

ISO 8000 (大)数据质量标准及应用

张坦¹, 黄伟^{1,2}, 石勇³

1. 西安交通大学管理学院, 陕西 西安 710049;
2. 昆士兰科技大学, 澳大利亚 昆士兰州 布里斯班市 4702;
3. 中国科学院大学管理学院, 北京 100190

摘要

阐述了ISO 8000数据质量标准产生的背景, 将ISO 8000与已有的产品质量标准ISO 9000进行对比, 分析了实施ISO 8000可以为企业带来的优势, 并举例介绍了ISO 8000的应用方式。分析了我国数据质量现状, 并对我国在金融和医疗领域实施ISO 8000 数据质量标准提出了实施建议。旨在帮助我国在大数据、数据质量等前沿研究和工业领域找到国际化标准契合点, 推进大数据和数据质量建设。

关键词

大数据质量 ; ISO 8000 ; 大数据质量标准

中图分类号 : C931.6

文献标识码 : A

doi: 10.11959/j.issn.2096-0271.2017001

ISO 8000 (big) data quality standard and application

ZHANG Tan¹, HUANG Wei^{1,2}, SHI Yong³

1. School of Management, Xi'an Jiaotong University, Xi'an, Shanxi 710049, China
2. Queensland University of Technology, Brisbane, Queensland 4702, Australia
3. School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China

Abstract

Background of ISO 8000 was introduced, and ISO 8000 was compared with existing product quality standard ISO 9000. After that, the advantages that ISO 8000 can bring to the enterprise were analyzed. Then, the application of ISO 8000 with a simple example was demonstrated. The current situation of data quality in China was analyzed, and some suggestions on implementing ISO 8000 data quality standards in financial and medical fields were put forward. It aims to promote the construction of large data and data quality.

Key words

big data quality, ISO 8000, big data quality standard

1 引言

如今,人们正处于大数据时代。随着计算机技术的发展,个人和组织需要处理越来越多的数据。这些海量数据产生自人、物以及他们之间的相互作用。大数据不是指数据量大,而是搜索、聚合和交叉引用的能力大。组织在跨业务、跨系统或跨组织利用信息时经常会遇到困难,当组织相信其拥有的数据开展某项业务而事实上却不可以开展时,就会产生挫败感。公司希望进行某些趋势分析,或者与顾客或合作者建立更紧密的关系,信息技术部门却无法为顾客提供其所需的整合信息,无法在顾客要求的时间范围内为其提供所需信息,公司因此可能会失去利用这些已经收集和整理信息的机会。高质量的数据有助于管理者高效地做出最优决策,而低质量数据(如不完整的、冗杂的、错误的信息)将会极大地影响决策者的判断,造成低效的资源分配和利用,为个人或组织带来巨大的损失。什么样的数据才是高质量的数据?研究者从不同角度对数据质量做出了描述和定义。从信息技术角度来看,数据质量可以从数据的精确性、完整性、一致性、及时性等维度来衡量,高质量的数据精确度高、数据完整、与其他数据有相互印证的一致性,并能及时更新,保证数据的及时性;从用户角度来看,数据质量的高低取决于用户是否可以高效地使用数据,不同用户对数据质量的要求不同,需要针对其需求制定适合用户使用的数据,适合用户使用的数据即高质量的数据,不适合用户使用的数据为低质量数据,从这一角度来看,数据质量不仅包括信息技术领域讨论的精确性、完整性、一致性、及时性等方面,也包括数据的表现形式是否可以被用

