



MK<sup>®</sup>  
MORGAN KAUFMANN



# 数据质量工程实践

——获取高质量数据和可信信息™的  
十大步骤

〔美〕 Danette McGilvray 著  
刁兴春 曹建军 张健美 等译



**Executing Data Quality Projects**  
**Ten Steps to Quality Data and Trusted Information™**



电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>

Executing Data Quality Projects  
Ten Steps to Quality Data and Trusted Information™



# 数据质量工程实践

——获取高质量数据和可信信息™的十大步骤

“我尊敬的同事Danette McGilvray描述了一种非常实用的规划和管理信息质量的方法。我建议您阅读、领会和运用书中的这些知识。”

——Larry P. English, 信息影响国际股份有限公司的总裁和首席专家, TIQM®质量体系的创始人, 国际信息和数据质量协会的构想者和共同创办人

“在这个人们讨论时间很长, 但可行性建议很少的话题上, Danette McGilvray是一个给人们带来清新感觉的例外。如果您想了解如何实施数据质量项目, 请阅读本书, 您需要的一切尽在其中。”

——David Plotkin, 加州汽车协会数据质量经理

信息就是金钱。研究表明, 数据质量问题每年都使企业损失数10亿美元, 劣质数据造成浪费和低效, 破坏客户与供应商之间的信誉, 甚至使组织无力做出明智决策。在这本既重要又及时的新书中, Danette McGilvray提出了她的信息质量“10步法”——一种被证实了的、用于在企业中了解和创建信息质量的方法。她提出的方法适用于各种数据, 以及各种类型和规模的组织。Danette McGilvray使用这种方法已培训了一些“财富500强”客户以及数百名研究人员。

## 精彩内容

- 许多模板、详细实例以及实施“10步法”中每一步的实用建议;
- 以易于使用的格式突出显示了有关基本概念和定义、重要检查点、沟通活动以及一些最佳做法, 方便快速参考;
- 一个可链接到众多数据质量资源的配套网站, 其中包含文中勾勒出的许多规划和信息收集模板, “10步法”基本思想的快速汇总, 以及其他技术和信息。

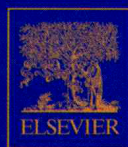
## 作者简介

Danette McGilvray是Granite Falls咨询公司的总裁。该公司专门研究信息质量管理和数据管治, 以支持客户满意度、决策支持、供应链管理及卓越运营方面的关键业务流程。

本书的在线支持资料网址为:  
[www.books/elsevier.com/companions/9780123743695](http://www.books/elsevier.com/companions/9780123743695)



策划编辑: 赵平  
责任编辑: 徐云鹏  
封面设计: 孙焱津



ISBN 978-7-121-12160-9



9 787121 121609 >

定价: 59.80元

本书贴有激光防伪标志, 凡没有防伪标志者, 属盗版图书。

# 数据质量工程实践

获取高质量数据和可信信息™的十大步骤

[美] Danette McGilvray 著  
刁兴春 曹建军 张健美 等译

电子工业出版社

**Publishing House of Electronics Industry**

北京·BEIJING

## 内 容 简 介

本书将信息质量的概念框架与改善信息质量的技术、工具、操作指南相结合,提出并详细描述了获取高质量数据和可信信息的10个步骤,为提高信息质量提供了一套与业务无关的系统方法论(简称“10步法”)。本书在简述信息质量方法和介绍“10步流程”赖以建立的基本原理及概念的基础上,系统阐述了“10步流程”所包含的流程、指南、建议、实例和模板,讨论了信息质量项目的有关方法建议、时间安排和团队组建,介绍了方法论中的具体应用技术,还以简明易懂的形式对书中的重要概念、模板、术语等精华部分进行了归纳整理。

本书内容丰富,理论和实践结合紧密,易读性和可操作性强,既可作为信息质量的入门和进阶用书,又可作为数据工程、信息技术等领域工程技术人员的参考用书。

Executing Data Quality Projects,  
Ten Steps to Quality Data and Trusted Information™

Danette McGilvray

ISBN-13:978-0123743695

Copyright © 2008 by Elsevier. All rights reserved

Authorized Simplified Chinese translation edition published by the Proprietor.

