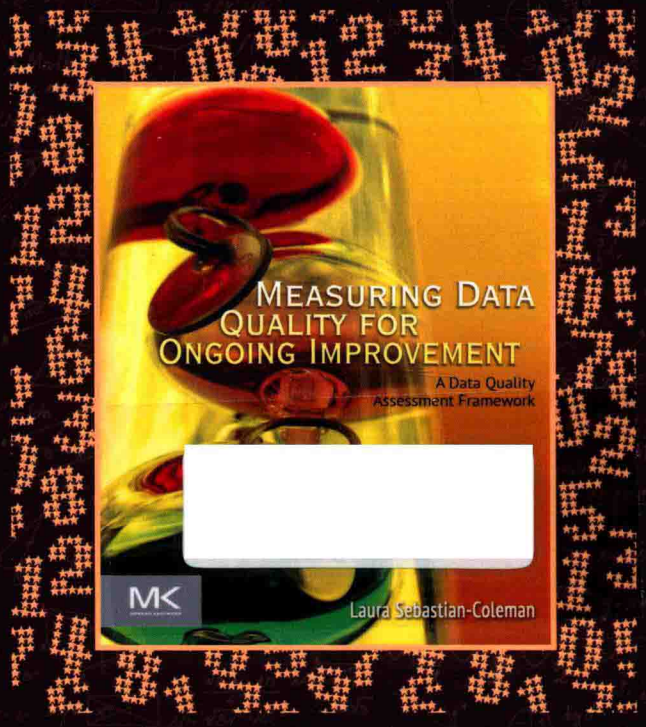
HZ BOOKS 华章IT
ELSEVIER 爱思唯尔
数据科学与工程技术丛书

数据质量测量的持续改进

[美] 劳拉·塞巴斯蒂安-科尔曼（Laura Sebastian-Coleman）著
卢涛 李颖 译



MEASURING DATA QUALITY FOR ONGOING IMPROVEMENT

A DATA QUALITY ASSESSMENT FRAMEWORK

机械工业出版社
China Machine Press

数据科学与工程技术丛书

MEASURING DATA QUALITY FOR ONGOING IMPROVEMENT

A DATA QUALITY ASSESSMENT FRAMEWORK

数据质量测量的持续改进

[美] 劳拉·塞巴斯蒂安-科尔曼（Laura Sebastian-Coleman）著

卢涛 李颖 译

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

数据质量测量的持续改进 /（美）塞巴斯蒂安 - 科尔曼（Sebastian-Coleman, L.）著；卢涛，李颖译 . —北京：机械工业出版社，2016.4
（数据科学与工程技术丛书）
书名原文：Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework

ISBN 978-7-111-53239-2

I. 数… II. ① 塞… ② 卢… ③ 李… III. 数据管理－质量管理－研究 IV. TP274

中国版本图书馆 CIP 数据核字（2016）第 053606 号

本书版权登记号：图字：01-2013-7843

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码：100037）

责任编辑：秦　健　　责任校对：殷　虹

印　　刷：北京瑞德印刷有限公司　　版　　次：2016 年 4 月第 1 版第 1 次印刷

开　　本：185mm × 260mm　1/16　　印　　张：17

书　　号：ISBN 978-7-111-53239-2　　定　　价：79.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：（010）88378991　88361066　　投稿热线：（010）88379604

购书热线：（010）68326294　88379649　68995259　　读者信箱：hzjsj@hzbook.com

序　言

我第一次认识劳拉·塞巴斯蒂安－科尔曼是在2007年。在麻省理工学院信息产业质量研讨会的全体会议上听她发表意见时，我注意到她是一个非常能说会道的人。随着时间的推移，这第一印象并没有改变，反而加强了。2008年在麻省理工学院，当我们都应邀出席会议的时候，我们直接见面了。她的思虑周详且旗帜鲜明地表达自己想法的能力再次给我留下了深刻印象。我们继续在麻省理工学院和IAIDQ（信息和数据质量国际协会）随后的会议中进行交流。每次我都期待听到她在医疗部门数据质量方面取得的成就的报告。这本书的问世，使得现在我们所有人都有机会向她学习。

