

# 网络环境下面向过程的数据管理

杨小惠 王 芸

(江苏省信息中心, 江苏南京 210013)

**摘 要:** 首先探讨网络环境下的数据质量控制, 阐述现代管理理念和数据管理的发展、现代数据管理下的业务流程再造, 强调管理向服务转型。然后分析数据管理现状与发展趋势, 给出数据管理中过程管理的思想, 指出未来的数据质量控制将以服务为导向的数据质量控制。最后对数据质量问题的成因进行了分析, 并针对存在的问题, 从软件改造、优化数据结构、加强过程管理、建立有效管理体制 4 个方面提出建议。

**关键词:** 数据管理; 数据质量控制; 过程管理

**中图分类号:** G352 **文献标识码:** A **DOI:** 10.3772/j.issn.1674-1544.2009.03.014

## 1 引 言

数据是信息的基本组成, 通过对数据处理、组合、分析、挖掘得到信息, 从而辅助领导层正确决策、管理层科学管理、执行层高效工作。因此, 组织的数据管理工作是非常重要的, 对于依赖数据进行服务的信息中心或数据中心, 其数据的管理、数据质量的控制、数据的充分利用更是重中之重。国家信息中心针对现存的数据质量和安全问题, 曾呼吁拯救数据, 强调“数据”在社会经济活动中的地位越来越重要。West-World 的报告指出: “100 兆数据的价值超过 100 万美元。”<sup>[1]</sup> 由此可见, 数据在未来的信息服务中心具有重要的地位与价值。这就要求我们必须加强对数据的管理, 必须重视对数据中心的建设。

网络环境下的数据管理任务更加繁重, 采集途径增加了, 处理手段更加复杂, 服务需求更加多样化, 服务形式也越来越多。这些变化促使数据管理手段的更新, 必须引入新的管理手段来提高数据质量, 确保满足用户对数据的各类需求。数据管理的直接目标是保证数据的质量和安全, 良好的

数据质量是数据充分利用的前提, 因此数据质量控制是数据管理的核心。根据美国克劳士比质量管理学院的研究成果, 质量管理实质是过程管理, 过程管理是产品高质量的保证<sup>[2]</sup>。20 世纪 90 年代中期, 过程管理的理念从对产品的质量管理逐步引入其他管理领域。对数据管理而言, 过程管理的思想更加适合对数据生产过程的质量控制, 过程管理将数据质量管理的内涵从传统的生产流程反馈, 经过质量控制、质量保证、质量管理等发展阶段, 演化到适应组织革新、营造组织文化和辅助决策支持。

因此, 为了提高数据的质量, 加强对数据的生产、管理和服务, 我们借鉴新的管理理念, 针对数据管理的特征, 提出了数据管理中的过程管理思想, 力图为信息服务和数据中心等机构提高管理水平和服务质量引进新的管理理念与方法。

## 2 指导数据管理的理念

20 世纪初诞生于美国的“泰罗制”开启了组织的现代管理之门, 宣告“科学管理”开始取代“经验管理”<sup>[3]</sup>。今天, 人类进入了知识经济时代, 组织的

第一作者简介: 杨小惠(1959-), 女, 江苏省信息中心计算机应用开发部主任, 高级工程师, 研究方向是数据处理与电子政务。

收稿日期: 2009 年 2 月 17 日。

万方数据

管理又迎来了第二次革命,即进入了以知识管理为代表的崭新时代。理想的知识管理就是要实现“在最恰当的时候、能够将最恰当的知识、提供给最需要的人”。然而,如何将现代管理理念渗透到数据管理之中,确保提高数据质量、实现数据共享、提升数据效用,这是我们必须深入研究的课题。现代化的管理理念强调资源数字化,同时将难以量化的东西进行量化,使得可以对其进行准确的价值估算,而数据恰恰是实现这一理念的基础。因此,在数据管理上我们必须接收现代管理思想,做到数据数字化、管理计量化、产品多样化、获取多途径、传递网络化,并确保数据科学组织、过程合理控制、质量应有保证、制度要有保障、管理面向服务,只有用现代管理理念来指导数据管理,进行业务重组来适应数据管理,以服务为导向来完善数据管理,才能满足现代管理对数据的要求。

