

2008 年 11 月 23 日,第三届科技信息资源共享促进国际研讨会在京召开。与会代表针对“信息质量”展开了深入的讨论。本刊曾对“信息质量管理”作了相应报道,引起了较大反响。从信息系统的角度看,信息质量(IQ)的研究从本质上属于应用学科,无论对 IQ 组织的参与者,还是对想要解决该问题的研究者都极其重要。目前,信息质量研究正与其他研究领域呈现互相融合之势。本期专题报道的 4 篇文章分别从信息质量的研究综述、信息内容安全性、管理者的 IQ 意识以及在外科治疗过程中的应用等方面进行分析研究,为研究人员提供基本理论与方法。

——编者

# 信息质量研究的现状与 未来研究方向

葛某智 Markus Helfert

(爱尔兰都柏林城市大学计算机学院,爱尔兰都柏林 5442)

苏颖 [校译]

(中国科学技术信息研究所,北京 100038)

**摘要:**从信息质量评估、信息质量管理和情景信息质量 3 个方面进行了综述分析。信息质量评估从质量问题、维度和评估方法 3 个方面进行了分析。信息质量管理从质量管理、信息管理和知识管理 3 个角度进行了分析。情景信息质量主要对信息质量在信息系统和决策支持方面进行了研究。分析结果揭示出潜在的研究领域以及当前在信息质量方面研究的局限性,旨在弥补研究差距。文章通过提供未来信息质量研究存在的问题以及经验应用给予的启示做出总结。

**关键词:**信息质量研究;信息质量评价;信息质量管理;情景信息质量

中图分类号: G203 文献标识码: A DOI: 10.3772/j.issn.1674-1544.2009.01.002

## 1 引言

大部分信息质量(IQ)研究始于多个机构出现的信息质量问题。大量的案例研究记录了很多与信息质量相关的问题。很多业务方案和项目被延迟或是没有成功,其主要原因就是信息质量较差。比如,某重要金融机构由于错误的输入引

发了大问题,其错误的执行订单高达 5 亿美元<sup>[1]</sup>。与此类似,某电信公司由于发送票据时出现信息质量问题,导致其在客户票据托收方面损失 300 万美元<sup>[2]</sup>。此外,“挑战者”号航天飞机爆炸以及海军“文森斯”号导弹巡洋舰误射伊朗“空中客车”客机事故都是信息质量管理故障的结果<sup>[3]</sup>。通过上述以及其他类似的真实案例,我们可以看到,信息质量问题是普遍存在的<sup>[1]</sup>,而且花费巨

第一作者简介:葛某智(1983-),男,博士研究生,研究方向是信息质量评估和管理以及信息质量对决策质量的影响。

基金项目:国家自然科学基金资助项目(70772021,70831003)。

收稿日期:2008 年 11 月 3 日。

大<sup>[4]</sup>,有时甚至造成灾难性后果<sup>[3]</sup>。为了解决信息质量问题,并防止较差信息质量造成的负面影响,在过去20年中,信息质量研究得到了极大的重视。

“质量”一词可定义为“适于应用<sup>[5]</sup>”。此定义在质量文献中得到大量应用<sup>[6]</sup>。与此定义相呼应,文献<sup>[6]</sup>将信息质量定义为“适于信息用户使用的信息”。他们认为,最终还是由客户判定信息产品是否适用于特定的用途。但是,信息用户并不擅长发现信息里的错误,也不擅长改变信息的应用方式<sup>[7]</sup>。因此,从数据视角看,信息质量可定义为“满足规范或要求的信息<sup>[8]</sup>”。根据以上信息质量的两种主要定义,信息质量研究可分为两大群体:管理与数据库<sup>[9]</sup>。数据库派是技术导向型,研究储存在数据库内的数据。管理派是业务导向型,研究发送给信息用户的信息。近来很多研究人员正在综合以上两种研究方向,以提高信息质量<sup>[10]</sup>。

本研究的目标是回顾信息质量研究的当前状况并指明未来研究的方向。下面从信息质量评估、信息质量管理与情景信息质量三大领域介绍当前信息质量研究的最新动态以及未来信息质量研究的发展。

本文的组织结构是:首先是本研究的概述与范围。随后的三节分别探讨信息质量评估、信息质量管理与情景信息质量。最后一节总结了信息质量研究的难点并深入研究了其发展方向。

## 2 探讨方法

大部分信息质量研究都会着眼于解决如下研究问题:(1)如何评估信息质量?(2)如何管理信息质量?(3)信息质量如何影响组织背景。因此,本文集中探讨了信息质量研究的3个领域:信息质量评估、信息质量管理与情景信息质量。信息质量评估包含3个关键部分:信息质量问题、信息质量维度与信息质量评估方法学。信息质量管理融合了管理的3个领域:质量管理、信息管理和知识管理。情景信息质量关注信息质量对各种组织背景,如电子商务、医疗卫生与会计的影响。本文重点讨论两个引用最多的背景:信

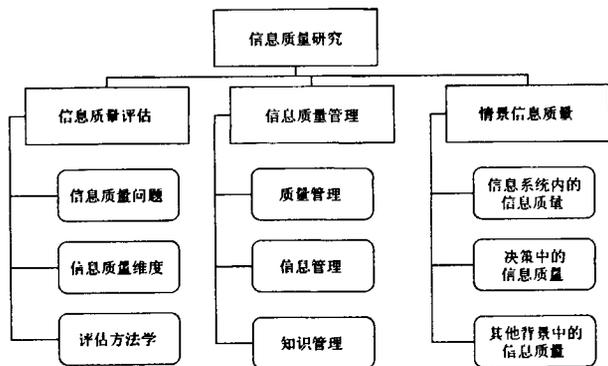


图1 本文基本结构

息系统和决策。图1是本文的基本结构。

本文综述范围划分为:(1)信息质量问题的确认与分类;(2)信息质量范围的确认、定义、分类与相关性;(3)信息质量评估方法学的评估;(4)从质量管理、信息管理与知识管理角度概述信息质量管理;(5)不同背景下的信息质量研究总结;(6)信息系统的信息质量研究;(7)决策中的信息质量研究。概述以上各个部分后,我们将介绍当前信息质量研究的问题及其未来研究方向。

