technology is appropriate for your project. Through the DQR and SRP processes the business provides direction over data selection, data preparation, data quality, decommissioning etc., and takes ownership not just of the end point but of the process of getting there. All this is completed within an integrated set of linked activities.

Lack of specialist skills

PDMv2 provides the tools and techniques that are all you need to perform low-risk data migrations. You can use it either as a checklist against likely internal resources or seek training in specific skills or even the whole methodology.

For those partners, like implementation service suppliers, who will perform the final 'lift and shift' and who have their own preferred approach, *PDMv2* has the DMZ concept to insulate them from having to change their approach, which might well be optimised from their knowledge of their own technology and the target systems' load requirements.

Underestimating

PDMv2 provides two ways of managing estimates. Firstly, there is the LA module that can be run separately on a fast pass through basis to generate the understanding on which estimates of scale can be arrived at prior to setting the budget for the

rest of the migration. Secondly, when the project is in-flight there is an assumption (Golden Rule 3) that more issues will be generated than can be solved, but there is the DQR process to manage the prioritisation needed to get the appropriate data of an appropriate level of quality to the right place at the right time. As I have shown, knowing there is an issue and solving an issue is not the same thing, especially if the issue is a business-side or semantic problem. Only *PDMv2* has built in the controls of these decision-making processes via DQR that reach beyond the boundaries of the project. Spotting the challenges early while working in a collegiate, virtual team, with the business taking a leading role, means that time, quality or budget can be flexed in a dynamic but controlled manner. Put together as a coherent whole you have a set of processes that allow sensible decisions to be made that will deliver the data you need at the time you want it within a budget that you can accept.

Uncontrolled recursion
Using *PDMv2* from the start of the project will build the single virtual team across business and technical stovepipes that will ensure that the responsibility gap never has a chance to develop. Via the DQR and SRP processes, you retain a tight control on your migrations using common metrics across multiple different data types, data sources, geographical locations etc.

Technology
Within *PDMv2* there is an understood place for technology and a clear decision point for deciding which technology is appropriate to your migration. However, technology is not divorced from the rest of the migration processes. Technology on its own, in any walk of life, rarely solves any problems without being wrapped in some form of meaningful best practice. Technology is embedded in the modules that make up *PDMv2* and is therefore still directed by business towards optimally aiding your data migration. If necessary, and to get the maximum benefit from your software investment, encapsulation in the DMZ means that you do not expect to have to destabilise your chosen implementation partner's favoured approach, which will optimise the use of their toolset's salient features.

To achieve all of this, however, you need to make sure that you have set up your project in the right way in the first place: in other words that you are setting off from the right starting point. For that, you need to get your data migration strategy right in the first place. This is the subject of Chapter 4.

CHAPTER REVIEW

In this chapter I have introduced you to the various modules of *PDMv2* and showed their interrelated nature. I also looked at the use of technology and you saw how it overlays the activities within the modules. Finally, you saw how the use of *PDMv2* mitigates all the risks, both technical and non-technical, that I identified in previous chapters.

INDEX

PRACTICAL DATA MIGRATION

Second edition

Johny Morris

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.

For any practitioner faced with the challenge of delivering a successful data migration, this book is an absolute necessity.

Dylan Jones, Founder
Data Migration Pro/Data Quality Pro

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)

Management;
Information technology

ISBN 9781906124847