**教据治**理

**工业企业数字化转型之道**

祝守宇蔡春久等著

工业大数据应用技术国家工程实验室

**中工信出版集团** 







rr

**工业企业数字化转型之道**

祝守宇 蔡春久 等著

工业大数据应用技术国家工程实验室

重 庆 工 业 大 数 据 创 新 中 心

F y

**内** **容** **简** **介**

本书是一本全面关注工业企业数据治理方面的工具书，主要内容分为概述篇、体系篇、工具篇、实施 篇及案例篇。其中概述篇主要介绍工业企业数据治理的基础概念、主流数据治理标准及框架、数据治理的 发展趋势等；体系篇主要介绍数据管控、数据战略、数据架构、主数据管理等的基本原理与管理体系；工 具篇主要介绍主数据管理工具、数据模型管理工具、数据资产运营工具等；实施篇主要介绍具体实施策略 及路径选择、顶层架构规划与设计、数据资产运营实施等；案例篇主要介绍电力、能源化工、钢铁、制造、 战略投资等行业的数据治理案例，为读者提供专业、丰富、可信的数据治理实施范例。

本书是工业大数据应用技术国家工程实验室多年潜心研究的重要科研成果的总结和凝聚，既具有理论 高度，也具备面向中国工业企业的可实操性。参与本书编著的作者均为国内各企业的数据治理专家，所有 案例均来自这些企业的实践。

对企业的基层管理者或初入职场的人士来说，本书是充分认识数据治理意义、组织进行数据治理的具 体方案和工具手册；对企业中层管理者来说，本书是一本配合企业数据治理的纲领性指南；对企业高层管 理者来说，本书是一本推动企业数据治理的方法论。本书还适合作为高校的MBA 、EMBA教材。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

**图书在版编目** **(CIP)** **数据**

数据治理：工业企业数字化转型之道/祝守宇等著.一北京：电子工业出版社，2020.11 ISBN 978-7-121-39597-0

I.① 数…Ⅱ.①祝…Ⅲ.①数字技术一应用一工业企业管理 IV.①F406-39

中国版本图书馆 CIP 数据核字(2020)第180278号

责任编辑：王 静

印 刷：北京京师印务有限公司

装 订： 北京京师印务有限公司

出版发行： 电子工业出版社

北京市海淀区万寿路173信箱 邮编：100036

开 本： 787×980 1/16 印张：35.75 字数：775千字

版 次： 2020年11月第1版

印 次： 2020年11月第1次印刷

定 价：158.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系， 联系及邮购电话：(010)88254888,88258888。

质量投诉请发邮件至zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。 本书咨询联系方式：010-51260888-819, faq@phei.com.cn。

**本书编委会**

**编委会成员(排名按姓氏拼音):**

蔡春久 曹禹 陈彬 陈争胜 程华军 黄谷 蒋蕊 蒋文英 金凯伟 李剑锋

褚幼鸿 杜斌 甘腊梅 高伟 胡林海 胡小林 李朋 李然辉 李天白 李学东 刘博 娄山

骆阳 聂学明 彭秀东 乔慧 宋清波 孙晓峰 孙艳芝 谭海华 陶建辉 万婵 王睁 王建峰 王轩 王自红 吴超 肖鹏 肖青华 辛华 邢镔 杨春荣 杨更更 杨桂锋 于洪伟 于万钦 张华 张康华 张磊 张云 祝守宇

**案例编者及所属单位：**

案例01(第38章) 娄山 陈彬 中国南方电网有限责任公司

案例02(第39章) 于洪伟 张华 国家电力投资集团有限公司

|  |  |
| --- | --- |
| 案例03(第40章) | 陈争胜 甘腊梅 陕西延长石油(集团)有限责任公司 |
| 案例04(第41章) | 王自红 张磊 酒泉钢铁(集团)有限责任公司 |
| 案例05(第42章) | 胡林海 黄谷 重庆长安汽车股份有限公司 |
| 案例06(第43章) | 金凯伟 张云 中核核电运行管理有限公司 |
| 案例07(第44章) | 杨春荣 曹禹 乔慧 中国航天科工集团有限公司 |
| 案例08(第45章) | 张康华 吴超 中国航空综合技术研究所 |
| 案例09(第46章) | 李学东 中国一重集团有限公司 |
| 案例10(第47章) | 宋清波 中国外运股份有限公司 |
| 案例11(第48章) | 杜斌 孙晓峰 中建材信息技术股份有限公司 |
| 案例12(第49章) | 杨桂锋 无锡威孚高科技集团股份有限公司 |
| 案例13(第50章) | 肖青华 蒋蕊 聂学明 国投智能科技有限公司 |
| 案例14(第51章) | 孙艳芝 新兴际华集团有限公司 |

编辑组：袁传奇 乐雯娇

美工组：吴海燕 唐泓 梁琴

**致谢**

(排名不分先后)

腾讯云计算(北京)有限责任公司

京东数字科技集团

中国航天科工集团有限公司

北京数语科技有限公司

重庆斯欧信息技术股份有限公司

广州信安数据有限公司

上海企源科技股份有限公司

重庆长安汽车股份有限公司

国投智能科技有限公司

中国航空工业集团有限公司

中建材信息技术股份有限公司

新兴际华集团有限公司

国家电投集团科学技术研究院

中核核电运行管理有限公司

无锡威孚高科技集团股份有限公司

华为技术有限公司

中国南方电网有限责任公司

北京数博智云信息技术有限公司

广州市华矩商业信息科技有限公司

杭州安恒信息技术股份有限公司

北京涛思数据科技有限公司

重庆斯欧智能科技研究院有限公司

中国外运股份有限公司

陕西延长石油(集团)有限责任公司

中国一重集团有限公司

石化盈科信息技术有限责任公司

重庆工业大数据创新中心

酒泉钢铁(集团)有限责任公司

北京发那科机电有限公司

**推荐语**

本书较好地满足了读者理解工业大数据治理框架的需要，系统地阐述了工业大数据治理的 体系、工具、实施等，提出和分析了主要类别工业大数据治理的解决方案；通过对典型行业工 业大数据治理实践的考察，深入浅出地介绍了当今主流的工业大数据技术与平台。本书具有可 参照性、可操作性和可读性，是工业大数据治理领域少见的参考书，值得一读。

中国工程院院士、中国工业互联网研究院技术专家委员会主任，高金吉

数据是企业的核心资产，要加大数据治理工作的力度，建立数据资产化管理体系，明确数 据采集和管理职责，制定统一的数据标准，搭建集成、统一的数据管理平台，实现数据的资产 化、集中化、平台化管理，确保数据的及时性、准确性和完整性，提高数据集成共享能力，充 分挖掘数据资产价值，夯实数字化转型基础。本书结合工业大数据、区块链、移动互联、人工 智能等前沿技术，在数据治理和数据共享交换等方面为企业提供相关的建议和技术指南，值得 有志于数据治理技术研究与应用的企业和个人一读。

中国工程院院士、中国科学院计算技术研究所研究员，倪光南

本书借助国内工业大数据领域的国家级工程技术研究平台——工业大数据应用技术国家工 程实验室在工业大数据、工业数据治理等方面的技术积累和研究能力，对工业企业数据治理体 系进行了系统、深入的阐释，展示了工业企业大数据治理、大数据应用、大数据管理的基础理 论和实践体系，是一本值得工业界同仁认真研读的专业性指导图书。

国家信息中心信息化和产业发展部主任、国家大数据发展专家咨询委员会秘书长，单志广

VI 数据治理——工业企业数字化转型之道

随着信息化技术的演进，网络技术的进步和应用的普及，产生了海量的数据。这些海量数 据的生成为智能化技术的发展奠定了基础，也是经济、社会发展的新的契机，同时也带来了新 的挑战，比如如何利用与开发数据?如何保护个人隐私?数据的所有权与数据共享的平衡点到 底在哪里?数据作为资产如何确权?数据治理是一个难题也是一个迫切需要解决的问题。而当 前工业互联网发展方兴未艾。本书的出版就像一场及时雨，它系统地回答了人们关心的一些问 题，比如工业企业数据治理的核心价值、概念，以及主要内容、标准和框架等，是值得认真阅 读和研究的图书。阅读本书可以给人以启发和思考，有利于指导我们的实际工作，有效提高工 业企业数据治理的水平，促进工业互联网和智能制造产业的发展。

中国互联网协会副理事长、国家计算机网络与信息安全管理中心原主任，黄澄清

国际上的普遍共识是：数据是数字经济和第四次工业革命的新生产要素。智能制造、工业 互联网、互联工业等各类名词，虽然阐述的角度有所不同，但共同点都是数据驱动与传统行业 机理、知识相结合而形成的智能化发展，而实现这一智能化转型的重要基石是数据治理。从全 球看，数据治理还处在起步的阶段，从技术、商业到法律都还面临一系列的挑战，而工业数据 的治理更是任重道远，亟需实现从理论到实践的突破。本书对工业企业的数据治理进行了系统 阐述，几乎涵盖了数据治理的各个方面，既有理论层面的分析，也有实战经验的总结，可以指 导工业企业建立健全工业数据治理体系，为更好释放数据生产要素的价值，实现数字化、智能 化升级奠定坚实基础。

中国信息通信研究院副院长、工业互联网产业联盟秘书长，余晓晖

数据不仅是新生产要素，也是新生产力——数据生产力。数据生产力是在用“数据+算力+ 算法”定义世界，是知识创造者借助智能工具，基于能源、资源及数据这一新生产要素，构建 的一种认识、适应和改造自然的新能力。今天人们对数据的价值创造、运行规律和本质特征的 认知才刚刚开始。本书是对工业数据治理的理念、路径、方法进行的一次全面系统的探索和研 究，有很多独到的见解，对数字化工作者具有非要重要的启示意义。

阿里研究院副院长、数字化企业研习社副理事长、中国信息化百人会执行委员，安筱鹏

推荐语 VII

本书对工业数据治理的对象、主题、框架和方式等进行分析，使读者认识到数字经济时代 数据流动的重要性和巨大意义。从国内、国际标准、工具、最佳实践等多方面进行阐述，详细 介绍工业数据治理的方式、方法及数据治理策略，通过典型的实际案例，分析验证工业数据治 理体系，使得该书具有较大的理论意义和应用推广价值。

中国科学技术大学计算机科学与技术学院院长、IEEE Fellow、 ACM Fellow 、ACM 杰出科学家，李向阳

本书以国际视角对数据治理的发展、演变及工业领域的实际应用进行阐述，涵盖了数据治 理的方方面面，可以指导工业企业从无到有地建立健全工业数据治理体系，全面支撑高质量的 工业数据分析与应用。本书对企业数据治理具有指导作用。

美国密歇根州立大学计算机系系主任及 大学基金讲席教授、IEEE Fellow 、ACM Fellow, 刘云浩

工业互联网是5G 最主要的应用领域之一，它将开创一个万物互联的时代，物均流量将是 消费互联网人均流量的几百倍，连接数量也会从50亿发展到千亿级水平。数据量的爆增将会带 来云计算、大数据技术架构的新一轮转变，同时现有的数据治理模式也会受到严重挑战。本书 从体系架构到治理工具，从实施路径到具体案例，给我们提供了不可多得的经验分享与专家意 见。希望读者能够从中得到启发，制定出适合工业企业数据治理的发展规划。

中国联通大数据首席科学家，范济安

在数字化工业中，如果说数据是原油，那么数据治理则是输油管、炼油厂、储油库。数据 治理实现了数据服务的互通、互信、互惠，使数据真正发挥价值。本书从工业企业数字化转型 的战略和需求出发，系统地阐述了工业数据治理的要素、体系、工具和实施，通过大量案例介 绍实战经验。本书视野宽广、立足前沿、内容翔实、深入浅出，是工业企业数字化转型的决策 者、实施者和研究者难得的参考书。

哈尔滨工业大学人工智能研究院院长、 IEEE Fellow 、ACM 杰出科学家，刘劫

**VI** 数据治理——工业企业数字化转型之道

本书融合了国内外数据治理的权威理论和技术体系，涵盖了工业企业数字化转型过程中所 需要的数据治理技术架构、实施路径和参考案例，可以有效指导工业企业全方位开展高质量的 数据治理，是企业培训、员工培养的优选图书。

华中科技大学电子信息与通信学院院长、IEEE Fellow,邱才明

本书是《DAMA 数据管理知识体系指南(原书第2版)》在泛工业企业领域实践落地的典 范图书，具有很强的实操性和指导性，是数据管理从业人员的优选图书。

国际数据管理协会 (DAMA) 中国分会主席，汪广盛

**序—**

70多年来，伴随着信息革命和信息化的飞速发展，计算机数据量的急剧增长，数据利用和 管理的重要性与日俱增，数据逐渐在信息化这个大舞台上扮演着越来越重要的角色。

早期，数据处理(data processing)解决的是利用计算机技术对数据进行采集、存储、加工、 转换和传输等的技术问题，目的在于将原始的、看似无序的和非结构化的数据，通过格式化的 方法，使其转换为结构化的数据，并存储于计算机系统之中，以便于数据的高效检索、管理和 利用。其后，随着数据的不断增加和重要性的凸显，计算机中的数据管理(data management) 和数据管治(data administration) 成为焦点。前者解决的是对于计算机数据的存储、检索、控 制的管理，包括文件和数据库的接入、数据处理系统的管理等；后者关注的则是一个组织机构 的计算机系统中所存储的数据、信息作为组织机构的资源的管理，包括数据的分析、分类、维 护、流动、应用等。

进入21世纪以来，基于互联网的企业信息系统(企业内部网和外部网)的发展，企业数据 的管理和管治更为复杂，不仅包含了企业内部的各种产品设计、生产、管理数据，还包含了与 企业外部运行环境和竞争环境相关的一切数据；数据不仅要支撑企业的运行和管理，更要为企 业对环境的把握和决策服务。在这样的背景下，企业数据治理(data governance)的概念，及其 理论、方法和工具等应运而生，目的在于对企业所需数据的可获得性、相关性、可用性、整体 性、安全性等，实现全面、有效的管理，将数据作为企业的战略资产加以重视和综合利用，为 实现企业长期的发展战略和增长目标服务。

不过，值得注意的是，数据治理是一个宽泛的概念，在国际、国家、地区、企事业单位， 乃至个人等层面，都存在着内涵各不相同的数据治理问题。



**X** 数据治理——工业企业数字化转型之道

在数据时代来临之际，本书的适时推出，无论是对推动中国企业的数据管理和利用水平的 提高，还是对推动工业大数据的应用发展，无疑都是一件非常有意义的大事。

本书对于工业企业数据治理的讨论非常全面而系统。正如书中所介绍的，完整的数据治理 包括战略、组织、制度、流程、绩效、标准、工具，以及数据价值、数据共享、数据变现等许 多方面。全书既介绍了工业企业数据治理的概念和内涵、标准和框架(特别是主流数据治理的 标准及框架);也从系统工程的角度，介绍了工业企业数据治理体系的各个关键环节，以及现有 的、可获得的各种数据治理工具；而且，书中所给出的大量中国工业企业数据治理的实践和经 验，非常具有启发性、实践性和可操作性。鉴于本书的编著者之一工作于中国工业大数据领域 唯一的国家级工程技术研究平台，对于工业大数据的应用技术、工业大数据的管理和治理，有 着长期、深入的研究和丰富的实践经验，因此，本书对于工业数据治理体系的顶层设计提出的 一系列推进中国大数据应用和治理的建议，特别具有创新性和指导性，值得中国工业界相关领 域的同行认真研究和讨论。

数据治理是现代企业在信息化和全球化的大环境下，谋求竞争优势和向高端发展进程中难 得的一个机遇，也是一个无可回避的挑战。对中国企业更是如此。根据国际数据公司(IDC)

2018年年末的测算，2025年，中国将成为全球五个分区'中，最大的数据资源拥有地区(占比 为28%,数据总量为49ZB), 其数据总量将是美国(排名第四，占比为18%)的1.56倍。实际 上，2019年，中国的数据总量已经超过了美国。但是，本书的研究指出，中国工业企业的数据 资源存量普遍不大，宝贵的数据资源由于缺乏科学的数据管理而随意流失；工业企业数据总量 低下，与企业规模极不相称；半数以上的工业企业仍在使用纸质或更原始的方式进行数据的存 储和管理；数据孤岛几乎是所有工业企业都面临的困境。此外，无论是数据管理还是数据治理， 中国工业企业的状况也不容乐观。调查显示，仅有37.84%的大型工业企业、46.67%的中型工业 企业、13.64%的小型工业企业开展了数据管理工作；大多数工业企业缺乏专门的数据管理部门 投入数据管理的人、财资源也非常有限，更谈不上顶层规划和战略管理。凡此种种都说明，中 国工业企业的数据拥有量、数据管理和治理水平，甚至落后于许多其他行业。这些，都从侧面 证明了中国工业企业在数据管理和数据治理方面亟待迎头赶上。否则，中国制造业的转型升级 将无从谈起。

中国企业与发达国家的企业对标，所显露出来的差距并不可怕。正是这些差距，向我们揭

1 这五个分区是：1)中国；2)欧洲、中东、非洲地区(EMEA);3) 亚太国家，指除中国之外的、包括日本 在内的亚太地区所有国家(APJxC);4) 美国；5)世界其他地区。

序一 Xl

示了大多数中国企业进一步发展和努力的方向，告诉我们中国企业数字化转型的方向和道路何 在。“欲致鱼者先通水，欲致鸟者先树木”。中国企业的数字化转型，只有充分利用“大数据、 人工智能、全联网、云计算”等新一代信息技术提供的条件，以新的形态实现企业业务活动的 数字化和网络化，并且在这个过程中不断认识和强化企业的数据治理，向着智能化的方向进发， 才有可能走上一条与时俱进的发展快车道，跨入现代企业的行列。

衷心期盼本书的出版，能让中国工业界从中获益，有效推动工业企业数据治理的发展，并 促使企业的信息化扎扎实实迈向一个数据驱动的、新的发展阶段。

周宏仁

国家信息化专家咨询委员会常务副主任

2020年7月8日

**序二**

我们正处在一个激动人心的变革时代。马克思于1848年在《共产党宣言》中写道：“资产 阶级在它的不到一百年的阶级统治中所创造的生产力，比过去一切时代创造的全部生产力还要 多，还要大。”这是第一次工业革命带来的巨大变革。今天，我们处在第四次工业革命的前沿。 这次革命同样也将极大地释放人类前所未见的生产力，改变人们的生产、生活方式。第四次工 业革命的本质是信息革命。计算能力、存储能力、网络传输的巨大发展，使得数据的产生、存 储得以呈指数级增长，奠定了大数据普遍应用和改变世界的基础。国家将数据和土地、劳动力、 资本、技术并列为新的生产要素，强调要高度重视数据资源的利用和产业发展。

我在多年前即从事通信大数据的研究和产业化的开创性工作。当时移动通信的主要目标还 是网络的扩容和优化，让更多的人打更多的电话，从而获取更多的收入。同时，企业也认识到 网络服务的形式会快速变化，语音和电话服务的价值将快速下降甚至最终免费，企业必须进行 根本性的业务模式转型。网络服务沉淀了海量的数据，很多人一直在探索如何更好地实现新的 商业模式，发挥它的价值。然而，任何革命式的创新绝不是一帆风顺的，都面临着巨大的困难 和挑战，也需要相当长的时间。数据在本质上和网络一样，需要达到一定的量才能产生价值。 而为了达到这个临界点需要很大的投资，但早期缺乏回报，商业模式也不清晰。同时，数据的 所有权和隐私问题始终是数据采集和应用的关键障碍之一。但是，实践的需要往往推动着历史 的发展。在这次突如其来的“新冠肺炎疫情”中，在反应时间就是一切的情况下，我们看到了 通信网络的位置数据所起到的关键作用，各方数据难以整合的顽疾也在逐步被攻克。因此形成 的精准防控大大降低了防控的成本，提高了防控的效率，总体效果超过了一些西方发达国家。

制造业是立国之本、兴国之器、强国之基。以浏览器为先导、以智能手机为代表的网络革 命实际上是信息和知识传播的革命。它打破了人类社会几千年来的信息垄断和知识获取壁垒。 制造业和互联网、大数据的结合产生了工业互联网、工业大数据。制造业的数字化进程将持续

序二 X

产生海量数据，将来人类社会中的绝大部分数据将是工业数据。5G 网络在制造业中的应用将加 速数据传输，降低数据传输的应用成本。当前，世界各国对工业大数据非常重视。欧盟委会员 在2020年发布的《欧洲数据战略》中认为欧洲拥有强大的工业基础，强调首先要建设欧洲工业 (制造业)公共空间。预计到2027年，欧洲制造业中非个人数据的使用价值可达到1.5万亿欧 元。自2016年中国杭州G20 峰会上提出“数字经济”发展理念之后，在2019年日本大阪 G20 峰会上，各国达成的 DFFT(Data Free Flow with Trust)—— “可信的自由流通的数据”,就是 指工业领域的数据流通和应用。这是世界强国把工业大数据作为全球新技术产业发展方向而达 成的共识。

我国历来重视工业大数据的发展。2017年，国家即设立了工业大数据应用技术国家工程实 验室，作为领域中重要的研究和产业化平台，由中国航天科工集团公司航天云网承建。2020年， 工信部出台了《关于工业大数据发展的指导意见》等一系列政策，进一步明确了工业大数据发 展的目标、方向和具体措施。

发展工业大数据，我国具有独一无二的优势。我国拥有世界上规模最大的制造业、最全面 的工业门类和制造业产业链；同时，经过多年的发展，我国新兴的互联网和大数据产业也位居 世界前列。虽然在很多高端核心技术和高端产品上仍需要追赶，但是发挥好我国的独特优势， 做好工业和大数据的融合发展，利用好世界上数量最大、范围最广的工业数据资源，通过新的 工业大数据帮助实现制造业的知识化、共享化、智能化，是制造业转型升级和引领发展的重要 途径。

当前，工业大数据仍处于产业发展的早期，面临的问题很多、困难很大。工业数据资源的 应用，首先是要聚集、整理企业自身的数据，做好基本的数据治理。要做好工业大数据产业发 展，需要充分发挥国有企业引领的优势，团结业内各个领域，大力协同，构建互利共赢的产业 生态。我们有幸汇集了几十位业务一线的专家，把大家多年研究和实践的经验凝聚并提炼在这 本书里，其中涉及企业数据治理的标准、理论、工具和实施方法，并涵盖了绝大部分工业行业的 实践案例。希望此书的出版发行能为工业大数据的产业发展和企业的数字化创新发展做出贡献。

感谢各位专家和社会各界的大力支持和无私奉献!

