

# 数据治理过程浅析

张一鸣

(中国铁建股份有限公司信息中心 北京 100855)

**摘要：**数据是一种参与企业经济活动的经济资源，是企业无形、有价值的核心资产，维护数据质量是一个长期而艰巨的任务，然而企业有可能花费大量精力和金钱在保护某些没有价值的数据上。有效的管理和使用数据可以减少和消除企业经济活动中的风险。数据综合治理是提升数据资产管理与应用水平的关键举措，是进一步梳理业务流程及其数据标准、加强跨业务部门的协同、提升信息系统应用收益的过程，是帮助企业强化标准、提高数据质量、控制成本的重要工作内容。

**关键词：**数据治理 数据资产 数据质量

## 1 数据现状概述

国际金融危机以后，全球的竞争格局呈现出新战略、新思路，信息化已经成为当今的时代特征，是争夺竞争优势的关键。大部分领先企业都陆续建设了ERP、EAM、人力资源、供应链系统、物流系统、电子商务系统、合同系统、风险管理、安全、集成门户、协同办公、决策分析等信息化系统。根据国际数据公司IDC作出的估测，企业的数据一直都在以每年50%的速度增长，换言之，也就是每两年就增长一倍。企业面临数据整合的挑战，难度不断增长，数据类型更具有多样性，散落在不同的系统中，零散的且不一致的数据越来越成为一个引人注目的瓶颈。数据标准、安全性、数据质量、如何利用数据、获取有效的决策分析结果等，已经成为在信息化与业务深度融合过程中的关键制约因素。

数据已经成为现代企业最大的价值来源，是支持企业发展战略的重要工具。合理的利用企业数据来寻求竞争优势，挖掘利用海量数据的潜力已经成为发展的关键推动力。数据应用是企业进行理性分析和决策的基础，利用数据创造价值、实现决策分析对提升业务效率、提升综合竞争实力具有重要的作用。

当前数据资产的价值和潜力还远远没有被挖掘出来，数据资产一旦处于混乱无治理的状况，其重要性就会降低，价值会大打折扣，更甚者会严重影响企业的利益和决策。如何对数据进行治理已经成为困扰企业CTO乃至CFO和CEO一项巨大的挑战，数据治理已经不仅仅是信息化部门所面临的问题，数据质量已经提升到企业的核心战略层面，成为一项复杂而艰巨的系统工程。

总体来说，目前国内企业所开展的数据治理工作还处于初级阶段，很多企业的资产都或多或少都面临着各方面的问题。

(1) 数据不完整：缺少关键基础数据，部分辅助数据缺失或不全面，历史数据丢失严重。

(2) 数据分散、不一致：企业内的数据入口众多，同一类数据采用的标准、规则都不一致。

(3) 数据质量低：由于大量数据基本上是“堆积”在一起，缺少必要的管理，集成数据的可用性差，质量比较低。

(4) 数据共享集成成本高：数据标准不统一、分散、可用质量差，数据核对、清理、映射的工作量巨大，导致共享集成和数据分析的成本非常高。

(5) 数据经济效益不显著：数据决策分析的结果可靠性差，投入与产出不匹配。

经权威机构调研，大型企业对于数据治理有强烈的需求，调查数据显示，他们迫切的期望通过数据治理有效的提高数据质量和优化数据仓库，分别占62%与59%。数据的应用与数据质量是一个相辅相成、相互推动的关系，对数据资产进行治理，是提升公司数据管理与应用水平的关键举措。企业应该着眼于长期持续有效的数据治理，建立行之有效的数据治理体系，挖掘数据的潜力，从而发挥数据资产在企业中的核心价值。

IBM的数据治理过程卓有成效。1992年之前，IBM在数据治理方面存在很多问题，没有明确的可依赖的数据源，没有明确的数据所有人，数据质量低下。1995年，IBM在ERP里面做了业务数据标准，将所有的业务定了15大类业务标准、79个分类子业务标准，使全公司看到的是一个统一的业务定义。2004年IBM成立了数据责任人论坛，2005年成立了数据治理委员会，之后又成立数据审核委员会。1992年，IBM全球有128个CIO、155个数据中心、80个Web拓展中心、31个不同的网络、16000个应用。通过数据治理，IBM简化了基础架构，并降低了管理的复杂度。2007年，IBM全球只有一位CIO，主数据中心变成

了6个, Web拓展中心变成全球统一的网络, 16000多个应用变成4000个左右。IBM的数据治理策略包括四大领域11个要素的数据治理框架和方法。框架包括产出领域: 数据风险管理、价值创造; 驱动领域: 组织机构/流程、管理制度、数据责任人; 核心领域: 数据质量管理、信息生命周期管理、安全/信息披露/合规; 支撑领域: 数据模型/数据架构、元数据/主数据/数据标准、质量审计与报告。

越来越多的国内企业也已经充分认识到企业最重要的核心资产是数据, 越来越希望通过进行数据治理来创造更多的业务价值。包括能源、金融在内的诸多行业已经将数据治理列为信息科技的“十二五”规划中, 已经有许多来自银行、电力、航空等行业率先开展了数据治理工作。