## 3 信息质量评估

与常规的评估定义相呼应<sup>[11]</sup>,信息质量评估可定义为“根据给定的设置,为信息质量维度分配数字值或类别值的过程”。根据与信息质量评估有关的文献,我们将信息质量评估分为3个层次:度量层、维度层和方法层。度量层包括表示不同信息质量问题的信息质量度量。这些信息质量问题可根据“2个背景视角×2个评估视角”模型分类。维度层是由信息质量维度组成,体现了信息的特征。这些信息质量维度与对应的信息质量度量相关联。一个维度可以与多个度量相关联,而且一个度量也可以与多个维度相关联。比如,准确性(信息质量维度)可以与错误的信息以及过时的信息相关联(信息质量度量)。过时的信息(信息质量度量)可以与准确性(信息质量维度)和时间性(信息质量维度)相关联。如果一个信息质量度量与多个信息质量维度相关联,则应定义这些信息质量维度的相关性。方法层包含信息质

量评估模型、框架与方法学。本层的组成部分采用一组信息质量维度来测量信息质量。整体来说,信息质量评估方法使用的一组信息质量维度与不同的信息质量度量相关联。图 2 总结了上述讨论。

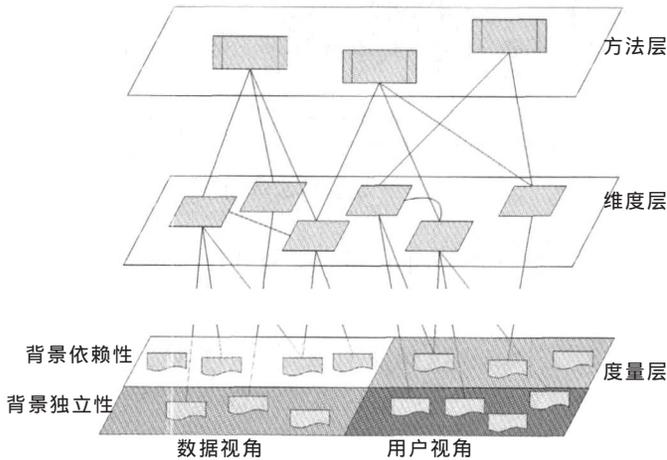


图 2 信息质量评估框架

有关信息质量问题的确认可以参阅若干文献。比如,文献 [12] 就指出了 3 种类型的信息质量问题:偏见信息、过时信息与修改信息。偏见信息是指信息内容不精确或在转换过程中失真。过时信息是对相关任务而言不是足够新的信息。修改信息是指对同样的信息有不同的体现。文献 [13] 将信息质量问题按产品与过程角度分类。产品角度集中在信息自身缺陷上,比如信息不完整、不一致,而过程角度则将重点集中在信息产生与分配过程中引起的问题。根据与信息质量问题有关的文献,我们将信息质量问题按  $2 \times 2$  概念模型分类。矩阵的列捕捉数据视角和用户视角的信息质量问题,而行则捕捉背景相关性与背景独立性的信息质量问题。利用此模型,在表 1 中将常见的信息质量问题进行分类。此表列出了大部分相关的研究文献,这些文献全面探讨了信息质量问题。

上表中 4 个象限说明:

(1) 数据视角/背景独立性象限说明数据库内的信息质量问题。这些信息质量问题适用于所有数据组。

(2) 数据视角/背景相关性象限说明违反业务规范的信息质量问题。利用背景规则可以发现—7—万方数据

表 1 信息质量问题分类

	数据视角	用户视角
背景独立性	拼写错误 <sup>[9,14]</sup> 丢失数据 <sup>[9,14]</sup> 复制数据 <sup>[9,14]</sup> 错误值 <sup>[9,12,14]</sup> 不一致数据格式 <sup>[9,13-14]</sup> 过期数据 <sup>[9,12,14]</sup> 不完整数据格式 <sup>[9,13]</sup> 违反句法 <sup>[9]</sup> 违反唯一值 <sup>[9]</sup> 违反完整性约束条件 <sup>[9]</sup> 文本格式 <sup>[9,12]</sup>	信息不可用 <sup>[2]</sup> 信息不安全 <sup>[2]</sup> 信息难以检索 <sup>[2]</sup> 信息难以聚合 <sup>[2]</sup> 信息转换中的错误 <sup>[2]</sup>
背景相关性	违反域限制 <sup>[9,13]</sup> 违反组织业务规则 <sup>[9,13]</sup> 违反公司与政府法律法规 <sup>[9,13]</sup> 违反数据库管理员提出的限制条件 <sup>[9]</sup>	信息不以事实为基础 <sup>[2,12]</sup> 信息存在可信度问题 <sup>[2]</sup> 信息与工作无关 <sup>[2]</sup> 信息包含不一致的含义 <sup>[2,12-13]</sup> 信息难以利用 <sup>[2,14]</sup> 信息难以理解 <sup>[2]</sup>

这些信息质量问题。

(3) 用户视角/背景独立性象限说明信息处理过程中可能发生的信息质量问题。

(4) 用户视角/背景相关性象限说明不适于信息用户特定用途的信息质量问题。

很多解决信息质量问题的方法已经得到应用。从数据视角来看,可以通过数据清洗算法<sup>[15]</sup>、数据挖掘规则<sup>[16]</sup>、统计过程控制<sup>[17]</sup>或字典匹配惯例<sup>[18]</sup>等方法解决信息质量问题。从用户视角来看,自动化流程不能解决信息质量问题<sup>[14]</sup>。这些问题需要优化资源配置<sup>[19]</sup>、分析业务问题<sup>[20]</sup>、重新设置流程<sup>[17]</sup>,或者将信息流与相应的信息生成系统匹配<sup>[21]</sup>。

### 3.1 信息质量维度

很多研究表明,信息质量属于多维度概念<sup>[2,6,17,22-23]</sup>。在过去 20 年中,从数据库与管理角度已经确认了多组信息质量维度。现在我们从以下方面探讨信息质量维度:确认、定义、分类与相关性。