户理解、数据是否容易获取、数据量是否足量或过量等。Lee Y等人深度调查了多个企业的数据库质量项目,并总结出数据库质量低下的问题,具体如下。

- 多数据源。当同样一个信息拥有多个来源时,会产生不同数值,这包括数值在某个给定的时间点是准确的,如研究某个股票的价格,需要明确该股票在对应时间点上的数值,脱离了时间限定的股票价格无法讨论其准确程度。
- 信息产生过程中的主观判断。信息产生的过程中如果有主观判断存在,则会导致结果含有偏见因素信息。
- 有限的计算资源。缺乏足够的计算资源,将限制相关信息的可访问性。
- 安全性和可访问性的权衡。信息的可访问性会与安全、隐私和保密的要求发生冲突。
- 跨学科的编码数据。来自不同的部门和学科的编码数据的识别和理解很困难,这些代码之间也可能会发生冲突。
- 复杂数据的表示。到目前为止,还没有一种算法可以自动分析同时具有文本和图像信息的样例,非数字信息很难用索引的方式来定位。
- 数据量过大。当一个数据库的信息存储量过大时,用户难以在一个合理的时间内获取其所需的信息。
- 输入规则过于严格或被跳过。如果输入规则过于严格,则会出现不必要的控制并导致某些重要数据的丢失。数据录入人员可能会跳过某些数据的输入(造成信息缺失),或擅自改变这些数值,使其符合录入规则并通过编辑检查(造成信息错误)。
- 数据需求的改变。当用户的任务和组织环境(如新的市场、新的法律要求、新的趋势)发生变化时,相关的有用信息也会随之改变。

● 分布式异构系统。没有适当整合机制的分布式异构系统,会导致其内部数据的定义、格式、规则和数值的不一致。数据的原本含义可能会在数据的流动过程中丢失或扭曲,随后在不同的系统、时间、地点、用户使用情况下,因为相同或不同的用途而被检索。

目前,不论在制造业、金融业、医疗卫生领域,还是其他各个工业领域,不论企业规模大小,信息或数据质量问题均是工业界普遍认同的问题。面对快速发展的计算机技术以及爆炸式增长的数据量,企业需要关注数据质量问题,赢得先动者优势,从而帮助公司增加市场份额,维持市场领导者地位,对新的商业机遇迅速采取行动,对组织生存的潜在威胁率先采取行动。反过来会转变成企业总收入、净利润的增长和企业形象的提升。由于很难达到挑战的要求,进入壁垒较高,因此企业能更好地维持竞争优势。工业界和学术界对数据质量进行了大量的应用性和学术性研究,其中一个重要的领域就是对数据质量标准的研究与规范,通过制定统一化的数据质量标准,国家之间、组织之间、部门之间可以进行更高效的数据质量统一管理,从而利于数据存储、传递和共享,减少由于数据不兼容、数据冗杂、数据缺失等各种数据质量问题造成的判断和决策失误。国际标准化组织(International Organization for Standardization, ISO)为此也进行了大量努力,专门制定了ISO 8000数据质量标准对国际化数据质量进行规范。

2 ISO 8000数据质量标准的介绍

ISO 8000数据质量标准是针对数据质量制定的国际标准化组织标准,它是

由ISO工业自动化系统与集成技术委员会(TC 184)SC4小组委员会开发的,ISO TC 184/SC4是负责工业数据的国际标准组织,这一标准以一系列文件的形式发布,每个文件被ISO称为“部分”。该组织开发和维护ISO标准,在产品的整个生命周期中描述和管理工业产品数据。它在自动化系统领域制定标准,并在相关的设计、采购、制造和交付、支持、维护和处置产品及其相关集成服务方面进行标准的制定。标准化领域包括信息系统、用于工业和特定非工业环境中的固定和移动机器人技术、自动化和控制软件以及集成技术。

ISO 8000数据质量标准致力于管理数据质量,具体来说,包括规范和管理数据质量活动、数据质量原则、数据质量术语、数据质量特征(标准)和数据质量测试。通过ISO 8000的标准规范,可以保证用户在满足决策需求和数据质量的基础上,在整个产品或服务的周期内高质量地交换、分享和存储数据,从而保证用户可以依托获取的数据高效地做出最优化的安全决策。根据ISO 8000数据质量标准的定义,数据质量高低程度由系统数据与明确定义的数据要求进行对比而得到。ISO 8000数据质量标准的目的是使其更容易定义质量数据,并识别可以提供高质量数据的公司和软件应用程序。满足ISO 8000质量标准的数据是“满足规定要求的便携式数据”,便携式数据是指可以与软件应用程序分离的数据,这一特征非常重要,因为如果数据只能被特定的许可软件应用程序使用或读取,那么数据也必须受许可证的条款约束,这就导致企业自己的数据不属于企业自身,企业对数据的操作可能会受软件许可证条款的限制。使用ISO 8000数据质量标准,企业可以不受某个软件应用的约束,独立地购买或销售高质量的数据,若企业想要长期保存数据,将数据与

