

信息化时代企业数据质量管理研究

陈映西¹, 刘解放²

(1. 武汉市城乡建设局 武汉市海绵城市和综合管廊建设管理站, 湖北 武汉 430079;

2. 湖北交通职业技术学院 交通信息学院, 湖北 武汉 430079)

摘要:在信息化飞速发展的时代,数据是企业至关重要的资源,是企业的核心,而数据质量是企业提升业绩和盈利能力的关键。为了确保企业的数质量,从数据维度及其定义、质量标准及其定义和评定方法等方面提出了企业数据质量框架。该框架可用于指导企业获取高质量的数据,同时针对企业不同用户和任务对数据质量的不同要求,提出了数据质量管理模型,为企业开展数据质量管理提供了参考依据。

关键词:数据质量;数据质量框架;数据质量管理模型

中图分类号:F272.1

文献标志码:A

文章编号:1671-9891(2019)03-0020-04

0 引言

随着企业管理和信息技术的迅速发展,信息系统扮演着越来越重要的角色,涉及企业大大小小的业务。然而,在企业信息化建设过程中,由于业务人员操作不当、系统缺陷等原因,出现了种种数据质量问题。为帮助企业解决数据质量管理问题,提高企业生产效率,本文将从数据质量管理的意义入手,提出数据质量框架和数据质量管理模型,为企业开展数据质量管理提供参考依据。

1 企业数据质量管理的意义

良好的数据是企业发展的关键,也是业务发展战略的重要组成部分。数据在所有业务的各个方面都无处不在,例如管理供应链、制造、营销和客户服务。^[1]因此,数据是企业的核心,而数据质量是企业业绩和盈利能力的关键。^[2]企业将数据视为生成多种信息资产的原材料^[3],良好的数据能够使企业获得业务的成功,并从中获取最大的利益和价值,而劣质的数据对企业造成的损失是巨大的,直接影响业务信息系统的实用化水平和系统的决策支持能力,最终影响企业的管理能力,损害企业的经济利益和社会效益。因此,在当今数据驱动的知识经济中,随着数据的不断发展,相关领域(如商务数据交换)数据体量的不断扩大,如何确保理想的数据质量及解决与数据质量相关的问题,亦即数据质量管理,对企业来说至关重要。^[4]

2 数据质量框架构建

本文所构建的综合性数据质量框架包括三个层面:数据质量的维度及维度定义、各维度下相应的数据质量标准及标准定义、各个标准评定方法,如表 1 所示。为了便于信息管理者 and 不同利益相关方的理解和利用,该框架将具有相似定义的 16 个标准分组到 8 个维度集群,每个维度及相应的标准都从数据本身、数据与实际情况、数据用户及数据任务之间关系等不同角度对数据质量进行定义。

各个标准的评定方法包括定量测量和定性测量两种。对于某些标准,既可以用定量评定方法,也可以用定性评定方法来确定数据质量水平。定性评定多以数据用户针对某一标准的反馈来决定,是主观的评定方法。虽然定量评定可以得到一个固定的百分比,但仍应视作主观评定,因为针对不同系统任务,不同的利益相关方对同一类数据的质量期望或要求可能不同。对多数企业而言,数据质量的关键点和难点在于企业内部会同时存在多个信息系统、多个利益相关方和多个数据任务,且同一类数据有可能运用于多个系统中的不同任务,每个任务对数据质量都有不同的要求。因此,文章将针对这种多用户、多任务、多数据的数据质量

收稿日期:2019-06-25

基金项目:教育部科技发展中心高校产学研创新基金重点项目(2018A02005)