祝守宇

工业大数据应用技术国家工程实验室主任

2020年6月23日



**XXIV** 数据治理——工业企业数字化转型之道

**读者服务**

微信扫码回复：39597

● 获取线上直播、技术分享等免费资源

● 加入本书读者交流群，与作者更多读者互动

● 获取博文视点学院在线课程、电子书20元代金券

**目录**

**第1篇** **概述篇**

**第** **1** **章** **工业企业需要数据治理** [2](#bookmark1)

1.1 工业革命的演变与发展趋势 [2](#bookmark2)

1.2 工业大数据是第四次工业革命的核心

基础 [4](#bookmark3)

1.3 各国的工业大数据战略 [6](#bookmark4)

1.4 工业企业数据的核心价值 [7](#bookmark5)

1.5 我国各行业数据治理现状 [10](#bookmark6)

1.6 数据治理是工业大数据的基础 [12](#bookmark7)

1.7 工业企业数据治理面临的挑战 [12](#bookmark8)

本章精要 [14](#bookmark9)

**第** **2** **章** **工业企业数据治理概述** [**15**](#bookmark10)

2.1 数据治理的相关概念和定义 [15](#bookmark11)

2.2 数据的分类 [17](#bookmark12)

2.3 数据治理的顶层架构 [20](#bookmark13)

2.4 数据治理的核心内容 [21](#bookmark14)

本章精要 [22](#bookmark15)

**第** **3** **章** **主流数据治理标准及框架**

**介绍** [23](#bookmark16)

3.1 国际标准 [23](#bookmark17)

3.2 国内标准及模型 [24](#bookmark18)

3.3 专业组织 [26](#bookmark19)

3.4 国内外数据治理体系的对比分析 28

本章精要 [30](#bookmark20)

**第** **4** **章** **数据治理的发展趋势** [31](#bookmark21)

4.1 国内外数据治理的演变与发展 [31](#bookmark22)

4.2 数据隐私保护政策 [32](#bookmark23)

4.3 区块链与数据共享 [33](#bookmark24)

4.4 5G技术与数据安全 [38](#bookmark25)

4.5 新技术与数据治理 [39](#bookmark26)

4.6 数据文化与伦理道德 [40](#bookmark27)

4.7 工业企业数据的运营 [41](#bookmark28)

本章精要 [43](#bookmark29)

5.4 数据治理已在诸多行业成功实施……46

目录 XV

**第** **5** **章** **本书阅读导引………………… 44** 5.3 实施数据治理有路线可循……………45

5.1 数据治理是一个系统工程 [44](#bookmark30)

5.2 工具是数据治理的保障 [45](#bookmark31)

参考资料 [47](#bookmark59)

**第2篇** **体系篇**

**第** **6** **章** **数据管控** 51 7.3 数据战略实施…………………………[82](#bookmark32)

7.3.1 实施策略 [83](#bookmark33)

6.1 数据管控概述 51 ………………………

7.3.2 实施路径 [83](#bookmark34)

6.2 组织架构 [53](#bookmark35)

7.3.3 实施步骤 [83](#bookmark36)

6.2.1 数据治理组织架构 53

本章精要 [87](#bookmark37)

6.2.2 数据治理组织模式 55 ……………………………………

6.2.3 数据治理职责分工 58 **第** **8** **章** **数据架构**……………………… [88](#bookmark38)

6.3 制度规范 [61](#bookmark39)

6.3.1 数据治理制度框架 61 8.1 数据架构概述…………………………[89](#bookmark40)

6.3.2 数据治理制度修订 64 8.2 框架设计………………………………[90](#bookmark41)

6.4 执行流程 65 8.2.1 数据分布………………………[90](#bookmark42)

6.4.1 数据治理总体流程框架 [65](#bookmark43)

8.2.2 数据主题域 [92](#bookmark58)

6.4.2 数据治理典型场景的流程 67 8.2.3 数据关联关系…………………[93](#bookmark44)

6.5 设计机制 70 8.3 数据建模………………………………[98](#bookmark45)

6.6 绩效体系 72 8.3.1 概念数据模型…………………[99](#bookmark46)

6.7 标准体系 74 8.3.2 逻辑数据模型………………[100](#bookmark47)

本章精要 76 8.3.3 物理数据模型…………………[101](#bookmark48)

8.3.4 数据模型开发方法 [102](#bookmark49)

**第** **7** **章** **数据战略** 77 本章精要……………………………………[105](#bookmark50)

7.1 数据战略概述 77 **第** **9** **章** **主数据管理** ……………………[106](#bookmark51)

7.2 数据战略规划 [77](#bookmark52)

7.2.1 愿景和目标 78 9.1 主数据和主数据管理 …………………[106](#bookmark53)

7.2.2 基本原则 79 9.1.1 主数据的特征……………… [106](#bookmark54)

7.2.3 战略举措选择 80 9.1.2 主数据管理的基本概念…… [107](#bookmark55)

7.2.4 模型工具 81 9.2 主数据标准管理………………………[108](#bookmark56)

9.3 主数据全生命周期管理 [109](#bookmark57)

**XVI 数据治理——工业企业数字化转型之道**

9.4 主数据应用管理 [110](#bookmark60)

9.5 企业常用的几类主数据 [112](#bookmark61)

9.5.1 物料主数据 [112](#bookmark62)

9.5.2 设备主数据 [113](#bookmark63)

9.5.3 资产主数据 [114](#bookmark64)

9.5.4 财务主数据 [115](#bookmark65)

9.5.5 组织机构和员工主数据 [116](#bookmark66)

本章精要 [116](#bookmark67)

**第** **1** **0** **章** **元数据管理** [117](#bookmark68)

10.1 元数据的定义 [117](#bookmark69)

10.2 元数据分类 [117](#bookmark70)

10.2.1 业务元数据 [118](#bookmark71)

10.2.2 技术元数据 [119](#bookmark72)

10.2.3 操作元数据 [120](#bookmark73)

10.3 元数据核心能力 [120](#bookmark74)

10.4 元数据的价值 [123](#bookmark75)

本章精要 [124](#bookmark76)

**第** **1** **1** **章** **数据指标管理** [**125**](#bookmark77)

11.1 数据指标管理概述 [125](#bookmark78)

11.1.1 数据指标应用和管理中的

挑战 [125](#bookmark79)

11.1.2 设计目的 [126](#bookmark80)

11.1.3 设计思路 [126](#bookmark81)

11.2 体系框架 [128](#bookmark82)

11.2.1 典型的数据指标定义

框架 [128](#bookmark83)

11.2.2 指标选取原则及方法 [129](#bookmark84)

11.2.3 指标体系层级设计 [130](#bookmark85)

11.2.4 指标体系评价方法 [131](#bookmark86)

11.3 找指标 [132](#bookmark87)

11.4 理指标 [134](#bookmark88)

11.5 管指标 [136](#bookmark89)

11.6 用指标 [137](#bookmark90)

本章精要 [137](#bookmark91)

**第** **1** **2** **章** **时序数据管理** [**138**](#bookmark92)

12.1 时序数据管理概述 [138](#bookmark93)

12.2 时序数据的特点 [139](#bookmark94)

12.3 时序数据的应用 [141](#bookmark95)

12.3.1 技术挑战 [141](#bookmark96)

12.3.2 典型的技术架构及特点 142

12.3.3 系统核心功能 [143](#bookmark97)

本章精要 [143](#bookmark98)

**第** **1** **3** **章** **数据质量管理** [144](#bookmark99)

13.1 数据质量需求 [144](#bookmark100)

13.2 数据质量检查 [145](#bookmark101)

13.3 数据质量分析 [146](#bookmark102)

13.4 数据质量提升 [147](#bookmark103)

13.5 数据质量评估 [149](#bookmark104)

13.5.1 数据质量问题的起因 [150](#bookmark105)

13.5.2 数据质量管理技术指标……151

13.5.3 数据质量管理业务指标 ……152

本章精要 [153](#bookmark106)

**第** **1** **4** **章** **数据安全管理** [155](#bookmark107)

14.1 数据安全管理概述 [155](#bookmark108)

14.2 数据安全体系框架 [156](#bookmark109)

14.3 数据安全防护策略 [159](#bookmark110)

14.4 数据安全审计 [161](#bookmark111)

14.5 数据安全风险评估 [162](#bookmark112)

14.6 数据应急保障 [164](#bookmark113)

本章精要 [165](#bookmark114)

**第** **1** **5** **章** **数据交换与服务** [166](#bookmark115)

15.1 数据交换与服务的意义 [167](#bookmark116)

15.2 数据交换与服务技术演进 [168](#bookmark117)

15.2.1 文件共享技术 [168](#bookmark118)

15.2.2 数据库中间表技术 [168](#bookmark119)

15.2.3 点对点接口技术 [168](#bookmark120)

15.2.4 消息队列技术 [170](#bookmark121)

15.2.5 企业服务总线交换技术 171

15.2.6 ETL 数据交换技术 [173](#bookmark122)

15.2.7 物联网数据采集交换

技术 [173](#bookmark123)

15.3 工业企业数据交换与服务标准

体系架构 [175](#bookmark124)

15.3.1 CPS 信息交换模型 [176](#bookmark125)

15.3.2 设备互联总线 [177](#bookmark126)

15.3.3 应用互联总线 [178](#bookmark127)

15.3.4 数据总线 [179](#bookmark128)

15.3.5 开放互联API 网关 [181](#bookmark129)

本章精要 [182](#bookmark130)

第 3篇

**第** **1** **8** **章** **数据治理工具概述** [**203**](#bookmark131)

**第** **1** **9** **章** **数据资产运营工具** [**207**](#bookmark132)

19.1 数据资产目录 [207](#bookmark133)

19.1.1 总体概述 [208](#bookmark134)

19.1.2 数据资产目录系统构建……208

19.1.3 数据资产目录能力评估

模型 [210](#bookmark135)

**目录** **XVI**

**第** **1** **6** **章** **数据共享与开放** [**183**](#bookmark136)

16.1 共享与开放概述 [183](#bookmark137)

16.2 数据资源目录 [185](#bookmark138)

16.3 数据资源准备 [186](#bookmark139)

16.3.1 数据采集 [186](#bookmark140)

16.3.2 数据加工 [187](#bookmark141)

16.3.3 数据保密 [187](#bookmark142)

16.3.4 数据装载 [189](#bookmark143)

16.3.5 数据发布 [189](#bookmark144)

16.4 数据服务 [190](#bookmark145)

16.5 共享与开放评价 [190](#bookmark146)

本章精要 [191](#bookmark147)

**第** **1** **7** **章** **数据管理成熟度评估** [**192**](#bookmark148)

17.1 数据管理成熟度评估模型 [192](#bookmark149)

17.2 数据管理成熟度等级定义 [195](#bookmark150)

17.3 开展数据管理成熟度评估 [198](#bookmark151)

17.4 数据管理成熟度评估实施 [199](#bookmark152)

本章精要 [200](#bookmark153)

参考资料 [200](#bookmark154)

**工具篇**

19.2 数据资产价值评估 [213](#bookmark155)

19.2.1 总体概述 [213](#bookmark156)

19.2.2 数据资产价值评估模型……214

19.2.3 数据资产价值评估工具……223

本章精要 [224](#bookmark157)

**第** **20** **章** **数据模型管理工具** [**225**](#bookmark158)

20.1 数据模型管理工具概述 [225](#bookmark159)

**XVIH 数据治理——工业企业数字化转型之道**

20.2 企业级数据模型管控 [226](#bookmark160)

20.3 数据标准管控 [228](#bookmark161)

20.3.1 标准的发布和工具访问 228

20.3.2 模型设计中的应用数据

标准 [228](#bookmark162)

20.3.3 数据标准应用情况的自动

检核 [229](#bookmark163)

20.3.4 自定义标准的发布管理 229

20.4 数据字典的质量检核 [230](#bookmark164)

本章精要 [230](#bookmark165)

**第** **21章** **数据指标管理工具** [**231**](#bookmark166)

21.1 指标库管理 [231](#bookmark167)

21.2 指标体系管理 [232](#bookmark168)

21.3 指标评价管理 [233](#bookmark169)

21.4 指标应用管理 [234](#bookmark170)

本章精要 [235](#bookmark171)

**第** **22章** **主数据管理工具** [**236**](#bookmark172)

22.1 主数据提取与整合 [236](#bookmark173)

22.2 主数据模型管理 [237](#bookmark174)

22.3 主数据清洗管理 [238](#bookmark175)

22.3.1 主数据清洗的内容 [239](#bookmark176)

22.3.2 主数据清洗的一般过程 [239](#bookmark177)

22.4 主数据全周期管理 [242](#bookmark178)

22.5 主数据质量管理 [244](#bookmark179)

22.6 主数据发布与共享 [246](#bookmark180)

本章精要 [248](#bookmark181)

**第** **23章** **元数据管理工具** [**249**](#bookmark182)

23.1 元数据管理工具概述 [249](#bookmark183)

23.2 元数据在数据架构管理中的

应用 [250](#bookmark184)

23.3 元数据在数据资产目录中的

应用 [251](#bookmark185)

23.4 元数据在主数据管理中的应用 [251](#bookmark186)

23.5 元数据在数据交换和共享中的

应用 [251](#bookmark187)

23.6 元数据在大数据平台中的应用 252

本章精要 [253](#bookmark188)

**第** **2** **4** **章** **时序数据处理工具** [**254**](#bookmark189)

24.1 通用大数据处理工具的不足 [254](#bookmark190)

24.2 时序数据处理工具应具备的功能

和特点 [255](#bookmark191)

24.3 时序数据的采集 [257](#bookmark192)

24.4 时序数据处理工具 [258](#bookmark193)

本章精要 [260](#bookmark194)

**第** **25章** **数据质量管理工具** [**261**](#bookmark195)

25.1 数据质量管理工具概述 [261](#bookmark196)

25.2 数据质量稽核规则设置 [262](#bookmark197)

25.3 数据质量任务管理 [263](#bookmark198)

25.4 数据质量报告 [264](#bookmark199)

本章精要 [264](#bookmark200)

**第** **26章** **数据交换与服务工具** [**265**](#bookmark201)

26.1 数据交换与服务工具概述 [265](#bookmark202)

26.2 数据采集 [266](#bookmark203)

26.3 数据交换 [268](#bookmark204)

26.3.1 前置交换子系统 [268](#bookmark205)

26.3.2 交换传输子系统 [269](#bookmark206)

26.3.3 交换管理子系统 [269](#bookmark207)

26.4 数据加工服务 [269](#bookmark208)

26.5 数据共享服务 [271](#bookmark209)

26.6 工业大数据技术平台 [272](#bookmark210)

26.6.1 工业大数据的采集 [272](#bookmark211)

26.6.2 工业大数据的交换 [274](#bookmark212)

26.6.3 工业大数据的处理 [275](#bookmark213)

本章精要 [277](#bookmark214)

**第** **2** **7** **章** **数据安全管理工具** [**278**](#bookmark215)

27.1 数据安全管理工具概述 [278](#bookmark216)

27.2 数据采集安全管理工具 [279](#bookmark217)

27.2.1 数据分类与分级工具 [279](#bookmark218)

27.2.2 采集内容及策略 [279](#bookmark219)

27.2.3 数据采集人员管理工具 280

27.2.4 数据源鉴别及记录 [280](#bookmark220)

27.3 数据传输安全管理工具 [280](#bookmark221)

27.3.1 加密算法 [281](#bookmark222)

27.3.2 对称加密 [281](#bookmark223)

27.3.3 非对称加密 [282](#bookmark224)

27.4 数据存储安全管理工具 [282](#bookmark225)

27.4.1 数据存储介质管理 [283](#bookmark226)

27.4.2 数据存储安全 [283](#bookmark227)

27.4.3 数据备份和恢复 [283](#bookmark228)

27.4.4 等级划分 [284](#bookmark229)

27.5 数据处理安全管理工具 [285](#bookmark230)

27.6 数据交换安全管理工具 [286](#bookmark231)

27.6.1 数据导入/导出的安全

保障 [287](#bookmark232)

27.6.2 数据交换安全 [287](#bookmark233)

27.6.3 数据销毁安全管理 [288](#bookmark234)

27.7 统一的身份认证系统 [289](#bookmark235)

目录 XIX

本章精要 [290](#bookmark236)

**第** **28章** **数** **据** **中** **台** [291](#bookmark237)

28.1 数据中台的概念和定位 [291](#bookmark238)

28.2 数据采集 [293](#bookmark239)

28.2.1 数据采集方式 [293](#bookmark240)

28.2.2 通用数据采集 [293](#bookmark241)

28.2.3 流式数据采集 [293](#bookmark242)

28.3 数据存储 [294](#bookmark243)

28.3.1 分布式数据存储 [294](#bookmark244)

28.3.2 NoSQL 数据存储 [294](#bookmark245)

28.4 数据计算 [294](#bookmark246)

28.4.1 分布式查询 [295](#bookmark247)

28.4.2 分布式计算 [295](#bookmark248)

28.4.3 数据建模 [295](#bookmark249)

28.4.4 数据分析 [296](#bookmark250)

28.5 数据服务 [296](#bookmark251)

28.5.1 API 网 关 [297](#bookmark252)

28.5.2 API 生成 [298](#bookmark253)

28.5.3 API 发布 [298](#bookmark254)

28.5.4 API 调用申请 [298](#bookmark255)

28.5.5 API 调用审核 [298](#bookmark256)

28.5.6 API 信息支持 [298](#bookmark257)

28.5.7 API 服务监控 [299](#bookmark258)

28.6 从 ETL 向 ELT转 变 [299](#bookmark259)

本章精要 [300](#bookmark260)

参考资料 [300](#bookmark261)

**XX** **数据治理——工业企业数字化转型之道**

**第4篇** **实施篇**

**第** **29** **章** **数据治理实施策略和路径**

**选** **择** [303](#bookmark262)

29.1 实施内容 [303](#bookmark263)

29.2 路径选择 [304](#bookmark264)

本章精要 [306](#bookmark265)

**第** **30** **章** **数据治理顶层架构规划与**

**设** **计** [307](#bookmark266)

30.1 实施内容 [307](#bookmark267)

30.2 步骤和方法 [309](#bookmark268)

30.2.1 顶层设计总体思路 [309](#bookmark269)

30.2.2 数据治理顶层设计要点 311

30.3 成熟度评估 [322](#bookmark270)

本章精要 [324](#bookmark271)

**第** **31章** **数据资产运营实施** [**325**](#bookmark272)

31.1 实施内容 [325](#bookmark273)

31.2 步骤和方法 [327](#bookmark274)

本章精要 [328](#bookmark275)

**第** **32章** **主数据管理实施** [**329**](#bookmark276)

32.1 实施内容 [329](#bookmark277)

32.2 步骤和方法 [329](#bookmark278)

32.2.1 实施步骤 [329](#bookmark279)

32.2.2 实施方法 [331](#bookmark280)

本章精要 [336](#bookmark281)

**第** **33章** **元数据管理实施** [**337**](#bookmark282)

33.1 实施内容 [337](#bookmark283)

33.2 步骤和方法 [337](#bookmark284)

本章精要 [340](#bookmark285)

**第** **3** **4** **章** **数据指标管理实施** [**341**](#bookmark286)

34.1 实施内容 [341](#bookmark287)

34.2 步骤和方法 [342](#bookmark288)

34.3 模板 [344](#bookmark289)

34.3.1 数据指标项定义 [344](#bookmark290)

34.3.2 形成指标卡片及指标模板 345

34.3.3 数据需求规划 [346](#bookmark291)

本章精要 [347](#bookmark292)

**第** **3** **5** **章** **数据质量管理实施** [**348**](#bookmark293)

35.1 实施内容 [348](#bookmark294)

35.2 步骤和方法 [349](#bookmark295)

35.2.1 数据剖析 [349](#bookmark296)

35.2.2 数据质量诊断 [350](#bookmark297)

35.2.3 数据处理规则 [351](#bookmark298)

35.2.4 数据质量优化 [351](#bookmark299)

35.2.5 数据质量监管 [352](#bookmark300)

35.2.6 实施数据质量管理时需注意

的问题 [353](#bookmark301)

本章精要 [354](#bookmark302)

**第** **3** **6** **章** **数据安全管理实施** [**355**](#bookmark303)

36.1 实施内容 [355](#bookmark304)

36.2 实施步骤 [355](#bookmark305)

36.3 实践模式 [358](#bookmark306)

36.3.1 数据安全管理的建设策略.. 358

目录 XXI

36.3.2 数据安全管理的切入方式 …359

36.3.3 工业互联网云平台的数据

安全 [359](#bookmark307)

本章精要 [360](#bookmark308)

**第** **3** **7** **章** **数据治理常见误区** [361](#bookmark309)

本章精要 [363](#bookmark310)

参考资料 [363](#bookmark311)

**第5篇** **案例篇**

**第** **3** **8** **章** **电力行业：夯实数字化转型**

**基础——南方电网数据资产**

**管理行动实践** [**365**](#bookmark312)

38.1 背景介绍 [365](#bookmark313)

38.2 项目实施 [367](#bookmark314)

38.3 项目成果 [377](#bookmark315)

38.4 项目亮点和洞察 [378](#bookmark316)

38.5 数据治理愿景 [379](#bookmark317)

**第** **3** **9** **章** **电力行业：支撑集团产业**

**数字化转型——国家电投**

**集团数据治理实践** [**380**](#bookmark318)

39.1 背景介绍 [380](#bookmark319)

39.2 数据治理工作实践 [382](#bookmark320)

39.2.1 五凌电力数据治理实践——

水电领域 [386](#bookmark321)

39.2.2 黄河公司数据治理实践——

光、风、水领域 [390](#bookmark322)

39.2.3 云南国际数据治理实践——

风电领域 [392](#bookmark323)

39.3 经验总结 [396](#bookmark324)

39.4 总结与展望 [397](#bookmark325)

**第** **4** **0** **章** **能源化工行业：数据治理**

**助百年油企数字化转型** **……398**

40.1 背景介绍 [398](#bookmark326)

40.2 工作概况 [400](#bookmark327)

40.3 组织保障 [404](#bookmark328)

40.4 主要成果 [405](#bookmark329)

40.5 炼化公司智能工厂数据治理实践

案例 [408](#bookmark330)

40.6 建设主要内容 [409](#bookmark331)

40.7 总结与展望 [411](#bookmark332)

**第** **4** **1** **章** **钢铁行业：酒钢集团数据**

**治理实践** [413](#bookmark333)

41.1 背景介绍 [413](#bookmark334)

41.2 项目目标 [414](#bookmark335)

41.3 项目实施 [414](#bookmark336)

41.4 项目总结 [417](#bookmark337)

41.5 未来展望 [419](#bookmark338)

**第** **4** **2** **章** **汽车行业：数据驱动长安**

**汽车数字化转型** [420](#bookmark339)

42.1 背景介绍 [420](#bookmark340)

**XXI 数据治理——工业企业数字化转型之道**

42.2 工作概况 [422](#bookmark341)

42.3 组织保障 [426](#bookmark342)

42.4 项目成果 [427](#bookmark343)

42.5 工作价值 [431](#bookmark344)

42.6 交流分享 [432](#bookmark345)

**第** **4** **3** **章** **核工业：物料主数据治理**

**助力核电智慧运营** [433](#bookmark346)

43.1 背景介绍 [433](#bookmark347)

43.2 工作概况 [435](#bookmark348)

43.3 组织保障 [437](#bookmark349)

43.4 项目成果 [439](#bookmark350)

43.5 项目总结 [440](#bookmark351)

**第** **4** **4** **章** **航空行业：军工企业的**

**“三位一体”数据治理**

**体系建设实践** [441](#bookmark352)

44.1 背景介绍 [441](#bookmark353)

44.2 工作历程 [443](#bookmark354)

44.3 组织保障 [447](#bookmark355)

44.4 实施效果 [449](#bookmark356)

44.5 项目总结 [450](#bookmark357)

44.6 未来展望 [453](#bookmark358)

**第** **4** **5** **章** **航空行业：面向航空装备**

**研制生产的数据治理研究**

**与实践** [**456**](#bookmark359)

45.1 背景介绍 [456](#bookmark360)

45.2 两级数据管控模式 [459](#bookmark361)

45.3 信息分类与编码标准 [460](#bookmark362)

45.4 数据应用场景 [462](#bookmark363)

45.5 总结与展望 [466](#bookmark364)

**第** **4** **6** **章** **重型装备制造行业：数据**

**标准，装备中国——中国**

**一重的数据标准化管理**

**项目** [468](#bookmark365)

46.1 背景介绍 [468](#bookmark366)

46.2 数据治理概况 [472](#bookmark367)

46.3 数据治理成果 [474](#bookmark368)

46.4 总结与成效 [475](#bookmark369)

**第** **4** **7** **章** **交通物流行业：主数据**

**治理助力中国外运数字**

**化转型** [477](#bookmark370)

47.1 背景介绍 [477](#bookmark371)

47.2 项目实施 [479](#bookmark372)

47.3 主要成果 [482](#bookmark373)

47.4 未来展望 [489](#bookmark374)

**第** **4** **8** **章** **建材行业：中国建材集团**

**工业大数据应用实践** [490](#bookmark375)

48.1 背景介绍 [490](#bookmark376)

48.2 工作实施 [491](#bookmark377)

48.3 应用框架与技术路线 [493](#bookmark378)

48.4 工作成果 [494](#bookmark379)

**第** **4** **9** **章** **制造行业：威孚集团基于**

**斯欧应用互联平台建设数**

**据通道** [501](#bookmark380)

49.1 背景介绍 [501](#bookmark381)

49.2 项目建设技术方案 [505](#bookmark382)

49.3 项目实施步骤 [509](#bookmark383)

49.4 项目实施效果 [511](#bookmark384)

49.5 项目价值及特点 [512](#bookmark385)

**第** **50章** **战略投资行业：国投集团的**

**数据标准化管理实践** [**514**](#bookmark386)

50.1 背景介绍 [514](#bookmark387)

50.2 工作概况 [516](#bookmark388)

50.3 组织保障 [519](#bookmark389)

50.4 项目成果 [520](#bookmark390)

50.5 工作价值 [528](#bookmark391)

50.6 经验分享 [529](#bookmark392)

**目录** **XXI**

**第** **51章** **多元化集团：数据治理助力**

**多元化企业集团管控** **…… 532**

51.1 背景介绍 [532](#bookmark393)