#### 3.1.1 信息质量维度确认

文献 [6] 建议采用 3 种方法研究信息质量:直觉法、理论法与经验法。我们利用这些方法分析信息质量维度的确认。直觉法依靠研究人员的经验或特定案例的要求得出信息质量维度。此方法是根据具体的应用背景确认信息质量维度。比如,文献 [24] 利用信息的可得性、准确性、特性、时间性、相关性以及信息总量来评估决策背景下

的信息质量。文献[22]应用准确性、时间性、完整性和连续性建立多输入、多输出信息系统的信息质量缺陷模型。理论法根据数据生成过程的数据缺陷生成信息质量维度。比如,文献[23]利用本体论方法,通过观察现实世界系统与信息系统的不一致来获得信息质量维度。经验法将重点集中在数据是否适于数据用户,从而得出信息质量维度。比如,文献[6]获得了15个对数据用户至关重要的信息质量维度。文献[8]选择了16个信息质量维度,以便向数据用户发送高质量信息。通过以上讨论,我们可以看到使用不同的方法可以得到不同组的信息质量维度。

### 3.1.2 信息质量维度定义

上述3种方法也可以理解为定义信息质量维度的3个方面。直觉法从数据视角定义了信息质量维度。比如,文献[22]将完整性定义为记录下特定变量的所有数值。理论法从现实世界角度定义了信息质量维度。比如,文献[23]将完整性定义为信息系统能够表现所表现的现实世界系统里的每一个有意义的状态。经验法从用户视角定义信息质量维度。比如,文献[23]将完整性定义为数据对于当前任务具有充分的广度、深度和范围。图3表示定义信息质量维度的角度。

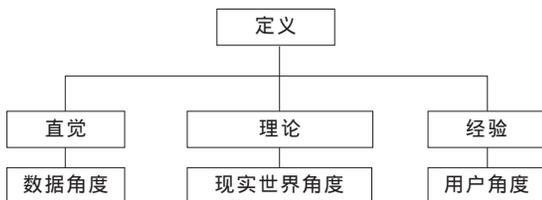


图3 定义信息质量维度的角度和方法

使用数据视角的优势是可以自动评估并客观地控制信息质量。使用现实世界角度的优势是,从理论角度讲,参考质量标准是最佳的选择。但是,以上两种角度都不能达到数据用户的期望值。从用户视角讲,可以使用全组维度测量信息质量。同样,在这个角度,可以通过特定的用途来提高信息质量。但是,此角度不能自动测量信息质量。因此,处理大量数据很困难。此外,收集并综合来自不同数据客户的不同评估结果也有困难。

万方数据

### 3.1.3 信息质量维度分类

根据信息质量维度的确认和定义,推荐使用不同种类的方法对信息质量维度进行分类。文献[6]推荐由4种信息质量类别组成的分级框架,即内在信息质量、情景信息质量、表象信息质量和可访问信息质量。内在信息质量的重点是数据质量本身。情景信息质量强调背景下的信息质量要求。表象信息质量关注信息使用,比如可解释性与可理解性。可访问信息质量是指可以获得安全的信息。文献[23]利用本体论方法获得信息质量维度,并从内外两个角度对其进行分类。内在角度与用户无关,包含一组可在各种应用中进行比较的信息质量维度。外在角度与信息系统的使用与效果有关,该系统代表现实世界系统。文献[25]利用影响信息质量的用户感觉、信息本身以及信息访问过程3个主要因素组织信息质量维度。以上3个要素可视为主题、目标和过程。文献[26]利用符号学以及质量的设计质量与一致性质量两个方面对信息质量维度进行分类。符号学包括句法、语义与语用3个层次。句法层涉及信息的基本表达。语义层强调与现实世界目标有关的信息。语义层解决信息流程与信息用户的问题。文献[8]开发出2×2概念模型,用于描述信息质量维度。其中两行为产品质量与服务质量,两列为符合规范要求以及满足并超过客户期望值。因此,信息质量维度可在4个象限内考虑,即合理、可靠、有用与可用。文献[27]根据信息使用顺序提供了信息质量维度分类。此顺序包括如下4个步骤:获得信息(可访问性)、理解信息(可解释性)、信息适于给定背景(相关性)以及信息没有错误(完整性)。表2对上述探讨进行了总结。

表2 信息质量维度分类

文献	分类方法	类别			
		内在	背景	表象	可访问
文献[6]	分级法	内在	背景	表象	可访问
文献[23]	本体论法	外在	内在	系统相关	数据相关
文献[25]	信息质量元数据源	主题		对象	过程
文献[26]	符号学	句法		语义	语用
文献[8]	产品与服务	合理	可靠	有用	可用
文献[27]	数据使用顺序	可访问性	可解释性	相关性	完整性

### 3.1.4 信息质量维度的相关性

很多文献资料对信息质量维度的相关性进行了分析。文献[28]推荐一种框架,用于研究在

决策背景内准确性与时间性的此消彼长。文献 [17] 指出, 时间性对准确性有一定影响。文献 [29] 为完整性与一致性之间的此消彼长制作了模型。文献 [15] 暗示了准确性与完整性之间的关系, 并声称一致性是准确性的一部分。文献 [30] 分析了多通道信息系统内与时间相关的准确性及完整性。文献 [31] 推荐使用数据驱动方法, 分析句法准确性与时间性之间的相关性以及完整性与时间性的相关性。图 4 显示出上述文献讨论的信息质量维度与结构的重要相关性。

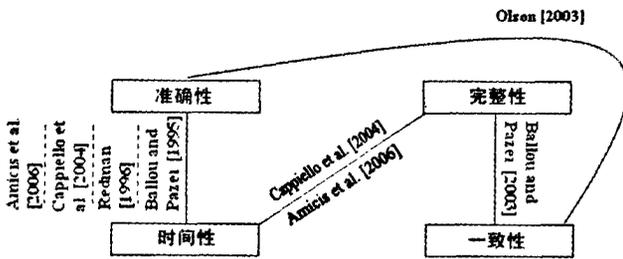


图 4 信息质量维度的相关性

根据上述文献, 我们将信息质量维度的相关性分为两类: 负关联与正关联。负关联是指一个信息质量维度的提高可能导致另一个维度目标值的降低(常称为信息质量维度的此消彼长)。比如, 通过引入新信息提高完整性, 则新引入的信息可能与现有信息不一致。在此方式中, 完整性与一致性为负关联。对于负关联, 我们发现了信息质量维度之间的两种折衷: (1) 信息发送得越快, 检查其他信息质量维度的可用时间就越少; (2) 如果使用新信息提高某信息质量维度, 则新信息可能导致其他维度目标值的降低。正关联是指两个信息质量维度呼应, 并共有一组信息质量问题。比如, 如果时间性与准确性共有的信息质量问题是过时数据, 则提高时间性可能导致准确性值的提高。在此方式中, 时间性与准确性为正关联。根据上述讨论, 我们在表 3 中总结了信息质量维度的关联性。