其实，自从我听说她有计划出版本书后，我就一直在热切地等待着她这本书。熟悉我的《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》一书的读者都知道，我的方法填补了我们的知识体系中的高层概念和数据质量大饼中特定片段的深入细节之间的空白。我的十大步骤中的第9步被称为“实现控制”。有了劳拉的书，我们现在有了实现控制的深入细节。

这本书是做联机测量的首选手册，这是一种结合在数据处理中进行的测量。她研制数据质量评估框架（DQAF）的最初目的是解决以下问题：“如何建立一个进行数据质量测量的方法，它将跨多个数据存储系统工作，提供有意义的测量结果，并有助于努力提高数据质量？”正如在这本书中介绍的，DQAF已成功地解答了上述问题。

劳拉曾是Optum Insight公司最初创建和实现此框架的团队中的一员。她认识到市面上还没有解决持续测量数据质量问题的书，这是数据质量从业者最大的挑战之一。很多书都写了测量的必要性以及与剖析、数据发现和检查相关的做法，但都没有写如何持续地监控数据，以确保它继续符合要求。提高数据质量取决于持续测量数据是否符合业务期望的能力。劳拉从数据质量测量的上下文开始，并最终转到实现所必需的细节上。她的实践经验，连同她的教育背景，使她完全有资格写这本书。本书是数据质量文献的一个重要补充，我相信它必定会成为数据专业的标准参考资料。

从五年前我第一次听说劳拉，到现在她作为值得信任的同事和朋友，我一直都在注意她的言论。随着很多年前得到商业上的肯定，每个人都应该听劳拉·塞巴斯蒂安－科尔曼怎么说。这里是你的机会！学习和享受吧！

Danette McGilvray

《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》作者

Granite Falls Consulting公司主席和首席顾问

弗里蒙特，加利福尼亚，2012

致　谢

“主呵，您赐给我生命，请再赐给我充满感激之情的胸怀吧！”

——威廉·莎士比亚

《亨利六世》，第二部分（1591）

写书是一个既非常私人，又非常社会化的体验，说它私人，是因为你花了很多时间在自己的头脑中构思想法，通过对这些想法进行归类和字斟句酌，赋予这些想法一个形态，说它社会化，是因为你知道，有多少你了解、理解并完成的东西都要依赖于其他人。

我很感谢 Morgan Kaufmann 公司的 Andrea Dierna 和 Heather Scherer 使我有机会做这个项目，并感谢他们使它在不现实的最后期限前成为现实。同时感谢 Alan Studholme 和 Anitha Kittusamy Ramasamy 在完成最终产品中所做的工作。

我要感谢在 Optum 和 UnitedHealth 集团为这个框架提供了直接和间接投入的那些人：Dave Albright、Jeff Alexander、Derryl Bell、Morgan Berkus、Danita Born、Toni Bozada、Linda Briere、Tanya Bruzek、Tom Carey、Nancy Couture、Laura Cullen、Karen Davis、John Drake、Sharon Ehrlich、Jerry Enright、Dave Fallert、Bill Franzen、Celia Fuller、Lisa Groothausen、Rick Hauritz、Joe Herron、Lisa Hodne、Rose Knott、Shawna Jarzabek、Jon Lindquist、Jim Locke、Michelle Love、Peggy Magness、Eric Mellum、Vinnie Mercer、Cheryl Middlekauff、Roselle Monsalud、Kathi Mohr、Bob Naughton、Steen Poulsen、Jim O’Connell、Pam Opulski、Omkar Patel、Krishna Reddy、Jim Reitter、Sara Rekow、Janice Richardson、Ansh Sarkari、Roy Scott、Varjesh Shah、John Shin、Dave Stumpf、Ralph Tartaglione、Barb Turner 和 Diana Walter。包括当前的 UDW 和 UCG-DQAF 实施团队的 Jim Aronson、Debasis Acharyya、Vignesh Asaithambi、Sreeni Barla、Nagender Bethy、Mary Ellen Cash、Igor Gitlevich、Kishore Kolanu、Bhuvanesh Kumarsomu、Dan Pahl、Anand Ramasamy、Bhanu Rayapaneni、Kanwar Singh、Kurt Skepper、Ron Smith、Vijay Tadepalli、Bindu Uppuluri 和 Bill Wood。我衷心希望我没有遗漏任何人。