作者简介:陈映西(1992—),女,湖北武汉人,武汉市城乡建设局武汉市海绵城市和综合管廊建设管理站助理工程师,硕士。

管理问题,提出一种基本数据质量管理模型。

表 1 数据质量框架

维度	维度定义	标准	标准定义	评定方法
准确性	系统中的数据与事实描述相符	正确性	实体与信息系统中某类数据正确对应	用户用于执行任务的数据无差错;也可计算无错误数据的数量与总数据量的百分比
		及时性	数据应在预期的时间内完成处理和交付	核对不同用户的数据需求以及他们对数据实际反馈的情况
		无歧义	不应存在现实中有两类实体对应系统中的同一类数据的情况	可计算出一一对应的数据量占总数据量的百分比,也可考虑数据用户对无歧义数据的反馈情况
		有意义	信息系统中的每一个数据都应代表现实中的一个实体	根据数据用户对数据是否有意义的反馈,可计算匹配实体的数据量占总数据量的百分比
完整性	数据应能涵盖相应任务的所有相关方面	完整性	信息系统的数据库应能体现实际要求的所有方面	可计算非空值数据量占总数据量的百分比;也可根据数据用户对数据广度深度的反馈进行评定
		相关性	数据应从客观、实际及认识层面被数据用户视作与特定的数据任务相关	根据数据用户对特定任务涉及数据相关性的反馈进行评定
易得性	用户应能从授权的渠道获得其所需数据	易得性	用户可通过授权,不受障碍地获得所需要的数据	根据数据用户对数据访问要求的反馈进行评定
一致性	数据应于实体的所有特性保持一致	一致性	信息系统中数据的格式、定义及数值应于实体保持一致	可计算与实际格式、定义及数值一致的数据量占总数据量的百分比;或根据数据用户对数据一致性的反馈进行评定
		符合度	采用国际或国家标准生成数据	可计算标准化数据占总数据量的百分比
无冗余性	数据应利用最少化的信息资源表现实体	无冗余性	实体在信息系统中不应与超过一个以上的数据单元对应	可计算非重复数据占总数据量的百分比,或根据数据用户对无冗余性的反馈进行评定
可读性	数据应能易于理解且提供足够的信息	易理解性	数据易于用户理解	根据数据用户对数据易理解程度的反馈判断
		可展示性	数据应按合适的方式呈现给用户及相应任务	根据用户对数据展示情况的反馈进行评定
		灵活性	数据能在不同数据层级之间变换展示	根据用户对数据操作情况的反馈进行评定
		透明度	数据应有完整的元数据来记录数据生成、处理有关的信息	可计算有元数据的数据量占整体数据量的百分比
有用性	数据应能为用户带来好处或优势	有用性	数据应能为实现某一目标提供帮助	根据用户对数据是否能帮助其完成某一任务的反馈进行评定
可信度	数据及来源应权威可靠	安全性	数据应足够安全,能防范数据丢失及非授权访问	根据用户对数据安全及可信度的反馈进行评定

3 数据质量管理模型及框架在企业中的应用

3.1 决策支持系统自适应方法

本文所提出的数据质量框架包括若干维度、标准、定义及评定方法,但最终重要的是如何将此框架在企业中得到应用。不同的企业必须做好数据质量配置和治理工作以适应一系列突发事件。^[5-6]同样,不同的业务部门必须重新配置通用的数据质量框架以解决他们的数据质量问题,并且这个过程将延伸至每个可能的业务部门和数据用户。企业设计开发决策支持系统(Decision Support Systems, 简称为“DSS”),并将其与评估和维护数据质量框架进行类比,可更直观地感受企业在不同任务中定制数据质量维度的过程。

为确保信息系统正常运行,必须把数据用户对数据质量标准的反馈和对现实世界的感知放在同等重要的地位,因此,如果没有包含数据用户及其任务,就无法开展数据质量管理。DSS 能通过学习和进化的自适应过程来开展数据质量管理。DSS 用户通过使用来理解系统,并通过探索新见解、新功能实现对 DSS 的优

化。因此,研究人员 Keen 为 DSS 设计了一个自适应框架,其中定义了系统、用户和系统构建者三者之间的关系。DSS 用户根据自身个性化需求来试用系统,了解其运行机制以及系统与用户需求的差距,从而进一步更好地细化其需求,系统构建者则据此进一步完善系统并快速交付用户。用户学习和系统的个性化试用推动了系统和相应机制的完善,通过不断完善用户的功能需求,从而实现整个系统的升级。

3.2 数据质量管理模型

类似 DSS 自我完善的过程被称为自适应方法,同样可以应用于企业数据质量管理,因为维护和控制数据质量的过程同样可视为一种自适应过程。数据质量被认为具有“适用性”,意味着数据质量取决于用户对数据的使用。当我们描述数据质量时,决策任务和决策者的选择等背景因素对数据质量起着重要作用,因此,决策者和相关任务的结合确定了决策环境,换言之,由于决策环境具有动态性,我们需要从决策者及其任务的角度考虑 DSS 的数据质量。此外,由于来自第三方和社交媒体的潜在错误数据将不断涌现,数据质量不是一次性问题,引入错误数据的过程或事件变化并不是唯一的,总会有新的更危险的方法来破坏数据。

数据用户会预先指定他们对数据质量的要求,但当实际执行其数据任务时,也会形成新的或改进对现有数据质量的认知,这些数据也必然会由于各种原因或环境的动态性发生变化。不同用户有不同的需求,从而产生不同的质量标准,因此数据质量管理需要数据质量顾问与数据用户密切合作,以了解他们的数据质量要求,使数据质量达到用户的标准,并得到质量团队和数据用户的持续监控。

本文提出了数据质量管理模型,如图 1 所示。它描述了数据用户、数据质量顾问、数据任务和数据质量之间的关系。数据质量是使用中需要的一项功能,数据用户通过个性化应用执行数据任务了解自身的数据质量要求。数据质量顾问团队应与用户密切合作,了解他们的要求和数据质量维度、标准,据此调整数据质量。数据用户通过个性化应用质量要求促使质量顾问团队根据用户对数据质量认知的变化进行调整。根据所提数据质量框架和数据管理模型,本文给出了数据质量动态优化调整的方法,如图 2 所示。