### 3.2 信息质量评估方法学

过去 10 年出现了很多信息质量评估方法学。我们选择其中 5 种常见的方法学<sup>[2, 10, 17, 32-33]</sup>, 从信息质量维度定义、信息质量维度分类、模型、工具与案例研究等方面进行评估。维度定义是指确定信息质量维度的视角。信息质量维度

表 3 信息质量维度相关性

负关联	正关联
<p>信息发送得越快, 检查其他信息质量维度可用的时间越少: 时间性与其他信息质量维度之间以及准确性与其他信息质量维度之间的此消彼长。如果使用新信息提高某信息质量维度, 则新信息可能导致其他维度目标值的降低: 完整性与其他维度、可访问性与其他维度、安全性与其他维度以及相关性与其他维度之间的此消彼长。</p>	<p>信息质量维度 1 信息质量维度 2</p> <p>信息质量问题</p> <p>如果提高信息质量维度 1, 则信息质量维度 2 可能提高或保持相同的质量值。这取决于我们是否解决了共有的信息质量问题。</p>
负关联模型	正关联模型

分类用于比较每种方法学的维度分类。模型可以显示方法学的理论基础。工具用于确认方法学的实施。案例研究的重点是这些方法学的实际可行性。利用上述标准, 我们可以获得每种方法学的特征。

如果方法学只适用于一个信息质量群体, 则可视其为特定的方法学。如果方法学可用于两个信息质量群体, 则其属于通用方法学。如果文献提供了案例研究, 则我们视其为实际研究, 否则为理论研究。表 4 对 5 种方法学及其特征进行了总结。

文献 [10] 将信息质量评估分为客观评估与主观评估。客观信息质量评估揭示数据库内的信息质量问题, 而主观信息质量评估反映数据用户的需要与体验。我们按此分类, 从客观与主观角度探讨信息质量评估。

客观信息质量评估就是测量信息符合质量规范以及相关参考要求的程度。我们将客观信息质量评估分为两类: 内在信息质量评估与现实世界信息质量评估。内在信息质量评估从数据视角出发, 强调数据库内数据的质量。比如, 文献 [16] 开发了项目频率规则与正则表达式模式, 便于实施内在信息质量自动评估。现实世界评估从本体论角度出发, 其重点是系统设计与数据生成过程中可能发生的信息质量缺陷。比如, 文献 [23] 发现了现实世界状态与信息系统的

表 4 信息质量评估方法学比较

	文献[17]	文献[2]	文献[32]	文献[10]	文献[33]
定义	定义了数据库领域 12 个信息质量维度	定义了管理领域 16 个信息质量维度	定义了两个领域的 15 个信息质量维度	定义了两个领域的 16 个信息质量维度	定义了两个领域的 22 个信息质量维度
分类	概念观点、数据值与表示	文献 [6] Wand 和 Wang 的分类	文献 [8] Kahn 等人的分类	没有分类	文献 [6] Wang 和 Strong 的分类
模型	根据数据统计流程控制编制的循序渐进步骤	采用文献 [23] Wand 和 Wang 的缺陷模型	采用文献 [8] Kahn 等人的 PSP/信息质量模型	此模型将主观与客观评估融合在一起。	模型包括活动类型、信息质量问题与信息质量分类学
工具	DCI 系统	信息质量评估调查	信息质量评估调查	信息质量评估软件	信息质量评估调查
案例研究	Telstra 有限责任公司	电器公司		Global ConsumerGoods 公司 数据产品制造公司	Simple Dublin Core 公司 英语维基百科
总结	特殊、实用	特殊、实用	通用、理论	通用、实用	通用、实用

表达之间的数据映射缺陷。客观信息质量综合评估可以作为比较当前数据值与高质量理想数据值的步骤。

主观信息质量评估就是测量信息适于数据用户使用的程度。数据用户根据其需要和期望值评估信息质量。主观信息质量评估从用户视角出发,其重点是信息的当前质量与用户期望值之间的差别。为了说明客观与主观信息质量评估之间的区别,我们在表 5 中对两者进行了比较。

表 5 客观与主观信息质量评估比较

方法	客观评估	主观评估
项目		
工具	软件	调查
测量目标	数据	信息
标准	规则 模式	用户满意度
过程	自动	用户相关
结果	单个	多个

客观信息质量评估根据一套质量规则,并利用软件自动测量数据库内的数据,而主观信息质量评估则采用调查或访问手段测量数据用户的情景信息。客观信息质量评估可以提供单个评估结果。但是,我们可以从不同的用户获得不同的评估结果。随着客观与主观信息质量评估的发展,研究人员建议将这两种评估方法综合使用。文献 [10] 提出一种框架,综合了客观与主观信息质量评估。文献 [8] 推荐使用 PSP/信息质量模型,在此模型中他们指定了两个质量角度:符合规范要求(客观)以及满足或超过用户的期望值  
万方数据

(主观)。从最近的文献中,我们可以看出客观与主观信息质量评估融合的趋势。但是,相关研究中尚有两个问题有待解决:(1) 如何协调并管理多个主观评估结果?实际上,不同的数据用户极有可能产生不同的评估结果。比如,数据管理者可以获得某些数据,但是经理却无法获得。因此,数据管理者与经理可能得出完全不同的评估结果。(2) 如何协调并管理客观与主观评估结果的差异与不一致。客观与主观结果可能不一致。比如,信息可能具有完整的结构,但其内容却不完整。即从客观角度评价的高质量信息,换成主观角度评价却是低质量信息。