特别感谢 Tim Holt 鼓励我把这些想法制定成一个框架，并为变为这本书的白皮书提供了宝贵意见，感谢最初的 DQAF 团队：Rich Howey、Greg Kozbauer、Gary Mersy 和 Susan White，感谢 Donna Delvecchio 在 Galaxy 的实施项目中的工作，她让我看到了我以前根本没有意识到的框架的各个方面，还要感谢 Kent Rissman 对书稿的见解、反馈和对这个项目的持续鼓励。特别感谢 Eric Infeld 启动了我的数据质量之旅，提供了关于书稿的反馈和见解，并在过去九年中与我一起从事有关数据质量的持续对话。

Optum Insight 的慷慨使我能够参加并接触到数据质量领域的思想领袖和同行会议。本书已经从在麻省理工学院的信息质量和产业研讨国际会议，以及 IAIDQ 和 DGO 会议中播下的种子成长起来。通过这些活动，我从使数据质量变得更好的人们那里学到了很多。感谢 David Loshin、Danette McGilvray、Tom Redman 和 Lwanga Yonke，他们都非常慷慨地为书稿提供了建议和反馈。我非常感谢 Danette 对整个项目的鼓励与支持，她对书稿仔细和周到的阅读，以及她的友谊。本书中的任何错误或遗漏都是我自己的责任。

我也很幸运地得到了朋友和家人的鼓励和支持，妈妈、爸爸、Karen、Maura、Lisa、Amanda，感谢他们适时的电话和鼓励。感谢 Virginia Janzig 和她的丈夫 Dick Janzig，他们仔细阅读了书稿，并把写作专业知识以及在信息技术行业几十年有价值的经验分享给我。我的丈夫 George、我的孩子 Janet 和 Richard，无怨无悔地忍受着我在深夜和清晨工作，还占用了许多周末，并完全信任我能写完本书。感谢他们。

作者简介

劳拉·塞巴斯蒂安-科尔曼（Laura Sebastian-Coleman），Optum Insight公司数据质量架构师，自2003年以来，一直在大型医疗保健数据仓库从事数据质量方面的工作。Optum Insight专门通过提供分析、技术和咨询服务来改善医疗保健系统的绩效。劳拉已实现数据质量指标和报表，发起并推动Optum Insight的数据质量社区，促进数据消费者的培训项目，并领导建立数据标准和管理元数据的工作。2009年，她带领一队来自Optum和UnitedHealth集团的分析师，研发了最初的数据质量评估框架（DQAF），这是本书的基础。

作为一名活跃的专业人士，劳拉曾在麻省理工学院的信息质量会议、信息和数据质量国际协会（IAIDQ）以及数据治理组织（DGO）主办的会议上发表论文。在2009年与2010年，她曾担任IAIDQ会员服务总监。

加入Optum Insight公司之前，劳拉在商业保险行业从事了八年的内部通信和信息技术工作。她拥有IAIDQ颁发的IQCP（信息质量认证专家）证书，这是麻省理工学院的信息质量领域的一种证书，她在富兰克林和马歇尔学院取得了英语和历史学士学位，并在罗切斯特大学（纽约州）取得了英国文学博士学位。