51.2 数据治理概况 [534](#bookmark394)

51.3 组织保障 [539](#bookmark395)

51.4 项目成果 [540](#bookmark396)

51.5 工作价值 [543](#bookmark397)

51.6 项目总结 [543](#bookmark398)

**附录** **A 工业英文缩写术语表** **…………545**



**第** **1** **篇** **概述篇**

第1章

工业企业需要数据治理

如今，人类已经进入以智能化为核心的第四次工业革命， 工业企业依靠智能感知、信息挖 掘、网络协同、认知决策、优化调度的智能化系统来解决规模化生产与定制化、效率提升与成 本控制的问题。智能化的基础是数据，数据变得越来越重要，国家和很多企业已经把数据提升 到战略性资源层面来对待。

**1.1** **工业革命的演变与发展趋势**

按照经济史教科书的说法，第一次工业革命发生于18世纪60年代，人类由手工劳动进入 “蒸汽时代”,由农业文明进入了工业文明。第二次工业革命发生于19世纪70年代，以1866 年德国人西门子研制成发电机为标志，使人类从此由蒸汽时代进入电气时代。第二次工业革命 对世界政治、经济产生了巨大的影响，生产力的迅猛发展改变了社会经济结构和世界秩序。第 三次工业革命发生于20世纪四五十年代，这一时期的主要成就是在原子能技术、航天技术、电 子计算机等方面，人类由电气时代进入信息时代。

目前学术界并没有对“第四次工业革命”的权威解释。各个国家对第四次工业革命的定义 也不同，德国将其命名为“工业4.0”,欧洲其他国家及中国也采用这一概念，美国将其称为“再 工业化”,日本则称其为“工业智能化”。

每一次工业革命的直接驱动力都是日益增长的物质需求和落后的生产力之间的矛盾，每一 次生产力的变革都是缓解这一矛盾的过程。第四次工业革命首先要解决的是规模化生产和定制 化需求的矛盾。消费者越来越追求个性化，如何低成本地满足个性化需求?就是要解决个性和 共性之间的矛盾。再次要解决大规模生产与定制化生产产生的巨大成本差异，要解决设备和工

第1章 工业企业需要数据治**理** **3**

艺的多样性造成技术的普适性和特殊性难以兼顾的矛盾，其核心是要解决自适应和柔性制造问 题。第三要解决宏观和微观之间的矛盾，要解决个体活动和集体活动之间的矛盾，其核心是要 解决协同优化。随着第四次工业革命的深入开展，世界上发达国家在实践中呈现出了不同的技 术路线和实施策略。

**1.** **代表性国家的技术路线**

第四次工业革命是以智能化为核心的工业价值创造革命，为了满足客户个性化的需求，必 须将传统的刚性生产模式转变为柔性生产模式。因此，仅仅靠在现有的控制与信息系统基础上 升级满足不了智能化的要求。德国工业4.0工作项目包括“智能工厂”“智能生产”“智能物流” “智能电网”“智能建筑”等项目。其中“智能工厂”部分根据德国自身在制造业中的优势，从 产品的制造端提出了智能化转型方案。其核心是利用物联网和网络物理生产系统 (Cyber-Physical Production System,CPPS), 等技术，为生产过程中的每个环节建立信息化的链 接，实现“人、机、物、法、环”信息的高度透明。其相应的技术如表1-1-1所示。

**表1-1-1** **智能工厂相应技术**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 对 象 | 客户需求 | 商业流程 | 生产过程 | 产 品 | 设 备 | 人 员 | 供 应 链 |
| 目标 | 定制化、可重  构的生产线 | 动态快速响应 | 透明化 | 生产全流程的 可追溯 | 相互链接、监控、 智能化 | 高效配置 | 按需配给 |
| 技术 | 3D打印、智 能设备 | ERP等管理信 息系统 | 生产线监控、 可视化 | RFID、产品数 据库 | 监控系统、PLC控 制、实时控制技术 | 人员追溯和  通信系统 | 供应链管 理系统 |

在实施方案的规划上，德国提出了“二维战略”的发展思路，从纵向和横向两个维度推进 工业体系的智能化。纵向指的是企业内部“端到端的信息融合”,实现从底层的驱动器和传感器 信号到最高层的战略决策层的无缝连接；横向指的是企业内部价值链及上下游产业链的整合和 协同优化。

不同于德国在制造业和控制领域具有深厚的功底，美国人在互联网技术及应用上有先天的 优势，因此其选择的方向是智能互联。美国国家科学基金会(NSF) 最早于2006年就提出了信 息物理系统(Cyber-Physical Systems,CPS)的概念，即从实体空间的对象、环境、活动中进行 大数据的采集、存储、建模、分析、挖掘、评估、预测、优化、协同，并与对象的设计、测试 和运行性能表征相结合，产生与实体空间深度融合、实时交互、互相耦合、互相更新的网络空 间，从而通过自感知、自记忆、自认知、自决策、自重构和智能支持促进对象的全面智能化。 以CPS 为核心的智能化体系，正是根据工业数据环境中的分析和决策要求设计的，其特征是智

**4** 数据治理- 工业企业数字化转型之道

能感知、数据到信息的转换、网络的融合、自我的认知、自由的配置。

根据这5个特征，美国构建了5层技术模型。第一层是智能感知层，如何高效和可靠地采 集数据。第二层是数据到信息的转换层，也就是信息的挖掘层，如何把各传感器、控制器及管 理信息系统的数据抽取出来转换成信息。第三层是网络层(Cyber), 这里的网络不同于信息系 统中的网络，其指的是网络化的协同管理，是面向设备集群及整个公司的运营及经营活动的横 向数据挖掘。第四层是认知层，也就是识别与决策层。通过CPS 的网络感知，根据健康状况的 历史性分析及通过某种特定的算法预测潜在的故障，为决策提供依据。第五层是配置层，也就 是执行层。由于可以追踪机器的健康状况并做出智能分析，可以根据历史数据及实时采集到的 数据动态调整参数，从而动态优化配置，达到自适应。

**2.** **中国的技术路线**

为了推动第四次工业革命，中国也相继推出了一系列政策，重点强调了推进信息化和工业 化深度融合，推动工业互联网创新发展。2016年2月在工业和信息化部(下文简称工信部)的 领导下，工业互联网产业联盟于2016年发布了《工业互联网体系架构(版本1.0)》,经过3年 的修订和完善，于2019年发布了《工业互联网体系架构(版本2.0)》。该体系共分业务、功能、 技术、实施四大部分，从产业层、业务层、能力层、场景层给予指导。

工信部于2016年发布了《智能制造发展规划(2016—2020年)》。智能制造系统架构从生 命周期、系统层级和智能特征3个维度对智能制造所涉及的活动、装备、特征等内容进行描述， 主要用于明确智能制造的标准化需求、对象和范围，指导国家智能制造标准体系建设。工信部、 国家标准化管理委员会在2015年共同组织制定了《国家智能制造标准体系建设指南(2015年 版)》。智能制造标准体系结构包括“A 基础共性”“B 关键技术”“C 行业应用”3个部分，主要 反映标准体系各部分的组成关系。

中国电子技术标准化研究院于2016年9月发布了《智能制造能力成熟度模型》。该模型由 维度、类、域、等级和成熟度要求等内容组成。

**1.2** **工业大数据是第四次工业革命的核心基础**

第四次工业革命以智能化为特征，以CPS 作为实施核心技术。CPS 的主要特征表现为：

(1)智能的感知：从信息来源、采集方式和管理方式上保证了数据的质量和全面性，建立 支持 CPS 上层建筑的数据环境基础。

第1章 工业企业需要数据治理 **5**

(2)数据到信息的转换：可以对数据进行特征提取、筛选、分类和优先级排序。

(3)网络的融合：将机理、环境与群体有机结合，构建能够指导实体空间的网络环境，包 括精确同步、关键建模、变化记录、分析预测等。

(4)自我的认知：将机理模型和数据驱动模型相结合，保证数据的解读符合客观的物理规 律，并从机理上反映对象的状态变化。

(5)自由的配置：根据活动目标进行优化，进而通过执行优化后的决策实现价值的应用。

物理世界和数字世界的同步必须依赖数据的传递，没有数据就没有 CPS,更谈不上智能化。 另外，从应用的角度来看工业数据为工业企业带来的主要价值体现在：

(1)支持用户直连制造定制( Customer-to-Manufacturer,C2M), 以较低成本满足用户个性 化需求。

(2)使制造过程的信息透明化，从而提升效率、提高质量、降低成本和能耗。

(3)提供设备的全生命周期健康管理，使设备的使用更加高效、节能，提高设备的使用 寿命。

(4)实现全产业链的信息整合，使整个生产系统协同优化，让生产系统变得更加智能，进 一步提高生产效率并降低生产成本。

第四次工业革命另一个重要的技术体系就是工业互联网。工业互联网是新一代信息技术与 工业系统全方位深度融合所形成的产业和应用生态，是工业智能化发展的关键综合信息基础设 施。其本质是以机器、原材料、控制系统、信息系统、产品及人之间的网络互联为基础，通过 对工业数据的全面深度感知、实时传输交换、快速计算处理和高级建模分析，实现智能控制、 运营优化和生产组织变革。网络、平台及安全是构成工业互联网的三大体系，其中网络是基础， 平台是核心，安全是保障。而工业数据是工业互联网核心的核心，是驱动工业智能化的关键要 素和原动力。

企业的信息化和工业物联网中机器产生的海量时序数据，以及与企业运营相关的外部数据 是工业数据的主要来源，且规模巨大。工业数据的综合就是工业大数据，包括企业信息化数据、 工业物联网数据及外部跨界数据，是第四次工业革命的核心基础。

工业大数据是工业领域产品和服务全生命周期数据的总称，包括工业企业在研发设计、生 产制造、经营管理、运维服务等环节中生成和使用的数据，以及工业互联网平台中的数据等。 随着第四次工业革命的深入展开，工业大数据日渐成为工业发展中最宝贵的战略资源，是推动

**6** 数据治理——工业企业数字化转型之道

制造业数字化、网络化、智能化发展的关键生产要素。

我国的工业大数据资源极为丰富。近年来，随着新一代信息技术与工业融合的不断深化， 特别是工业互联网的创新发展，工业大数据应用迈出了从理念研究走向落地实施的关键步伐， 在需求分析、流程优化、预测运维、能源管理等环节，数据驱动的工业新模式、新业态不断涌 现。但相比于互联网服务领域大数据应用的普及和成熟，工业大数据更加复杂，还面临数据采 集汇聚不全面、流通共享不充分、开发应用不深化、治理安全短板突出等问题，总体上仍处于 探索和起步阶段，亟待拓展和深化。

未来3至5年，随着5G、工业互联网、人工智能等的发展，工业大数据将从探索起步阶段 迈入纵深发展阶段，迎来快速发展的机遇期，全球主要国家和领军企业将向工业大数据聚力发 力，积极发展数据驱动的新型工业发展模式，全球工业大数据的竞争也将变得更为激烈。

**1.3** **各国的工业大数据战略**

美国是全球较早关注大数据的国家之一。2009年美国即开通政府数据网站，要求各联邦机 构将需依法公开的数据和文件按照统一标准分类整合，上传至该网站，实现了政府信息的集中、 开放和共享，为启动国家大数据战略奠定了思想基础、技术基础和数据基础。2019年美国发布 《联邦数据战略和2020年行动计划》,描绘了美国联邦政府未来十年的数据愿景，并初步确定了 各政府机构在2020年需要采取的关键行动计划，涵盖了一系列支持美国工业大数据的举措，涉 及大型企业、中小企业、创新型初创企业、研究中心、服务供应商和社会公共机构等。

欧盟委员会认为日益增长的数据对经济和社会发展越来越重要，数据能重塑生产、消费和 生活方式。欧盟委员会于2020年发布了《欧洲数据战略》,积极推进数字化转型工作，打造欧 盟单一数据市场，强化技术主权，提升企业竞争力，以期在新一轮的数字化革命中先发制人。

日本领导人在2019年达沃斯会议上强调“经济增长的动力不再是汽油，而是数字数据”, 并首次提出“要建立针对 DFFT(Data Free Flow with Trust,值得依赖的自由数据流通)的体制”。

2015 年中国首次提出“国家大数据战略”,并发布了《促进大数据发展行动纲要》。2017 年工信部发布了《大数据产业发展规划(2016—2020年)》,对工业大数据建设制定了加快工业 大数据基础设施建设、推进工业大数据全流程应用和培育数据驱动的制造业新模式三大重点工 作规划。2020年2月工信部发布了《工业数据分类分级指南(试行)》,旨在指导企业提升工业 数据管理能力，促进工业数据的使用、流动与共享，释放数据潜在价值，赋能制造业高质量发

第1章 工业企业需要数据治理 **7**

展。2020年4月，工信部又发布了《关于工业大数据发展的指导意见》,明确指出：为促进工

业数字化转型，激发工业数据资源要素潜力，需加快工业大数据产业发展。

从各国工业大数据战略上可看出，发展工业大数据是一项复杂的系统工程：既要构建工业 大数据采集、汇聚、流通、分析、应用的价值闭环，推动创新发展，也要提升数据治理和安全 防护能力，保障发展安全；既需要在宏观层面加强体系化布局，建立全面、系统的工业大数据 生态，也需要在微观层面务实着力，提升企业的数据管理能力；既要重视在需求侧促进大数据 与实际业务深度融合，也要在供给侧推动大数据技术和产业创新发展。但有一点是明确的，即 各工业发达地区和国家对此都非常重视，都是上升到国家战略层面规划布局和发展工业大数据

的 。

**1.4** **工业企业数据的核心价值**

**1.** **以价值创造为核心的工业企业转型**

在前三次工业革命中，以硬实力为代表的技术概念成为价值创造的重要源泉，而第四次工 业革命价值源泉正在向软实力倾斜。老子说过“有之以为利，无之以为用”,意思是说既要看到 有形实体的价值，也要看到真正隐藏在背后的无形价值所在。这种思想在工业企业的商业模式 设计和产品设计上非常有用。比如在制造业，依靠工业互联网平台，不仅能实现客户个性化定 制，同时能通过后续的服务提升价值，由单一的产品销售向产品+服务模式转变。在建筑行业， 依托于 BIM 、CIM 技术，由单一的工程建设向投资、基建、运营转型。通过设计、建设、运营 过程中积累的大量数据，并对这些数据的分析利用，可以优化相关标准，通过向其他企业输出 标准的方式实现数据资产变现。在采矿业，核心企业可以通过产业互联网平台，实现采矿机械 的产业链协同，把整机生产、供应、备件销售、设备维护，金融租赁等相关企业链接到平台上，

从而衍生出新的商业模式。

**2.** **数据驱动下的智能制造**

2012年11月美国通用公司(下文简称 GE) 发布了《工业互联网：打破智慧与机器的边界》 白皮书，在书中首次提出了工业互联网的概念，其代表的是一个开放的、全球化的，将人、数 据和机器连接起来的网络。其核心三要素包括智能设备、先进的数据分析工具以及人与设备的 交互接口。利用智能设备产生的海量数据是工业互联网的一个重要功能，只有从联网的智能设 备中获得数据，才有可能利用大数据及人工智能的分析工具得到“智能信息”(即工业互联网的

**8** 数据治理——工业企业数字化转型之道

第一层级),供决策者使用， 工业互联网的数据循环如图1-4-1所示。



安装仪器仪表的

工业机器

专有机器数据流的 提取和存储

**物理和人际网络**

**PATA 工业数据系统**

安全的，基于云计算

的网络

与正确的人

和机器分析数据

**DATA**

远程和集中式

数据可视化

基于机器的算法 和数据分析

数据流返回机器

**大数据分析**

**图1-4-1** **工业互联网的数据循环图**

工业互联网的第二个层级是智能系统，包括整合广泛的机器、仪器仪表及系统网络上部署 的软件，以及运营网络优化、预测性维护、系统快速恢复、机器自学习等。工业互联网的第三 个层级是智能决策，当智能设备和系统收集到足够的信息以促进数据驱动的学习时，智能决策 就出现了。数据是核心，没有数据的采集和分析、利用，就没有智能化和通过客户订单来驱动

供应链、生产等一系列活动。

**3.** **数据是企业的核心资产**

数据给工业企业带来的价值可以从企业内部经营管理和外部市场两个方面来分析。企业内

部经营管理的核心要素是如何降低运营成本和提高科学决策水平。

比如对煤电行业来说，60%以上的成本是燃料成本，能否提高能效直接决定了企业成本的 高低。因此需要构建影响能效的知识图谱，找出影响能效的相关性因子，通过大数据分析确定 能效和相关因子的函数关系，制定科学的燃煤掺烧比。

对钢铁、机械制造、油气管网等重资产行业来说，如何延长设备的寿命，减少因为非计划 性停机而产生的损失则是企业经营者重点考虑的问题。如何通过数据分析由传统应急性维护迈 向预测性维护?首先找出产生零部件和整机缺陷的相关因子，并构建临界缺陷和相关因子的函 数关系。通过对使用传感器实时采集的状态数据和历史维修记录进行综合分析，预测设备可能 出现的故障时间和故障点，可以提前维护。国内某特大型加工企业主要给苹果等品牌手机做代 加工，其中3C 加工刀具的磨损是影响企业成本的重要因素。因此减少刀具磨损、延长刀具的 使用寿命能显著提高经济效益。工程师们知道刀具切削面的磨损会导致切削阻力变大，相应地

第1章 工业企业需要数据治理 **9**

会使机床主轴电机电流、电压上升。通过构建刀具切削面的磨损和电流、电压变化的函数关系， 在测量机床电流电压的变化后就能知道刀具切削面的磨损，然后反馈给控制系统，由控制系统 发出指令让全自动数控机床调整刀具切削面，从而做到刀具的各个切削面均衡使用，达到延长 刀具使用寿命的目的。单这一项每年就节约成本达数千万元。

家电行业的痛点是销售预测不准而造成产品和原材料库存周转率低，从而大大提高了企业 经营成本。因此，按订单生产和零库存是企业追求的目标。通过消费互联网打通与消费者的连 接，客户通过消费互联网进行个性化定制和提交订单。客户的个性化定制订单传到产业互联网 平台，然后企业通过设计、生产排程、原材料采购到柔性制造等一系列复杂的流程把产品生产 出来，最后再通过消费互联网平台把产品交付给客户。这一人、机、物的高度协同必须依赖数 据驱动，以订单这一数据来驱动后面所有的一系列活动。

数据给工业企业带来的价值在外部市场的体现是能提高产品的附加值。GE 生产的航空发动 机的市场占有率超过50%,在工业互联网被提出前，他们单纯地销售发动机给飞机生产商。对 航空公司来说，减少因发动机故障导致的航班延误及降低安全风险是其核心诉求点。GE 利用他 们在行业的经验开发了一套智能运营系统，帮助航空公司监控发动机的运行情况，不仅做到了 发动机的预测性维护，而且通过优化发动机的燃油消耗，把燃油成本降低了5%。GE 通过后续 的这些附加服务的收费，由单纯的销售产品，变成了产品+服务的收费模式，大大提升了附加值。 类似的国内案例有宁德时代，其建立了一套电池监控平台，通过监控电池的使用状况进行数据 分析不仅可以优化设计，而且能给整机厂提供及时的维修服务，同时把平台开放给4S 店，让 4S店进行众包维修服务。

数据给工业企业带来的价值甚至能改变企业的商业模式，例如由传统的产品销售模式变为 服务模式，由重资产模式转为轻资产运营模式。GE 自发布了《工业互联网：打破智慧与机器的 边界》白皮书后就开启了数字化转型之路：2013年推出工业互联网平台产品 Predix,2015年推 出 Predix2.0,监测分析来自全球各地的超过5000万项数据；2015年成立了 GE Digital,Predix 是核心资产，着力GE 数字化转型；2018年12月，GE以 Predix 平台及数字资产成立全资子公 司，单独运营。GE 以 Predix 工业互联网操作系统为基础，开放公有云平台和帮助客户构建私 有云平台，打造云生态，发力云 SaaS 应用。

**1.5** **我国各行业数据治理现状**

不同行业信息化发展水平不一样，对数据的依赖程度不一样，数据治理水平也不一样，金

**10** 数据治理. 工业企业数字化转型之道

融行业及电信行业是我国数据治理起步较早并且发展较好的行业，具有很好的代表性。下面重 点介绍这两个行业的数据治理情况。

**1.** **金融行业的数据治理**

说到大数据应用所带来的颠覆性变革，没有一个行业比金融行业更加明显。从客户画像到 精准营销，从风险管控到运营优化，几乎所有的业务环节都与大数据息息相关。

2018年5月，中国银行保险监督委员会(简称银保监会)发布《银行业金融机构数据治理 指引》,从数据治理架构、数据管理、数据质量控制、数据价值实现、监督管理等方面规范银行 业金融机构的数据管理活动。这次是银保监会首次将数据治理提高到银行常规管理的战略高度， 明确要将银行数据治理工作常态化、持久化，标志着我国银行业数据治理新时代的正式启幕。

近年来，金融行业在数据治理方面具体以下特点。

(1)强调顶层设计，把数据规划好。

数据治理是一项长期、复杂的系统工程，要在组织、机制和标准等方面加强统筹谋划。

(2)健全治理体系，把数据管理好。

建立全局数据模型和科学合理的数据架构。

通过数据交换机制实现数据的有序流转和安全应用。

(3)加强安全管控，把数据保护好。

加强数据全生命周期安全管理，严防用户数据的泄露、篡改和滥用。

(4)强化科技赋能，把数据应用好。

数据治理的核心环节是数据应用，要提升数据洞察能力和基于场景的数据挖掘能力，为数 据插上翅膀。

**2.** **电信行业的数据治理**

近年来，随着国家大数据发展战略的加快实施，大数据技术创新与应用日趋活跃，产生和 聚集了类型丰富多样、应用价值不断提升的海量网络数据，成为数字经济发展的关键生产要素。 与此同时，数据过度采集与滥用、非法交易及用户数据泄露等数据安全问题日益凸显，做好电 信行业网络数据的管理尤为迫切。为此，电信行业展开了以下几项重要工作。

(1)加快完善网络数据安全制度标准。针对《电信和互联网用户个人信息保护规定》等法

第1章 工业企业需要数据治理 **11**

律规定的要求，出台《网络数据安全标准体系建设指南》,建立网络数据分类分级保护、数据安 全风险评估、数据安全事件通报处置、数据对外提供使用报告等制度，完善网络数据安全标准 体系。

(2)开展合规性评估和专项治理。通过出台网络数据安全合规性评估要点，针对物联网、 车联网、卫星互联网、人工智能等新技术新应用带来的重大互联网数据安全问题，及时开展评 估工作。

(3)推进App 违法违规收集使用个人信息专项治理行动，深化 App违法违规专项治理，强 化网络数据安全监督执法。

(4)强化行业网络数据安全管理。通过明确企业网络数据安全职能部门的职责，实施网络 数据资源“清单式”管理。通过明确企业网络数据安全职能部门的职责，创新推动网络数据安 全技术防护能力建设。通过加强网络数据安全技术手段建设，推动网络数据安全技术创新发展， 完善数据防攻击、防窃取、防泄露、数据备份和恢复等安全技术保障措施，提升企业网络数据 安全保障能力。

其他行业也在陆续开展数据治理工作，但相对而言，金融行业和电信行业做了更多的尝试， 积累了大量的经验，已初步显现出数据治理带来的效益和前景。

**1.6** **数据治理是工业大数据的基础**

工业大数据区别于其他行业大数据是由智能化时代需要人机协同的特点所决定的。它不仅 有企业经营数据、人的行为数据，更重要的是还有来自传感器采集的设备海量数据。其主要特 点如下：

(1)数据来源的多样性，既有经营管理的数据，也有客户行为画像的数据，更有多种设备 状态、控制数据。

(2)数据的实时性，生产现场的数据具有连续性、实时性、数据海量的特点。这就要求数 据的采集、清洗、存储和处理的技术不一样，尤其是需要实时分析。

(3)工业机理的复杂性对知识图谱的构建提出了很高的要求，相应地，数据之间的相关性 分析非常重要。要围绕产品全生命周期、企业全价值链甚至产业链去构建。

(4)从技术层面讲，工业大数据中以非结构化数据、时序数据居多，对存储和处理能力有 更高的要求。

**12** 数据治理——工业企业数字化转型之道

工业大数据的特点需要工业企业有更好的数据治理体系，无论是数据治理的机制还是数据 治理的手段，构建完善的数据治理体系是工业大数据发挥价值的基础和前提。

**1.7** **工业企业数据治理面临的挑战**

工业领域信息化起步相对较晚，工业数据也更为复杂，涉及研发、生产、管理、运维、服 务等多个环节，因而数据管理工作的推进也相对滞后。随着工业互联网发展的不断深化，在工 业领域加强数据管理的重要性日益突出。根据行业信息化发展的现状，结合当今行业数据治理 的要求，工业企业现阶段在数据应用和治理方面依然存在诸多挑战，和金融行业、电信行业相 比还有较大的差距。