## 4 信息质量管理

由于信息质量评估是信息质量管理的基础,因此信息质量管理的目标是提高信息的有用性与有效性<sup>[14]</sup>。信息质量管理融合了管理的 3 个领域:质量管理、信息管理和知识管理。图 5 表示出 3 个研究领域的融合趋势。

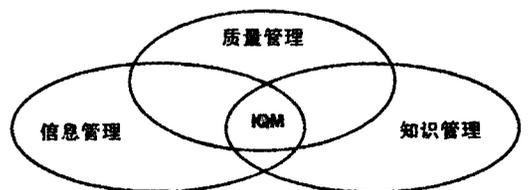


图 5 信息质量管理研究领域的融合

为了分析信息质量管理的当前研究情况，我们选择了 3 种常被引用的文献材料。这些文献分别将质量管理、信息管理与知识管理融入了信息质量管理。

**质量角度：**根据“将信息作为产品来管理”的原则，文献 [20] 建议使用全面数据质量管理 (TDQM) 方法学。该方法包含以下 4 个阶段：定义、测量、分析与提高。TDQM 的目标是为信息用户提供高质量信息产品。

**信息角度：**根据“整合、验证、背景化与激活”的原则，文献 [14] 推荐使用一种包含识别、评估、分配与应用 4 个步骤的框架。该框架的目标是构筑信息质量处理和增值活动。

**知识角度：**根据“知其本、知其法、知其理”的原则，文献 [2] 推荐使用包含提高信息质量、揭示隐含知识、提供有组织的信息 3 个流程的框架。此框架的目标是将高质量信息转化为有组织的知识。

根据信息质量与其他研究领域之间的关系，上述 3 个文献从不同的角度研究了信息质量管理。文献 [20] 将信息作为产品来考虑，从而将信息质量管理与质量管理联系起来。文献 [14] 采用信息使用周期法，将信息质量管理与信息管理联系起来。文献 [2] 利用信息与知识之间的关系，将信息质量管理与知识管理联系起来。每个不同的角度都有其优势。但是，如果将信息质量作为独立的研究领域，则未来研究中需要全面的信息质量管理框架。

## 5 情景信息质量

认识到信息质量的重要性后，研究人员将信息质量理论应用到各种组织背景中。表 6 总结了 1996 - 2006 年间每种应用背景下的代表性出版物。

在上述应用背景中，我们选择了两个引用最多的背景加以分析，即信息系统和决策。

### 5.1 信息质量与信息系统

大部分有影响的信息质量研究都起始于信息系统研究。信息系统研究人员最初发现并使用一组维度来解决信息系统内的信息质量问题。我

表 6 信息质量应用背景(1996 - 2006)

应用背景	文献
数据库	文献 [17]
信息管理系统	文献 [34]
会计	文献 [35]
营销	文献 [36]
数据仓库	文献 [37]
决策	文献 [38]
医疗卫生	文献 [39]
企业资源规划	文献 [40]
客户关系管理	文献 [41]
金融	文献 [42]
电子商务	文献 [43]
万维网	文献 [44]
供应链管理	文献 [45]

们选择了 12 篇引用最多的文章。在图 6 中，12 篇文章沿时间轴排列。根据文章的研究目标和探讨重点来区分有关信息系统与信息质量的作品。比如，文献 [46] 使用 5 个信息质量标准——即准确性、时间性、相关性、聚合性与格式，以评估信息系统的价值。文献 [47] 利用准确性和外观来评估信息系统。随着信息质量意识提高并认识到信息质量要求，研究人员开始把重点放在信息质量框架<sup>[22, 48]</sup>、信息质量维度<sup>[6, 10]</sup>、信息质量评估<sup>[23, 37]</sup>、以及信息质量管理<sup>[2, 20]</sup>等领域。如图 6 所示，随着信息质量研究作为独立研究领域的成熟，信息质量理论又返回到信息系统背景应用中。

图 6 用图解法表示引用数量（比如，点越大，文章被引用的次数越多）。我们可以看到，这里选择的所有信息质量文章都引用了文献 [22] 的作品。作者 Ballou 与 Pazer 对信息质量研究做出了开拓性的贡献，而且两人的文章对信息系统研究领域里的信息质量研究产生了巨大影响。文献 [49] 的文章探讨了信息系统文献资料，并被以后的信息质量研究引用。作者 Delone 和 Mclean 完成了信息质量与信息系统之间的联系工作。此外，在 1995 年前后也显示出信息质量研究扩大的端倪。

### 5.2 信息质量与决策

在考虑影响决策的其他因素时，很多研究人员指出了信息质量对决策的影响。文献 [50] 考虑了信息过载问题，推荐了一种模型，可以确定信息质量以及信息数量如何影响决策效果。文献 [51] 采用社会相互影响与决策辅助，暗示社会相