软件分离是非常重要的。

ISO 8000数据质量标准由以下4个部分组成,每个部分独立发布,该标准是受版权保护的,不可免费使用。

- 一般原则:第0~99部分。
- 主数据质量:第100~199部分。
- 交易数据质量:第200~299部分。
- 产品数据质量:第300~399部分。

3 ISO 8000数据质量标准的优势

在ISO 8000数据质量标准发布之前,各个国家、地区、不同行业提出过针对自身特点的一些数据质量标准。例如,英国政府出台了政府数据质量的管理标准,该标准文字叙述了各阶层政府工作人员在数据质量管理过程中需要遵循的准则和流程。英国国家医疗服务体系(national health service, NHS)提出了医疗人员在突发事件发生时,需要遵从数据质量标准进行报告,该标准用文字描述和流程评分方式规范了突发事件报告的流程以及报告内信息的准确性和及时性。另外,在空间数据质量、金融数据质量等领域均有相关的数据质量标准提出。然而,并没有一个跨学科、

跨行业的国际化数据质量标准,用详细的技术性解决方案规范信息时代的数据质量问题。而随着经济全球化的发展,跨国家、跨组织的信息交流日益频繁,对数据质量的要求也越来越高。

ISO 8000数据质量标准是一个国际认可的全球性数据质量标准,正是为了统一规范国际性数据质量而产生的。在ISO 8000数据质量标准发布之前,存在部分涉及数据质量标准的较成熟的产品质量标准管理体系——ISO 9000。1987年,国际标准化组织首次发布了ISO 9000,并在之后的多年间进行了多次改进,ISO 9000系列标准与质量管理体系相关,旨在帮助组织确保满足客户和其他利益相关者的需求,同时满足与产品相关的法律法规要求。ISO 9000是一个质量管理的一般性要求和指南,可以帮助组织有效地实施和运行质量管理体系。

ISO 9000主要关注产品的质量,而ISO 8000标准主要关注数据的质量。ISO 8000标准是ISO 9000标准的扩展,旨在改善以数据为主的质量管理体系,ISO 8000标准弥补了ISO 9000标准和数据产品之间的差距。ISO 8000标准保证了用户在质量管理体系,可以达到标准的数据质量要求。此外,虽然已有较成熟的ISO 9000标准来进行产品质量管理,ISO 8000标准仍是必不可少的,因为在没有高质量数据支持的情况下,很难通过单独实施ISO 9000标准来实现高质量的目标。

图1展示了ISO 8000标准与ISO 9000标准以及其他数据产品标准之间的关系。数据产品标准(如ISO 22745)可以用于交换数据模型和格式。ISO 8000标准是在这些标准的基础上开发的,并增加了关于使用这些标准的要求,以确保高质量的数据交换。可以看到,ISO 8000标准填补了ISO 9000标准和数据产品之间的空白。

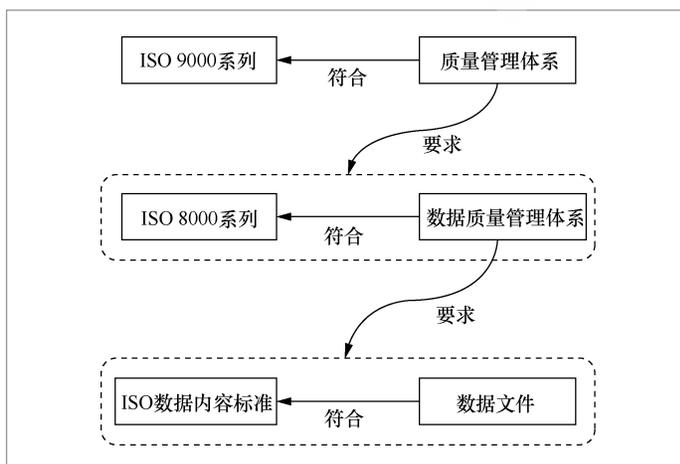


图1 ISO 8000 数据质量标准、ISO 9000 数据质量标准和
其他数据的产品标准之间的关系

与未实施ISO 8000标准的企业相比,实施ISO 8000标准有提高运营效率、优化运营过程中的数据可用性、降低风险、助于改善企业之间的关系、强化品牌形象等优势。具体来说,ISO 8000标准的优势可以在以下4个技术方面的规范中体现。