**1.** **数据基础薄弱**

我国工业企业的数据资源存量普遍不大，调查显示，66%的企业数据总量都在20TB 以下； 管理手段比较落后，51%的企业仍在使用纸质或更原始的方式进行数据管理。数据孤岛几乎是 所有企业都面临的困境。从单一企业内部来看，存在着不同时期由不同供应商开发建设的客户 管理、生产管理、销售采购、订单仓储、财务人力等众多IT 系统，可谓烟囱林立。而要深度推 进智能化，不仅信息系统要横向互通，还要进一步纵向打通IT(Information Technology)和 OT (Operation Technology) 两界的数据，难度非常大。而且，企业越大，管理和技术包袱越重。

从产业链来讲，工业企业的上下游供应链之间缺少数据的互联互通。大部分企业并没有实 现供应链协同，销售订单和采购订单还依赖于传统的电子邮件或者纸质传递。这种传统的方式 很难做精准的销售预测，更不用说做个性化的定制。

**2.** **数据治理滞后**

随着工业企业信息化的普及以及工业互联网的快速发展， 工业企业对于数据治理的重要性 认识正逐步提高，但实际进行的数据治理工作却不容乐观。调查显示，在大型企业中，已经开 展数据管理工作的企业占37.84%;在中型企业中，已经开展数据管理工作的企业达到了46.67%; 在小型企业中，已经开展数据管理工作的仅占13.64%。

大多数工业企业缺乏专门的数据管理组织，投入数据管理的人力也有限，而且大部分做的 是数据操作基础工作，缺少顶层规划和管理的组织架构和人员。有部分企业建立了数据管理的 相关制度、标准、流程及绩效管理机制，但很多企业在这方面都是缺失的。

第1章 工业企业需要数据治理 **13**

数据管理的技术手段也相对落后。有部分大型企业实施了主数据管理平台，只有极少数企 业实施了数据治理平台，而绝大部分企业都没有任何数据管理的工具，对元数据、主数据、分 析数据等缺乏标准管理、质量管理、安全管理、全生命周期管理的手段。

因为数据管理缺失，技术手段落后，导致企业数据质量难以得到保障，数据共享困难，数 据的价值不能得到充分的挖掘和变现。

**3.** **数据交易法规尚不完善**

工业企业的数据价值体现在跨企业、跨行业的数据流通和共享，数据在流动过程中才能产 生价值。从全行业看，发展工业互联网，实现从单一企业内的局部优化，到整个产业链的全局 优化的跨越，必然要实现整个供应链上下游企业的数据流通，实现产业链上的企业数据的共享。

目前，虽有多家活跃于市场中的数据交易中心，但数据流通的合法、合规性仍未得到应有 的重视，现行法律对于数据流通的很多问题都没有明确，许多工作仍以行业自律的模式开展。 工业企业的数据流通需求与日俱增，规范数据的共享和开放刻不容缓。德国工业4.0 已经计划 把数据流通作为重点议题，在构建工业数据空间方面进行模式上的探索。我国工业企业如何打 破数据孤岛，促进工业数据流通，需要有关方面的高度重视，并有待相关的法律规定和政策的 完善。

**4.** **数据价值难以量化评估**

尽管有很多人意识到数据是企业的核心资产，但是对无形资产的评估比较困难，尤其是数 据资产的量化和评估。首先缺乏财务量化模型，不知道如何评价数据价值；其次数据要在交易 过程中才能变现，而在内部流通的过程中却不能折算成财务意义上的价值，因此其在企业内的 价值无法体现在财务报表上。

数据在工业领域的流动，横向可以跨越设计、采购、生产、销售、售后服务等价值链，纵 向可以跨越战略层与设备控制层。工业企业的特点是数据量大，不同类型、不同层级数据产生 的价值不一样。如何准确地评价这些数据的价值，需要相关机构尽快研究和解决。

数据治理投入大，在短期内很难看到成效，而数据价值的评估又很难量化。因此，很多企 业投入数据治理的意愿不大，这反过来又影响了企业数据的使用。

**14** 数据治理——工业企业数字化转型之道

**本章精要**

本章主要讲述了四次工业革命的演变，主要发达国家的数据战略，国内金融、电信行业的 数据治理现状。第四次工业革命是以智能化为特征，以数据驱动的工业生产，因此数据的重要 性和价值越来越大。但目前工业企业的数据治理滞后，数据管理的组织、制度、流程及技术手 段离智能化的要求差距还很大，制约了工业企业智能化发展，需要引起高度重视。

第2章

工业企业数据治理概述

对工业企业来说，数据越来越重要，但大部分企业的数据治理水平有待提高。本书涉及数 据治理的内容多、范围广，其中概念和术语众多，为了便于读者理解，避免产生歧义，并且对 数据治理的内容有总体的认识，本章先简单介绍一下数据治理的相关概念及主要内容。

**2.1** **数据治理的相关概念和定义**

**1.** **数据与数据管理**

本书中所指的数据是指所有能输入计算机并被计算机程序处理的符号的介质的总称，是用 于输入计算机进行处理，具有一定意义的数字、字母、符号和模拟量等的通称，是组成信息系 统的最基本要素。

数据管理的概念是伴随20世纪80年代数据随机存储技术和数据库技术的使用，计算机系 统中的数据可以方便地存储和访问而提出的。2015年，国际数据管理协会在《DAMA 数据管理 知识体系指南(原书第2版)》中将其扩展为11 个管理职能，分别是数据治理、数据架构、数 据建模与设计、数据安全、数据存储与操作、数据集成与互操性、文件和内容管理、参考数据 和主数据管理、数据仓库和商务智能、元数据管理、数据质量管理。

数据管理是数据资源获取、控制、价值提升等活动的集合，具体指通过规划、控制与提供 数据和信息资产职能，包括开发、执行和监督有关数据的计划、政策、方案、项目、流程、方 法和程序，以获取、控制、保护、交付和提高数据和信息资产价值。

**16** 数据治理——工业企业数字化转型之道

**2.** **狭义数据治理与广义数据治理**

(1)数据治理。

不同的机构对数据治理的定义不一样。

●IS/IEC、TRO38505-2:2018 对数据治理的定义：数据治理是关于数据采集、存储、利用、 分发、销毁过程的活动的集合。

●GB/T4960.5-2018 对数据治理的定义：数据治理就是数据资源及其在应用过程中相关管控 活动、绩效和风险管理的集合。

●国际数据管理协会(DAMA) 对数据治理的定义：数据治理是指对数据资产管理行使权 力和控制的活动集合(规划、监督和执行)。

●国际数据治理研究所(DGI) 对数据治理的定义：数据治理是一个通过一系列信息相关 的过程来实现决策权和职责分工的系统，这些过程按照达成共识的模型来执行，该模型 描述了谁能根据什么信息，在什么时间和情况下，用什么方法，采取什么行动。

(2)狭义的数据治理。

狭义的数据治理指数据资源及其应用过程中相关管控活动、绩效和风险管理的集合，保证 数据资产的高质量、安全及持续改进。在本书第6章“数据管控”中所谈的数据管控即取其狭 义。狭义的数据治理的驱动力最早源自两个方面：

●内部风险管理的需要，风险包括数据质量差影响关键决策等。

●为了满足外部监管和合规的需要，比如萨班斯-奥克斯利法案、巴塞尔I/巴塞尔协议、健 康保险流通与责任法案( HIPAA )等。

但随着全球越来越多的企业认识到信息资产的重要性和价值，在过去几年中，数据治理的 目标也在发生一些转变。除满足监管和风险管理外，如何通过数据治理来创建业务价值备受 关注。

(3)广义的数据治理。

广义的数据治理的含义大于狭义数据治理，包括数据管理和数据价值“变现”,具体包含数 据架构、主数据、数据指标、时序数据、数据质量、数据安全等一系列数据管理活动的集合。

本书取“广义的数据治理”的概念，后面所用到的“数据治理”的概念都指广义的数据 治理。

第2章 工业企业数据治理概述 **17**

**3.** **数据资产与数据资产管理**

(1)数据资产。

数据资产是指由企业拥有或者控制的，能够为企业未来带来经济利益的，以物理或电子的 方式记录的数据资源，如文件资料、电子数据等。在企业中，并非所有的数据都构成数据资产， 数据资产是能够为企业产生价值的数据资源。在这个定义中包含3个要素。

●拥有或者控制：除企业内部的数据外，通过各种渠道合法获取的外部数据也属于企业数 据资产。

●带来经济价值：体现了资产的经济属性，未来能给企业带来经济利益。

●数据资源：数据资产包括各种以物理或电子方式记录的数据、软件、服务等。 (2)数据资产管理。

数据资产管理是指规划、控制和提供数据及信息资产的一组业务职能，包括开发、执行和 监督有关数据的计划、政策、方案、项目、流程、方法和程序，从而控制、保护、交付和提高数 据资产的价值。数据资产管理需要充分融合业务、技术和管理，以确保数据资产的保值、增值。

**2.2** **数据的分类**

工业企业数据的分类维度有很多种，目前业内还没有特别通用的标准，通常可以按照数据 对象划分，也可以按照数据结构来划分，还可以按照数据库类型来划分，等等。下面介绍的是 常见的4种分类方式。

**1.** **按照数据对象划分**

按照数据对象， 工业企业数据可以被分成如下5类。

(1)参考数据。

参考数据是指对其他数据进行分类和规范的数据，如国家、地区、货币、计量单位等产业通 用的数据及各产业特色基础配置数据。为了简化，有的企业称这类数据为配置型主数据，也有 的企业称这类数据为通用基础类数据。它是相对稳定、静态的数据，基本上不会变化，往往通 过系统配置文件给予规范并固化在信息管理系统中。

(2)主数据。

主数据是指满足跨部门业务协同需要的、反映核心业务实体状态属性的基础信息。

**18** 数据治理——工业企业数字化转型之道

主数据是用来描述企业核心业务实体的数据，是企业核心业务对象、交易业务的执行主 体，是在整个价值链上被重复或共享应用于多个业务流程、跨越多个业务部门和系统、高价 值的基础数据，也是各业务应用和各系统之间进行数据交互的基础。从业务角度看，主数据 是相对固定、变化缓慢的，但它是企业信息系统的神经中枢，是业务运行和决策分析的基础， 如供应商、客户、企业组织机构和员工、产品、客户、供应商、物料等主数据(见《数据资产 管理实践白皮书(4.0版)》)。

(3)业务活动数据。

业务活动数据(又称交易数据)是指在业务活动过程中产生的数据，是企业日常经营活动 的直接体现，也是围绕主数据实体产生的业务行为和结果型数据，如采购订单、销售订单、发 票、会计凭证等数据。业务活动数据存在于联机事务处理系统中(OLTP 系统),具有瞬间生成 和动态的特点。

(4)分析数据。

分析数据(又称统计数据、报表数据或指标数据等)是组织在经营分析过程中衡量某一个 目标或事物的数据， 一般由指标名称、时间和数值等组成。

(5)时序数据。

时序数据是指时间序列数据。它是按时间顺序记录的数据列，在同一个数据列中的各个数 据必须是同口径的，要求具有可比性。在工业企业中，实时数据是时序数据的一种，如设备运 行监测类数据、安全类监测数据、环境监测类数据。

**2.** **按照数据的存储形式划分**

按照数据的存储形式， 工业企业数据可以被分为结构化数据、非结构化数据、半结构化数 据(下面的名词解释来自《管理科学技术名词》)。

(1)结构化数据。

结构化数据是指数据元素之间具有统一且确定关系的数据。它由明确定义的数据类型组成。 结构化数据的一般特点是数据以行为单位， 一行数据表示一个实体的信息，每一行数据的属性 是相同的。结构化数据的分析更为便利，且存在成熟的分析工具。

(2)非结构化数据。

非结构化数据是指数据元素之间没有统一和确定关系的数据。它是具有内部结构，但不通

第2章 工业企业数据治理概述 **19**

过预定义的数据模型或模式进行结构化的数据，如各种格式的图片、视频等。非结构化数据占 企业全部数据的80%以上，但直接分析非结构化数据得有很强的专业性。

(3)半结构化数据。

半结构化数据是指数据元素之间的关系介于结构化数据和非结构数据之间的数据。它是非 关系模型的、有基本固定结构模式的数据，例如日志文件、XML 文档、JSON 文档、E-mail等。

**3.** **按照数据库的类型划分**

按照数据库的类型，工业企业数据可以被分为关系型数据库、非关系型数据库、图数据库、 时序数据库。

(1)关系型数据库。

关系型数据库是采用关系数据模型的数据库系统。关系数据模型实际上是表示各类实体及 其之间联系的由行和列构成的二维表结构。 一个关系数据数据库由多个二维表组成。表中的每 一行为一个元组(或称一个记录),每一列为一个属性。属性的取值范围被称为域。对关系型数 据库进行操作通常采用结构化查询语言(SQL) (见《管理学大辞典》)。

(2)非关系型数据库。

非关系型数据库是对不同于传统的关系数据库的数据库管理系统的统称。和关系型数据库 相比，两者存在许多显著的不同点，其中最重要的是非关系型数据库使用NoSQL 而不使用SQL 作为查询语言。其数据存储可以不需要固定的表格模式，也经常会避免使用SQL的 JOIN 操作， 一般有水平可扩展性的特征(来自维基百科)。

(3)图数据库。

图数据库是以图结构来表示和存储信息的数据库(见《计算机科学技术名词》)。

(4)时序数据库。

时序数据库是指时间序列数据。它是按时间顺序记录的数据列，在同一个数据列中的各个 数据必须是同口径的，要求具有可比性。时序数据可以是时期数，也可以是时点数。

**4.** **按照权属类型划分**

按照权属类型，工业企业数据可分为私有数据和公有数据。

(1)私有数据。

**20** 数据治理——工业企业数字化转型之道

私有数据是指有明确归属的数据，归属方为可决定数据使用目的的自然人、法人或其他组 织。如私人数据、企业数据等。

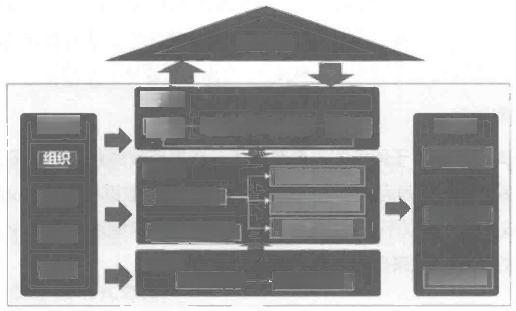
(2)公有数据。

公有数据指具有公共财产属性且可被公众访问的数据，如天气数据、人口数据等。

**2.3** **数据治理的顶层架构**

企业中不同层级的人对数据治理的关注点不一样，因此各自的视图也不一样。下面从管理 者视图来看数据治理的顶层架构。

数据治理的管理者视图可以概括为“五域模型”,即管控域、过程域、治理域、技术域、价 值域，如图2-3-1所示。



**数据战略**

**过程域**

[分析 设计]一实施 评值

|  |  |
| --- | --- |
| **治理域**  标准体系  质量][女全  **技术域**  **数据架构** | 数据指标治理 交易数据治理 主数据治理  **治理工具** |

**价值域**

数据价值

**数据共享**

**数据变现**

**制度**

**流程**

**绩效**

**管控域**

**图2-3-1** **五域管理视图**

企业在开展数据治理之前，首先要基于企业战略和IT 战略制定数据治理的战略目标，在明 确战略目标的基础上再细化“五域模型”内容。

(1)管控域：在数据治理战略指导下制定企业数据治理组织，明确组织的责、权、利、岗 位编制及技能要求。 一般在大中型企业中会设立由企业高层领导及相关专家组成的数据治理委 员会，审批数据治理相关的重大决策，并制定数据治理的相关制度、流程，建立数据认责及绩 效考核机制，以支撑数据治理活动。

(2)治理域：是数据治理的主体，明确数据治理的对象和目标。根据数据资产的构成，企

业数据治理又分为主数据治理 、 交易数据治理和数据指标治理 。

第2章 工业企业数据治理概述 **21**

(3)技术域：数据治理的支撑手段，提供数据治理所需的数据架构、治理工具平台，包括 元数据管理、主数据管理、数据指标管理、数据模型管理、数据质量管理、数据安全管理等功 能模块。

(4)过程域：是数据治理的方法论。数据治理过程包括评估与分析、规划与设计、实施的 PDCA 循环(即 Plan 、Do 、Check 和 Action 循环，也称戴明环)。在评估与分析阶段，要评价 现有数据治理的成熟度、风险及合规性，业务对数据治理的需求。在规划和设计阶段，要明确 数据治理的目标和任务，制定数据治理的相关制度和流程，设计数据标准、数据模型、数据架 构及数据治理的实施路径。在实施阶段，要制定数据治理的相关制度、流程细节，选择合适的 数据治理工具并通过定制化开发来满足数据治理要求。

(5)价值域：数据治理的目标就是通过对数据资产的管控，挖掘数据资产的价值，并通过 数据的流动、共享、交易，实现数据资产的变现。具体包括以下3个方面：

●数据价值：对数据价值进行财务建模及数据价值评估的过程。

●数据共享：通过实现信息整合和分发机制，支持跨业务、跨部门、跨行业、跨企业的信 息流通和共享。

●数据变现：通过数据的共享和交易，将数据转变成财务意义上的资产。

**2.4** **数** **据** **治** **理** **的** **核** **心** **内** **容**

完整的数据治理包括战略、组织、制度、流程、绩效、标准、工具及数据价值、数据共享、 数据变现。其中数据价值、数据共享及数据变现等不是本书的讨论重点，这里就不赘述了。

**1.** **战略**

数据治理的首要任务是制定数据治理战略目标，否则缺乏目标和行动纲领，数据治理难以 开展。企业的信息化是为了服务于业务，因此，企业的信息化战略必须匹配业务战略。数据战 略是信息化战略的重要组成部分，企业要清晰地定义企业数据治理的使命、愿景，中长期目标 及行动计划，用以指导企业数据治理。企业数据战略一般根据 IT 战略的制定而制定，随着 IT 战略的修订而修订，由企业的信息化负责人及业务负责人共同主导制定。

**2.** **组织**

建立合适的数据治理组织是企业数据治理的关键。数据治理的组织建设一般包括组织架构

**22** 数据治理——工业企业数字化转型之道

设计、部门职责、人员编制、岗位职责及能力要求、绩效管理等内容。

**3.** **制度**

企业的数据治理必须要有相关制度，否则无法可依，再好的技术工具也没有用。因此，建 立完善的数据治理制度很重要。企业的数据治理制度通常根据企业的IT制度的总体框架和指导 原则制定，往往包含数据质量管理、数据标准管理、数据安全管理、数据绩效管理等制度，以 及元数据管理、主数据管理、交易数据管理、数据指标管理等办法及若干指导手册。

**4.** **流程**

制定数据治理的流程框架及流程也是数据治理的重要工作。数据治理流程主要包括从数据 的生产、存储、处理、使用、共享、销毁全生命周期过程中所遵循的活动步骤，以及元数据管 理、主数据管理、数据指标管理等流程。

**5.** **绩效**

要使数据治理的体系运转好，必须要有好的激励体系。数据绩效管理包括数据管理指标、 数据认责机制、数据考核标准、数据管理的奖惩机制，以及绩效管理过程的一系列活动集合。

**6.** **标准**

数据标准是实现数据标准化、规范化的前提，是保证数据质量的必要条件。数据标准一般 分为元数据标准、主数据标准、交易数据标准、数据指标标准、数据分类标准、数据编码标准、 数据集成标准等内容。数据标准管理是规范数据标准的内容、程序和方法的活动，分为标准制 定、标准实施和控制、标准修订等。

**7.** **工具**

数据治理管理工具包括数据架构工具、元数据管理工具、数据指标管理工具、主数据管理 工具、时序数据管理工具、数据交换与服务工具、质量管理工具和安全管理工具等。

**本章精要**

本章主要介绍了数据治理相关的重要概念和内容，其中简要介绍了数据治理的核心内容， 让读者对这些概念有清晰的了解，并对数据治理内容框架有初步的认识。

建议和计划

第3章

主流数据治理标准及框架介绍

**3.1** **国际标准**

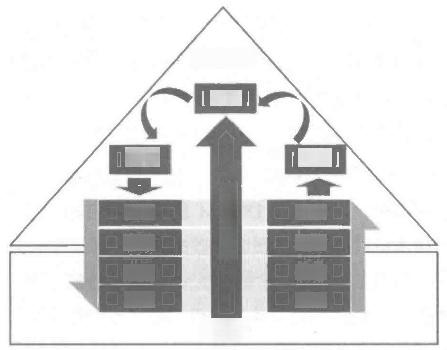
2015年，国际标准化组织IT 服务管理和IT 治理分技术委员会制定了ISO/IEC 38500系 列

标准，提出了IT 治理的通用模型和方法论。

在 ISO/IEC 38505 标准中，阐述了基于原则驱动的数据治理方法论，提出了通过评估现在

和将来的数据利用情况，指导数据治理的准备及实施，并监督数据治理实施的符合性等。其数

据治理的框架模型如图3-1-1 所示。

**治理主体**

评估

**指导**

**监督**

**战略**

策略 控制 过程

战略

策略

控制

过程

管理者

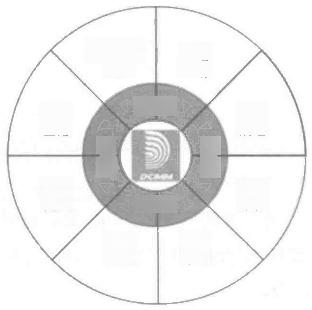
图3-1-1 数据治理框架模型图

**24** 数据治理- 工业企业数字化转型之道

**3.2** **国内标准及模型**

**1.** **数** **据** **管** **理** **能** **力** **成** **熟** **度** **评** **估** **模** **型** **(DCMM)**

DCMM(Data Management Capability Maturity Model, 数据管理能力成熟度评估模型)是在 工信部、国家标准化管理委员会的指导下，由全国信息技术标准化技术委员会大数据标准工作 组组织编写的国家标准，也是我国首个数据管理领域国家标准。DCMM 借鉴了国内外数据管理 的相关理论思想，并充分结合了我国大数据行业的发展趋势，创造性地提出了符合我国企业的 数据管理框架。该框架将组织数据管理能力划分为8个能力域：数据战略、数据治理、数据架 构、数据标准、数据质量、数据安全、数据应用和数据生存周期，如图3-2-1 所示。

**数据**

**数据生存**

**周期**

**战略**

**数据** **应用**

长 塔

**数据** **安全**

**组** **织**

烟 露

**流程**

**数据**

**治理**

**数据**

**架构**

**数据**

**数据** **标准**

**质量**

**图3-2-1** **数据管理能力成熟度评估模型**

DCMM 新增了数据生存周期管理功能域，这是一个进步，它考虑到了原始数据转化为可用 于行动的知识的整个过程，包括数据需求、数据设计与开发、数据运维直至数据退役。只有让

数据治理工作贯穿数据的整个生存周期，才能彻底将数据治理到位。

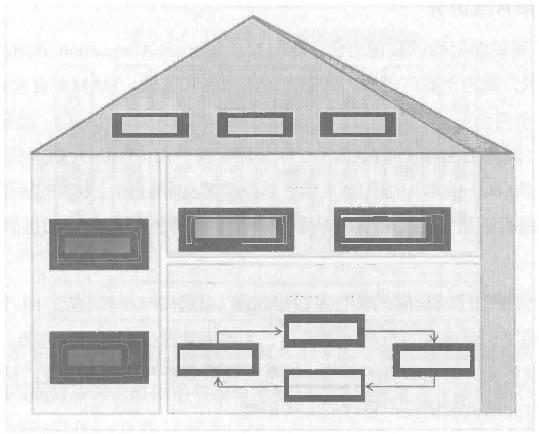
DCMM 的优点在于它不只是理论和知识体系，而是可以直接应用的。而且 DCMM 已经在 工业企业中有过很多应用案例。为了推进 DCMM 国家标准的落地实施，指导相关组织提升数 据管理能力，全国信息技术标准化技术委员会大数据标准工作组在全国范围内组织开展了数据 管理能力成熟度评估试点示范工作，涵盖金融、能源、互联网和工业等多个领域的30余家企事 业单位，其中就包括7家工业企业。DCMM 的缺点也很突出：通过数据管理能力成熟度评估只 能了解组织数据管理现状，包括已取得的成果和不足，但是并不提供能力提升的方法，还需要