目　录

概　述

“有两次我被问及，‘请问，巴贝奇先生，如果把错误的数字输入机器，它会输出正确的答案吗？’……我不能正确地理解什么样的混乱想法才可能产生这样的一个问题。”

——查尔斯·巴贝奇[⊖]

《哲学家的人生旅程》(Passages from the Life of a Philosopher)(1864 年)

数据质量测量：我们正在努力解决的问题

对于数据质量从业人员来说，最大的挑战之一是如何定义测量数据质量，特别是当数据的用途在不断发展，并且我们所依靠的数据量随时间的推移不断增长时。本书的目的是帮助人们理解测量数据质量的方法，使他们能够提高所负责的数据的质量。首先假设大多数人，甚至那些在信息质量和数据管理领域工作的人，都发现数据质量的测量是困难或令人困惑的。对于数据，我们还没有诸如卡尺和千分尺等制造业的物理工具，也没有诸如温度计和带式血压计的医学诊断工具来测量数据质量的基础。我们甚至对这些基础是什么都没有达成一定的共识。从 20 世纪 90 年代初开始，关于数据质量方面的持续讨论就出现了。同时在数据分析的相关工具和概念方面也出现了进步。但如何把这些概念和范畴应用到数据质量测量和监控上，却并非总是有明确的方法。如果没有一个方法来指导“到底怎么进行”持续测量，那么就可能很难持续开展提高数据质量的工作。

本书将尝试通过描述数据质量评估框架（Data Quality Assessment Framework，DQAF）来减少上述那种困难，这个框架包括一套共 48 种的通用测量类型，这些类型基于数据质量的五个方面，即完备性[⊖]、及时性、有效性、一致性和完整性。每种 DQAF 测量类型都是数据质量维度内的类别，它允许针对任何适合于该类型的标准要求的数据来执行一种可重复的测量模式，而不管具体的数据内容如何（例如，检查文件的完整性，处理的及时性，验证栏目内容的有效性，相关栏目总体的一致性）。这里提出的想法脱胎于在 Optum Insight 重新制定“到底怎么进行”持续的数据质量测量的工作。

DQAF 测量类型将在第二部分介绍。每个测量类型都由六个方面或特征集来定义：一个详细的定义，该测量类型解决的一组业务问题，一个测量方法，一组工程或编程的考虑，对该测量类型所需的支持过程的一个描述，以及确定特定指标并存储测量结果所需的一组逻辑属性（模型）。每种类型的各方面都将在第六部分进行深入描述。

DQAF 最初被开发来解决数据质量的联机测量问题，这种测量发生在结合了数据处理的数据存储或其他应用程序中。例如，作为提取、转换和加载（ETL）过程的组成部分的测量。

⊖ 查尔斯·巴贝奇（Charles Babbage），英国数学家、发明家兼机械工程师。他提出了差分机与分析机的设计概念，被视为计算机先驱。——译者注

⊖ completeness 也有译成完整性的，把它译成完备性是为了与后面的 integrity（完整性）区分。——译者注

但它也可以应用到最初的评估中，如用于特定分析目的的数据分析和检查或评估，或用于定期测量活动，如确定现有的数据存储库中的数据的完整性。这些应用将在第三部分展开讨论。一旦框架已定义，很明显，我们就可以用它作为需求定义过程的一部分，以便更好地定义有关数据的期望。这样，我们就可以找到提高数据的生产、采集、加工、储存和使用质量的方法。这些应用将在第四部分进行讨论。

在数据质量的上下文中反复出现的挑战

DQAF 的特点中有许多可以被理解为是面向技术和过程的。事实上，我的部分意图是指导组织如何把数据质量测量建立到其流程中。但是，正如经常发生的，在一个领域努力做出改善会引出能够影响一个组织实现数据质量测量的能力的相关挑战。这些挑战为大部分已写入本书的思想提供了背景，并在第一部分中进行描述。