ISBN-13:978-9812726216

Copyright © 2010 by Elsevier (Singapore) Pte Ltd. All rights reserved.

Printed in China by Publishing House of Electronics Industry under special arrangement with Elsevier (Singapore) Pte Ltd.

This edition is authorized for sale in China only, excluding Hong Kong SAR and Taiwan. Unauthorized export of this edition is a violation of Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书简体中文版由 Elsevier (Singapore) Pte Ltd. 授予电子工业出版社在中国大陆(不包括香港、澳门特别行政区以及台湾地区)发行与销售。专有版权受法律保护。

本书封底贴有 Elsevier 公司防伪标签,无标签者不得销售。

版权贸易合同登记号 图字:01-2010-0517

### 图书在版编目(CIP)数据

数据质量工程实践——获取高质量数据和可信信息™的十大步骤/(美)麦吉利夫雷(McGilvray, D.)著;刁兴春等译. —北京:电子工业出版社,2010.11

书名原文:Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information

ISBN 978-7-121-12160-9

I. ①数… II. ①麦… ②刁… III. ①信息管理:质量管理 IV. ①G203

中国版本图书馆CIP数据核字(2010)第213769号

策划编辑:赵平 责任编辑:徐云鹏 特约编辑:张燕虹

印刷:北京市顺义兴华印刷厂

装订:三河市双峰印刷装订有限公司

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编100036

开本:787×980 1/16 印张:19.25 字数:443千字

印次:2010年11月第1次印刷

印数:3000册 定价:59.80元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zltts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

## 致中国读者

欢迎进入信息和数据质量领域！

几年前,我访问中国时,立刻就爱上了这个美丽的国家及其人民。我所接触的中国人都是那么善良、诚实、慷慨、大方。自那时起,中国在我心中就占据了一个特殊的位置。因此,当决定将这本书译成中文时,我感到由衷的高兴。

很高兴本书能有中译版加盟传播信息质量的思想。每一个以营利、非营利、教育和卫生保健等为目的的组织,都依赖信息以及该组织所能提供的一切来满足其用户的需求。试想一下,假如某组织的信息每次都能作为可靠信息使用,即该信息作为正确的信息,在正确的时间和地点供合适的人用来做出正确的决策和采取正确的行动,那么这一组织该是多么成功啊!再想象一下,如果将资源用于提供服务、辅助决策、开发产品、卫生保健、培养下一代等活动,而不是用来处理劣质数据造成的各种损失,那该多好啊!

本书阐述了一种可将基本概念与提高或创建高质量信息的技术、工具以及操作指南结合起来的方法论。无论是个人实施者、从业人员、项目组成员,还是项目经理都能从书中找到自己所需的知识。我希望本书能够使读者对信息质量的必要性和重要性的理解更加深入、全面。如果读者刚涉足这一领域,可将书中的基本知识作为事业的起点;如果读者已具备了一定的经验,可使用书中的一些附加知识作为目前工作的补充。读者可将书中所学应用于任何行业的所有信息类型,并且我相信无论是工商企业、慈善机构、学校还是医疗机构,所有组织都能从中受益。

本人不谙中文,因此只能依赖本书的翻译人员将书中"高质量信息"提供给读者。在此,我衷心地感谢本书译者刁兴春研究员、曹建军博士、张健美高级工程师为本书的翻译工作所付出的辛勤劳动。

希望本书描述的概念和流程能够帮助读者解决高质量数据和可靠信息方面的具体难题。我期望听到读者应用本书中的思想帮助所在组织提高成功几率的好消息!

