

<<数据质量工程实践>>

图书基本信息

书名：<<数据质量工程实践>>

13位ISBN编号：9787121121609

10位ISBN编号：7121121603

出版时间：2010-11

出版时间：电子工业出版社

作者：麦吉利夫雷

页数：284

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## <<数据质量工程实践>>

### 前言

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，ICIS）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，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。

## <<数据质量工程实践>>

### 内容概要

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”）。

《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》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，还以通俗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。

## <<数据质量工程实践>>

### 作者简介

Danette McGilvray , 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。

## &lt;&lt;数据质量工程实践&gt;&gt;

## 书籍目录

第1章 概述 信息和数据质量的影响 关于方法论的概念和步骤 工程中的数据质量方法 付诸管理  
第2章 基本概念 引言 信息质量框架 信息生命周期 数据质量维度 业务影响技术 数据分类  
数据规范 数据管治和数据统管 信息和数据质量提高周期 10步流程 最佳实践及指导方针第3章  
10步流程 引言 步骤1 定义业务需求和方法 引言 步骤1.1确定业务问题优先次序 步骤1.2项目规划  
步骤2分析信息环境 引言 步骤2.1了解相关需求 步骤2.2了解相关数据和规范 步骤2.3了解  
相关技术 步骤2.4了解相关流程 步骤2.5了解相关人员/组织 步骤2.6定义信息生命周期 步骤2.7  
制订数据捕获和评估方案 步骤3 评估数据质量 引言 步骤3.1数据规范 步骤3.2数据完整性准则  
步骤3.3重复 步骤3.4准确性 步骤3.5一致性和同步 步骤3.6及时性和有效性 步骤3.7易用性和可  
维护性 步骤3.8数据覆盖 步骤3.9表达质量 步骤3.10可理解性、相关性和可信度 步骤3.11数据衰  
变 步骤3.12效用性 步骤4 评估业务影响 引言 步骤4.1事例 步骤4.2用法 步骤4.3业务影响的5  
个“为什么” 步骤4.4费效矩阵 步骤4.5排序和优先级划分 步骤4.6流程影响 步骤4.7低质量数  
据的代价 步骤4.8费效分析 步骤5 确定根本原因 引言 步骤5.1根本原因的5个“为什么” 步  
骤5.2跟踪和回溯 步骤5.3因果/鱼骨图 步骤6 制订提高方案 步骤7 预防未来数据错误 步骤8 纠  
正当前数据错误 步骤9 实施控制 步骤10 沟通行动和结果 10步流程小结第4章 构建个人项目 项目  
和10个步骤 数据质量项目角色 项目时间安排第5章 其他技术和工具 引言 信息生命周期方法  
数据捕获 结果分析和归档 度量 数据质量工具 10个步骤和六西格玛第6章 结尾 附录A快速参考  
信息质量框架 详细的POSMAD交互矩阵 POSMAD的阶段和行动 数据质量维度 业务影响技术  
10步流程概述 数据分类的定义 术语表 参考文献 图、表和模板列表

## <<数据质量工程实践>>

### 章节摘录

插图：多数工具都使用开放性算法，但需要调整到能适合具体数据。

需将业务需求转换成工具所需的规则和算法：

- 确定要比较的字段和匹配标准。

- 确定标准化规则、消除重复算法、权重和阈值。
- 要求几轮测试以使标准化和匹配程序达到可接受的水平。

预测数据准备和跨语言匹配方面的困难。

在像瑞士这样的一些国家中存在多种语言，很难程序化地判断使用哪种语言才能将数据标准化，因此，因不同语言和地址格式不得不使用不同算法和阈值。

还必须处理数据的录入方式、参考点和数据录入人员知识上的差异。

比如，在法国的某人录入了一个法国地址，而在德国的某人通常采用十分不同的方式录入同样的地址。

准备实施匹配时，在分析行动上要花足够的时间，对获得成功结果非常关键。

5.设计重复评估流程包括以下几点：

- 感兴趣的总体和相关的选择标准是什么（业务和技术两方面）

- 将由谁以及何时抽取数据，需要什么样的输出格式。
- 将由谁在消除重复工具中操作数据，何时进行。
- 将由谁审查消除重复的结果，何时进行。
- 将由谁报告测试和分析结果，何时进行。
- 将收集什么测量标准，需要什么报表。

6.抽取数据参考第5章“数据捕获”一节，以确保抽取正确的数据集。

7.测试重复数据使用最能满足需求的工具。

对于初步评估，或许只需查看匹配结果。

直到全面测试和定义了这些结果后才可自动进行数据变更。

从工具提供商那里得到适当培训，并使用可得到的最佳实践。

根据需要，管理和调整标准化程序、阈值和匹配算法。

在最终满意之前，不得不多次检查结果 / 调整算法，这是正常的，不足为奇。

对那些显示重复但可接受的记录，不做处理，但需对它们做标记。

报告所使用的测量标准，并讨论其影响。

如果想自动运行保留流程，请在安排时间表时多规划些时间。

即便保留流程由人员通过检查结果和选择保留记录来人工完成，也必须有记录或字段要优先的指导准则。

将规则归档，并提供培训，以便获取更一致的保留结果。

根据经验，建议将标准化数据保存在与原始数据字段不同的字段内。

标准化数据用于帮助匹配。

如果研究组对标准化程序进行了调整，就可使用更新算法对数据进行重新标准化。

如果没有原始数据，这样做将不太可能。

## <<数据质量工程实践>>

### 媒体关注与评论

“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

## <<数据质量工程实践>>

### 编辑推荐

《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》：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 Mc Gilvray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 Mc Gilvray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。



<<数据质量工程实践>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>