### 2.1 业务流程再造与数据管理

现代数据管理强调及时生产数据、多渠道采集数据、分布式处理数据、集中式控制管理、全面共享数据和全方位提供服务。虽然传统的业务流程是经过长时间的经验积累而形成的,但它适应的是手工式的数据处理,完全模拟手工过程的计算机数据管理在计算机应用的初期发挥了应有的作用,产生了重要影响。而随着网络化的信息社会到来,过去的数据处理流程显然不能适应现代数据管理的要求,必须进行业务流程的再造。业务流程再造强调的是,优化资源配置,实现信息共享,提供强有力的信息技术支撑<sup>[4]</sup>。因此,数据成为这一问题的核心。所以,现代数据管理下的业务流程再造应当注重多渠道的采集和分布式的数据处理,加强集中管理,严格过程控制,提供多种形式和多方位的服务,全面体现数据在为政府服务和社会发展中的作用。

### 2.2 以服务为导向的数据管理

数据管理步入智能阶段,数据管理理念延伸更侧重服务。目前,政府行政正处在一个由管理向服务的转型时期,作为以信息提供与服务为主的数据处理中心更是以“服务用户”为出发点和落脚点。以服务为导向的数据管理要求:信息部门针对用户需要采集数据,把握社会所关注的问题;用更加科学的方法组织数据,也就是说将数据有机地

联系起来,将数据变成具有语义关系或相互关联的数据链,使数据中包含知识;数据处理是指能够从数据中提炼出内在知识和所涉及的主题;数据分析就是能够利用现有的信息分析工具(如数据仓库与联机分析 OLAP 等)进行多维数据分析,为决策提供帮助;数据提供是指能够提供个性化的、主动推送的服务方式和信息的多种表达方式(如信息可视化的表示)。

## 3 数据质量管理的发展阶段

现代信息技术的发展,使数据的生产能力大大提高。而面对大量的数据,如何保证数据的正确性;如何有效地组织数据使之便于采集;如何科学地构造数据结构使之便于统计分析、发现知识,辅助领导决策,提高组织内部员工的工作效率和知识水平等。这一切对数据管理工作提出了更高的要求,我们必须解放思想,与时俱进,突破过去囿于具体业务、部门内部、行业内部等数据采集、组织和应用,应扩大数据共享与交流的范围。质量管理已由简单原始渐渐变得复杂成熟,人们深刻认识到质量管理在组织经营中的重要性。在数字化的数据管理中,其质量问题更为重要,质量问题逐渐演变为过程问题。我们可以把数据质量管理的发展分成 4 个阶段:

第一阶段,等待式纠正阶段。数据生产出来后,等待用户发现问题,然后再去解决。这是在数字化数据生产初期普遍采取的做法。

第二阶段,成品数据质量检查阶段。数据提交服务之前,进行质量检查,挑出并纠正存在质量问题的数据,这是目前普遍采用的做法。

第三阶段,过程控制阶段。即对数据生产的每一过程进行质量控制<sup>[6]</sup>,把问题扼杀在摇篮里。这是目前国外倡导的先进的质量管理理念,也是我们今后应采取的手段。

第四阶段,服务改进过程阶段。有许多问题只有在服务中才能发现,单纯从过程控制根本无法发现,其问题的出现,可能由于数据生产过程的不合理所造成的,必须对过程进行改进,并加以过程控制,从而消除这些问题。这是现代数据管理应遵循的方法。

## 4 数据管理的环节

过程管理强调对“过程”进行管理,而不是对“流程”进行管理,这和管理的方法有着显著差异。数据管理的任务应包括数据的采集、加工、组织、分析与服务等多个环节,各环节之间密切相关,如建立一个企业技术改造融资项目管理数据库,其各环节的基本任务是:

(1)数据采集。采集与企业融资项目有关的各种数据,包括企业基本信息、技术改造项目信息、资金来源信息、项目完成后新增经济收益预测信息等。采集渠道有企业、金融机构、政府部门,包括公开信息资源、网络信息资源以及与其他行业进行交换的信息。