数据管理专家给出提升建议、方法论和实施路线图。

**2.GB/T** **34960** **数据治理规范**

GB/T34960《信息技术服务治理第5部分：数据治理规范》(以下简称《数据治理规范》) 是我国信息技术服务标准(ITSS) 体系中的服务管控领域标准，该标准根据GB/T34960.1-2017 《信息技术服务治理第1部分：通用要求》中的治理理念，在数据治理领域进行了细化，提出 了数据治理的总则、框架，明确了数据治理的顶层设计、数据治理环境、数据治理域及数据 治理的过程，可对组织数据治理现状进行评估，指导组织建立数据治理体系，并监督其运行和 完善。

《数据治理规范》将数据治理划分为顶层设计、数据治理环境、数据治理域和数据治理过程 4大部分，如图3-2-2所示。

战略规划

**数据治理环境**

内外部环境

**预层设计**

组织架构 架构设计

**数据治理域**

数据管理体系 数据价值体系

**数据治理过程**

统筹和规划

改进和优化

促成因素

**构建和运行**

监控和评价

**图3-2-2** **数据治理框架**

顶层设计包括制定数据战略规划、建立组织机构和机制、建立数据架构等，是数据治理实

施的基础。

数据治理环境包括分析业务、市场和利益相关方需求，适应内外部环境变化，营造企业内 部数据治理文化，评估自身数据治理能力及驱动因素等，是数据治理实施的保障。

数据治理域包括数据管理体系和数据价值体系，是数据治理实施的对象。

数据治理过程包括统筹和规划、构建和运行、监控和评价、改进和优化，是数据治理实施

**26** 数据治理 工业企业数字化转型之道

的方法。

《数据治理规范》开创性地把数据价值实现作为数据治理的核心目标，并通过数据价值体系 明确了数据价值实现的方式，帮助企业实现数据驱动业务的战略转型。

在《数据治理规范》附录中对数据治理涉及的核心治理域提出了明确的管理要求，为数据 治理实施提供参考，为评估数据治理成效提供评价依据，通过正文和附录的结合，有利于数据 治理的落地实施。

**3.3** **专业组织**

**1.** **国际数据管理协会**

成立于1988年的国际数据管理协会(Data Management Association International,DAMA) 是一个非营利组织，致力于推广信息和数据管理的概念和实践。DAMA 在全球设立了40多个 分会，拥有7500余名会员，在数据管理领域中累积了丰富的知识和经验，是全球公认的数据管 理权威组织之一。其先后出版了《DAMA 数据管理字典》和《DAMA 数据管理知识体系指南》 (简称 DAMA-DMBOK) 的第1版和第2版，该指南集业界数百位专家的经验于一体，是数据 管理业界最佳实践的结晶，已被公认为从事数据管理工作的经典参考和指南，在全球范围内 广受好评。

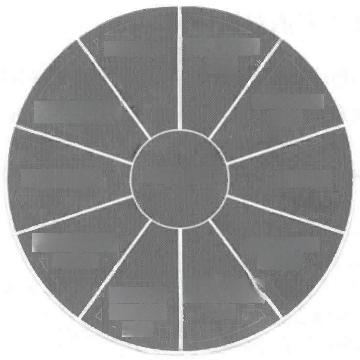
DAMA 的数据管理理论框架的核心是数据治理(见图3-3-1),通过10个数据治理的职能 域建立一个能够满足企业需求的数据决策体系，为数据管理提供指导和监督。其优点在于充分 考虑了功能与环境要素对数据本身的影响，但考虑到数据资产化成为企业的核心竞争力，这10 个职能域尚不能全面覆盖数据资产管理的业务职能。

**2.** **数据资产管理实践白皮书**

为了落实国家大数据战略，中国信息通信研究院联合相关知名企业共同编写了《数据资产 管理实践白皮书(4.0版)》。

《数据资产管理实践白皮书(4.0版)》基于 DAMA-DBMOK 中定义的数据管理理论框架， 弥补了数据资产管理特有功能的缺失，并结合数据资产管理在各行业中的实践经验，形成了数 据资产管理的8个管理职能和5个保障措施，如图3-3-2所示。其中管理职能是指落实数据资 产管理的一系列具体行为，保障措施是为了支持管理职能实现的一些辅助的组织架构和制度体 系。

第3章 主流数据治理标准及框架介绍 **27**



数据建模和

设计

**数据存储和** **操作**

**元数据管理**

数据仓库和商

务智能

**文件和内容**

参考数据

和主数据

**管理**

**数据集成和** **互提作**

**数据质量管理**

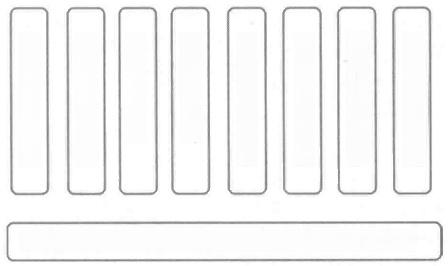
**数据治理**

**数据架构**

**数据安全**

**管理**

图3- 3- 1 DAMA 的 数 据 管 理 理 论 框 架



数据共享管理

数 据 价 值 管 理

数据安全管理

数据质量管理

主数据管理

元数据管理

数 据 模 型 管 理

数 据 标 准 管 理

战略规划、组织架构、制度体系、审计制度、培训宣贯

图3 - 3 - 2 《 数 据 资 产 管 理 实 践 白 皮 书 ( 4 . 0 版 ) 》 的 数 据 管 理 理 论 框 架

在 DAMA 的数据管理理论框架中并没有把数据标准单独作为一项重要的数据管理功能，而 《数据资产管理实践白皮书(4.0版)》将数据标准管理放在第一位，体现了“标准先行”的管理

思想。另外，其中还增加了数据价值管理和数据共享管理两项内容。

数据价值管理是对数据内在价值的度量，包括数据成本和数据应用价值。

数据共享管理主要是指开展数据共享和交换，实现数据内外部价值的一系列活动。数据共 享管理包括数据内部共享(企业内部跨组织、部门的数据交换)、外部流通(企业之间的数据交

换)和对外开放。

**28** 数据治理——工业企业数字化转型之道

**3.4** **国内外数据治理体系的对比分析**

国内外各相关机构都有自己的数据治理体系，但各体系之间存在一定的差异。表3-4-1 从

主流数据治理体系的构成要素、治理特点、优势和不足4个方面进行了分析。

**表3-4-1** **主流数据治理体系对比分析**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **体系** **分类** | **体系名称** | **构成要素** | **治理特点** | **优** **势** | **不** **足** |
| 国际 标准 | ISO 8000 数据质量的 国际标准 | 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。  数据质量标准由以下4个部分组成 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 | 基于国际协议； 对于使用哪些数 据特征来定义和 衡量其质量得到 了世界各地的行  业专家认可；适用  于各类工业数据，  包括主数据、交易  数据和产品数据 | 在数字供应链  方面能够发挥重  要作用，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 | 主要关注数据的质 量，只是整个数据治 理体系的一部分功能 |
| 国际 标准 | ISO  38500 IT治 理国际标准 | 包括IT治理的目标、原则和模型。 5项基本原则是指职责、策略、采购、 绩效合规和人员行为；在模型方面，  认为企业的领导者应该重点关注3 项核心任务：  一是评估现在和将来的IT利用情 况 ；  二是对治理准备和实施的方针和 计划做出指导  三是建立“评估一领导一监督” 的循环模型 | 第一个IT治理 的ISO标准；  跳 出 了 P D C A 等生命周期的概  令，采用了Direct Evaluate-  Monitor模型 ( D E M 模 型 ) | 确保所有IT风  险和活动都有明  确的责任分配，尤  其是分配和监控  IT安全责任、策略  和行为，以便采取  适当的措施和机  制，对当前和计划  的IT建立报告和  响应机制 | 聚焦于更广泛层次 上的IT治理，虽然其 号称也适用于数据治 理，但是相对来讲针 对性并不强；  如果企业将IT业务 外包，则无法把一些  针对IT管理者的特 殊要求强加给公司管 理者  介绍了好的治理所 需要的一些特征和治 理流程，但是离真正 的实施还有距离，需  要COBIT、ITL或者 其他的标准或框架的 补充 |

**第3章** **主流数据治理标准及框架介绍** **29**

**续表**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **体系** **分类** | **体系名称** | **构成要素** | **治理特点** | **优** **势** | **不** **足** |
| 国内 标准 | DCMM  数据管理能 力成熟度评 估模型 | 定义了8个能力域：数据战略、 数据治理、数据架构、数据标准、 数据质量、数据安全、数据应用和 数据生命周期管理 | 企业可以清楚 地定义企业内数 据的能力水平并 以模型为标准确 定组织内数据改 进方向 | 量化评估企业 数据管理能力水 平 ；  指明企业数据 管理能力缺陷；  通用性较高 | 只提出了数据管理  应该具备什么能力，但  是并未指明应该怎么  做，落地效果不明显 |
| GB/T  34960.5-201 8 《信息技 术 服 务 治 理 第 5 部 分：数据治 理规范》 | 包含顶层设计、数据治理环境、 数据治理域和数据治理过程4大部 分。数据治理域是数据治理的对  象，包括数据管理体系和数据价值 体系两个部分。  其中数据管理体系包含数据标  准、数据质量、数据安全、元数据 和数据生存周期；数据价值体系涵 盖数据流通、数据服务和数据洞察 | 通过正文和附 录相结合的方式， 解决了国际数据 治理标准不易落 地的问题 | 增加了数据价 值体系，提出了 面向数据价值实 现的治理目标； 也对数据治理体 系的实施路径提 出了要求，解决 了治理与管理脱 节的问题 | 高屋建瓴地指出了 数据治理体系应该包  含的内容和落地实施 的路径，但是缺乏具体 实施办法，在具体工作  开展中仍需要大量细 化的工作 |
| 专业 组织 | D A M A 数 据管理知识 体系 | 总结了数据治理、数据架构、数 据建模和设计、数据存储和操作、 数据安全、数据整合和互操作、文 档和内容管理、参考数据和主数据、 数据仓库和商务智能、元数据、数 据质量11个职能领域，以及目标和 原则、活动、主要交付物、角色和 责任、技术、实践和方法、组织和 文化7大环境要素，并建立了11个 职能领域和7大环境要素之间的对 应关系 | 以数据管理为  主导，数据管理的 核心是数据治理 解决了数据治理 各项功能与环境 要素的匹配问题 | 充分考虑到功 能与环境要素对 数据本身的影 响，并建立对应 的关系 | 11个职能领域，全  面建设，复杂度较高；  全面实施企业级数  据治理，可落地性、难 度较高 ；  11个职能领域尚不 能满足未来数据治理， 尤其是数据资产管理 的需求 |
| 《 数据资 产管理实践 白皮书(4.0 版 ) 》 | 包括数据标准、数据模型、元数 据、数据安全等8个方面 | 是一套针对数  据资产的管理体  系，引入了数据资  产价值管理和运  营等内容，并囊括  了数据资产管理  在过程当中的一 些管理工具 | 偏重数据资产 管理方面的国家 标 准 ；  实 践 案 例 丰 富，可参考价值 较高 | 偏重行业实践案例  研究与风险，理论指导 性稍弱。提出的时间较 短，所以经过的企业验 证还很少 |

**30** 数据治理——工业企业数字化转型之道

不同的数据治理框架和标准适用于不同的行业、企业，企业应根据自身特点，选择适合自 己的标准体系。数据治理体系也不是一成不变的，因为数据治理是一个动态过程，过于僵化的 体系不仅不会给工作带来便捷，还会增加应用的复杂度。

**本章精要**

本章介绍了数据治理相关的国际、国家标准，以及行业组织和协会的数据治理框架，并对 各个框架的优、缺点及适用场景做了比较，便于读者根据企业的实际情况进行比较。

数据治理框架是指为了实现数据治理的总体战略和目标，将数据治理域所蕴含的基本概念 (如原则、组织架构、过程和规则)组织起来的一种逻辑结构。本章先从国际、国内相关标准和 专业组织3个方面，全面介绍数据治理框架，之后对各个框架进行对比分析，以便于企业进行 选择。

第4章

数据治理的发展趋势

数据治理的发展是伴随着各国及不同行业对数据资源资产化，数据确权与合规，数据价值 创造与共享，隐私保护的认识、研究和实践的一个演进过程。目前，随着数据治理理论体系的 逐步完善，技术方法和工具的日趋成熟，数据治理被越来越多的企业学习、了解和实际应用。

**4.1** **国内外数据治理的演变与发展**

**1.** **国外数据治理的发展**

自“数据治理”的概念被首次提出以来，数据治理已在政府、企业、公共服务性机构等中 得到广泛的关注、研究和实践。根据国际数据管理协会发布的调研报告，数据是企业最为宝贵 的资产之一，已成为企业界的共识。然而，目前企业的数据状态与数据管理水平并不匹配，普 遍存在着“重创造轻管理、重数量轻质量、重利用轻增值利用”的现象，在数据质量、服务创 新、开放共享、安全合规、隐私保护以及数据伦理道德规范等方面面临着越来越严峻的挑战。 数据管理中出现的问题，究其根源是由于在更深的层面——数据治理中出现混乱或缺失。

事实上，以欧洲、北美洲、日本和澳大利亚等为代表的地区和国家，都高度重视数据治理 工作，自2012 年以来都密集出台了多项政策予以支持和指导。从各国的举措来看，政策着力 点主要在于3个方面：其一是开放数据，给予产业界高质量的数据资源；其二是在前沿及共性 基础技术上增加研发投入；其三是积极推动大数据的共享与应用。

**2.** **中国数据治理的发展**

在中国，大数据已上升到国家战略层面，大数据技术和应用对国家的数字化治理和企业的

**32** 数据治理——工业企业数字化转型之道

智能化升级，都产生着深刻的影响。

近年来数据治理在中国的蓬勃发展更偏重于实践，更聚焦于数据治理工程项目的落地实施 和技术工具的设计开发。在金融、电信、能源、互联网等信息化较为成熟的行业中，多年前就 已经积极开展相关的数据治理工作，并积累了一定的数据资产管理经验。这些经验的总结对于 补充及完善数据治理理论体系、推进数据资产管理在各个行业的普及和发展有着重要的指导 意义。

中国信息通信研究院、中国通信标准化协会大数据技术标准推进委员会(CCSATC601) 及 众多的技术研究社团和机构，通过白皮书、技术论坛和研讨会的形式，就数据治理相关的内容 和目标、实施步骤、实践模式、技术工具及需要注意的要素等问题进行研究和讨论，为政府和 企业开展数据治理工作提供参考，为相关服务商和工具产品开发提供指导。

各级政府和中央大型企业在数据治理方面担任了关键主体，其既是重要推动者也是执行者。 在推进数据治理和大数据应用的过程中，有一批数据服务型公司脱颖而出，形成完整的技术方 案和工具产品；除大型企业外，也有许多中小企业开始关注和实践数据治理，成功实现数字化 转型。可以说中国的数据治理理论体系正逐步完善，实施工具越来越丰富，工业企业的数据治 理热潮已经到来。

**4.2** **数据隐私保护政策**

无论是企业还是政府或其他机构，在数据应用层面，首先面临的问题就是数据的隐私保护。 数据的隐私保护与数据的安全保障的侧重点不同，但都是倡导在数据共享互通的前提下对数据 使用权限、范围和目的限制。随着数据资料传递越来越容易，获取企业数据信息和个人影音资 料的渠道变多，这些数据在被应用时的目的，可能已与当初记录资料、采集数据时完全不同了。 如何保障数据隐私，成了企业、政府及相关机构关注的问题。在执行数据共享的同时，如何防 止滥用数据的情况发生，各国政府，都在积极寻求新的立法，以保障数据的流通、共享与合理 使用。

关于保护隐私数据的立法，各国的法律法规虽然不同，但大致有3种模式：第一种是以美 国为代表的行业自律模式，这一模式依靠企业的自我约束及行业协会的监管，以避免过度保护 形成对行业正常发展服务的阻碍，也避免国家立法限制信息科技在社会的应用；第二种是以欧 盟为代表的纯粹的立法规制模式，是由国家和政府主导的一种模式，在法律层面确立隐私信息 保护制度，进而将这些制度作为建立司法和行政措施的基础；第三种模式是以日本为代表的将

第4章 数据治理的发展趋势 **33**

行业自律模式与立法规制模式相结合的模式，既通过法律全方面地保护隐私信息，也注重行业 自律和协会的监管。

中国早在2013年就开始进行隐私信息保护的立法研究工作，并已逐步形成了隐私信息保护 法规体系。但在数据应用过程中出现的数据完整性、准确性、 一致性、及时性等质量问题，造 成数据可信度和可用性不高，价值难以充分发挥。如何在数据隐私保护的框架下，提升数据质 量，推动数据共享，深度挖掘数据价值，是中国各级政府机构、各企事业单位正在积极探讨和 践行的课题。

纵观上述几个国家和地区的情况，隐私保护立法已成为一个趋势，目前已有90个左右的国 家和地区制定了隐私信息保护法律，且信息越来越趋于严苛。总体上看，现有数据法规难以适 用因诸如人脸识别、区块链等创新技术不断涌现导致的隐私信息保护和数据流通应用的现象。 主流解决思路是从两端出发， 一是行业主动自律，以“科技向善”的理念进行自我约束，明确 技术不能突破的底线；二是立法与监管也需要不断创新，以创造性提出解决方案。未来，对数 据主权隶属、数据跨境流动与共享、数据保护合规等环节的深入研究和相应法规颁布，将是隐 私保护和数据安全政策领域的发展趋势。

**4.3** **区块链与数据共享**

区块链技术的出现和应用，为提升工业企业的数据治理能力提供了新契机。工业企业中的 数据孤岛、垄断现象比较普遍，而数据共享和流通中的安全与隐私保护，时时面临着威胁和攻 击的风险。区块链是实现信息数据资产流通、共享和隐私保护的新技术，为企业基于数据治理 促进生产经营转型发展带来了新的机遇。

**1.** **区块链与大数据的融合**

如今对区块链的技术研究和实际应用已如火如荼地展开。同时，以大数据作为载体，将区 块链技术作为新的技术融入大数据技术中，已逐步实现了区块链与数据管理的融合。主要表现 在以下两个方面。

一方面，区块链技术可以有效打破企业数据孤岛现象。区块链技术的本质是一种分布式存 储技术，属于计算机底层技术。在现有的计算机技术能力下，只要基于大数据平台开发出一套 符合区块链技术的程序与接口，就可把分散、孤立的数据源整合并汇集到大数据平台中，形成 分布式数据资源库，进行统一管理。

**34** 数据治理——工业企业数字化转型之道

另一方面，区块链技术可以提升企业数据的安全性。将区块链数据作为企业大数据平台的 数据源，依靠区块链技术的私钥签名验证形式可以有效保护数据在存储、传输过程中的安全。

区块链和大数据的结合，不仅能解决目前工业企业普遍存在的数据孤岛现象和数据安全问 题，更能促进数据流通，数据只有流通才能产生更多的价值。区块链的加密技术和分布式账本 技术使得互联网的去中心化成为可能，在链上的每一步操作都公开可见，想要获得链上数据也 更加容易。由于具有不可篡改和时间戳的特性，区块链还能保证数据的真实性，极大地降低信 用成本，促进更多陌生商家之间的合作。此外，大数据具有的海量存储技术和灵活、高效的分 析技术可以弥补区块链数据存储量小、分析能力弱的缺点，从而极大地提升区块链的数据处理 能力。

**2.** **区块链促进数据共享**

传统的数据共享模式是通过统一的数据仓库或大数据平台集中采集、处理、存储并应用数 据。这一模式在中心化的部门内运转良好，为提升部门的运营效率、推动部门数字化转型带来 很大帮助。但如果是关系对等的部门之间需要共享数据，采用中心化的共享模式就会带来一系 列问题。各部门之间共享数据困难，主要是因为数据的校验、保存和同步是难点。在工业企业 内部各部门之间以及企业与企业之间推动数据共享、开放、市场化利用的过程中，企业面临数 据不敢共享、不愿开放、不知如何共享和开放等难题。破解这些难题需要一种足够安全、可靠、 灵活的技术来支撑。

各个利益相关方在访问和共享数据时会存在严格的限制，需要花费大量的资源和时间用在 共享数据的权限审查、质量保障、真伪校验、产权界定上。造成这种现象深层次的原因主要来 自以下4个挑战：

(1)主导权之争。如果由多个机构共同建设数据集中平台，则该平台的主导方会很难界定。 任何一方主导平台都会带来其他参与机构的争议或者不满。

(2)隐私保护难。数据共享不可避免地会产生数据隐私泄露风险，如何在充分保障用户数 据隐私的前提下实现数据共享是一个必须尽快解决的难题。

(3)数据确权难。数据在共享及流通过程中很容易被复制。如果不能对数据确权，明确数 据的产生者、使用者、管理者及受益者，那么将无法很好地实现数据的精准授权，严重阻碍数 据的共享及流通。

第4章 数据治理的发展趋势 **35**

(4)无激励机制。传统的数据集中方式很难量化每个数据贡献者的实际贡献大小，因此没 有很好的共享激励机制。参与方无论共享的数据是多还是少，数据质量是好还是坏，获得的收 益都是一样的。如果没有合理的贡献激励机制，每个参与方都不会积极共享自己的数据或干脆 不共享。

根据《中国区块链技术和应用发展白皮书》的定义，区块链是分布式数据存储、点对点传 输、共识机制、加密算法等计算机技术的新型应用模式。区块链不仅包括加密货币等公有链， 还包括应用区块链技术解决商业及社会问题的联盟链。联盟链在本质上是一个分布式账本，其 通过共享账本实现多方业务协同和数据共享。

基于区块链技术的分布式账本、密码学技术、智能合约及激励机制这4个特点，可有针对 性地解决多组织数据共享的诸多难题，例如：

(1)分布式账本技术解决主导权问题。

分布式账本技术(Distributed Ledger Technology,DLT) 是一种在网络成员之间共享、复制 和同步数据的技术。记录网络参与者之间的交易，如资产或数据的交换等。在传统数据集的共 享模式下，关系对等的参与方不可避免地会遇到系统主导权问题。在基于分布式账本技术搭建 的共享账本模式下，因每个参与方都拥有相同的账本，主导权问题迎刃而解。分布式账本技术 具有以下优点：去中心化，每个参与方都有一份账本，没有一个中心化的机构进行集中维护； 不可被篡改，通过数学算法保障账本数据不可被篡改，增加账本的可信度并降低审计成本；大 幅降低交易及记账成本，因为是数字账本，记账的边际成本为零。

(2)密码学技术解决数据隐私保护问题。

国内外立法机构对数据隐私保护都有明确的规定。对此，区块链使用加/解密授权、零知识 证明等密码学技术。

(3)区块链可溯源解决数据确权问题。

利用区块链分布式账本、不可被篡改等特点，可以有效进行数据确权。数据的产生者及终 端使用者作为节点被加入区块链网络中，利用区块链详细记录数据的产生、流转、交易等全部 环节，通过节点标示每笔数据对应的产生者及使用者身份。区块链不但记录数据本身，而且记 录数据的上传方，以及数据被访问的全部历史，实现数据确权及精准授权，从而促进数据的共 享和流通。

**36** 数据治理——工业企业数字化转型之道

(4)区块链数学算法解决终端用户授权问题。

利用区块链的数学加密算法，如哈希、非对称加密、CA 签名等，可以实现终端用户对自 身数据的掌控：在未获得用户授权时，机构不得共享用户的数据；只有在获得用户授权时，机 构才可以共享用户的数据。相比查询征信报告的授权机制，基于区块链的授权可以做到颗粒度 更细，如支持字段级别的授权。

(5)基于零知识证明技术实现数据共享。

零知识证明是指证明者能够在不向验证者提供任何有用信息的情况下，使验证者相信某个 论断是正确的。应用零知识证明技术，可以在密文情况下实现数据的关联关系验证，在保障数 据隐私的同时实现数据共享。例如，可以实现对于那些用同一标的物在不同银行重复融资的企 业或个人的预警。