数据质量的定义

数据质量由两个相关的因素定义：它满足数据消费者的预期的程度（它能够把预定的使用目的或用途完成到什么程度），以及它在多大程度上表示了创建它的对象、事件和概念。为了测量数据是否符合期望或“适合使用”，需要对期望和用途进行定义。我们通常把预期与针对特定用途的需求等同起来，而且，在大多数项目和系统中，它们是同义词。但是，我们也期望数据有更广泛的意义。生活在信息时代，我们是受数据的有关主流假定影响的，这些数据只是在那里存在，等待被使用（或误用），它们用真实和简单的方法表示了现实。早期的系统开发者以不同的方式看待数据，将其作为（总是从一个特定的角度进行的）观察和测量的结果。

纵贯全书，我会一直提醒读者，数据提供符号学功能。数据是由人们构建的抽象表述。[⊖] 虽然数据不是现实——它并不以一个简单的或没有问题的方式来表示现实——但在努力提高其质量时，我们应该认识到与数据有关的现实。数据以专门定义的方式表示现实的切片。数据是通过回答使组织能够发挥作用的具体问题而产生的。因为对用电子形式来采集和存储数据的依赖性，所以我们收集的数据的特征与用来采集它的系统有关。虽然数据可以从它的初始系统中分离出来，但它仍然会承载一些形成它的一个特定系统内的特征。它的起源仍然与它同在。数据作为现实中的某些方面的表示，与使人们能够访问和使用数据的系统设计之间存在紧密的联系。系统的设计对于理解数据表示什么，以及它如何影响其表示十分关键，因此，它对创建高质量的数据和理解数据的质量都至关重要（Wand & Wang，1996 年，第 88 页。Ivanov，1972 年）。

有关数据的期望

为了测量有关数据质量的期望，就必须知道这些期望是什么。我们在谈论期望时，通常将其当作由数据消费者定义的，但这样做，就假定了一个比我们实际所处的环境要简单得多的生产和消费的模式。在大多数情况下，数据最初不是为数据消费者创建的，相反，它是由数据“生产者”出于自身的原因或作为其他过程的副产品而产生的。一旦它存在了，它就可

⊖ 无论是在拉丁语还是英语中，数据（data）都是一个复数名词。通常情况下，我会按照数据的这种当代的用法将其用作一个集体名词，所以它在语法上是单数的。不过，必要时为清晰起见，我也将数据用作复数，比如，在“数据是抽象的表示”这句话中。

以被用于多种用途。随着时间的推移和新用途的出现，也就出现了新的期望。

虽然大多数组织都会记录对数据的需要，但很少有与期望的状况或数据的质量相关的明确期望。要阐明期望是很困难的，因为我们很少质疑自己有关数据如何产生、储存和使用的假设。对有关数据的假设和期望进行解读的过程往往揭示了妨碍满足这些期望的障碍。障碍的产生有一系列的原因。它们可能与数据的完备性有关，例如，一个事务处理系统甚至可能不收集所需的数据。或者它们可能涉及数据的结构，例如，一个系统可能无法以数据消费者所要求的粒度、原子性或精度来采集数据。或者它们可能与受产生该数据的选择的影响的其他因素有关。在第四部分中，我将介绍如何使用 DQAF 类别以一致和可理解的方式来定义数据质量需求，这样就可以确定这些需求已被满足的程度。测量数据质量的问题是一个如何测量抽象概念的问题。为了保证测量的有效性，人们需要了解它们代表什么，以及为什么它们很重要。所有测量都涉及被测量和它所针对的测量对象之间的比较。数据质量是针对与数据状况有关的某种形式的期望来测量的。

数据的风险

阻碍数据质量的其他障碍可被描述为产生、传输、转换，或存储数据的业务或技术过程所带来的风险。如果这些过程产生意想不到的结果或对数据做了意想不到的事情，该数据可能就不再符合需求。换句话说，我们对数据的预期不仅包括数据应该表示什么和被用来实现这个表示的结构，而且还包括发生在数据上的事情：它是如何收集、存储和移动的。所以在对有关数据的假设进行解读时，识别与数据处理或移动相关的潜在风险是重要的。风险可以被测量。事实上，测量在一个数据流中的风险点往往是防止下游数据问题的一种方式。在第六部分中，我将描述如何用 DQAF 测量类型来测量数据的风险和对数据的期望。