Danette McGilvray

2010年9月于美国加州旧金山

## 作者简介

Danette McGilvray 是 Granite Falls 咨询公司的总裁和首席咨询师。Granite Falls 咨询公司是一个专门从事信息质量管理和数据管治的公司,该公司支持包括客户满意度、决策支持、供应链管理 and 卓越运营在内的关键业务流程。

10 多年来, Danette 在惠普 (Hewlett-Packard) 和安捷伦科技有限公司 (Agilent Technologies) 中倡导数据质量。作为从业者、计划和项目经理以及内部顾问,她获得了实施数据质量项目的第一手经验。除了挑战和机遇之外,她看到了数据质量带来的好处。她的工作从最初支持销售和经销职能的客户信息,一直扩大到包括所有类型的信息以及 ERP 系统和数据仓库中的数据集成。

作为 Granite Falls 的首席咨询师, Danette 不遗余力地利用其专业知识来帮助许多行业的客户。她已经认识到,即便组织了解信息质量的好处,但在实施侧重于数据质量的项目以及将重要的信息质量行动集成到其他项目和方法论中时仍然面临困难。她也认识到这些观点对于所有数据以及所有类型和规模的组织来说都是通用的。所有这些,结合她以前的经验,促使其开发了 10 步法——一种将概念与可靠指导建议相结合的灵活实用的方法。

美国和欧洲的很多会议都邀请 Danette 做讲座。她的照片登上了《个人电脑周刊》(PC Week) 和《惠普测量杂志》(HP Measure Magazine), 并作为受邀代表到中国讨论女性在计算机领域的角色和机遇。她是数据仓库研究所 (The Data Warehousing Institute, TDWI) 的教学人员和 DMReview.com 咨询专家小组的成员 (在此, 商务智能和数据仓库专业人士可向行业领导者提出问题)。Danette 是国际信息和数据质量专业协会 (International Association for Information and Data Quality Professional, IAIDQ) 的创办人和国际数据管理协会 (Data Management Association, DAMA) 的活跃分子。她是 Larry P. English 在其《改进数据仓库和企业信息质量》(Improving Data Warehouse and Business Information Quality) 一书中特别推崇的最佳创作者, 此书也是一本受到高度关注的关于信息质量和价值的书籍。

Danette 毕业于犹他州州立大学 (位于美国犹他州洛根地区), 她获得了商务信息系统学士学位并副修了企业管理。

目前, Danette 与丈夫 Jeff 居住在美国旧金山, 可以通过网址 [www.gfalls.com](http://www.gfalls.com) 与她取得联系。

## 译者序

最近 20 年,在麻省理工学院的全面数据质量管理计划(Total Data Quality Management program, TDQM)的引领下,信息质量(数据质量)已成为一门发展迅速的新兴学科:由 TDQM 发起的信息质量国际会议(International Conference on Information Quality, ICIQ)自 1996 年起已召开 14 届,面向应用的 MIT 信息质量工业研讨会(MIT IQ Industry Symposium Program, IQIS)自 2007 年起已召开 4 届。2006 年,在 TDQM 的帮助下,美国阿肯色州州立大学设立了世界首个信息质量硕士和博士学位授予点;2009 年,TDQM 创办了信息质量学术期刊——ACM Journal of Data and Information Quality(JDIQ),当年共出版 3 期,刊登学术论文 9 篇,其中首篇论文系统讨论了信息质量的研究内容,将信息质量分为 4 大类 19 个研究主题,从而形成信息质量的完整研究框架体系。

美国已基本完成企业级数据集成,并将信息质量问题列为企业信息化面临的首要问题。尽管国内有多位学者不断呼吁要加强对“数据”的重视程度,也有少数学者较早进入信息质量研究领域,但整体上来看,当前国内对信息质量仍以零星分散研究为主,缺乏系统性成果和规模性组织,尤其缺乏针对国内信息化特点的信息质量研究。我国企业数据集成环境比美国企业更加复杂,据 IDC 调查(2008),国内企业在数据集成相关项目建设中,面临的首要问题是业务需求不明确,其次才是信息质量问题;但事实上,国内 72% 的企业存在相似重复数据,60% 的企业存在不完整数据,信息质量问题相当严重,对信息质量研究的迫切需求不容忽视。