(6)基于智能合约的积分激励机制。

积分激励机制是激励用户为获得更多有价值积分而共享更多数据的机制设计。当用户上传 数据至共享账本时，积分增加，上传的数据越多，获得的积分越多；当用户通过共享账本查询 数据时，消耗积分，查询的次数越多，消耗的积分越多；当用户上传的数据被其他参与方查询 时，积分增加，被查询次数越多，增加的积分越多，建立共享数据后评价机制，根据数据质量 的高低以及价值变现的大小设置不同的积分。

**3.** **区块链与数据治理**

工业企业数据治理的目的是盘活企业的数据资产，挖掘和利用数据的潜在价值，取得社会 化利用和变现。区块链具有的自治、匿名、可溯源及不可篡改的分布式账本技术，非常适合应 用于多主体参与、多流程的复杂过程的机制重塑及流程改造，以增强各参与方的信任度，提高 系统效率。区块链的这些技术特性，对工业企业在数据治理中涉及的提高数据质量、保障数据 安全、促进数据共享等工作可以起到很好的支撑作用，具体介绍如下。

(1)分布式可追溯数据区块链系统有助于提高数据的质量。在传统的工业企业数据治理体 系中，企业主体之间的商业行为，需要工商、税务、行业协会、银行、会计师事务所等充当中 介组织进行协调或出具证明，导致企业数据以孤岛的形式散落在不同性质、相互毫无业务关联 的机构的数据库中，或集中在行业龙头企业中。企业的数据无法自行在社会或行业中直接流通， 会影响数据的实时获取和真实性，制约数据资产的共享、流通和变现。区块链在不同企业主体 之间(甚至企业内的业务部门之间)构建了点对点分布式数据系统。各企业主体通过访问此数

第4章 数据治理的发展趋势 **3 7**

据系统，将各项生产经营和交易活动记录入区块链，保证数据资产交易、转移事件及生产、经 营信息能快速被广泛传播、校验和确认，保证数据的真实性与完整性，提高了数据的质量。

(2)非对称加密与分散式数据库技术可提升数据的安全性。数据的私密性与数据的完整性 是数据安全的重要内容。区块链运用数据的非对称加密、哈希算法等技术可以实现数据安全和 隐私保护。非对称加密算法能验证数据来源，保护数据在存储和传递过程中的安全。哈希算法 等匿名算法能保护数据隐私，防止数据泄露。非对称加密技术是指在加密和解密过程中使用公 钥和私钥两个密钥来验证身份、加密和解密信息，以满足信息所有权的验证和签名，提高数据 的安全性。数据的使用者和监管机构用相应的公钥访问数据库，解密并读取数据。由于时间戳 记录读取数据的时间，当任何一方发现不合理时，可随时随地通过区块数据和时间戳来追溯历 史数据。

(3)在使用区块链的分布式链接结构来存储加密信息时，当这些信息被访问时，将当前块 与存储的每个前一个块进行比较。如果链被正确地验证，则用户将获得对该文件的访问权。否 则，用户将得到一个不可读的加密文件。每次共享一个文件时，链就会增长，这就增加了共享 文件的真实性和安全性。如果文件落入其他人手中，他们将缺少验证链的信息，从而阻止他们 访问数据。通过将区块链技术应用于分散式数据库，会要求数据库中的每个节点对信息进行身 份验证。要想窃取数据，每个节点都必须同时受到攻击，这几乎是不可能的，因为节点中的任 何更改都将取消该节点的身份验证。这种数据处理方式使得数据安全性得到了极大的提高。

(4)点对点技术与智能合约有助于实现数据共享。作为一种“去中心化”的分布式账本系 统，区块链中的每个参与主体都能单独地写入、读取和存储数据，并在全网迅速广播和及时查 证。经全体成员确认及核实后，数据作为某一事件的唯一、真实的信息在区块链全网中实现共 享。点对点技术可以实现全网共同拥有数据，解决数据共享需要中介的问题。另外，利用区块 链的信息技术和智能合约可以打破各自为政的数据统计标准和方法，取代传统的数据协议。通 过在区块链的智能合约中写入指定、统一的代码，系统将根据代码推断合约的实现条件，保障 合约执行。当政府部门统计数据时，区块链会自动根据智能合约上的写入代码对照相关数据的 类型、标准、范围、数量等内容和电子签名进行核对和验证，方便数据统计及共享，提高效率。 当某些部门数据残缺或者未及时上报时，智能合约自动在全网发送实时警告，并将警告记录和 相关部门的答复记录在区块链上，便于追溯问责。由此，点对点技术与智能合约会扩大数据共 享的范围和程度，提高数据共享的及时性和标准化程度。

**38** 数据治理——工业企业数字化转型之道

**4.4** **5G** **技术与数据安全**

5G 技术加速了万物互联时代的到来，使得互联网的范围和容量等级都得到大幅提升，整个 互联网开始成为真正意义上的物联网，更可能推动许多工业场景的真正落地：万物互联、边缘 计算、多接入技术、数字生态系统、智能工业……5G 技术不仅使网速大幅提高，更使信息量呈 指数级增长。而随着大数据的价值越来越大，数据安全问题将更加严峻。随着5G 技术的普及 应用，黑客攻击、恶意代码感染、数据窃取等风险也会随之而来，数据安全防范范围也已被扩 展到更广泛的领域。

**1.5G** **带来新的数据安全风险**

工业企业数据治理中的主要任务之一就是保障数据的安全性。5G 网络在工业领域的广泛应 用所带来的数据安全风险主要体现在以下两个方面。

(1)5G 网络采用了更加强大、高速、灵活和开放的网络架构设计，但同时也将面临更多的 安全威胁。5G 核心网和 MEC 边缘网络将包含大量的诸如虚拟机和物理机、虚拟网络和物理设 备、虚拟存储和物理存储等基础设施，这些基础设施一旦被入侵，将可能使重要的工业数据被 窃取和篡改。传统的数据威胁治理战略、单边的信息安全工具都应重新规划及设置。

(2)5G 网络使企业可以把更多的设备和传感器、控制器连接起来，完成更为复杂、高效的 操作，使工业场景下的新应用层出不穷。各种连接设备数量的增加，使得数据的多源异构现象 更为突出，这也意味着黑客拥有更多的攻击目标和发射台。

**2.** **基** **于** **5G** **技术的数据安全防护**

虽然5G 技术带来了一系列新的安全问题，但作为商业化电信网络，5G 通过 SIM/USIM 等 身份标示、认证授权、信道与承载加密、访问控制方式保障了网络接入的安全性，为用户提供 了更健壮的业务安全保护、更严密的数据保护及更强的用户隐私保护。

5G 网络继承了4G 网络分层分域的安全架构，采用传送层、归属层/服务层和应用层的安全 分层，各层之间相互隔离。采用接入域安全、网络域安全、用户域安全、应用域安全、服务域 安全、安全可视化和配置安全6个安全分域，与4G 网络安全架构相比，5G 网络增加了服务域 安全。这些都在技术层面提升了数据安全保障。

另外，为应对5G 技术带来的新的数据安全风险，选择边缘计算与5G 技术结合的布局是一 个重要举措和趋势。边缘计算的基础架构布置在云端，5G 技术实施网络“切片”,即将网络资 源为特定应用进行“切片”,在同一个物理基础架构下，提供各种专门服务。网络“切片”旨在

第4章 数据治理的发展趋势 **39**

提供隔离中断的流量分段，用统一的安全防护措施减小网络受攻击的风险。

总之，5G 技术带给数据安全保障的优势和风险并存，相应的数据治理战略也将随着网络通 信技术的发展不断面临新的挑战，需要持续更新数据安全管理策略。

**4.5** **新技术与数据治理**

目前正处于第四次工业革命时期，新技术的发展比以往任何时候都快。除基础科研方面的 突破外，物联网、人工智能、云计算、增强现实、虚拟现实、深度学习、机器学习、机器视觉 和图像识别等新兴技术亦取得了跳跃式的进步。新技术之间的相互组合及快速发展与应用，促 使制造业生产效率提高、产品质量改善、产品成本和资源消耗降低，并将传统工业提升到智能 工业的新阶段。新技术的应用发展，使得数据迎来爆发式的增长。要有效利用数据，让数据变 得更有价值，就需要建立完善的数据治理体系。

**1.** **新技术与大数据**

目前，我国在物联网+工业、人工智能+物联网、深度学习、机器学习、图像识别、增强现 实和虚拟现实等新兴技术领域中已经取得突破性成就，并在市场上也得到了广泛的应用和推广。 新兴技术成为当前最重要的投资发展领域之一。新技术的广泛应用，使得各种数据信息呈爆发 式增长，比如传统单体智能无法协调大规模智能设备共同完成实时感知和决策等工作，但随着 物联网协同感知新技术、5G 高速通信新技术的发展，多智能体之间的协同合作已成为可能。大 规模智能交通灯调度可以实现动态实时调整；无人仓储多机器人可以将完成货物分拣的高效协 作；无人驾驶车可以感知全局路况；群体无人机可以协同，并且可以与地面多机器人协同。它 们都会产生超大规模的数据交换与处理。又如深度学习需要采用规模非常庞大的网络，存储很 多参数及完成大量计算，同时，在这些计算过程中，会生成大量的数据。如果没有有效的数据 治理，这些大量的数据要么使得系统崩溃或者变慢，要么就安静地躺在数据服务器里变成一堆 毫无用处的“废弃品”。随着新技术的不断发展，在未来的十年、二十年中，数据呈爆炸式地增 长的趋势是不变的，特别是随着大数据、云计算及5G 技术的发展，万物互联成为一个现实， 在这个时代下大家产生的数据只会逐渐增多，这就更加迫切地需要数据治理。

**2.** **新技术下的数据治理**

在大数据时代，数字经济极大地改变了人类的生产及生活方式，给人类带来了诸多的便利。 人类的生活与数字经济的发展已密不可分，而数据治理则是数字经济发展绕不开的话题。

**40** 数据治理 敬字化转型之道

随着工业大数据的发展，为满足数据治理的需求，采用与机器学习相关的数据分析模型和 算法可实现对工业数据的精准类别的识别、划分；采用线性回归、逻辑回归、支持向量机、决 策树等算法可实现工业设备、产品等数据预测模型的建立，对设备和产品的运行趋势或故障状 况进行科学预测分析，支撑设备和产品的预测性维护，以提高工业企业设备、产品等运营管理 水平；采用图像识别处理、计算机视觉等算法，可实现对工业生产制造中产生的非结构化的图 形图像大数据进行结构化知识发现和分类分析，为工业数字孪生、数字仿真等数据应用提供重 要的技术支撑。

在大数据时代，数据治理能够更好地辅助新技术的推广与应用，促进技术的不断创新发展， 而新兴技术的广泛应用又能带动数据治理体系的建立和完善。

**4.6** **数据文化与伦理道德**

**1.** **数据文化与伦理问题**

在大数据时代数据文化和技术同等重要。其中，数字鸿沟是其主要特征。大数据技术的迅 速发展对社会变革的驱动力越来越大，而各国的数据伦理制度和规范相对滞后。因此，要实现 数字鸿沟的有效伦理治理，就必须要不断完善相关的伦理制度和规范，并努力弘扬公平参与及 协作精神、共享精神、契约精神和人文精神。

DAMA 在《数据管理知识体系》中表示，数据道德文化是一项社会责任，不仅要保护数据， 还要管理数据质量。由于数据会影响决策，数据的完整和准确尤为重要，所以应当避免数据被 滥用、误解的风险，这项工作贯穿数据全生命周期。

数据伦理问题包括以下5个方面。

(1)当数据被共享时需要妥善保护和处理的隐私和保密问题。

(2)使用数据方与供给(被采集)数据方可能在时间和地点上都离得比较远，使用数据方 是否会顾及供给数据方的意愿，即是“道德距离”(moral distance) 问题。

(3)当数据的使用场景还不明晰时，数据的采集能否获得供给方认同的问题。

(4)当数据泄露后，发生的与社会公正有关的诸如污名和歧视的问题。

(5)数据在未来使用时，对公众信任的影响及对未知事件、流程、规则的影响问题。

因此，大数据的整体框架不仅包括物质层、行为层、制度层，也包括精神层，其使大数据

第4章 数据治理的发展趋势 **4 1**

具有思维属性。大数据的互通共享与资产化，带来了诸多使社会、商业及亿万人受益的创新， 但同时也带来风险和新的文化与伦理道德问题，我们必须能鉴定风险，平衡创新带来的利益与 风险。

**2.** **数据的伦理道德治理**

数据的伦理道德治理应提供所有利益相关者关于数据价值实现的意见，应由数据应用者在 开展数据需求、数据质量、数据安全和数据价值实现时进行有关伦理道德的影响和风险评估。 企业负责人应清醒地认识到：

(1)越坚持规范数据道德的企业，越具有商业竞争力优势。

(2)在数据道德规范下处理和使用数据，可以提升企业的可信度，建立更好的客户关系。

(3)企业要通过设置必要的岗位制度和职责权限来控制数据道德风险，防止未经授权或不 当的数据访问及操作。

总之，企业对于数据处理的监督，要上升到伦理道德规范和法律范畴，要制定基于伦理道 德及规范的政策、标准、工作流程，且严格监督执行。

**4.7** **工业企业数据的运营**

**1.** **数据运营的定义**

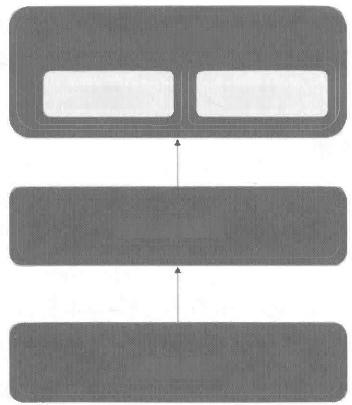
数据运营也被称为数据化运营。它是一种将数据作为新型经济资产类别来经营和管理的能 力和行为，以数据存储和分析挖掘应用作为核心支撑，广泛采集和融合组织内外部各类数据， 通过数据逻辑关联和分析应用，挖掘其中蕴藏的业务价值，并将其内涵信息和潜在价值作为服 务或商品，提供给组织内外部的数据消费者。

**2.** **数据运营的定位**

数据运营包括两个重要方面， 一是数据运营的核心活动职能，二是确保这些活动职能能够

落地实施的保障措施(包括组织架构、制度体系)。数据运营的定位如图4-7-1所示。

**42** 数据治理——工业企业数字化转型之道

**数据运营**

活动职能 保障措施

数据资产管理

**大数据平台**

**图4-7-1** **数据运营的定位**

**3.** **数据运营的目标**

数据运营可以为企业运营带来以下5个方面的价值。

(1)客户管理。

通过数据运营，可以融合客户通过社交网络、电子商务、终端媒介等产生的非结构化数据，

构建客户360°视图。

(2)产品管理。

通过数据运营，可以获取客户的反馈信息，及时获取客户需求，再经过深入分析，可以合

理设计企业产品与服务。

(3)营销管理。

借助数据运营，通过挖掘、追踪、分析客户数据，可以提升精准营销水平。

(4)风险管理。

数据运营能够帮助企业了解客户的自然属性和行为属性，结合客户情绪分析、行为分析、

客户价值度分析、客户风险分析及客户的信用情况，可以及时识别风险，建立完善的风险防范

体系。

第4章 数据治理的发展趋势 **43**

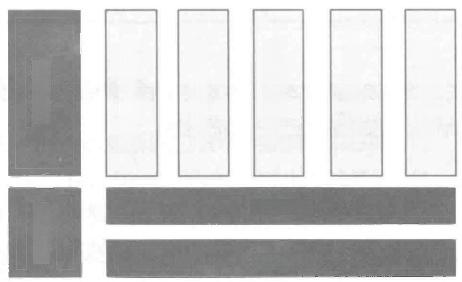
(5)组织管理。

数据运营能增强企业内部机构的透明度，使得企业上下级之间的信息流通更加顺畅。同时 数据运营可以优化企业内部的各种流程，提高企业的运营效率。

**4.** **数据运营的主要内容**

数据运营管理框架包含5个职能活动和2个保障措施，如图4-7-2所示。职能活动是指落 实数据运营管理的一系列具体行为，保障措施是为了支持活动职能实现的一些辅助的组织架构

和制度体系。



数据效益评估

数据共享开发管理

数据运维管理

数据服务管理

数据需求管理

职能活动 保障措施

制度体系

组织架构

**图4-7-2** **数据运营管理框架**

**本章精要**

本章对主要发达国家的数据治理情况进行了介绍，对区块链、5G 等新技术对数据治理的影 响做了分析，并倡导企业要重视数据的相关伦理道德和隐私保护。

第5章

本书阅读导引

为了便于有不同关注点的读者阅读本书，本章对本书各章的主要内容进行了概述，编成阅

读导引，读者可以根据导引直接阅读自己感兴趣的内容。

**5.1** **数据治理是一个系统工程**

数据治理知识体系涉及管理、技术等多个学科领域，是一个非常复杂的系统工程，如何全 面而系统地构建较为完整的数据治理体系，是企业实施数据治理的关键课题。

本书第2篇“体系篇”从企业数据治理的全视角系统性地论述了数据治理的知识领域，通 过11个维度，科学地构建了数据治理体系框架。同时，为了帮助企业更好地评估数据资产管理 能力和水平，提出了可行的数据资产管理能力成熟度评估方法。

体系篇共分12章，包括数据管控、数据战略、数据架构、主数据管理、元数据管理、数据 指标管理、时序数据管理、数据质量管理、数据安全管理、数据交换与服务、数据共享与开放、 数据管理能力成熟度评估，如表5-1-1所示。

**表5-1-1** **第2篇各章一览表**

|  |  |
| --- | --- |
| **章** | **主** **要** **内** **容** |
| 第6章数据管控 | 数据治理组织、制度、流程、设计机制、绩效体系及标准体系等 |
| 第7章数据战略 | 数据战略规划、数据战略实施等 |
| 第8章数据架构 | 数据架构概述、框架设计、数据建模等 |
| 第9章主数据管理 | 主数据标准管理、主数据全生命周期管理、主数据应用管理及企业常用的几类主数据 |
| 第10章元数据管理 | 元数据定义、元数据分类、元数据核心能力等 |
| 第11章数据指标管理 | 指标体系框架、找指标、理指标、管指标和用指标等 |

**第5章** **本书阅读导引** **45**

**续表**

|  |  |
| --- | --- |
| **章** | **主要内容** |
| 第12章时序数据管理 | 时序数据的特点、时序数据的应用等 |
| 第13章数据质量管理 | 数据质量需求、数据质量检查、数据质量分析、数据质量评估等 |
| 第14章数据安全管理 | 数据安全体系框架、数据安全防护策略、数据安全审计、数据安全风险评估、数据应 急保障等 |
| 第15章数据交换与服务 | 数据交换与服务的意义、数据交换与服务技术演进、工业企业数据共享交换服务标准 体系架构等 |
| 第16章数据共享与开放 | 数据资源目录、数据资源准备、数据服务等 |
| 第17章数据管理成熟度 评估 | 数据管理成熟度评估模型、数据管理成熟度等级定义、企业如何开展数据管理成熟度 评估、数据管理成熟度评估实施等 |

**5.2** **工具是数据治理的保障**

工业大数据治理需要多种数据治理工具软件的支撑。本书第3篇“工具篇”聚焦在数据治 理不同领域的常用工具的介绍，内容包括以主数据为核心的套装软件、以数据资源目录为核心 的数据资源管理工具、以元数据和数据模型为核心的数据中台，此外还有针对时序数据、数据 交换处理的工具。这些工具各有侧重，需要根据实际需求进行取舍。

主数据管理工具可以实现对数据治理组织、数据标准、主数据的有效管理。数据指标工具 可以使企业系统梳理的统一数据指标标准落地，以及规范企业业务统计分析语言。数据交换平 台属于中间件，IT 技术人员要熟练驾驭。时序数据记录了工业过程，支撑对工艺与质量的控制。 数据中台是基于数据治理过程的成果面向数据共享与开发的各种应用。

**5.3** **实施数据治理有路线可循**

工业企业实施数据治理是有路线可循，有方法论的。本书第4篇“实施篇”,通过对国际数 据管理协会、数据治理学院和国际商业机器公司(IBM) 数据治理理事会的方法论进行分析、 研究及大量实践，总结出数据治理施实策略，并就数据资产运营实施、主数据管理实施、元数 据管理实施、数据指标管理实施、数据质量管理实施、数据安全管理实施及数据治理常见误区 进行介绍，为工业企业的数据治理实施，提供了包括内容、方法、路径、模板等方面可参考的 意见。

**46** 数据治理——工业企业数字化转型之道

**5.4 数据治理已在诸多行业成功实施**

工业企业的数据治理需求迫在眉睫。为了促进企业有序开展数据治理工作，进一步厘清工业企 业转型升级的主要痛点和关键需求，本书第5篇“案例篇”总结了数据治理解决方案在重点行业、 典型企业中落地应用的经验，并梳理了应用成效和价值，包括电力、能源化工、钢铁、重型装备制 造、汽车、建材、制造、航天、核工业、战略投资、交通物流、多元化集团等有代表性的14 家企 业的数据治理实践案例。其中每个案例主要包括企业简介、数据治理中的痛点及难点、实施步骤、 应用效果及交流分享等，具体如表5-4-1所示。

**表5-4-1案例名称一览表**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 章 | 行 业 | 单位(简称) | **包含业务领域** | **案例名称** |
| 第38章 | 电力行业 | 南方电网 | 负责投资、建设和经营管理南方区域电网 参与投资、建设和经营相关的跨区域输变电 和联网工程，服务广东、广西、云南、贵州、 海南五省(自治区)和港澳地区 | 夯实数字化转型基础—— 南方电网数据资产管理行动 实践 |
| 第39章 | 电力行业 | 国家电投集团 | 火电、水电、风电、光伏发电等全部发电 类型 | 支撑集团产业数字化转型 ——国家电投集团数据治理  实践 |
| 第40章 | 能源化工 行业 | 延长石油 | 油气探采、加工、储运、销售，石油炼制， 煤油气综合化工，煤炭与电力，工程设计以及 施工、新能源、装备制造、金融服务等领域 | 数据治理助百年油企数字 化转型 |
| 第41章 | 钢铁行业 | 酒钢集团 | 钢铁、电力能源、铝冶炼加工、生产性服 务业、现代农业、装备制造产业 | 酒钢集团数据治理实践 |
| 第42章 | 汽车行业 | 长安汽车 | 汽车开发、制造、销售 | 数据驱动长安汽车数字化 转型 |
| 第43章 | 核工业 | 中核集团 | 核电、核燃料循环、核技术应用、核环保 工程等领域的科研开发、建设和生产经营 | 物料主数据治理助力核电 智慧运营 |
| 第44章 | 航空行业 | 航天科工 | 航天、建筑、医疗、汽车、食品、化工、 石油装备 | 军工企业的“三位一体” 数据治理体系建设实践 |
| 第45章 | 航空行业 | 航空工业 | 航空客运、航空货运及物流两大核心产业， 涉及航空配餐、航空货站、地面、机场服务 等相关产业 | 面向航空装备研制生产的 数据治理研究与实践 |
| 第46章 | 重型装备  制造行业 | 中国一重 | 重大成套技术装备、高新技术产品和服务、 国际贸易 | 数据标准，装备中国—— 中国一重的数据标准化管理 项 目 |
| 第47章 | 交通物流  行业 | 中国外运 | 海运、空运、国际快递、公路和铁路运输、 船务代理、仓储及配送、码头服务等 | 主数据治理助力中国外运 数字化转型 |

第5章 本书阅读导引 **4 7**

续表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 章 | 行 业 | 单位(简称) | 包含业务领域 | **案例名称** |
| 第48章 | 建材行业 | 建材集团 | 建材产业、新材料开发和综合服务 | 中国建材集团工业大数据 应用实践 |
| 第49章 | 制造行业 | 威孚集团 | 零部件设计与制造 | 威孚集团基于斯欧应用互 联平台建设数据通道 |
| 第50章 | 战略投资 行业 | 国投集团 | 基础产业(电力、交通、矿产),制造业、 大健康、城市环保、生物质能源，金融及服 务业和国际业务 | 国投集团的数据标准化管 理实践 |
| 第51章 | 多元化集  团 | 新兴际华 | 冶金、轻纺、装备、医药、应急、服务 | 数据治理助力多元化企业 集团管控 |