(1) 实施ISO 8000标准可以提高运营效率

由于组织之间有统一的数据标准,在信息沟通和传递过程中可以进行高效的匹配和识别,减少因为数据不可识别等问题造成的时间成本、编码成本和人工成本。使用清晰的语义编码(如ISO 22745)可以有效地在组织中记录各项条目,这比使用混合和不一致编码标准更有效。根据ISO 8000标准的指导方针,系统中所有项目都使用唯一的代码进行编码。这种做法可以提升添加、删除和搜索某个项目信息的效率。此外,统一编码和语法使计算机更容易存储和处理数据。

(2) 实施ISO 8000标准可以优化运营可用性

通过将ISO 8000标准应用于组织内部,可以对组织内数据进行规范化整合和管理,对各个部门的数据进行统一识别和管理,从组织的整体层面进行资源与信息的协调管理,从而减少因为信息沟通不畅带来的运营成本。如果在合作公司之间或整个行业采用ISO 8000标准,数据或信息将更有可用性。例如,如果部件供应商和汽车制造商都使用ISO 8000标准作为其存储库存和库存信息的数据标准,当汽车制造商需要采购3 000个特定型号的轮胎时,制造商可以采用与供应商共同遵守的ISO 8000数据质量标准发布订单,使采购材料变得更加清楚、简洁、准确,避免由于术语、测量单位等不一致造成的订单错误。如果整个行业都应用ISO 8000标准,

将使信息更加可用,使通信和信息价值链更加畅通无阻。

(3) 实施ISO 8000标准可降低风险

在为公司做出决策时,正确的数据或信息的重要性不言而喻。组织可以基于高质量的数据和信息进行风险评估和决策分析,同时,可以高效地获得标准化的行业数据、市场数据,便于整合多种信息来源,做出更全面的分析和决策。ISO 8000标准为公司提供了一系列说明,使公司能够通过数据出处、准确性、完整性和质量管理框架的标准,生产和维护高质量的数据,使接收者能够跟踪主数据并评估数据的可信度。ISO 8000标准使组织能够保持来自权威数据源的高质量数据,使决策比以前风险更小。

(4) 实施ISO 8000标准有助于改善企业之间的关系,强化品牌形象

应用ISO 8000标准进行数据质量规范和管理的企业之间可以更顺畅、高效地进行沟通,并给对方更好的国际化标准形象,降低了企业间沟通的成本,有利于维护企业之间的关系,强化品牌形象。

4 ISO 8000数据质量标准的组成

ISO 8000数据质量标准包括一般原则、主数据质量、交易数据质量和产品数据质量的标准。ISO 8000标准的结构见表1。本部分主要对ISO 8000标准进行概要性介绍。下文包含了整个ISO 8000数据质量标准的声明,明确了数据质量原则和ISO 8000标准的高级数据架构,描述了ISO 8000数据质量标准的结构,对其他部分内容进行了总结,并描述了ISO 8000标准和其他标准之间的关系。最后,解释了整个ISO 8000数据质量标准中使用的术语,方便读者的阅读和参考。

表1 ISO 8000 数据质量标准结构

ISO 8000数据质量标准	标准内容
一般原则	第1部分 (ISO 8000-1): 简介 第2部分 (ISO 8000-2): 术语
主数据质量	第100部分 (ISO 8000-100): 简介 第110部分 (ISO 8000-110): 主数据的语法、语义编码、符合数据规范 第120部分 (ISO 8000-120): 主数据的数据来源 第130部分 (ISO 8000-130): 主数据的准确性 第140部分 (ISO 8000-140): 主数据的完整性 第150部分 (ISO 8000-150): 主数据的质量管理框架
交易数据质量	未出版
产品数据质量	未出版