信息质量研究在全球范围内日益活跃,但与此不相称的是,全球范围内相关专著不过 20 余种,而中国大陆市场迄今为止还没有该领域的书籍销售(含英文版书籍)。本书是全球第 1 本面向工程应用的信息质量专著,它将信息质量的概念框架与改善信息质量的技术、工具、操作指南相结合,提出并详细描述了获取高质量数据和可信信息的 10 个步骤,为提高信息质量提供了一套与业务无关的系统方法论。本书作者 Danette McGilvray 女士现为美国 Granite Falls 咨询公司总裁,该公司专门从事支持关键业务流程的信息质量管理和数据管治。10 余年来,作者在惠普(Hewlett-Packard)和安捷伦科技有限公司(Agilent Technologies)中倡导信息质量,曾多次在美国、欧洲和亚洲讲学。Danette 女士对本书的中译版非常关心,并专门为中译版拨冗作序,译者对此向 Danette 女士表示诚挚的谢意。

能有机会将《数据质量工程实践》这一信息质量理论和实践大餐呈现给国内读者,译者倍感荣幸,希望此书能为国内信息质量的发展发挥一定作用。

负责本书译校工作的人员有刁兴春、曹建军、张健美。

参加本书初译工作的人员还有王芳潇、严浩、李凯齐、翁年凤、谭明超、李戈。

在此,我们还要感谢在信息质量工作中付出努力以及提供过帮助的其他人员和同事,他们是崔之祜、杜鹃、吴建明、汪挺、瞿雷、朱爱平、江春、邓波、丁鲲、温俊、袁震、王俊、蒋国权、彭琮、黄宇、张慧、张潇毅。。

在本书翻译过程中,我们力求忠实、准确地把握原著,同时保留原著风格;但由于译者水平有限,书中难免有错误和不准确之处,恳请广大读者批评指正,并欢迎与译者直接交流。

信息质量研究组(IQRG)

E-mail: xinxizhiliang@163.com

2010年8月



## 对本书的赞誉

Danette 带给读者一套切实可行的方法,此方法可使某组织和机构内部的数据质量达到所预期的状态。对即将踏上数据质量之路的组织机构来说,这是一本“必读”之物。

——Susan Stewart Goubeaux, 联邦住宅贷款银行(FHL Banks)财务办商务智能主管

“数据质量”已成为业界普遍认可的常用术语之一,然而,只有少数几个组织机构达到了用数据质量推动其产业发展的境界。人们所需要的就是一本指导用书,以指导那些商务人士更好地获取他们所需的数据。该书就是这样一本指导用书。或许单个步骤不足为奇,但作者将这些步骤组织起来,就使其达到了一种其他书籍所无法达到的效果。总之,这本书的确值得从事数据质量工作的任何人阅读。

——David Hay,基础战略公司总裁

Danette 采纳了以前仅在抽象概念中见到的方法,对如何提高数据质量给予了非常好的具体指导。

——John Ladley,IMCue Solutions 公司总裁

使用本书介绍的方法,你将在数据质量项目中永远不会迷失方向!书中附有一些小贴士、指南、模板、交叉引用和插图图标,此外,还列举了许多数据质量项目中最常见的实例,易于遵循和模仿。

——Larissa T. Moss,Method Focus Inc. 总裁

这本书不仅为数据专业人士,也为爱好或负责建立、维护和/或提高数据和信息质量的项目经理及企业代表提供了颇有价值的参考。本书与该领域其他书的区别在于其评估和提高数据质量的业务影响驱动方式,以及为每个步骤所提供的具体步骤和技术。

——Mehmet Orun,财富 250 强公司,卓越数据服务中心高级经理和首席架构师

“全面”是我用来评价这本书的第一个词。它解决了数据质量评估和提高各个方面如此多的细微差别,这些细微差别是以往一些较浮浅的处理方法所未曾提及的。太棒了!