自2005年始, 光大银行在数据治理及应用方面进行了众多探索。首先推进业务系统建设及改造, 进行了企业级基础数据平台(即数据仓库EDW)的建设, 同时整合了核心系统、对公信贷系统(CECM)、个贷系统、网银系统、国际结算系统等41个源业务系统数据, 在有效整合全行数据的同时完成了银监会非现场稽核报表(1104)、人行大集中、信用卡BI、电子银行BI、信用风险集市、风险加权资产(RWA)、资产负债等19个数据集市和应用的建设和应用; 其次定义关键数据标准, 制定持续数据标准实施规划; 第三定义数据治理组织, 并逐步制定了数据治理相关管理规范, 通过制定数据标准和数据质量的绩效考核指标来推动相关管理规范和标准的落实; 第四建立数据治理长效机制, 定期重检数据治理相关制度、流程、规范; 规划与评价并重, 合理规划是复杂的数据治理过程有序进行的必要保证, 根据评价结果及时调整规划策略, 实现数据治理真正发挥价值。

在2011年芝加哥举行的企业数据世界大会上, 会议提出数据治理投资必须把目标定位在推进商业目标和提高底线。企业数据管理和数据治理协会主席 Michele Koch也指出: “有时数据治理方案是明确的, 但是在找出问题的原因并给与量化分析却是一个漫长的过程。”

随着全球经济一体化的迅速发展, 信息战日益激烈, 数据资产的优劣已成为企业之间竞争的重要砝码, 数据治理已被越来越多的企业所重视, 很多企业都将数据治理作为企业发展战略进行规划和管理。

## 2 数据治理的概念

IBM认为, 数据治理是根据企业的数据管控政策, 利用组织人员、流程和技术的相互协作, 使企业能将数据作为企业的核心资产(data as enterprise asset)来管理和应用。通俗来讲, 数据治理是指从使用零散数据变为使用统一主数据、从具有很少或没有组织和流程治理到企业范围内的数据治理、从尝试处理主数据混乱状况到主数据井井有条的一个过程, 并最终使企业能将数据作为企业的核心资产来管理。

数据治理是一种体系, 以实现数据标准化为目标, 以

健全的数据组织为保障, 以数据过程管控为手段, 实现全面、高效的数据管理。从技术支持范围来讲, 数据治理涵盖了从前端事务处理系统、后端业务数据库到终端的数据分析, 从源头到终端再回到源头形成一个闭环的负反馈系统; 从业务范围来讲, 数据治理就是要对数据的产生、处置、使用进行监管; 从控制范围来讲, 数据治理必须通过对人员、流程和系统的整体设计和调整满足数据与业务的全面结合。

随着信息技术的高速发展, 数据的重要性已被我们认知, 数据就好像一座待开采的含有丰富矿藏的矿山, 而数据治理则是具体的开采方法和手段。通过有效数据治理我们可以得到:

- (1) 完善的数据管控体系;
- (2) 统一的数据来源;
- (3) 标准化、规范化的数据;
- (4) 提高的工作效率;
- (5) 降低的数据管理、维护、集成成本。

加强数据治理是提升信息化能力、提升精细化管理水平、提高业务运营效率、增强企业决策能力和核心竞争的重要途径。

## 3 数据治理的路径

如今, 全球的各组织都认识到他们的数据资产的重要性和价值。与此同时, 高层管理人员并未充分利用这些信息, 原因在于缺乏数据的准确性、一致性、相关性和及时性。因此, 数据治理被推到了前线, 许多公司正在尽力研究如何有效设计和实现数据治理计划, 企业数据治理主要从以下方面来进行:

### 3.1 组织机构

组织机构在数据治理过程的重要性, 逐渐被企业认知, 是数据治理的关键。数据治理组织要实现由无组织向临时组织, 由临时组织向实体与虚拟结合的组织, 最终发展到专业的实体组织的发展。企业必须建立数据治理的组织机构, 设立各类职能部门, 加强数据治理的专业化管理, 并建立起专业化数据治理团队。

目前数据治理组织多以临时组织的方式存在, 这样的组织类似于项目部, 对企业来说组织构建和培养上没有连续性, 缺少数据管理经验和知识的有效传递; 而当今世界经济的发展要求在数据治理中建立有权威性、实体存在的组织机构, 且要求能够在企业一直存在并持续发展壮大。

组建数据治理组织机构一般包括定义数据治理规章和制度、定义数据治理的组织结构、建立数据治理委员会、建立数据治理工作组、确定数据专责人等内容。

### 3.2 岗位职责

伴随着组织机构的发展, 岗位的专业化是数据治理发展的必然趋势。在数据治理的要素之中, 人是数据治理工作的执行者, 即使组织机构设立再合理, 如果人的岗位职责不明确, 那么也会造成职责混乱, 工作者无所适从, 工