4.1 ISO 8000-110主数据的语法、语义和数据规范

即组织和系统之间主数据的通用规则、语法、语义编码和数据规范要求。ISO 8000-110规定了可由计算机检查组织和系统之间交换主要数据(由特征数据组成)的要求,包括关于符合主数据信息的正式语法的要求、主数据信息的语义编码要求以及关于与主数据信息的数据规范一致性的要求。

(1) 语法

ISO 8000标准要求必须有一种语法,语法必须在数据中引用,而且引用必须可解析为语法。语法可以是所有种类的,例如,EDI (ISO 9735)、ebXML、SWIFT

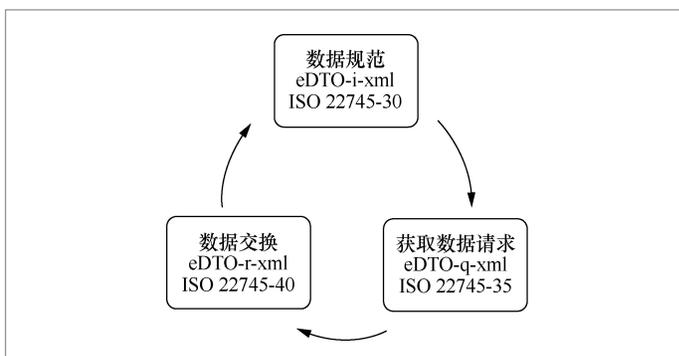


图2 ISO 22745 标准下的数据规范

MT、SWIFT MX、ISO 20022、eOTD-r-xml (ISO 22745) 都是可接受的语法。

(2) 语义编码

所有元数据必须在外部开源字典中显式定义,或者定义必须包含在数据中。任何定义的元数据都是可以接受的,这包括以电子形式、电子表格或数据库表格显式定义的字段、标题或属性。例如,ISO 22745通过使用符合ISO 22745的开放技术词典(例如ECCMA开放技术字典(eOTD))标记数据来创建便携式数据。eOTD就可用于语义编码,它是来自多源术语的注册表,其中每个概念被分配唯一和永久的公共域标识符,概念标识符可以彼此映射并根据它们的使用量来排列。ISO 22745-30是用于说明XML中的数据需求以及用于交换便携式数据的首选标准。

(3) 符合数据规范

语法和语义编码需要支持业务功能,允许访问计算机、网站或软件程序,只需要用所需的数据来提供正确的产品或服务。只要数据满足需求,就认为是高质量数据。ISO 22745可用于以简单的XML格式生成和交换数据请求,发送方和接收方可以自动创建一个集成的数据交换系统(如图2所示)。

4.2 ISO 8000-120主数据的数据来源

数据来源(provenance)对属性值对和数据集的来源信息在表示和交换方面进行了规范和要求,包括数据来源的背景、捕获和交换数据来源信息的要求以及用于数据来源信息的概念数据模型。它允许接收者跟踪已经交换的主数据,并追溯其源头,帮助接收者评估数据的可信性,并且帮助接收者在收到同一属性的不同值时做出值的选择。

4.3 ISO 8000-130主数据的准确性

该标准提出的是数据捕获和数据交换

精度方面的需求,并提出了以声明和担保的形式确保数据准确性的概念模型。数据准确性对属性值对、记录和数据集的准确性信息在表示和交换方面进行了规范和要求,这涉及主数据准确性的情况、捕获和交换主数据准确性信息的要求、主数据准确性信息的概念数据模型。

4.4 ISO 8000-140主数据的完整性

数据完整性对属性值对、记录和数据集的完整性信息在表示和交换方面进行了规范和要求。它包括主数据完整性的情况、捕获和交换主数据完整性信息的要求以及主数据完整性信息的概念数据模型。

4.5 ISO 8000-150质量管理框架

该框架规定了主数据质量管理的基本原则以及对ISO 8000标准的实施、数据交换和出处的要求。它还包含了一个信息框架,用于确定和识别数据质量管理的过程。该框架可以与质量管理体系标准(如ISO 9001)结合或独立使用。