——Michael Scofield, ESRI 公司数据资产开发部经理

这是业务与技术数据质量经理及从业人员所必须持有的一本书。Danette 清晰地阐述了其“10 步流程”将在何处将价值融入质量项目中,并成为成功从事数据集成工作的中坚力量。

——Robert S. Seiner, KIK Consulting & Educational 有限责任公司

The Data Administration Newsletter 有限责任公司

Danette 的写作风格很适合读者的口味,选材一流,她的“10 步法”简单明了,易于掌握,同时又很全面。这是一本非常优秀的书,我认为它将成为数据质量工作中最基本的一本参考书。

——Anne Marie Smith 博士,EW Solutions 公司教育主管和首席咨询师

Danette 编写了一本用于信息质量提高项目管理的颇有价值的工具书。她给出的清楚、简

明的概念定义也使此书成为一本有关信息质量原理方面的精美初级读本,可供数据专业人士、企业经理人士以及学生使用。我乐意向任何从事数据质量工作的人员推荐这本实用手册。

——Eva Smith, MSIM, CCP, CDMP, Instructor, Computer Information System 公司

没有哪两个数据质量项目是相同的。有些项目的重点是提高信息某方面的质量,有些则是其他项目的子项目,如数据迁移;还有一些由少数几个想有所作为的人承担,以期改变他们日常的工作模式。我之所以喜欢 McGilvray 的“10 步法”,是因为它能够适用于上述所有情况。本书提供了一个组织严密、易于理解和掌握的方法,你可将这种方法使用到你认为合适的程度。

——Gwen Thomas, 数据管治学会会长

谨以此书献给 Jeff

航行者,现在启航去探索、去发现吧!

——*Walt Whitman*

## 致谢

现在终于可以对众多需要感谢的人有所交代了——写书绝非仅靠一个人的努力就能完成的,本书也不例外。

感谢 Judy Kincaid,是她在不知不觉中让我走进信息质量领域。多年以前,Judy Kincaid 把我叫进她的办公室,让我和 Larry English 一起工作,当时 Larry English 即将进入 Hewlett-Packard 公司从事信息质量方面的咨询服务。Judy Kincaid 要求我即刻与 Larry English 一同工作,并尽快掌握有关知识,使得本公司的信息质量工作不因 Larry English 离开而中断。她的原话是“这周,你需满负荷地工作,以后可逐渐减少工作时间。”是 Judy 布置的这次任务开始了我的职业生涯,以至于 15 年以后我仍然能够全身心地致力于信息质量方面的工作。

非常感谢 Larry English,是他给我上了信息质量方面的第一课,在我的第一个项目中,他一直是我所信赖的、经验丰富的顾问,是他让我逐渐了解到这一领域的重要性。

特别要感谢的是 Mehmet Orun、Wonna Mark、Sonja Bock、Rachel Haverstick 和 Mary Nelson。当我写完这本书的初稿时,他们花费了大量时间,为我提供了宝贵的反馈意见和专业方面的建议。他们的丰富知识、启发性问题、建设性意见、独到的见解促成了该版本的形成,打下了本书的基础。没有他们的努力,就不会有本书的出版。

感谢那些审查最初方案或详细手稿的人,以及那些参与讨论并为本书具体章节提供输入数据的人,他们是 David Hay、Mehmet Orun、Eva Smith、Gwen Thomas、Michael Scofield、Anne Marie Smith、Lwanga Yonke、Larissa Moss、Tom Redman、Susan Goubeaux、Andres Perez、Jack Olson、Ron Ross、David Plotkin、Beth Hatcher、Mary Levins、Dee Dee Lozier,还有其他一些不愿意透露其姓名的人。他们为改进工作所提的中肯建议,以及对本人所做工作的肯定使得本书更加完善。

感谢那些多年来在不同组织中一直致力于实践或支持本书所提方法的资助者、项目经理、团队成员及实践者(可惜,在此没有更多的空间——列举)。感谢他们为获取信息质量知识所进行的相关实践,这些经验将为其他人从事信息质量探索提供帮助。