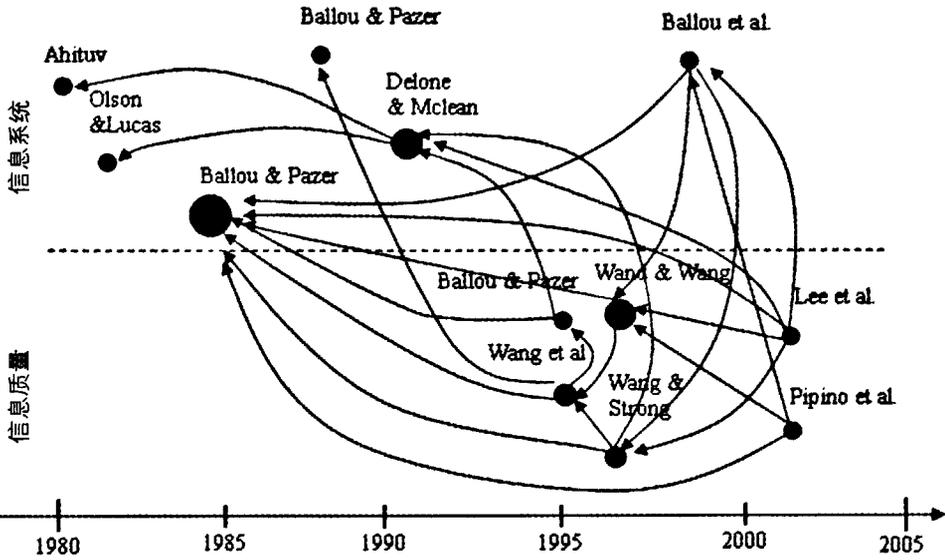


图6 信息质量研究与信息系统研究的引证关系

互影响与决策辅助的主要目的就是提高信息质量,并据此提高决策质量。文献[52]基于危机决策环境与决策辅助,推荐一个模型,可以表示信息质量与决策之间的关系。文献[28]从决策策略及其成本角度出发,分析了决策中两个信息质量维度(准确性与时间性)的折衷。文献[38]考虑任务的复杂性与决策策略,指出有关信息质量的信息可以影响到决策过程。文献[53]考虑时间、专业技能、任务复杂性以及决策策略的影响,开发出一种实验手段,可以解决信息质量信息在决策中的效用问题。文献[54]针对动态决策环境,推荐一种虚拟业务环境,用于解决信息质量管理在动态决策环境中的作用。文献[55]从任务复杂性与信息质量分类角度考虑,发现情景信息质量对决策表现作用巨大。通过以上文献,我们可以看到,在研究信息质量对决策的影响时,研究人员会考虑各种因素。也就是说信息质量影响到各种场景下的决策。通过控制各种影响因素,就可以开发出各种决策场景。根据对相关文献及其影响因素的说明,表7对上述探讨内容进行了总结。

通过上述表格,我们可以发现,在研究信息质量对决策的影响方面,存在以下两个研究空白:(1)需要控制信息质量实验研究中的无关变量;(2)相互作用与信息表达是影响决策的两个重要因素。在研究信息质量对决策的影响时,以上两点应作为独立变量进行考察。

表7 信息质量研究影响决策的因素

作者 \ 因素	信息质量	信息过载	决策辅助	决策策略	任务复杂性	专业技能	时间	环境
Keller 与 Staelin <sup>[50]</sup>	×	×						
Sage <sup>[51]</sup>	×		×					
Ballou 与 Pazer <sup>[28]</sup>	×	×						×
Belardo 与 Pazer <sup>[52]</sup>	×	×	×					×
Ahituv 等人 <sup>[56]</sup>	×					×	×	
Chengalur - Smith 等人 <sup>[38]</sup>	×			×	×			
Shankaranarayan 等人 <sup>[54]</sup>	×							×
Fisher 等人 <sup>[53]</sup>	×			×	×	×	×	
Jung 与 Olfman <sup>[55]</sup>	×				×			

## 6 总结与结论

本篇信息质量研究综述揭示了与信息质量评估、管理和情景信息质量相关的潜在研究问题。在本节中,我们将简要介绍信息质量研究中可填补的空白、研究方案以及未来发展方向。

我们分析了信息质量评估的3个部分:信息质量问题、信息质量维度与信息质量评估方法学。分析结果显示,一种背景中的高质量数据在另一种背景中可能就是低质量数据。某人认为的高质量数据可能被其他人视为是低质量的。符合一种规范的高质量数据根据其他规范的要求则可能是低质量数据。因此,大部分研究人员和研究机构需要面对的主要研究问题是:如何有效地评价和评估信息质量。为了解决这个问题,又产

生出这样的问题：如何发现可能的信息质量问题；如何定义并选择信息质量维度以及如何联系信息质量维度与信息质量问题。为了解决这些问题，我们提出以下有关信息质量评估的研究问题和方向(表 8)。

表 8 有关信息质量评估的研究问题

有关信息质量评估的研究问题
研究问题 1: 如何有效地评估信息质量?
研究问题 1a: 如何发现可能的信息质量问题?
研究问题 1b: 如何定义并选择信息质量维度?
研究问题 1c: 信息质量问题和信息质量维度的关系是什么?
研究问题 1d: 信息质量维度彼此之间的相关性如何? 如何处理信息质量维度的相关性?
研究问题 1e: 如何为相关机构选择最合适的信息质量评估方法学?

信息质量管理融合了管理的 3 个领域：质量管理、信息管理和知识管理。这种有关信息质量管理的泛泛观点存在缺陷，即难以在相关机构实施信息质量管理。因此，在信息质量管理的实证应用中，相关机构依然面对如何有效管理信息质量的问题。沿着这个问题的脉络，相关机构需要了解信息质量管理的成本和效益，需要了解如何实施信息质量管理以及如何构建相关机构的信息质量文化。因此，未来的研究需要信息质量管理的全面框架。在此，我们提出以下有关信息质量管理的研究方向和问题(表 9)。

表 9 有关信息质量管理的研究问题

有关信息质量管理的研究问题
研究问题 2: 如何管理相关机构内的信息质量?
研究问题 2a: 如何分析信息质量管理的成本和效益?
研究问题 2b: 如何评估信息质量管理的成熟程度?
研究问题 2c: 如何在相关机构内实施信息质量管理?
研究问题 2d: 如何培养相关机构内的信息质量意识以及建设信息质量文化?

信息质量理论已经在各种应用背景中得到应用。概述这些应用背景后，为了分析情景信息质量研究，我们选择两类引用最多的背景：信息系统和决策。由于信息质量研究大部分始于信息系统研究，因此了解信息质量与信息系统之间的关系对指导以后的信息质量研究很有价值。我们的分析显示出信息质量研究与信息系统研究的引证关系。但是，我们仍需要对以上两个领域进行全面的分析。在有关决策背景的信息质量研究中，我们提出研究人员要考虑不同的影响因素。

相关研究已经提出了信息质量与决策之间的关系。但是，我们现在需要解决的问题依然是如何在信息质量实证过程中控制这些影响因素。这个问题暗示，控制无关变量对信息质量实证设计至关重要。根据上述讨论，我们总结了有关情景信息质量的研究难点与问题(表 10)。

表 10 有关情景信息质量的研究问题

有关情景信息质量的研究问题
研究问题 3: 信息质量与应用背景的关系是什么?
研究问题 3a: 信息质量如何影响应用背景?
研究问题 3b: 信息质量研究和信息系统研究的关系是什么?
研究问题 3c: 如何控制信息质量实验中的无关变量?