5 ISO 8000数据质量标准的应用举例

以一家汽车制造商应用ISO 8000标

准为其信息系统进行规范为例,展示ISO 8000标准的实施过程。例如,汽车制造商在ISO 8000标准的指令下构建其信息系统。首先,公司选择eXML作为语法,构建初始信息系统。

其次,使用eOTD作为语义编码,将自然语言映射到eOTD,从而将库存数据存储在数据库中。这是一个工作量巨大的任务,因为在ISO 22745标准的要求下,每个项目都应该有一个独特的代码。例如,螺栓的信息可以进行如下编码,见表2和表3。

在建立数据库之后,可以方便地用XML格式编写检索请求。例如要检索某螺母的具体信息,就可用XML调出数据库中存储的ID为0161-1#01-014161#1的产品信息。如果需要查询每包价格在0.80美元以下的螺母数量,就可通过查询语句找到ID为0161-1#02-027378#1,值小于0.80的螺母ID,并得到所有螺母ID对应的产品数量(ID=0161-1#02-027375#1),

表2 螺栓信息的语义编码示例(原始信息)

类别名称	术语	数值	测量单位
螺栓	产品数量	3 225 020 037	个
	标准螺纹直径	1.0	英寸
	宽平面	1.450	英寸
	宽角	1.653	英寸
	头高	0.591	英寸
	每包数量	10	个
	每包价格	0.80	美元

表3 螺栓信息的语义编码示例(编码信息)

类别ID	属性ID	值	测量单位ID
0161-1#01-014161#1	0161-1#02-027375#1	3 225 020 037	
	0161-1#02-023822#1	1.0	0161-1#05-000798#1
	0161-1#02-010200#1	1.450	0161-1#05-000798#1
	0161-1#02-010196#1	1.653	0161-1#05-000798#1
	0161-1#02-004968#1	0.591	0161-1#05-000798#1
	0161-1#02-027376#1	10	
	0161-1#02-027378#1	0.80	0161-1#08-002569#1

并将每个产品数量相加得到螺母总数。该厂商如果发现ID为0161-1#01-014161#1的螺母数量较低需要采购,可以向螺母生产商发送请求订购,请求中引用螺母ID=0161-1#01-014161#1,订购量为ID-0161-1#02-027375#1,值为采购量。

6 推广和实施ISO 8000数据质量标准的现状与展望

如今,大多数组织意识到了内容开放标准的好处。为了建立一个真正的标准,需要允许所有利益相关组织和机构平等参与,参与标准开发是了解世界各地公司的最佳途径,同时可以与国际企业建立密切关系。ISO 8000数据质量标准触及各行业的大、中、小企业的各个业务,无论是高速发展的互联网行业,还是传统制造业、金融业、医疗产业、服务产业等。每个企业、每个人都可以参与并影响标准开发。目前,美国、英国发达国家对于ISO 8000数据质量标准的开发、发展和使用处于领先地位。例如,美国于1999年成立了电子商务编码管理协会(Electronic Commerce Code Management Association, ECCMA),它是一家非营利性国际主数据质量管理者协会,该协会的宗旨是开发和维护开放式解决方案,以更快更好地获取权威主数据。ECCMA成员可以获得数据质量相关的软件和管理培训、ISO 8000数据质量标准实施培训、ISO 8000数据质量标准专业咨询和评估以及对企业信息系统的ISO 8000数据质量标准认证。通过一系列的培训和认证,大大推广了ISO 8000数据质量标准的应用,同时企业可以最大程度地和国际信息质量标准接轨,获得高质量、可共享的数据。英国的MRO Insyte公司为企业提供专业的企业数据和资料管理咨询服务,并

为ECCMA提供咨询。MRO Insyte同时提供主数据管理咨询,并为制造业、零售业和终端用户提供数据质量培训。同时,该公司也为数据、系统和个人提供ECCMA ISO 8000数据质量标准的认证。