对于参加我的专题讨论会和课程的各位同仁,感谢他们的参与以及共同分析解决问题的思路、经验教训和期盼成功的良好愿望,他们的热情反馈是我写这本书的动力。

感谢在这一领域以及相关领域的先驱们,因为有他们的著作和教学让我们学习借鉴。参考书目中列出了我要感谢的其他人所做的工作,毫无疑问,我是这些工作的受益者。特别感谢 Tom Redman、David Loshin、Larissa Moss、Graeme Simsion、Peter Aiker、David Hay、Martin Eppler、Richard Wang、John Zachman、Micheal Brackett、John Ladley、Len Silverston 和 Larry English。

此外,我还要感谢那些为我的教学和出版提供场所,或者提供幕后建议和支持的部门负责人。尽管我不能列举所有这些人的姓名,但我仍然想提及 Tony Shaw 以及 Wilshire Conferences 和 TDWI 的所有人员,Jeremy Hall 以及 IRM UK 的所有人员,涉及 IAIDQ 和 DAMA 的相关国际

人员 Mary Jo Nott、Robert S. Seiner、Larissa Moss、Sharon Adams、Roger Brothers、John Hill、Ken Rhodes 和 Harry Zoccoli。

感谢我在 Morgan Kaufmann 的同事们,我要感谢他们卓越的工作,特别是感谢 Nate McFadden 的指导和独到见解,以及 Denise Penrose、Mary James、Marilyn E. Rash、Dianne Wood、Jodie Allen 和本书的写作团队,是他们使得本书的完成成为可能。

感谢 Logan City School District 和 Utah State University 的老师,是他们让我接受了教育并成就了我今天的一切。

感谢 Keith 和 Myrtle Munk,我非常庆幸有这样的父母,感谢他们对我严格的教育、对我行动的鼓励和始终的信任(甚至在不相信自己时)。作为一个大家庭和朋友圈的一员,我感到很幸运,因为他们为我带来了欢乐、笑声、关爱和支持——这些都是我努力工作的动力和生活的价值所在。

感谢我的女儿 Tiffani Taggart 和已故的儿子 Jason Taggart,尤其在艰难时期,他们是我不断前进的动力。

最要特别感谢的是我的丈夫 Jeff McGilvray,感谢他对我执著的爱以及鼓励和支持,没有他就不会有我今天的一切。

# 绪论

只有4类组织关注数据质量问题:那些关心客户的组织、关心赢利与亏损的组织、关心雇员的组织和关心其未来前途的组织。

——Thomas C. Redman (2006年)

## 著书目的

信息就是金钱。在全球通信便捷的当今时代,廉价、最新、可靠的信息是进行有效竞争的基本条件。在市场调查、专利申请、生产改进步骤、取得销售订单或接收付款中,信息对每一个业务的职能实现都是必不可少的,它可使企业具有竞争优势。

信息质量通过将正确的信息在正确的时间、正确的地点提供给正确的人来实现这一优势。根据有缺陷的、不完整的或误导性数据,人们无法做出有效的企业决策。人们需要以他们认为正确和及时的信息来推进其企业目标的实现。

正确的信息有助于库存管理员维持健康、有效的供应链,有助于CEO根据精确的和准确可靠的性能测量制订长期发展计划,有助于社会服务机构确认那些需要帮助的有潜在危险的少年,还能使选举人相信选举结果的正确性。通过了解信息的质量,决策者可以对劣质数据的影响给出解释。

错误信息可以将一个满意客户变为不满意客户,甚至可以引发数千个不满意客户。在某些情况下,不正确的信息可以改变生与死,即使利害关系不会如此之大,劣质数据也会对任何组织的健康状况造成影响。信息系统中遍布的错误信息增大了业务损失,能够使收益损失数百万美元。企业必须采取有效的措施来改善信息质量,减少错误以防止企业损失,并培养满意客户群。任何组织,无论是营利的、非营利的、政府部门还是教育机构,也不管它们供应什么,都要依赖信息对客户提供支持。