(2)数据加工。数据加工是将采集来的数据进行整理、筛选、甄别、标识,以便进入数据库。整理是将所收集来的“杂乱”信息归类、梳理成可比较形式;筛选是指去掉一些无用信息;甄别用于判别真伪信息;标识是将信息进行主题标引、数据分类和特征标识,确保数据被有效组织。

(3)数据组织。数据组织即数据的整序工作,主要将加工后的数据合理地存储在数据库或数据仓库中,使数据便于被检索、分析和挖掘。数据组织的内容包括数据标准化检验、数据库的设计、数据结构的定义、数据存储与表现形式的定义、关联数据的链接等。以保证数据的完整、完全、标准、少冗余、易共享等。

(4)数据分析。数据分析是指对已数字化的数据进行统计、分析和挖掘知识的过程。数据分析要求能够宏观和微观的兼顾、定性和定量的结合。可用一般性的统计完成定量分析,采用数据挖掘技术实现定性和定量相结合的分析,宏观分析来源于汇总数据,微观分析取自于具体数据。分析结果的表现形式必须多样化,如以文字、表格、图形等形式表现。

(5)数据服务。以上4个环节的最终目的落实在服务,现代数据服务必须改变传统的完全被动式服务,应根据客户需求提供多样化服务,如自助式服务(提供网络检索),信息公开(有关政策、法规、具体实行办法等),定题信息服务(根据客户要

求定期的向客户提供有关数据,如每个月向有关政府部门提交报告、向企业或个人发送他们所需要的数据),专项服务(用户提出的特殊要求的服务,可能采取包括数据挖掘技术的多种手段进行分析)推送式服务等。

数据是信息管理系统运行的“血液”,正常的血液是决定健康人的基本条件,因此,一个良好运行的信息系统,其数据的正确和高质量是基本保证。高质量的数据应满足以下要求:

(1)数据的完整性。数据的完整性是指数据属性(数据字段)设计的完全性和数据项的内容完备性。也就是说,尽可能地把数据各种特征全部表现出来,对字段中的内容不应有缺项。一般来说,被用于统计分析的属性、被用于挖掘或检索的数据项,不能缺失,否则将会影响对数据的统计分析。

(2)数据的标准化。标准化是数据质量的可靠保证,很难想象没有一定标准的数据具有高质量。数据标准主要应考虑代码的标准化,统计数据的标准化,数据流程的标准化,各类名称的标准化等。使数据质量得以控制,统计分析准确可靠,数据交流(交换)减少障碍。

(3)数据的少冗余。冗余即多余,数据的冗余主要指数据库中的多余字段,数据库之间的重复字段。数据的冗余极易造成数据的不一致性,如当冗余的字段在输入或修改时,很难保证两次或多次输入具有一致性,其差异带来的问题就是,不同的途径检索与统计分析得到的数据结果可能不一样。因此,应尽可能在整个数据库系统中保持字段的唯一性。

(4)数据的共享性。数据的共享是指在本系统内部或相关应用之间的数据共享。实现数据共享要求数据结构具有开放性,数据本身具有规范性。共享使数据得到了充分的利用,促进了组织间数据交流与交换,为数据库的共建、多方面的数据采集打下基础。数据共享还表现为行业间的数据交换和信息共享。

(5)数据的安全性。数据的安全性控制包括两个方面的内容:一是数据的存取权限的设定,为不同类型(级别)的用户赋予不同权限,确保每一用户都可以看到自己能够访问的数据。二是系统数



据本身的安全性,防止被攻击和破坏,定期备份数据。

为政府部门提供各类信息服务,需要与数据打交道,数据的正确与否关系重大。作为直接处理政府部门数据的信息中心,保证数据的畅通、数据的质量、数据的开发与利用,是其重要职责,科学的数据管理、深层的统计分析和挖掘利用,对辅助、指导领导层的决策有着重要意义。

## 5 数据质量问题

### 5.1 软件的缺陷造成的数据质量问题

计算机管理系统是数据管理的基础平台,现有的数据管理、数据质量问题,大多是由管理系统本身的问题引起的,在现有管理系统的设计时未能严格按照数据管理的基本要求,在数据字典的设计、管理、冗余控制、一致性检查等方面存在缺陷,从而导致数据的采集等环节出现质量失控的问题。同时,如果没有提供系统内部数据的软件稽核手段,往往会产生如下问题:

(1)软件缺乏完整性检验手段,造成数据输入的不完整。

(2)软件缺乏对数据的一致性保护,造成从不同角度查询、统计得不到一致的结果。

(3)软件的开放性不够,缺乏对代码字典库的维护,当编码字段出现新的类别时,程序无法加入。

(4)软件中具有大量重复输入信息,如一个字段数据是其他多个字段数据运算的结果,可还需继续输入这个字段,从而造成人力浪费和增加出错概率。

### 5.2 数据组织的不合理带来的数据质量问题

数据的组织应遵循少冗余和共享性等原则。数据组织的不合理,将会产生大量数据质量问题。

(1)数据库中具有大量的数据冗余,这是产生数据不一致的温床。主要表现在以下几个方面:一是相同数据在不同数据库中分设字段,这必然会造成数据不一致的情况。根据数据库设计要求,任何字段数据只能在数据库中出现一次,相关数据通过映射和链接关系将它们联系起来。二是两个

不同字段表达相同意义,如出生年月和年龄。三是为几个字段的运算结果再设置字段。这些数据冗余必然影响数据的质量。

(2)数据库关联结构定义不清晰,导致数据无法进行有效组织,不能为数据的深层挖掘提供支持。

(3)缺乏对数据字典的有效维护和管理,没有产生一套科学的字典编码体系。

### 5.3 数据流程缺乏过程控制

数据流程缺乏过程控制的原因可归纳为以下3点:

(1)业务流程不适应现代数据流程的过程管理要求,软件缺乏对数据流程的控制,数据不能够正确反映工作流程的状态和结果,影响后续或相关 workflows 的正确执行,不能满足数据的深层挖掘。

(2)组织内部对于数据流程的约束停留在行政管理阶段,主要是依据人的经验管理,没有具体的约束制度,从客观上导致了数据流程缺乏控制。

(3)数据质量手段缺乏革新,仍然停留在传统的质量保证的反馈控制上,仅仅是发现问题后,再处理问题,而不是预防问题的发生。

### 5.4 管理体制与制度不适应现代数据管理

一个组织的管理体制和其发展息息相关,因为它涉及具体的职能由谁去决定、执行,而制度是管理体制能够正确贯彻的保障和规范,同时制度实施的效果需要一个良好的组织氛围。目前相当多的数据处理中心所依赖的管理体制,并没有明确的定义其数据管理地位、界定与其他部门之间的数据接口。职能、权责相关制度不完备,导致数据处理中心缺乏对数据的统一监督管理职能、数据不能全方位实现共享、缺乏支持知识管理的组织氛围。其原因可归结为:

(1)管理体制不适应现代数据管理的要求,不能适应数据共享的要求,数据处理中心不能统管数据采集、校验及更新工作,导致部分数据有滞后情况发生,数据缺乏全方位共享。

(2)缺乏知识共享机制,各部门已经完成了内部数据管理的电子化,但是缺乏整体化建设意识,同时缺乏相应的制度支撑整个组织,缺乏支持知

识共享的数据资源管理环境,各部门之间缺乏电子协作,形成信息孤岛。

(3)缺乏支持知识管理的组织氛围,实行知识管理的主要障碍是企业文化,对于数据处理的认识停留在数据的数字化过程,认为数据的处理是进行数据收集、录入、存储、查询的事务性操作,缺乏数据挖掘和知识创新机制,尚未形成基于组织知识的决策体系。

## 6 解决问题的建议

### 6.1 软件应不断改造或升级

任何一个软件都存在一定的生命周期,需要不断地更新或升级,确保能够适应新的业务要求和新的环境。如操作系统 Windows、办公软件 Office 都在不断升级和改造。可以肯定地说,过去的软件对信息化建设发挥了重要作用,但我们还要看到,由于业务的发展,部门职能的增加,政府、企业、个人对数据的需求变化以及高速发展的网络,使得软件必须适应新的要求,必须进行改造或升级,确保软件能够适应各方面的需求和保证软件的可持续发展。

### 6.2 优化数据结构

优化数据结构应遵循以下数据管理的原则:

(1)减少冗余度:消灭表内不同名但同意义的数据项;减少可通过运算产生的数据项。这样才能实现最小的存储、最少的输入、最大的共享性和最严格的一致性。

(2)建立一套科学的字典编码体系:字典编码要科学规范,表达层次感。如所在地代码要表现出省市县,行业代码要表现出行业大中小类。有国标的最好采用国标编码。数据字典的建立是减少冗余的基础,也是设计开放性软件的基础。减少为适应新的数据要求而必须进行的更新软件。

(3)建立清晰的对应关系。冗余度小的数据库,表间数据少有相同,但密切相关。因此,必须建立清晰的对应关系,通过主键组织数据。

### 6.3 加强数据流程中的过程管理

过程管理的目标是保证数据的质量,质量保证主要来自于两个方面:一是数据流程的规范,没有规范的数据流程数据的质量是无法得到保证

的;二是规范的数据流程需要有监督和管理,没有对数据流的监督与管理,数据质量依然不能得到保证。因此,必须强调数据流过程的数据质量控制,这就是目前倡导的过程管理。从前面对数据质量问题的成因分析,可以看出,现有的数据质量问题在这两个方面都有所反映。所以,我们应当在规范数据流程、软件对数据控制方面加强规划和管理。

(1)规范数据流程。要求根据业务需求分析处理环节,刻画数据流。分析要全面,刻画要具体,给出每一个环节的具体任务、措施和细节,规定各环节之间的先后次序,规定前一个环节没有完成,数据不得流向下一环节,以保证数据的完整性。

(2)软件对数据控制。数据流程制定后,要靠软件对数据流进行约束,如利用软件控制前一个环节不完整的数据不能进入下一个环节、软件中的数据组织应保证数据项的唯一性、对每一项数据进行限定检查避免输入超出范围的数据、科学合理构造各类字典等。

### 6.4 建立有效的管理体制与制度

一个良好的数据管理系统,需要有一个良好的数据管理体制与制度给予支撑。数据需要采集、更新,数据管理中心就应被赋予数据采集、数据监管等职能。数据管理中心还应被赋予对本级及下属各级各个部门的软件系统统一管理的职能。其中包括系统整体化建设、统一开发平台、解决各部门之间的数据接口等。这样既能减少系统开发工作量,也使系统易于维护,又容易做到信息共享,以免形成信息孤岛。数据管理中心还应被赋予研究数据的职能。不能仅仅停留在加工数据上,而要进行深层次的挖掘和知识创新,形成基于数据的决策体系。

## 7 结 语

数据管理的过程是集过程管理、信息技术、体制及制度于一体的过程。过程管理是核心,信息技术是手段,健全的体制及制度是保障。将现代管理理念渗透到数据管理中,有利于提高数据质量、提升数据效用,充分发挥数据在为政府和社会服务中的作用。

## 参考文献

- [1] 胡喆. 拯救数据. 电脑报[J]. 2003(71).
- [2] 克劳士比. 零缺陷管理. <http://bbs.manaren.com/thread-11227-1-1.html>.
- [3] 德鲁克 彼得. 管理实践[M]. 毛忠明 译. 上海: 上海译文出版社, 1999.
- [4] 梅绍祖, 等. 流程再造: 理论、方法核技术 [M]. 北京: 清华大学出版社, 2004.
- [5] 路磊. 管理上的差距有多大[J]. 经理人, 1999(11): 63 - 65.
- [6] (美)乙韦恩( Wayne, B ). 过程控制( 英文版 ) [M]. 北京: 世界图书出版社, 2008.

# Data Management Oriented to the Process in the Network Environment

Yang Xiaohui, Wang Yun

( Jiangsu Information Center, Nanjing 210013 )

**Abstract:** This article has discussed the data quality control under the network environment, elaborated the development of the modern management idea and data management, and the restoration of operation process under the modern data management, emphasized the reforming of management to service. It has analyzed the present situation and the trend of development of the data management, giving the process management thought in the data management and pointing out the future data quality control will make service as the guidance. Finally, it has also carried on the analysis to the origin of data quality questions, and given the solution suggestion on the following four aspects: software transformation, the optimization of data construction, strengthening the process management and establishing effective management system.

**Keywords:** data management ,data quality control ,process management