我国在ISO 8000数据质量标准的实施和应用方面尚处于起步阶段,目前没有相关的权威推广、培训和认证机构。西安交通大学管理学院与美国麻省理工学院合作,于2012年成立了数据科学和数据质量研究中心,开始对ISO 8000数据质量标准等进行研究。国际化数据质量标准的重要性毋庸置疑,为了在大数据时代占领数据质量的高地,促进我国企业与国际接轨,我国需要加大对ISO 8000数据质量标准的重视程度,组织数据质量研究专家、工业界专家、成熟的数据质量研究中心或机构成立权威的数据质量推广和培训协会,帮助更多国内企业达到ISO 8000数据质量认证水平。另外,现阶段,我国可以首先针对不同行业进行细分推广和应用。例如,在医疗卫生领域,各个医疗机构的信息系统不能很好地兼容,导致同一病人在不同医院的信息无法快速共享和传递。通过全国范围内的应用ISO 8000数据质量标准,可以将病历信息与特定信息系统分离,病历的所有信息可以独立于医疗信息系统存在,并可被任意一个根据ISO 8000数据质量标准的信息系统读取,患者可以更加自主地选择就医医院,而不用担心由于自身的健康信息缺失导致的医疗误判。在金融领域,银行业是一个数据大量聚集的领域,这些来自银行账户交易、贷款申请、贷款还款以及信用卡还款等业务的海量操作的数据库,包含了对客户财务状况有价值的信息。对银行信息系统进行ISO 8000数据质量标准的规范可以帮助银行内部整合各部门之间的数据,并快速识别外部获得的个人信息,对

各渠道、各类型数据进行提取和整合,能够更准确、更全面、更实时地掌握借款人信息,并找到不同变量间的关联关系,使决策更加准确、统一和合理,从而大大提高银行风险管理能力。

总之,ISO 8000数据质量标准在各行各业都可以有广泛和有效的应用。我国应该加快该标准的推广和实施步伐,抓住经济全球化和大数据时代的机遇,参与数据质量标准的全球性应用,并在此基础上积极参与数据质量标准的制定,为我国企业赢得更大更好的国际发展空间。

参考文献:

- [1] MCAFEE A, BRYNJOLFSSON E, DAVENPORT T H, et al. Big data: the management revolution[J]. Harvard Business Review, 2012, 90(10): 61-67.
- [2] LIU L, CHI L. Evolutional data quality: a theory-specific view[C]//The 7th International Conference on Information Quality, November 8-10, Berkeley, California. [S.l.:s.n.], 2002: 292-304.
- [3] LEE Y W, PIPINO L, FUNK J D, et al. Journey to data quality[M]. Cambridge: MIT Press, 2009.
- [4] MADNICK S E, WANG R Y, LEE Y W, et al. Overview and framework for data and information quality research[J]. Journal of Data and Information, 2009, 1(1): 2.
- [5] 蔚赵春, 凌鸿. 商业银行大数据应用的理论、实践与影响[J]. 上海金融, 2013(9): 5.
- WEI Z C, LING H. The theory, practice and influence of big data application in commercial banks[J]. Shanghai Finance, 2013(9): 5.

作者简介



张坦(1989-),女,西安交通大学管理学院信息系统系博士生,主要研究方向为数据质量和文本挖掘。



黄伟(1964-),男,博士,西安交通大学管理学院教授、博士生导师、院长,教育部长江学者特聘教授,美国哈佛大学Fellow和俄亥俄大学商学院管理信息系统系终身正教授。主要研究方向为新一代信息技术的管理沟通(CMC)、群体支持系统(GSS)、大数据管理与数据质量、电子政务/电子商务、IT与服务外包、IT/IS管理。



石勇(1956-),男,中国科学院大学经济管理学院教授、博士生导师,发展中国家科学院院士,国务院参事,中国科学院虚拟经济与数据科学研究中心主任,中国科学院大数据挖掘与知识管理重点实验室主任。被列为Elsevier发布2014/2015中国高被引学者榜单之一、入选汤普生路透(Thomson Reuters)2016全球高被引作者(计算机科学领域)。现在担任《国际信息技术和决策杂志(IJITDM)》《国际数据科学年鉴(Annals of Data Science)》创立主编。先后荣获以集合论创始人命名的康拓学术奖、中国复旦管理学杰出贡献奖、教育部自然科学奖一等奖(2012年、2016年)。中国科学院百人计划学者、国家杰出青年科学基金获得者。

收稿日期: 2017-01-04