本书为提高和获取企业数据和信息质量提供了一套与业务无关的系统方法。它阐述了一套方法论,该方法论将一个认识信息质量的概念框架与提高和获取信息质量的技术、工具和操作指南结合起来。信息质量框架是概念框架,10个步骤是概念的实施流程。

该方法论可以通过选择适当的步骤、行动、流程和技术而应用于以下场合:

以信息质量为核心的工程,如某特定数据库的数据质量评估,或评价某个问题的业务影响,即以确认根本原因和实施提高为目的的劣质数据造成的后果。

在日常工作过程中随时会管理数据质量或影响数据质量。

将具体数据质量行动集成进其他项目和方法论,例如,企业资源规划(Enterprise Resource Planning, ERP)迁移、建立数据仓库或其他的应用开发和实施。

该方法论也可以作为获得自己的提高方法的基础,或将数据质量行动集成到一个公司标准的工程生命周期或软件/解决方案开发生命周期中。

各公司的情况都不相同,但本书描述的数据质量基本方法适用于所有公司和所有类型的数据,如财务、研究、开发、采购、制造、市场营销、订购管理、人力资源等。该方法论可以应用于各种类型的组织,如各种规模的企业和公司、教育机构、政府部门、非营利组织和慈善组织,因为这些组织都依靠信息通往成功。

## 面向对象

如果你是:

- 独立实施者。
- 某个项目组的成员。
- 项目经理。

如果你身处以下情形之一:

- 劣质数据正导致组织出现问题,但没人断定其真正影响,以及处理这些问题应该投入多少。
- 你的公司正在实施或已经实施 ERP 应用。劣质数据正影响项目的预定时间安排,并妨碍迁移和测试结果。一旦生产过程中对信息的信任度低,正被用于端到端流程的之前某业务职能使用的数据将会产生较差的结果。
- 你的新应用研发项目正在按期进行,或许已在预算中;用户认同度测试已经完成,解决方案已被采用。但是,几个月后开始出现一些抱怨。需要额外的职员去处理数据质量和需求协调。他们发现最终解决方案没有包括决策所需的所有信息,或者更糟糕的是,提供的错误信息导致决策者犯了代价高昂的错误。
- 因为合并和并购,公司正在启动一项集成并购公司数据的项目。项目组有紧凑的时间表,而你已知道待集成的数据有一些质量问题。
- 组织已出重资从外部信源购买数据,但它们的质量不能满足业务需求。
- 数据仓库已运转一年多的时间,但来自企业智囊团的用户对报告不信任,抱怨数据的质量,并调用他们自己的电子数据表进行核实。
- 公司已经投资了一个较大的主数据清洗项目,如用户、卖主、雇员或产品。因数据质量已下降和不断导致业务问题,几年之后启动了另一个清洗项目。
- 组织已购买了一个数据质量工具,但不知道如何对其进行最佳实施和高效使用。
- 你参加了公司的一个六西格玛(Six Sigma)项目,需要有关项目各方面信息和数据的更多帮助。
- 管理数据质量是你日常职责的重要部分。

如果你熟悉以上情形,那么这本书适合你! 无论是刚启动一个项目还是已经处于生产过

程中,你经常会发现信息质量问题影响该项目投入的全面效益实现。因此,企业不能在运作、决策和业务智能处理、用户满意度方面得到预期的改进。

希望还是有的! 尽管数据和信息质量问题看起来很迫切且势不可挡,一个迭代信息质量过程,包括可说明性、根本原因分析、预防、流程监测、持续改进和沟通可使问题得以解决。在此你将学到应对数据质量挑战的流程、技术和思想。本书提出的同一技术和工具也可用于系统研发和项目实施中,从一开始就预防或最小化数据质量风险。