元数据与显性知识的重要性

在描述知识管理方面的挑战时，数据被描述为更高层次的理解，即信息、知识和智慧的输入。对于数据的这种思维方式会干扰我们管理数据并提高其质量的能力，因为它没有认识到数据是在如何表示现实世界的决定的基础上创建的。没有知识就不可能有数据。

元数据是关于数据的显性知识。它被记录和共享，使得企业的数据存在一个共识，包括数据想要被用来代表什么（数据定义和业务规则），数据如何影响这种表示（数据定义和系统设计），这种表示的限制（数据不表示什么），在其生命周期中数据会发生什么，特别是，当它经过流程和系统（出处、谱系以及信息链）移动时会发生什么变化，数据是怎么被使用的，并且可以怎么用，以及它不应该怎么用。在本书中，我会强调，元数据对于数据的有效使用是必不可少的。

业务 / 信息技术鸿沟

组织必须消除高质量数据的文化障碍。在这些障碍中，最大的障碍之一是直接在信息技术（IT）的岗位工作的人们和那些在以业务为导向的岗位工作的人们之间的挑战性关系。在当今由数据驱动的世界中，大多数人认识到，数据不但是有价值的（即使他们不知道该怎么评估它的价值），而且对经营业务非常重要。但由于 IT 管理数据及容纳数据的系统，许多组织仍然把数据当作 IT 的领地。这种看法妨碍有关数据的业务需求的重要谈话，并使得能够生产高质量的数据的各种制度和流程难以建立。在本书中，我将强调在数据质量的提高中所需要承担的共同责任。这种关系将是第 2 章的重点。

数据质量策略

如果数据是当今组织的命脉[⊖]，并且测量数据质量的愿望不仅影响记录的做法也影响组织的关系、资金，以及在两者之间的一切东西的决定，那么它就不能被孤立地解决。要真正持久地改进数据质量只能通过精心制定的战略眼光采用组织性奉献来实现。第五部分将数据质量测量置于更广泛的数据质量策略和数据治理的范围内讲述。

DQAF：数据质量评估框架

虽然这本书将解决有关数据质量的常见挑战，但它的核心是对 DQAF 本身的描述。DQAF 讲解了完备性、及时性、有效性、一致性和完整性等各方面，以确定一种可重复的模式，特定的测量可通过这个模式以一致的方式来执行。DQAF 测量类型可视为执行可重复的各种测量的通用的业务需求，正如一个温度计可被认为是用来测量温度的一个通用装置。关于数据质量测量的讨论描述了数据质量的各个维度，但不一定描述如何应用它们。因此，许多负责测量数据质量的人仅关注属于自己数据的已知挑战的特定测量。他们不一定能看得到自己的测量之间的相似性，他们也不用相同的方法来收集和处理测量结果。DQAF 提供了这个机会，有了它这种数据质量测量和数据管理方法，将更好地使业务和 IT 人员共同工作，以提高数据质量。

一旦你有能力进行测量，那么对你可以测量的任何东西开始测量就是很有诱惑力的。该框架背后的一个工作假设是，数据质量测量应该是有目的和可操作的。它应该被用来确定异常、降低风险，并发现改进的机会。使用该框架包括建立关键测量的标准，并且在正在进行的测量某个系统或多个系统范围内一致地应用这些标准。它还包括为使用测量类型制定战略，这不仅用于直接的测量，还用于定义要求和工程过程，以产生更高质量的数据。