希望这些企业在数据治理方面的理论介绍及实践经验，能为众多工业企业在数据治理的研

究和实践提供参考和借鉴价值，以期达到让企业少走弯路、打好基础、快速取胜的效果。

**参考资料**

[1]李杰.工业大数据：工业4.0时代的工业转型与价值创造.邱伯华，等译.北京：机械工业 出版社，2015

[2]许正.工业互联网：互联网+时代的产业转型.北京：机械工业出版社，2015

[3]李克，朱新月.第四次工业革命.北京：北京理工大学出版社，2015

[4]克劳思·施瓦布.第四次工业革命.世界经济论坛北京代表处，李箐，译·北京：中信出版 社，2016

[5]通用电气公司(GE). 工业互联网：打破智慧与机器的边界.北京：机械工业出版社，2015

[6] DAMA 中国分会翻译组.DAMA 数据管理知识体系指南(原书第2版) .北京：机械工业

出版社，2020

[7]GB/T36073-2018 《 数据管理能力成熟度评估模型》

[8]GB/T34960.5-2018《 信息技术服务治理第5部分：数据治理规范》

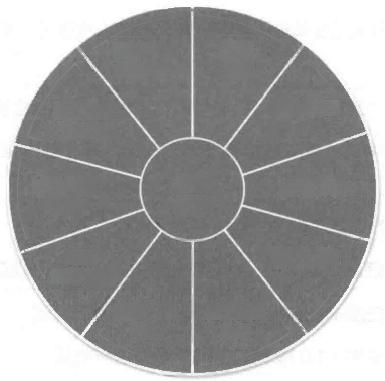
[9] 《数据资产管理实践白皮书(4.0版)》,中国信息通信研究院

[10]《主数据管理实践白皮书(1.0版)》,中国信息通信研究院



**第** **2** **篇** **体系篇**

本书在借鉴国际数据管理协会( DAMA International)数据管理知识体系的基础上，结合行 业标杆企业的最佳实践经验，形成了具有行业特色的工业企业数据治理体系，如下图所示。本 体系架构以数据管控为核心，通过数据管控统领数据治理的10大职能领域，包含数据战略、数 据架构、主数据管理、元数据管理、数据指标管理、时序数据管理、数据质量管理、数据安全 管理、数据交换与服务、数据开放与共享。



数据开放

与共享

数据交换

与服务

数据安全

管理

元数据管理

数据指标

管理

时序数据

管理

数 据

质量管理

数 据 管控

数据 战略

主数据管理

数据架构

**工业企业数据治理体系**

第2篇 体系篇 **49**

**1.** **数据管控**

数据管控是数据治理体系的基础。10大数据治理职能都需要数据管控管理，数据管控能力 在很大程度上直接影响数据治理目标的达成，以及制度、流程、绩效和相应的标准规范的落地 执行，所以数据管控也是整个数据治理体系的核心。

**2.** **数据战略**

数据战略是企业数据工作开展的目标指引，定义数据工作的方向、愿景和原则。从管理层 出发，自顶向下地全局部署数据管理规范，从而形成全面的标准规则体系和落地执行流程。

**3.** **数据架构**

数据架构是用于定义数据需求，整合和控制数据资产，与业务战略相匹配的一套整体构件 规范，主要包括企业数据模型设计和数据资源分布管理。

**4.** **主数据管理**

主数据是指满足跨部门业务协同、反映核心业务实体状态属性的基础信息。主数据相对于 交易数据而言，属性更稳定，准确度要求更高，且可唯一识别。

主数据管理主要通过一系列规则、应用和技术对主数据值进行控制，使得企业可以跨系统 地使用和共享一致的主数据，提供来自权威数据源的、协调一致的高质量主数据，降低运维成 本和管理复杂度。

**5.** **元数据管理**

元数据是描述数据的数据，用于对信息资源进行描述、解释、定位等，更方便检索、管理。 比如数据的责任人、存取路径、保密级别、访问权限等信息都是元数据。

元数据管理包括规划、实施和控制活动，以便能够访问高质量的集成元数据，包括定义、 模型、数据流和其他至关重要的信息。

**6.** **数据指标管理**

数据指标是衡量目标的方法，即预期中打算达到的指数、规格、标准， 一般用数据表示。

例如：销售收入活期存款金额、委托贷款余额等是数据指标。

数据指标管理是指通过对与企业若干个核心和关键业务环节相互联系的统计数据指标的全 面化、结构化、层次化和系统化的构建，满足企业找指标、理指标、管指标、用指标的需要。

**50** 数据治理——工业企业数字化转型之道

**7.** **时序数据管理**

时序数据是指时间序列数据，是按时间顺序记录的数据列，在同一数据列中的各个数据必 须是同口径的，要求具有可比性。时序数据可以是时期数，也可以是时点数。

时序数据管理主要通过对时序数据的采集、处理和分析来帮助企业实时监控企业的生产与 经营过程。

**8.** **数据质量管理**

数据质量管理包括规划和实施质量管理技术，以测量、评估和提高数据在组织内的适用性， 提高数据对业务和管理的满足度。

**9.** **数据安全管理**

数据安全管理是为了确保数据隐私和机密性得到维护，数据不被破坏，数据能被适当访问。 通过采用各种技术和管理措施，保证数据的机密性、完整性和可用性。

**10.** **数据交换与服务**

数据交换与服务是指与数据存储、应用程序和组织之间的数据移动及整合相关的过程。制 定相应的数据交换与服务标准，针对不同的场景合理采用相应的数据交换与服务技术，从而将 各业务系统的能力进行整合，为企业新的管理模式和业务模式提供灵活可靠的数据支撑。

**11.** **数据开放与共享**

数据开放与共享主要是指开展数据共享和交换，实现数据内外部价值的一系列活动。数据 共享管理包括数据内部共享(企业内部跨组织、部门的数据交换)、对外开放(企业之间的数据 交换)。

第6章

数据管控

**6.1** **数据管控概述**

数据治理对任何企业来说都是一项复杂且规模浩大的体系化工程，需要充分调动企业相关 的所有资源，只有形成全面、有效的管控体系，才能确保数据治理各项工作在企业内部得以有 序推进。

数据管控是一套以数据治理相关组织和人员为核心的，涵盖企业数据治理制度、流程、考 核等各个方面的执行保障机制，其本质是通过建立高质量的人才队伍和严明的制度体系来确保 数据战略被正确落实。

因此，数据管控是企业开展数据治理的重要基础性保障，为企业实施数据治理各项职能活 动提供人才团队、制度规范、文化氛围等基础资源，是企业数据治理得以开展的重要基石。

一般来说，企业的数据管控体系包括数据治理的组织架构、制度规范、执行流程、培训宣 传、设计机制和绩效体系等内容。

(1)组织架构：企业从事和涉及数据治理各项职能活动的人员的组织方式。由于数据治理 工作的重要性和复杂性，通常应该自上而下形成专业化且各司其职的团队，并在企业内部形成 顺畅的沟通、协调、合作机制。由于数据治理工作是跨部门、跨专业的，因此这个团队一般会 是虚拟的，但其执行力必须统一且高效，才能为数据治理各项工作的落地实施夯实基础。

(2)制度规范：为规范和约束企业数据治理各项职能活动的相关管理办法、实施细则、指 导意见、操作指南等制度性文件。覆盖全面并与实际工作结合良好的制度规范， 一方面有利于

**52** 数据治理 工 业 企 业 数字化转型之道

明确和固化数据治理团队组织内部的职责分工和协调机制，另一方面有利于理顺企业内部相关 部门和岗位之间的工作关系，也为开展数据认责及考核评价提供依据。

(3)执行流程：企业为落实制度规范相关管理要求，针对数据治理具体的职能活动场景， 结合企业自身的组织架构制定的一系列规范性和标准化的工作实施和流转过程。有了规范的执 行流程，企业内部相关部门和人员就可以按照统一的程序和方法进行数据治理的各项工作，有 利于促进相互之间的高效协作，避免出现凭个人经验办事、 一人一种做法、工作互不统一的混 乱状况。

(4)培训宣传：企业为培养数据治理相关专业人才，营造企业内部良好数据治理氛围的重 要措施。人才是企业实施数据治理的根本，缺乏数据治理专业人才会严重影响企业数据治理各 项工作的顺利推进。企业应为人才的成长搭建良好的平台，并逐步打造“金字塔”结构的人才 体系，满足管理、执行、监督等多个岗位角色的工作要求。除专业人才培养外，还应在企业内 部开展广泛的数据文化和知识传播，为数据治理工作的整体协作营造良好氛围。

(5)设计机制：为了达成数据治理的工作目标，在组织架构、制度规范和执行流程的基础 上形成的各类管理和保障过程。设计机制包括决策机制、监督机制、认责机制、沟通机制、激 励机制等，帮助企业数据治理的管理者有效掌控数据治理各项工作的具体执行情况，激发执行 人员的主观能动性。

(6)绩效体系：企业在既定的数据战略目标下，通过设定特定的衡量和评价指标，对团队 和人员已完成的数据治理工作行为及取得的工作业绩进行全面评价，并根据评价结果对团队和 人员就未来的工作行为和业绩进行正面引导的过程和方法。绩效体系的建立是数据管控构建完 成的重要标志之一，也是数据质量管控体系形成闭环，并可持续性、常态化执行的基础。同时，

设计一套科学、合理且让各方信服的绩效体系是企业数据治理管理者所面临的最大挑战之一。

需要充分认识到的是，任何企业的数据管控体系的构建都不是一蹴而就的，也不是一成不 变的。不同行业、不同企业都有其自身的特点，在充分尊重企业特点的基础上，构建行之有效 的数据管控是一项长期性、计划性和连续性的工作，必然也是一个不断持续改进的过程，其目 的是最终形成一套与企业运作机制完全匹配的管控模式。

由此可见，每个企业的数据管控体系都应该是独一无二的，本章内容只基于数据治理的职 能活动要求，提出通用性的体系。在具体落实过程中。需要结合企业的特点进行细化设计，确 保数据管控体系能与实际结合，保障企业数据治理工作落到实处。

明琐解

**6.2** **组织架构**

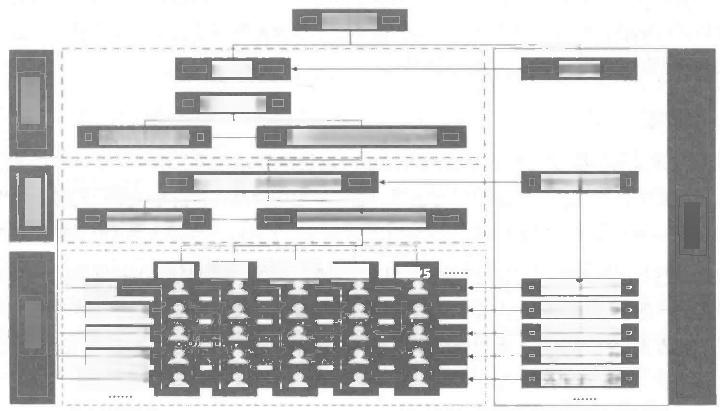
数据治理是一项需要企业通力协作的工作，而有效的组织架构是企业数据治理能够成功的 有力保障。为达到数据战略目标，企业有必要建立体系化的组织架构，明确职责分工。

数据治理组织架构的搭建需要充分考虑企业内部 IT 系统、数据资源、人力资源及业务应用 的开展现状。企业数据治理的管理人员应当基于通用组织架构的分层要求，细化设计一套有针 对性、符合企业运作机制的数据治理组织，力求与企业日常各项工作进行良好的衔接。

**6.2.1** **数据治理组织架构**

经过多年的数据治理实践，业界已经充分认识到，仅仅依靠技术部门来推动和开展数据治 理工作是无法取得项目立项时所期望的成果的，其原因在于数据治理涉及企业各个部门的业务 和资源，只有来自更高层管理者的驱动力，才能保证企业内部的高效协作。

因此，数据治理组织的通用架构需要自上而下形成完整的组织体系，从形式上看，这种组 织架构已经与企业的经营管理架构非常相似了。企业的数据治理组织架构主要分为决策层、管 理层、执行层和监督层4个层级，如图6-2-1所示。



**股东大会**

董事会

高级管理层

**数据治理委员会/领导小组**

**数据治理管理小组/办公室**

**归口管理邮门** 其他业务部门管理小组

专业1 **专业2** 2 **专业4** 专业

总 部

分支机构1

**分支机构2**

**分支机构3**

一分支机构4

执行监督小组 执行监督小组 执行监督小组 执行监督小组 执行监督小组

管理层 执行层

**管理监督小组**

**CTO/CIO/CDO**

决策层

监事会

**图6** **-** **2** **-** **1** **数** **据** **治** **理** **组** **织** **框** **架**

**1.** **决策层**

董事会作为企业经营管理的最高权力机构，同样应当作为企业开展数据治理各项工作的最

**54** 数据治理——工业企业 数字化转型之道

终决策机构。企业数据治理与经营管理相关的重大事项均应由董事会讨论决策。董事会有权将 企业数据治理相关事项的决策权全权委托或授权给数据治理委员会或领导小组。事实上，企业

的数据治理委员会或领导小组的负责人一般也是董事会成员。

高级管理层负责企业的日常经营管理，执行董事会的各项重大决策要求，以及负责对日常 实施数据治理的过程中需要高层支持的相关事项做出决策。这些事项虽很重要，但尚无提交董 事会决策的必要。因此，在实际开展数据治理的过程中，高级管理层的分管领导需要针对工作 的具体方向和目标进行决策。同样地，高级管理层的分管领导一般也会是数据治理委员会或领

导小组的重要成员。

企业的首席技术官(CTO)、 首席信息官(CIO) 和首席数据官(CDO) 是数据治理组织架 构中领导力的重要来源。首席数据官的设置目前尚未普及，这就需要在开展数据治理时，由首 席技术官或首席信息官来承担相应的职责。如果数据治理是由企业的首席技术官、首席信息官 或首席数据官直接发起并驱动的，则有利于后续各项工作的执行和协调。

数据治理委员会或领导小组代表董事会、高级管理层对数据治理日常工作中的大小事项进 行决策，其拥有在企业范围内对数据治理管理层和执行层的管理权力，通常由董事会、高级管 理层的分管领导担任负责人，企业各部门的经理担任成员。数据治理委员会或领导小组应当是 一个高效的即时决策机构，这不同于董事会或高级管理层需要通过定期的董事会或经营办公会 进行重大事项的决策，更具灵活性和实时性。

**2.** **管理层**

数据治理管理小组/办公室是企业内部组织开展日常数据治理工作并对整个过程进行管理 协调的专职机构，其成员一般由归口管理部门的领导担任负责人，各相关部门的数据治理负责

人或接口人担任成员。数据治理管理小组/办公室是主持企业数据治理日常工作的虚拟机构。

数据治理归口管理部门是企业内部牵头负责数据治理体系构建并组织开展数据治理实施的 责任部门。 一般情况下，企业内部的数据治理归口管理部门是单一部门，如技术部门、独立的 数据部门或者某个业务部门，但也存在多个部门同时作为数据治理归口管理部门的情况。这种

情况比较少见，从管理层面来看，容易出现职责不清、相互扯皮的问题。

对企业内部除归口管理部门以外的相关部门来说，每个部门内部应当成立相应的管理小组， 负责统一组织和协调本专业内部的数据治理工作。对存在分支机构的大型企业来说，各专业内 部的数据治理管理小组尤其重要，关系到决策层和数据治理办公室的相关要求是否能够在专业 内部得到很好的执行。各部门管理小组一般由部门分管领导担任负责人，数据治理负责人或接

第6章 数据管控 **55**

口人作为成员，如存在分支机构，则支机构的专业部门负责人也应作为小组的成员，参与本专 业数据治理实施管理工作。

**3.** **执行层**

数据治理组织的执行层是由“机构+专业”双维度构成的矩阵式组织体系。

从机构维度看，企业内部的每一个独立机构，包括总部，都应该根据管理层的要求构建本 机构内部负责执行数据治理各项工作的团队。这些团队可以由归口管理部门在本机构的下级单 位牵头，并纳入各专业部门的数据治理执行人员(甚至外部合作单位的相关人员),共同推动本 机构内部的数据治理工作。

从专业维度看，根据数据治理过程中业务专业性的要求，各专业自上而下在各分支机构建 立负责开展数据治理的执行小组。这些小组在专业内部接受本专业数据治理管理小组的指导和 管理，在本机构内部接受归口管理部门在本机构的下级单位的组织和安排。

**4.** **监督层**

监事会是代表股东大会对公司的业务活动进行监督和检查的法定必设和常设机构。数据治 理作为企业数据业务的重要组成部分，也必然接受监事会对决策层的监督，确保数据治理的决 策过程符合法律法规和企业制度的要求，保障决策内容符合企业的数据战略方向。

管理监督小组是对数据治理组织架构的管理层进行监督的机构。管理监督小组一般由企业 内部审计、法律、人力资源等部门的相关专业人员组成，其并不直接接受监事会的领导，而是 为了监督数据治理的管理层机构是否按照企业的数据战略组织开展科学管理数据治理工作，具 有相对的专业性和独立性。

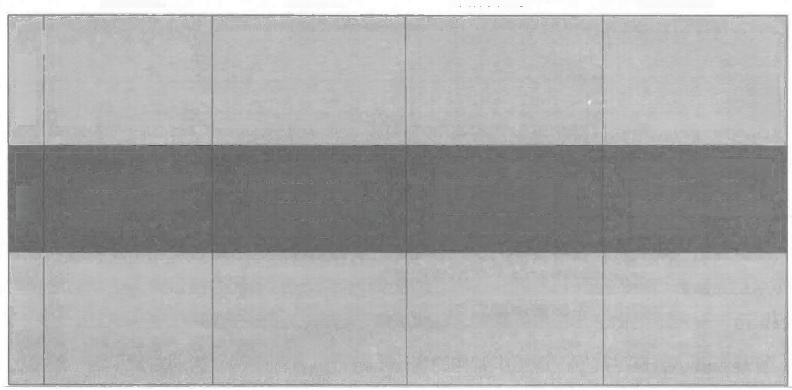
执行监督小组是针对数据治理组织架构的执行层进行监督的机构。对于不存在分支机构的 企业来说，其职能可以并入管理监督小组，即不必再设执行监督小组；对于存在分支机构的企 业来说，每个分支机构的执行团队都应设立相应的执行监督小组，其成员可以由本机构相关部 门人员担任，也可从企业总部层面进行统一安排。执行监督小组接受管理监督小组的指导和管 理，主要监督执行团队是否按照管理层要求落实执行，确保按时、保质完成工作。

**6.2.2** **数据治理组织模式**

数据治理组织架构与企业的组织架构存在着极强的关联性。例如，小企业的数据治理组织 架构和大型集团公司的数据治理组织架构必然存在着巨大的差异。即使同为大型集团公司，其

数据管理力度、范围、标准化程度的不同，也会对数据治理组织架构各层级的组成及职责分工 带来显著的差异。

经过多年发展和实践，得到业界普遍认可的数据治理组织模式主要有4种，基本可以包含 和匹配当前各行业、各类型企业的数据治理组织的构建需求。这4种模式分别为分散模式、归 口模式、半集中模式和全集中模式，如图6-2-2所示。



**分散模式** 归口模式 半集中模式 **全集中模式**

|  |  |  |  |
| --- | --- | --- | --- |
| 各职能部门，分支机 构自行组建数据团队 负责本业务领域，本 单位的数据治理工作 | 成立数据归口管理部门，各 业务部门数据园队仍然归属 业务部门，但同时报告给归 口管理部门，组成虚拟的企 业级数据治理专业图队 | 成立统一的企业级数据治理 回队，采用面向业务部门、 分支机构派驻团队的模式， 构建与业务部门的合作关系 业务单元保警少量专素资源 | 成立全集中的企业级数据治 理团队从各业务部门，分 支机构抽调敬据治理相关人 员，全部整合到集中后的专 职的数据部门中，统一管理 |

|  |  |  |  |
| --- | --- | --- | --- |
| 数据治理工作最大  限度满足本单位实  际需求  警专业、跑单位政据 问题难以解决  没有跨邮门的数据信 息路合、质员与管性 机制 | 兼顾数据治理的统筹 管理和业务自主性  统一的规划设计、财 务预算和绩效考核  对归口部门的专业能 力、组织能力、协调 能力和营理能力要求  较高  对人力资源的需求配 高 | **专业资源的高效利用**  **全面掌控各业务部门实** **际的故据治理工作进展** **有利于数据治理的统筹**  推进  企业组织机制需要调整转  黑  业务部门的自主性和参与 度较低 | **最大限度地共享和使用** **专业人力与技术资源**  **集中力量解决影响面广** **的重要问题**  末端需求响应速度的 下降  业务部门和分支机构 几乎露身事外  个业组妇机制的变展 |

情式内容 特点 桃 战

图6-2-2 数据治理组织模式

**1.** **分散模式**

分散模式是指企业内部各部门根据数据战略自行组建数据治理团队，负责各自专业范围内 的数据治理工作。团队之间相互独立，仅存在少量的沟通和协同，也没有总体统筹负责的组织 机构。

该模式的特点在于数据治理工作可以最大限度地满足本单位的实际需求，但是由于缺乏跨 专业协同和统筹管理，跨专业、跨单位的数据问题难以解决，也无法建立起企业级的数据整合 和质量管控机制。

**2.** **归口模式**

归口模式是指企业内部明确数据治理工作的归口管理部门，其他业务部门都成立数据治理

第6章 数据管控 **57**

团队，虽然其组织关系上仍归属本业务部门，但其数据治理相关的工作应同时向归口管理部门 报告，接受归口管理部门的工作指导和安排。两者共同组成数据治理的虚拟团队。

该模式的特点在于在一定程度上兼顾了数据治理工作的统筹管理和业务自主性，有利于进 行统一的规划设计、财务预算和绩效考核。但该模式对归口部门的专业能力、组织能力、协调 能力和管理能力均提出了较高要求，对人力资源的需求也相对较高。

归口模式是当前各企业构建数据管控体系时采用较多的模式，具备较强的可落地性和可复 制性。

**3.** **半集中模式**

半集中模式是指企业内部有明确、统一的数据归口管理部门，并且由归口管理部门向其他 业务部门派驻分团队或小组，再结合业务专家，构建与业务部门的合作关系。派驻的分团队或 小组接受数据归口管理部门的统一管理，并指导和负责业务部门的数据治理工作。

该模式的特点在于可以高效利用数据治理专业的资源，并全面掌控各业务部门实际的数据 治理工作进展，有利于整个企业数据治理工作的统筹推进。但该模式涉及企业组织机制的调整 转型，这一点在大型国有企业中尤难实现，并且业务部门的自主性和参与度较低，不利于充分 利用业务部门的专家资源。

**4.** **全集中模式**

全集中模式是指在企业内构建完全集中化运作的数据治理团队，除技术部门的数据治理人 员外，将各业务部门、各分支机构相关的业务专家和数据治理团队、人员全部整合到集中后的 专职的数据治理团队中，并由该团队全面负责整个企业的数据治理工作。

该模式的特点在于可以最大限度地共享和使用数据治理专业的人力资源和各类技术资源， 可以结合企业的业务实际需要统筹安排各项工作的优先级，并集中力量解决影响面广的重要问 题。但完全集中化必然会带来末端需求响应速度的下降，业务部门和分支机构几乎置身事外， 不利于数据治理的常态化管控，并且涉及企业组织机制的变更，挑战极大。

不管采用以上哪一种组织模式，企业都必须根据组织架构构建包含决策层、管理层、执行 层和监督层的团队，并进一步明确各层级、各岗位的职责分工，才能确保数据治理工作的有序 推进。

**58** 数据治理 · 工 业 企 业 数字化转型之道

**6.2.3** **数据治理职责分工**

根据数据治理组织架构的分层设计，各层级机构在数据治理工作中所承担的职责有所不同， 界面清晰、分工明确的职责定位是所有参与其中的角色各司其职、有条不紊地开展各项工作的 重要基础。

对于存在分支机构的大型企业来说，其数据治理的组织架构也应该是多级的。例如，某集 团公司在各省均设置了分公司，在集团统一开展数据治理工作的过程中，集团总部按照4层架 构从整个企业的角度构建了企业级数据治理团队，而各分公司也会从本公司需求出发，构建包 含4层架构的本地级数据治理团队。这两个团队的人员可能是复用的，但其职责和定位不应出 现较大偏差。