根据对信息质量研究的探讨，我们总结出以下若干结论。

(1) 相关机构需要有效的信息质量评估解决方案。根据对信息质量评估的探讨，我们发现信息质量评估有很重要的 3 个组成部分，即信息质量问题、信息质量维度与信息质量评估方法学。

(2) 相关机构需要全面的信息质量管理框架。我们的探讨涉及信息质量管理的 3 个方面。进一步的信息质量管理需要认识质量管理、信息管理和知识管理。

(3) 我们需要研究信息质量与组织背景的关系。本文概括介绍了情景信息质量研究。在此，特别强调两个被引用最多的背景：信息系统和决策。为了便于信息质量管理，相关机构需要了解信息质量对应用背景的作用。

总之，我们的研究表明，信息质量正成为重要的热门研究课题。它发源于各种研究领域。目前，与信息质量相关的研究正与其他研究领域相融合。我们的综述概括介绍了这个领域的来源与最前沿的研究，同时也介绍了未来研究可能面对的困难。过去 10 年中，研究人员在模型、框架、评估与管理手段等方面贡献巨大。但是，我们的研究表明，目前依然还有很多开放的研究问题。本文概括介绍了这些研究问题。相关结果指明了这个领域未来的研究方向，并对学术界和相关行业确定未来研究项目具有一定的价值。总之，我们的研究表明，信息质量问题和挑战十分普遍，意识到信息质量问题对学术界与相关行业的很多研究领域都很重要。

参考文献

- [1] Wang R Y ,Lee Y W ,Ziad M. Data Quality[M]. Springer. 2001.
- [2] Huang K T ,Lee Y W ,Wang R Y. Quality Information and Knowledge Management[M]. 1st edition. Prentice Hall, 1999.
- [3] Fisher C W , Kingma B R. Criticality of Data Quality as Exemplified in Two Disasters[J] . Information and Management, 2001 39 (2): 109 – 116.
- [4] Eppler M , Helfert M. A Classification and Analysis of Data Quality Costs[C]//Proceedings of the 9th International Conference on Information Quality. USA: MIT, 2004.
- [5] Juran J M, Gryna F M , Bingham R S. Quality Control Handbook[M] . 3rd edition. New York: McGraw – Hill, 1974.
- [6] Wang R Y , Strong D M. Beyond Accuracy: What Data Quality Means to Data Consumers[J] . Journal of Management Information System ,1996 12(4) 5 – 34.
- [7] Klein B D, Goodhue D L ,Davis G B. Can Humans Detect Errors in Data? Impact of Base Rates, Incentives, and Goals[J]. MIS Quarterly ,1997 21 (2) 169 – 194.
- [8] Kahn B, Strong D ,Wang R. Information Quality Benchmarks: Product and Service Performance[J] . Communications of the ACM 2002 45(4) 184 – 192.
- [9] Oliveira P, Rodrigues F ,Henriques P. A Formal Definition of Data Quality Problems[C] //Proceedings of the 10th International Conference on Information Quality. USA MIT, 2005.
- [10] Pipino L ,Lee Y W, Wang R Y. Data Quality Assessment [J] . Communications of the ACM ,2002 45(4) :211 – 218.
- [11] Gertz M, Ozsu T, Saake G, Sattler K. Report on Dagstuhl Seminar Data Quality on the Web[R]. SIGMOD Report. 2004.
- [12] Garvin D A. Managing Quality[M]. Free Press, 1988.
- [13] Lesca H Lesca E. Gestion de l'information: qualité de l' information et performances de l' entreprise[C] . LITEC, Les essentiels de la gestion, 1995.
- [14] Eppler M J. Managing Information Quality[M]. 2nd edition. Springer, 2006.
- [15] Olson J, E. Data Quality: The Accuracy Dimension[M]. Morgan Kaufmann, 2003.
- [16] Savchenko S. Automating Objective Data Quality Assessment[C] // Proceedings of the 8th International Conference on Information Quality. USA: MIT, 2003.
- [17] Redman T. Data Quality for the Information Age[M] . Artech House, 1996.
- [18] Strong D, Lee Y ,Wang R Y. Data Quality in Context[J]. Communications of the ACM, 1997 40 (5) 103 – 110.
- [19] Ballou D P , Tayi G K. Methodology for Allocating Resources for Data Quality Enhancement[J] . Communications of the ACM ,1989 32(3) 320 – 329.
- [20] Wang R Y. A Product Perspective on Total Data Quality Management[J]. Communications of the ACM ,1998 41 (2) 58 – 63.
- [21] Wang R Y, Lee Y W, Pipino L ,Strong D M. Manage Your Information as a Product[J] . Sloan Management Review ,1998 39(4) 95 – 105.
- [22] Ballou D Pazer H L. Modeling Data and Process Quality in Multi –input, Multi –output Information Systems[J]. Management Science ,1985 31(2) :150 – 162.
- [23] Wand Y ,Wang R Y. Anchoring Data Quality Dimensions in Ontological Foundations[J]. Communications of the ACM ,1996 39(11) 86 – 95.
- [24] O' Reilly III C A. Variations in Decision Makers Use of Information Source: The Impact of Quality and Accessibility of Information[J] . Academy of Management Journal ,1982 25 (4) 756 – 771.
- [25] Naumann F , Rolker C. Assessment Methods for Information Quality Criteria[C]//Proceedings of the 5th International Conference on Information Quality. USA: MIT, 2000.
- [26] Helfert M. Managing and Measuring Data Quality in Data Warehousing[C] //Proceedings of the World Multiconference on Systemics. Cybernetics and Informatics. Florida: Orlando, 2001.
- [27] Bovee M, Srivastava R P ,Mak B. A Conceptual Framework and Belief –function Approach to Assessing Overall Information Quality International[J] . Journal of Intelligent Systems 2003 18(1) 51 – 74.