本书组织结构

第一部分：概念和定义

第一部分讨论了一组术语，它们对数据质量和数据管理意义重大并将在全书中使用。

第 1 章描述了**数据**（data）的扩展定义，对数据的符号性功能，以及它作为可定义过程的产物，并作为分析的输入的存在进行了强调。数据表示与它自身不同的其他东西。它在这方面做得如何会影响对其质量的看法。

第 2 章定义了一组与数据和数据管理相关的角色，并讨论了涉及信息技术（IT）和业务人员之间的关系的挑战。

第 3 章给出了一组与数据管理相关的概念，它们与数据质量和数据质量测量的过程有直接关系。

第 4 章介绍了数据质量维度的概念，它可作为一种测量数据质量的手段，该章还定义了与数据质量评估相关联的若干一般概念和 DQAF 所使用的一些特定术语。

⊖ DAMA 知识全书首先断言“数据和信息是 21 世纪经济的命脉。在信息时代，数据被认为是一项重要的企业资产”（2009 年，第 1 页）。

第二部分：DQAF 的概念和测量类型

第二部分描述了创建 DQAF 的原因，概括了框架的假设、定义和管理思路，并给出了 48 种测量类型的简短描述。

第 5 章描述了 DQAF 的范围：基于质量维度的客观方面定义了一组测量类型，这使得数据的基础 IT 管理工作可以进行。该章定义了这些方面：完备性、及时性、有效性、一致性和完整性。还讨论了相关的概念：测量对象、评估类别和测量的功能（采集、计算、比较）。

第 6 章描述了数据质量评估框架最初是如何制定的。从高层次描述了 48 种测量类型。还包括几个基本流程的图解，帮助读者了解一个全面的数据质量测量系统的概况。

第三部分：数据评估方案

第三部分描述了数据评估这一更广泛的环境。

第 7 章描述了与这样的评估相关的目标和输入，并深入探讨“如何”剖析数据的详细信息（列剖析、结构剖析）。初始数据评估产生有关内容、结构和被评估数据的状况的宝贵的元数据。初步评估还应该产生一组针对数据质量的改善以及持续测量和控制的建议。

第 8 章说明测量与数据质量改进项目的关系。它包括一张问题根源和改进方法的表。

第 9 章描述了持续测量（联机数据质量测量、控制和定期测量）如何用于维持数据质量的一般原则。它讨论了定义持续测量所需的输入（对数据的重要性和风险的理解），对这种测量自动化的需要，并讨论了与持续测量的过程相关的交付成果。

第四部分：将 DQAF 运用到数据需求中

第四部分的目的是展示 DQAF 类别如何用于编制数据质量的需求，以便可以指定数据质量的联机测量、控制和定期测量。

第 10 章回顾了业务需求的上下文中的数据质量需求概念，并用它来识别数据可被测量的特征。它描述了如何对风险和重要性进行评估，以识别要通过持续的联机测量、控制和定期测量进行监控的特定数据和规则。

第 11 章提出了一系列的问题，可以帮助数据消费者表达自己的假设和有关数据质量特征的期望，并确定关键数据和处理风险。它们的答案可以被用来确定如何测量特定的数据元素和规则。从这个交流中，分析人员拥有了特定的数据质量指标定义的输入资料，这些特定指标包括想要执行联机测量和可能被定期测量的。

第五部分：数据质量战略

第五部分提供了定义数据质量战略的环境和方法。

第 12 章定义了数据质量战略的概念，使其能够相对于一个组织的整体战略，并有助于理解生产出更好的数据的其他功能（数据治理、数据管理、系统战略）。该章讨论了作为信息生命周期的一部分做出的不同类型的决策。它还提出了一套数据质量战略的总体考虑。

第 13 章提出了一套建立组织的数据质量战略的 12 个指令。它描述了如何评估这样的策略的组织准备。这些指令分成三组。第一组着重于企业内部数据的重要性。第二组将制造实物产品的概念应用到数据上。第三组的重点是为了建立一种质量文化，以回应满足战略数据管理的持续挑战。