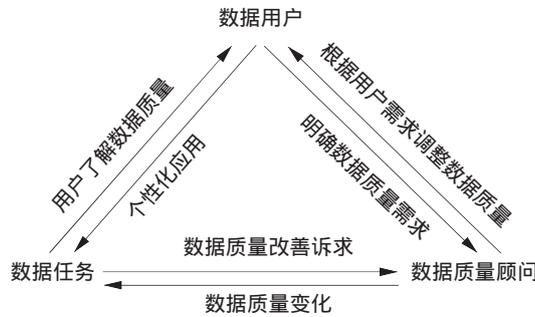


图 1 数据质量管理模型

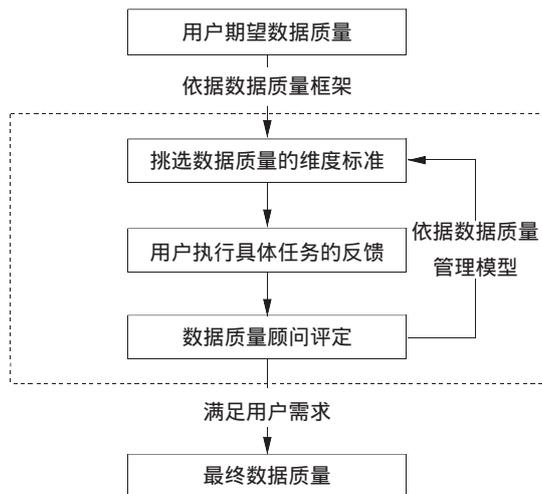


图 2 数据质量动态优化调整流程

由图1和图2可以看出,应更重视数据用户及其数据任务,这两个角色将直接定义并形成最终的数据质量需求。数据质量因企业而异,即使在同一企业的不同部门也是如此,因此用户根据他们对预期任务的数据质量需求来挑选他们认为重要的维度和标准。数据质量顾问团队根据用户最终确定的数据质量维度和标准,按照用户对数据质量的反馈做最终的数据质量评定,并进行相应的数据质量改进工作。

4 结束语

在信息化时代,数据是企业至关重要的资源,其质量是企业提升业绩和盈利能力的关键。为了解决企业数据管理质量问题,提高生产效率,需要从数据维度、质量标准、评定方法等方面全面构建企业数据质量框架,并形成与之相适应的数据管理模型及数据质量动态优化调整流程,以满足不同用户和任务对数据质量的不要需求。本文提出的数据质量框架和管理模型尚处于理论构建阶段,需在企业进行实际检验,做进一步的调整和完善,实现数据管理的自动化。

参考文献:

- [1]Experian.The 2016 global data management benchmark report [EB/OL].[2019-02-09].<https://www.edq.com/globalassets/whitepapers/2016-global-data-management-benchmark-report.pdf>.
- [2]Experian.Taking the pain out of data quality[EB/OL].(2018-07-12)[2019-02-09].<https://www.edq.com/uk/resources/papers/taking-the-pain-out-of-data-quality/>.
- [3]Eckerson,WW.Data quality and the bottom line[M].Seattle:TDWI Report,the Data Warehouse Institute,2002.
- [4]TDWI.Modernizing business-to-business data exchange[EB/OL].(2016-04-29)[2019-02-09].<https://tdwi.org/blogs/tdwi-blog/2016/04/modernizing-b2b-data-exchange.aspx>.
- [5]Batini C,Scannapieco M.Data and information quality:dimensions,principles and techniques[M].Switzerland:Springer,2016.
- [6]Weber K,Otto B,Österle H.One size does not fit all—a contingency approach to data governance [J].Journal of Data and Information Quality (JDIQ),2009(1):1-27.

(责任编辑 范可旭)

Study on Enterprises' Data Quality Management in Informationized Era

CHEN Ying-xi¹, LIU Jie-fang²

(1. Wuhan Sponge City and Integrated Pipe Gallery Construction Management Station, Wuhan Urban and Rural Construction Bureau, Wuhan 430079, China;

2. School of Traffic Information, Hubei Communications Technical College, Wuhan 430079, China)

Abstract: In the era of rapid development of information technology, data is a vital resource and the core for enterprises, while data quality is the key to improving enterprises' performance and profitability. In order to ensure data quality of enterprises, the data quality framework is proposed from the aspects of data dimension and its definition, quality standards and their definitions and assessment methods, etc. The framework can be used to guide enterprises to obtain high-quality data. Meanwhile, based on different requirements of different users and tasks for data quality, a data quality management model is proposed, which provides reference for enterprises to carry out data quality management.

Key words: data quality; data quality framework; data quality management model