- [28] Ballou D P ,Pazer H L. Designing Information Systems to Optimize the Accuracy –Timeliness Tradeoff[J] . Information Systems Research ,1995 6(1) 51 – 72.
- [29] Ballou D P ,Pazer H L. Modeling Completeness versus Consistency Tradeoffs in Information Decision Contexts[J] . IEEE Transactions on Knowledge and Data Engineering 2003 15(1) 240 –243.
- [30] Cappiello C, Francalanci C ,Pernici B. Time –Related Factors of Data Quality in Multichannel Information Systems[J] . Journal of Management Information Systems 2004 20 (3) 71 –91.
- [31] Amicis F D, Barone D ,Batini C. An Analytical Framework to Analyze Dependencies among Data Quality Dimensions[C]//Proceedings of the 11th International Conference on Information Quality. USA: MIT, 2006.
- [32] Lee Y, Strong D, Kahn B, Wang R Y. AIMQ: A Methodology for Information Quality Assessment[J] . Information & Management 2002 40(2) 133 –146.
- [33] Stivilia B, Gasser L, Twidale M B ,Smith L C. A Framework for Information Quality Assessment[J] . Journal of the American Society for Information Science and Technology 2007 58(12): 1720 –1733.
- [34] Ballou D P, Wang R Y, Pazer H ,Tayi G K. Modeling Information Manufacturing Systems to Determine Information Product Quality[J] . Management Science, 1998 44(4) 462 –484.
- [35] Kaplan D, Krishnan R, Padman R, Peters J. Assessing Data Quality in Accounting Information Systems[J] . Communications of the ACM ,1998 41(2) 72 –78.
- [36] Teflian M. Information Liquidity[M] . MA: Cambridge, 1999.
- [37] English L P. Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits[M]. Wiley, 1999.
- [38] Chengalur – Smith I N, Ballou D P ,Pazer H L. The Impact of Data Quality Information on Decision Making: An Exploratory Analysis[J] . IEEE Transactions on Knowledge and Data Engineering , 1999 , 11(6) : 853 –864.
- [39] Berndt D J, Fisher J W, Hevner A R , Studnicki J. Healthcare Data Warehousing and Quality Assurance [J]. Computer, 2001 34(12) 56 –65.
- [40] Xu H, Horn J, Brown N ,Nord G D. Data Quality Issues in Implementing an ERP[J] . Industrial Management & Data Systems 2002 102 (2) :47 –59.
- [41] Helfert M , Heinrich B. Analyzing Data Quality Investments in CRM –A Model –based Approach[C] //Proceedings of the 8th International Conference on Information Quality. USA: MIT, 2003.
- [42] Amicis F D , Batini C. A Methodology for Data Quality Assessment on Financial Data[J] . Studies, In Communication Sciences 2004 4(2) 115 –137.
- [43] Xu H ,Koronios A. Understanding Information Quality in E –Business[J] . Journal of Computer Information Systems 2004 45(2) 73 –82.
- [44] Knight S, Burn J. Developing a Framework for Assessing Information Quality on the World Wide Web[J] . Informing Science Journal, 2005 8 ) 160 –172.
- [45] S H Li, B S Lin. Accessing Information Sharing and Information Quality in Supply Chain Management[J] . Decision Support System, 2006, 42(11): 1641 –1656.
- [46] Ahituv N. A Systematic Approach toward Assessing the Value of an Information System[J] . MIS Quarterly , 1980 4(4) 61 –75.
- [47] Olson M H ,Lucas H C. The Impact of Office Automation on the Organization: Some Implications for Research and Practice[J]. Communications of the ACM ,1982 25 (11) 838 –847.
- [48] Wang R Y, Storey V C , Firth C P. A Framework for Analysis of Data Quality Research[J] . IEEE Transactions on Knowledge and Data Engineering, 1995 7(4) : 623 –640.
- [49] DeLone W H ,McLean E R. Information System Success: The Quest for Dependent Variables[J] . Information System Research, 1992 3(1) 60 –96.
- [50] Keller K L ,Staelin R. Effects of Quality and Quantity of Information on Decision Effectiveness[J] . Journal of Consumer Research, 1987 14(2) 200 –213.
- [51] Sage A P. Decision Support Systems Engineering[M] . New York: John Wiley and Sons, 1991.
- [52] Belardo S , Pazer H. A Framework for Analyzing the Information Monitoring and Decision Support System

- Investment Tradeoff Dilemma: An Application to Crisis Management[J]. IEEE Transactions on Engineering Management ,1995 ,42(4) :352 – 358.
- [53] Fisher C W, Chengalur – Smith I ,Ballou D P. The Impact of Experience and Time on the Use of Data Quality Information in Decision Making[J] . Information Systems Research 2003 ,11 (2) 853 – 864.
- [54] Shankaranarayan G, Ziad M ,Wang R Y. Managing Data Quality in Dynamic Decision Environments: An Information Product Approach[J] . Journal of Data Management 2003 ,14(4) :14 – 32.
- [55] Jung W ,Olfman L. An Experimental Study of the Effects of Contextual Data Quality and Task Complexity on Decision Performance, Information Reuse and Integration[C] // Conf, 2005. IRI –2005 IEEE International Conference on. USA: CA, 2005 :149 – 154.
- [56] Ahituv N, Igbaria M, Stella A. The Effects of Time Pressure and Completeness of Information on Decision Making[J] . Journal of Management Information Systems ,1998 ,15(2) :153 – 172.

## Status Quo and Future Agenda in Information Quality Research

Ge Mouzhi, Markus Helfert

( School of Computing, Dublin City University, Ireland 5442 )

Assessor: Su Ying

(Institute of Scientific and Technical Information of China, Beijing 100038)

**Abstract:** Recognizing the substantial development of information quality research, this review article analyzes three major aspects of information quality research: information quality assessment, information quality management and contextual information quality. Information quality assessment is analyzed by three components: information quality problem, dimension and assessment methodology. Information quality management is analyzed from three perspectives: quality management, information management and knowledge management. Following an overview of contextual information quality, this article analyzes information quality research in the context of information system and decision making. The analyzing results reveal the potential research streams and current research limitations of information quality. Aiming at bridging the research gaps, we conclude by providing the research issues for future information quality research and implications for empirical applications.

**Keywords:** information quality research, information quality assessment, information quality management, contextual information quality