本书包括了在各阶段进行信息改善所需的内容,从最初的评估计划到微调现有信息流程,再到制订和实施 IT 解决方案,以产生对使用最及时的、最相关的信息。

## 本书结构

本书描述了一个方法论——通往高质量数据和可信信息的 10 个步骤,包括一个框架、若干概念、提高和获取信息和数据质量的流程。

10 个步骤是执行信息和数据质量提高项目的具体指南。这些步骤将数据质量维度(用于定义、测量和管理数据的信息特征)与业务影响技术(分析数据质量问题影响的定性和定量技术)相结合。执行整个流程中的第 1 步,将会为你呈现出关于整个企业或组织中数据和信息质量当前状态的画面,并为寻求数据和信息质量提供有力的依据。接下来的几个步骤反映了信息质量当前状态画面中展现的企业需求。10 个步骤的流程描述包括:模板、实例和每一步的实施建议,目标是在未来信息质量和信息处理流程中达到更好的状态。

为了有效实施 10 个步骤,了解一些关于信息和数据质量的基本概念是十分必要的;鉴于此,首先介绍了几个概念。信息质量框架(The Framework for Information Quality, FIQ)为理解信息质量的基本组成提供了概念基础。框架包括非常重要的信息生命周期,可以从创建、应用、存储、维护和最终报废等流程来深入理解信息生命。通过定义其他概念,基础工作也得以确定,这些概念包括如数据质量维度、业务影响技术、数据分类(对有共同特征的数据进行分组,如参考数据、主数据、交易数据和元数据)、数据规范(标准、模型、业务规则、元数据和参考数据)、数据管治和数据统管、实施的最佳实践和准则。

## 主要章节

本书包括以下主要章节。

第 1 章 概述——信息和数据质量方法的概要叙述。

第 2 章 基本概念——作为方法论的必要组成部分以及 10 步流程赖以建立的基本原理和基本概念。

第 3 章 10 步流程——完成信息和数据质量提高项目的流程、指南、建议、实例和模板。

第 4 章 构建个人项目——数据质量项目有关方法、建议、时间安排、团队组建。此章有助于构建个人信息和数据质量提高项目。



第5章 其他技术和工具——在方法论的各种手段中可以应用的技术。

第6章 结尾——对其他章节的小结和几句勉励的话语。

附录A 快速参考——此部分用易读的格式汇集了本书所提供的资料。

术语表——按字母顺序列出了本书所讨论的术语及其释义。

参考文献——本书写作过程中所用到的书籍、文章、网站列表。

## 约定

下面列出了本书用到的插入图标和说明。



定义

定义关键词或短语。



基本概念

描述重要的基本概念。



最佳实践

面向卓越和成功的实施及建议经验小贴士。



警告

注意事项。



沟通

描述沟通行动,这些沟通有助于支持项目中发生的

事件,是关键的成功因素。



检查点

以提出问题的形式帮助确定是否完成每一个步骤,

以及是否准备好进入下一步骤的准则。

## 使用说明

本书的目标是阐述信息质量的基本概念,并提供足够具体的、循序渐进的指导,以便于改进和获得高质量信息。

在日新月异的当今时代,读者需能够获取信息并快速找到他们的所需内容。本书的形式不但易于由前至后阅读,也可以快速找到在任一时间节点上的特别兴趣点。希望本书能作为参考大纲,在新的数据质量问题出现时,供读者翻阅此书。以下是两种很常见的情况。

- “我们知道数据质量十分重要,但不知道如何开始。”
- “我们中的一些人知道数据质量的重要性,但我们不知道如何展现信息价值的价值。”

在这里将为任何负责数据和信息质量的人(程序管理员、项目经理、实践者、独立实施者、内部和外部顾问)提供帮助。实践者的工作头衔包括数据分析师、业务分析师、开发人员、数据库管理员、主题专家、信息用户或知识工作者、数据建模人员和数据管家(Data Steward)。本书也提供了能够说明信息质量重要性的企业影响评估技术。