作效率低下。数据治理需要整个治理团队的协同工作,每个岗位既要完成自己职责范围内的的工作,又需要与其他岗位进行良好的沟通和配合。

### 3.3数据治理标准

数据标准的制定是实现数据标准化、规范化,以及实现数据整合的前提,是保证数据质量的主要条件。标准不是一成不变的,它会因为企业管理要求、业务需求而变化,也会因为社会的发展、科学的进步而不断的推进,这就要求企业对标准进行持续的改进和维护。

在数据治理过程中的标准包括数据标准的制定和度量标准的制定两方面:

数据标准制定是数据标准化工作的核心,国外企业在进行数据治理时大多从数据标准管理入手,按照既定的目标、根据数据标准化、规范化的要求,整合离散的数据,定义数据标准。

数据治理执行过程中的度量标准也是不可缺少的,它用来检查执行过程中是否偏离既定目标,用来度量过程中成本以及进度。度量标准的制定是数据治理过程中,评估原有数据的价值、度量和监控组织的数据治理执行、有效度量数据治理效果的关键因素。原有数据的价值如何,企业需要花费多大的成本去做数据治理,这些问题都需要有能够度量数据价值的标准,按照度量后的原有数据的价值,确定数据的重要性优先级,以确定对数据治理的投入成本。

数据治理的效果如何也需要度量效果的标准。通过对治理效果的度量、分析,主动的采取措施去纠正,改善数据治理的工作。

### 3.4数据控制和使用

在数据治理发展的过程中,数据控制一直被企业所重视,它们明白提高数据的质量,数据控制是必不可少的。数据控制发展到现在,逐步形成了多重控制相互作用,共同管控的状况,数据控制主要有以下内容:

#### 3.4.1流程化控制

数据的流程化控制是最普遍的控制方式,发展至今,流程化的控制演变成为了多元化的流程控制。为了加强数据的流程化的管控,不仅有数据业务上的控制,也要有数据技术上的控制,而且还有数据逻辑上的控制。

#### 3.4.2合规性控制

当今世界进入了“大市场经济”时代,企业的数据是否符合国际、国家的法规,是否满足行业的标准,能否满足跨国度、跨行业的经济行为,都牵涉到合规性控制。数据合规性的控制是现在企业非常重视的控制方法,加强合规性控制是企业提高自身竞争力的必要手段之一。

#### 3.4.3工具化控制

当今信息化技术发展飞速,数据管理工具不断涌现,通过工具进行管理控制也是数据控制的一种方法,这种控制方法对既定的控制要求完全的严格执行。

数据的使用需要与数据控制建立必要的联系,即通过数据控制来管理数据使用,哪些数据可以被使用,可以被

哪些人访问,可以被哪些人维护。通过对这些方面的数据控制,以达到数据能够安全、顺利的使用。

### 3.5数据管理工具

当前,IT技术迅速发展,数据管理水平不断提高,对于数据管理工具,功能也越来越强大,使用越来越方便,更加人性化,并且持续有先进的管理思想融入其中。数据管理工具对数据治理是有效的支撑和辅助,采用一个成熟、先进、科学的数据管理工具也是数据治理工作成功的关键。

通过数据管理信息化工具应用,完成组织机构和岗位职责的定义,数据标准规范的应用,实现数据管控流程、规则,完成数据访问授权控制,提高数据的安全性,管理数据的产生、审核、使用、修订、消亡的全生命周期。

## 4 数据治理的工作展望

数据治理作为信息化过程中的必经之路,需要长期稳定、持续不懈的推进,企业高层决策者的战略支持和持续的资源投入是获得数据治理成效的重要保障。目前,国内大多数企业的数据治理工作仍然偏重于技术,主要工作还停留在现有业务环节的数据问题等方面,包括现有数据的清理、查重、映射、标准化等内容;而对更深层次的数据治理体系,例如涵盖企业业务数据总线的数据管控体系设计、数据规划设计、数据中心等方面还未有更多进展。期望国内企业会借鉴国外先进企业的数据集成、数据服务、数据管理的经验和优秀的数据治理构架策略,从组织、标准、流程、质量、运营等方面构建企业全生命周期的数据治理体系,不断探索建立数据治理机制的有效方式,切实提升企业信息化的应用水平,获得持续的核心竞争实力。

### 参考文献:

- [1] 许继楠,郭涛.大数据时代更需要数据治理.中国计算机报,2011-11-11
- [2] 杨兵兵,商业银行数据治理与应用——以光大银行为例.银行家,2012(1)
- [3] Mark Brunelli.解读数据治理的最佳实践.TechTarget,2011年5月
- [4] 李佳师.IBM:关注数据治理创造更多价值.中国电子报,2011-11-08

### 作者简介:

张一鸣,2008年毕业于北京航空航天大学软件学院,硕士学位,IT项目管理及产业信息化专业。专业领域:IT项目管理、IT行业应用。2011年1月至今任中国铁建股份有限公司信息中心任规划需求处处长,主要负责中国铁建的信息化整体规划和信息系统的实施,负责中国铁建人力资源系统、资本运营系统、信息化绩效评价体系、信息化管理制度体系、经营管理系统、施工调度系统等多项集团级信息化系统的建设工作。