**1.** **决策层职责**

决策层是企业数据治理各项重大事项的决策机构，负责制定企业数据战略，审批或授权审 批数据治理相关重大事项，全面协调、指导和推进企业的数据治理工作，督促管理层不断提升 数据治理有效性，对企业数据治理承担最终责任。

决策层的主要职责如下：

(1)贯彻落实国家有关数据治理的相关法律、规定、方针和政策。

(2)负责制定和发布企业数据治理及数据资产管理相关的战略规划。

(3)负责对企业数据治理相关的重大事项进行决策。

(4)负责审批和发布企业数据治理相关的管理制度、流程及相关标准规范。

(5)负责指导、督促管理层和执行层开展数据资产管理工作，组织、协调跨部门的数据治 理重大事项。

**2.** **管理层职责**

管理层主要负责建立企业数据治理的完整体系，制订企业数据治理的实施计划，统筹数据 治理资源配置，建立数据质量常态化控制机制，组织评估数据治理工作的有效性和执行情况， 制定并实施问责和激励机制，定期向董事会报告。

管理层的主要职责如下：

(1)贯彻落实决策层各项决策部署，行使日常数据治理工作管理职责。

(2)负责基于企业战略规划制订数据治理工作的发展规划，并报决策层批准、发布。

第6章 数据管控 **59**

(3)负责组织制订企业数据治理的年度投资计划和年度工作计划。

(4)负责制定、修订和发布数据治理相关的管理办法、实施细则和业务指导书，构建公司

的数据治理制度体系。

(5)负责组织制定企业级数据标准，并对数据标准执行情况进行监督和定期抽样检查。

(6)负责规划、构建和管控企业的数据架构，统筹开展数据模型、数据分布、数据目录和 元数据等管理工作。

(7)负责识别和确定企业主数据明细，并建立主数据应用管理机制。

(8)负责制定数据质量基础规则，明确管理要求，制定检查和考核指标，并提供数据质量 问题的指导和协调。

(9)负责组织制定数据安全基础策略，构建企业级数据安全管控体系，配合公司安全和审 计部门开展数据安全的监督和审计工作。

(10)负责受理跨部门、跨机构或重要的数据需求，组织开展数据需求的评审、实施和成果 交付工作。

(11)负责接受监督层管理监督小组的工作监督，配合开展定期监督检查。

(12)负责定期组织评价数据治理工作效果，制定考核制度。

**3.** **执行层职责**

执行层在管理层的统筹安排下，根据数据治理相关制度规范的要求，具体执行各项数据治 理工作。执行层主要负责落实数据治理体系建设和运行机制，推动数据治理各项流程与日常工 作相结合，并根据数据治理各职能域的管理要求承担具体执行工作。

各业务部门是本专业数据治理执行工作的责任主体，负责本专业领域的数据治理执行工作， 管控业务数据源，确保数据被准确记录和及时维护，落实数据质量管控机制，执行监管数据相 关工作要求。

执行层的主要职责如下：

(1)负责落实企业数据战略和数据治理规划的相关要求，根据战略规划目标组织具体工作 的开展。

(2)负责落实数据治理相关制度规范、数据标准和工作流程的要求，确保各项执行工作符 合企业要求。

**60** 数据治理 数字化转型之道

(3)负责根据管理层建立的数据架构落实本专业、本单位的数据模型、数据分布、数据目 录和元数据管理等工作。

(4)负责基于管理层明确的主数据管理应用机制开展主数据消费、维护等工作，保障主数 据的准确性、 一致性和及时性。

(5)负责执行本区域、本专业数据质量管理，制定数据质量提升工作方案，落实数据质量 管控机制，开展数据质量问题的识别、分析与整改。

(6)负责落实管理层制定的数据安全管理要求，配合开展数据安全监督与审计工作。

(7)负责处理本单位数据需求，开展需求分析、业务逻辑梳理等工作，对交付成果进行效 果验证。

(8)负责接受监督层执行监督小组的监督，配合开展定期检查工作。

(9)负责接受管理层数据质量相关的考核评价，并根据结果优化工作机制。

执行层一般除企业自有人员外，还会包括大量的外部合作单位人员，其职能主要是帮助自 有人员完成其岗位职责明确的工作内容，在此就不再作为单独的职能单位赘述。

**4.** **监督层职责**

监督层是相对独立的，主要行使除决策、管理和执行外的监督职责，负责对企业开展的数 据治理工作的战略符合度、行为合规性等进行内部审计、检查，可对重大事项出具监督意见。

监督层的主要职责如下：

(1)负责对决策层针对数据治理重大事项做出决策的过程进行监督，确保决策过程符合法 律及企业规章制度的相关规定。

(2)负责对管理层制订的工作计划、各项预算、管理制度等管理工作进行监督，确保管理 过程合法合规，且与企业数据战略方向一致。

(3)负责对执行层具体落实数据治理各项工作的过程进行监督，确保执行结果符合预期。

(4)负责定期配合审计部门对数据治理相关项目、操作行为、资金流向等进行全面审计，

及时发现违规行为或潜在风险。

(5)负责在考核评价过程中提出在监督过程中发现的问题，并参与实际的考核评价过程。

第6章 数据管控 **61**

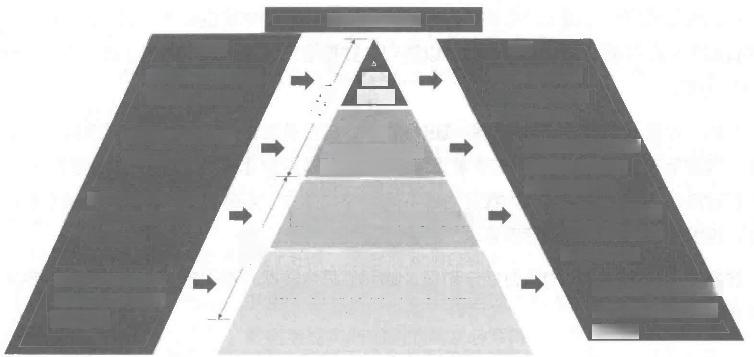
**6.3** **制度规范**

为了保障组织架构的正常运转和数据治理各项工作的有序实施，需要建立一套涵盖不同管 理颗粒度、不同适用对象，覆盖数据治理过程的管理制度体系，从“法理”层面保障数据治理

工作有据、可行、可控。

**6.3.1** **数据治理制度框架**

根据数据治理组织架构的层次和授权决策次序，数据治理制度框架分为政策、制度、细则、 手册4个梯次。该框架标准化地规定了数据治理的各职能域内的目标、遵循的行动原则、完成 的明确任务、实行的工作方式、采取的一般步骤和具体措施，如图6-3-1所示。



**故据治理制度框架**

**内容**

政策 说明数据治理的企

 企业级 业定位和战确目标，

# 明确其基本规则

制度

企业级，按数据治理 为各致据职能城内的活

的职能域分别细制 动开展而制定的一系

**列管理办法和流程**

**细则**

分支级，按数据治理的职能域分别细 为确保各做据治理制度

化，客业务部门制定本专业细则 **得到执行落实而派生**

**的实施细节规定**

手册

分支级，聚焦具体操作过程和方法，指导一 **针对某项具体工作制定的**

线执行人员标准化开展特定的工作事项 **操作过程指引，确保操作**

**标准准确**

**制定者**

决范层，酒常为数据治

理要员会/领导小组

普理层，通常为数据治

理B 口臂理解门

**执行层，通常为分支机** **构或业务解门的数据治** **理负责人员**

**执行屋，需由分支机构** **或业务部门数据治理负** **表单位批准**



**图6-3-1** **数据治理制度框架**

**1.** **数据政策**

数据政策是企业数据治理的纲领性文件，是最高层次的数据管理制度决策，是落实数据资

产管理各项活动必须遵循的最根本原则，描绘了企业实施数据战略的未来蓝图。

数据政策既贯穿了整个企业的组织和业务结构，也贯穿了企业数据创造、获取、整合、安 全、质量和使用的全过程，其内容包括数据资产管理及相关职能的意义、目标、原则、组织、 管理范围等，从最根本、最基础的角度规定了企业在数据方面的规范和要求。

数据政策应当符合企业的数据战略目标，数量不宜太多，内容描述应当言简意赅、直击要

点。

**62** 数据治理- 工业企业数字化转型之道

数据政策一般由企业决策层的数据治理委员会/领导小组发起，组织相关专业人员起草，并 在整个企业范围内进行广泛讨论、评审、完善。数据治理委员会/领导小组负责进行终审，并 正式发布执行。数据治理委员会/领导小组也可以授权委托数据治理归口管理部门组织执行以上 工作。

**2.** **管理制度**

数据治理管理制度是基于数据政策的原则性要求，是结合各企业组织和业务特点制定的数 据治理职能范围内的总体性管理制度。它的目的是确保数据治理的管理层对准备开展或正在开 展的数据治理各职能活动进行有效控制，并作为行为的基本准则为后续各角色的职责问责建立 依据。

数据治理管理制度清晰地描述了数据资产管理各项活动中所遵循的原则、要求和规范，各 级单位和机构在数据治理工作中必须予以遵守。数据治理管理制度从形式上包含章程、规则、 管理办法等。

数据治理管理制度一般根据职能域进行划分，与企业准备开展的数据治理实际工作相关。 例如“数据标准管理办法”“数据质量管理办法”“元数据管理办法”“主数据管理办法”“数据 安全管理办法”等。这些文件为数据治理不同职能域建立了规范性要求，内容一般包括目标、 意义、组织职责界面、主要管理要求、监督检查机制等。

数据治理管理制度中的所有规定和要求都必须符合数据政策规定，不应与数据政策所确立

的基本原则相违背。

一般情况下，企业数据治理的相关活动会早于数据治理管理制度的制定。因此，数据治理 管理制度更多地需要对已开展的数据治理活动从纷乱无章向统一有序引导。数据治理管理制度 的建立并不是推翻现有的工作机制，而是在标准化要求下对当前各项数据治理活动的规范化构 建和重组。

数据治理管理制度由数据治理归口管理部门负责组织编写。考虑到数据治理职能活动的差 异，应当成立一个专门的制度编制小组承担具体的编制工作。由于数据治理活动通常早于管理 制度的制定，不少的业务部门或分支机构的人员也广泛参与其中，所以制度编制小组的成员不 应该仅仅来源于归口管理部门或技术部门，企业应该更多地吸纳其他业务部门和分支机构人员 允许其代表本机构、本专业的利益对数据治理的管理制度提出相应的要求。但最终，数据治理 管理制度必须以整个企业的高度和角度来评判和衡量管理措施的有效性，目的是保证企业数据 质量符合数据需求方的使用要求。

第6章 数据管控 **63**

**3.** **数据治理实施细则**

数据治理实施细则是已有的企业级数据治理管理制度的从属性文件，用于补充解释特定 活动或任务中描述的具体内容，进一步确定后续步骤里的具体方法或技术，或管理制度相关要 求与不同业务部门、分支机构实际情况的结合和细化，以便促进特定领域或范围内具体工作的 可操作化。

数据治理实施细则一般是本地化的。但这并不意味着，对于组织结构比较简单、不存在分 支结构的企业来说，实施细则是不必要的。

数据治理实施细则可以分成两类。 一类是针对企业级数据治理管理制度在各业务领域落地 的细化要求，需要结合各业务领域的数据现状、组织架构、工作方式等，不同业务领域存在一 定的差异。这些细则是在企业统一要求的基础上由业务部门本地化定制的，是所有企业都应当 制定的。另一类是企业级数据治理管理制度在各分支机构的细化要求，同样是企业统一的管理 要求与各分支机构的实际情况结合后指导具体落地工作的文件。这些对于不存在分支机构的企 业来说是不需要考虑的。

从另外一个角度看，实施细则是管理制度的进一步细化，可依据实际情况而建立，不是所 有的制度都必须制定单独的细则。而且如果一个企业制定的管理制度足够详细，足以指导整个 企业具体的数据治理执行工作，那么实施细则并不是一定需要的。其衡量的标准还是现有的规 范性文件能否约束和指导实际的执行工作，是否需要通过实施细则来进一步补充细节。

数据治理实施细则一般由业务部门或分支机构的数据治理负责人组织编制，参与人为本单 位与数据治理相关的专业人员。数据治理实施细则的编制必须符合该领域管理制度的规定，各 种细化的、本地化的执行要求不应与管理制度确立的企业级要求相违背。

**4.** **数据治理操作手册**

操作手册是针对数据治理执行活动中的某个具体工作事项制定的，用于指导具体操作的文 件，是特定活动的执行中需要遵守的操作技术规范。

操作手册的内容和形式均不固定， 一般包括需要不同角色遵循同样的标准化要求的场景， 或多个制度执行活动中共同调用的相关标准。

操作手册的内容应当符合管理制度和实施细则的管理要求，可根据数据治理实际执行过程 中的标准化需求而不断新增、删减及持续优化完善。

**64** 数据治理 工业企业数字化转型之道

**6.3.2** **数据治理制度修订**

数据治理制度的制定并不是终点，只是对企业开展数据治理工作进行约束和管控的开始。 从这个意义上看，数据治理制度需要根据企业自身及数据治理工作的需求变化而变化，这就要 求企业对数据治理制度进行适时的修订，以符合实际工作的发展需要。所以，数据治理制度与 数据治理实施总是处在不断的匹配过程中，而且数据治理制度往往是滞后的一方。

**1.** **修订的时机**

数据治理制度的修订需要适当的时机，过于频繁的修订会对日常工作造成不良影响，而过 于滞后的修订会造成实际工作与制度不匹配，无法实现有效的约束。

通常比较合理的修订时机如下：

(1)当国家与数据管理相关的法律、规程废止、修订或新颁布，对企业数据治理工作产生 较大影响时。

(2)当企业组织结构和运营体制发生重大变化时。

(3)当内外部监督或审计单位提出相关整改意见时。

(4)当在安全检查、风险评价过程中发现涉及规章制度层面的问题时。

(5)当分析重大事故和重复事故原因，发现制度性因素时。

(6)其他相关时机。

**2.** **修订的原则**

企业在修订数据治理制度时，应遵循以下3个原则：

(1)辩证统一原则。坚持“稳”与“变”的辩证统一。企业数据治理制度在修订过程中既 要有针对新需求的内容新增，也要保持较强的一致性和稳定性。 一方面，企业要不断适时地用 最新、最适用的制度代替已不适应现状的制度；另一方面，数据治理制度的变化应当循序渐进， 尤其是层级越高的制度修改应当越谨慎，稳定性应当越强。

(2)先立后破原则。对企业数据治理制度的修订要采取“先立新，再破旧”的程序。在条 件尚不成熟，新制度尚未出台之前，应继续按原有制度执行，待新制度正式建立以后再废除旧 制度，以保持制度的连贯性、稳定性，保证企业数据治理活动的正常开展。

(3)消除例外原则。企业数据治理制度的修订要能准确识别“例外”和“偶然”事件。因 此，在出现“例外”和“偶然”事件的情况下，管理者要善于运用标准化原理，用管理制度来

第6章 数据管控 **65**

指导对“例外”与“偶然”事件的处理，并适时将“例外”和“偶然”事件纳入管理制度，使

其成为常规管理的一部分。

**3.** **修订的过程**

在实际工作中，在制度内容修订比例不大的情况下，数据治理制度的修订过程主要有以下5步：

(1)明确修订目标，即明确本次修订需要适应或解决当前制度存在的什么问题，通过修订 达到什么效果。

(2)补充必要数据及信息，针对本次修订的内容补充日常工作中积累的相关数据、材料或 信息，为修订提供基础。

(3)起草修订稿，并对制度修订前后的效果进行对比分析。

(4)征求意见，即在合理的范围内对修订的内容进行意见征集，采纳合理意见并进一步完

善修订稿。

(5)签审发布。

在起草修订稿时，需要特别慎重，充分考量修改部分的内容怎样才能与企业各方面的制度 保持协调，怎样避免出现顾此失彼的情况。如果一个制度的修订造成了同其他管理制度的矛盾， 则势必给企业数据治理工作带来混乱。

在特殊情况下，企业可随时对制度进行修订，但一般不宜过于频繁。如果无特殊情况，则 企业可在每年年末对现有制度进行年审，并根据年审结果考虑是否需要进行修订。

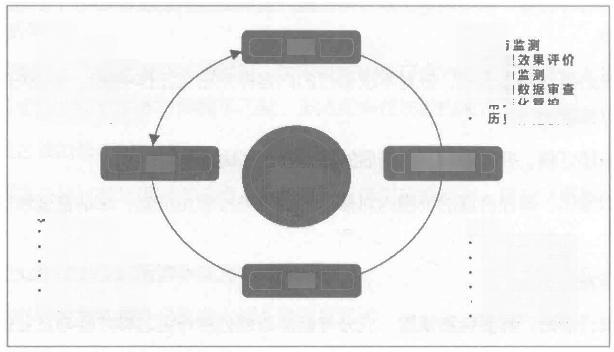
**6.4** **执行流程**

**6.4.1** **数据治理总体流程框架**

要真正把数据作为企业有价值的资产来管理，就必须像管理财务、人力资源等业务功能一 样进行数据治理，这就要求企业需要在明确数据治理管控目标的基础上，建立数据治理相关的 工作流程。例如，财务的日常工作由多个核心业务流程组成，包括应付账款、应收账款、工资 和财务计划等。数据治理核心工作流程的合理设计，将极大地帮助和规范各部门、各机构在数 据治理的各个阶段、各个领域的任务衔接和协调。

在具体的执行工作中，数据治理工作包含了众多的业务流程，包括标准和规则的制定、预

期的数据清洗、修复、保护、协调、授权等一系列工作流程。但从数据治理整体工作来看，这 些具体的工作流程又可以被归纳并划分为4个核心流程，包括定义、发现、实施、衡量与监测 4个环节。这4个环节可重复执行，形成一个闭环的数据治理流程体系，并且4个环节并不是 严格意义上的前后衔接关系，而是可以根据企业的不同和数据治理阶段的不同，可能包含一些 并行的活动，如图6-4-1所示。



定义

政据积累

**衡量与监测**

**数据治理**

实施

实施

定!义

业 务

。 数 据

·数 据:

数 据

关(键 痛禁

**发现**

发现

清单

清单

蜇位



理 动 向 态 史



景准级系理能 背标分关治性

定术动工行旁 制技自人执业

数 据

曲

图6-4-1 数据治理总体流程框架

(1)定义。定义环节首先定义数据与业务相关的背景、分类及相互之间的关联关系，其次 定义实施数据治理工作所需的政策、规则、标准、流程及评价策略。如上所述，此环节与发现 环节可阶段性迭代并行。

(2)发现。发现环节主要获取企业数据生命周期的当前状态、相关业务流程、组织和技术 支持能力，以及数据本身的状态，并根据定义阶段明确的数据治理策略、优先级、标准、规则、 架构等来对企业现状进行全面对比检查，发现数据治理需要解决的问题，形成问题清单，并通 过CRUD 分析方法初步定位数据治理问题的原因，为后续数据治理工作的实施建立基础。此流 程以“发现驱动定义”的形式与定义环节可阶段性迭代并行。

(3)实施。实施环节的目标是根据前两个环节的结果来执行具体的数据治理工作，并确保 数据治理与定义和发现环节确立的所有数据治理政策、业务规则、管理流程、工作流程、角色 职责相符，最终解决发现数据治理的相关问题，提升数据治理的水平。

(4)衡量与监测。衡量主要是获取并衡量数据治理和管理工作的有效性及价值；监测主要

是建立后向的实时管控机制，形成对数据治理过程及其后续各项工作的常态化闭环管理机制，

第 6 章 数据管控 **67**

使数据资产及其生命周期透明并可审核。

以上是企业开展各个领域数据治理工作的完整过程，通常情况下，这4个环节中如果有缺 失的环节，则数据治理工作是无法达到预期效果的。

以改进数据项(比如电话号码)质量或安全性的实验性数据治理项目为例，必须按照数据 治理总体流程框架的4个环节来制定相应的解决方案，并在实际工作中执行。其过程包含对“电 话号码”这个字段的业务定义、规则定义、关系建立、质量检查、问题发现、问题验证、清洗、 改进质量、常态化监测等一系列过程，最终确保对“电话号码”这个字段的数据质量有明显提 升。

**6.4.2** **数据治理典型场景的流程**

在实际工作中，根据涉及的数据治理职能领域、数据范围、应用类型、业务专业领域的不 同，数据治理存在着多样化的实施场景。这些场景表面上看起来差异很大，但其核心过程仍旧 符合数据治理总体流程框架。

本节选择数据治理的3个典型场景来详细说明其流程。

(1)数据标准管理流程：定义、审批、颁布、使用、反馈与维护全局数据标准的工作流程。

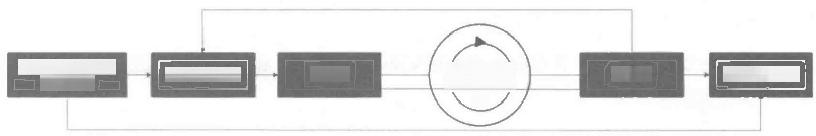
(2)数据质量管理流程：数据质量检查项制定，以及数据质量问题发现、分析、改善、反

馈的工作流程。

(3)数据安全分级和授权流程：数据安全分级和授权工作的标准制定、审批、落实、反馈 与维护的工作流程。

**1.** **数** **据** **标** **准** **管** **理** **流** **程**

数据标准管理流程是一套相对稳定的工作流程。下面从阶段级别的层次说明适用于数据标 准管理工作的流程。数据标准管理工作可以分为5大阶段，如图6-4-2所示。

1.收集，分析数据

数据治理委

虽会/领小

组

**2.制定与更新**

5.落实与反馈

3.审核

4.颁布

标准满求

**图6-4-2** **数据标准管理流程**

数据标准管理各工作阶段的具体说明如下。

(1)收集(定义、发现)、分析数据标准需求：数据标准管理单位收集和分析数据标准定义 的需求，判断是否需要新增或修改数据标准定义。

(2)制定与更新(实施):数据标准管理单位在各业务部门的参与下，初步制定或修改数据 标准定义，提交数据治理委员会/领导小组会议审核。

(3)审核(实施):由数据治理委员会/领导小组对新的数据标准定义进行审核；根据审核 中提出的意见，由数据标准管理单位进行相关定义的调整。

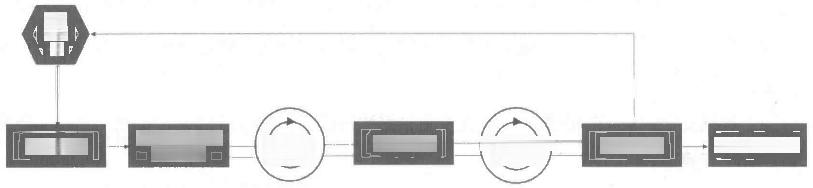
(4)颁布(实施):数据标准管理单位颁布更新后的数据标准定义。

(5)落实与反馈(衡量与监测):技术部门将数据标准定义落实到具体工作中，在落实过程 中发现并反馈存在的问题，在结束之后进入下一个循环。

**2.** **数据质量管理流程**

数据质量管理流程是一套以可衡量的数据质量标准为基准的数据质量问题处理流程。参考 领先实践的经验并经过多方讨论，此处定义通用的数据质量管理流程。

数据质量管理流程可以分为6大阶段，如图6-4-3所示。



**1.准**

**备工**

作

3.问题定位与优 先级划分

数据质量管理 协调会改

**数据质量** **营理部力**，

**6.评估改进成果**

4.制定方案

5.提升质量

**2.问题发现**

**图6** **-4** **-** **3** **数** **据** **质** **量** **管** **理** **流** **程**

数据质量管理各工作阶段的具体说明如下。

(1)准备工作(定义):数据质量管理部门牵头，协同业务部门和技术部门相关人员制定数 据质量指标、设置检查点并开展其他准备工作。

(2)问题发现(发现):通过设定的数据质量标准和规则，由技术部门和业务部门检查、发 现并反馈数据质量问题。

(3)问题定位与优先级划分(发现):由数据质量管理部门收集数据质量问题，与作为数据 用户的业务部门一起确定数据质量问题的实际情况，根据问题的影响程度等多方因素，初步划

第6章 数据管控 **69**

分问题的优先级。

(4)制定方案(实施):梳理需优先解决的数据质量问题，由数据质量管理部门制定初步解 决方案后，组织相关业务部门、归口管理部门、技术部门等共同评审，并确定最终解决方案。

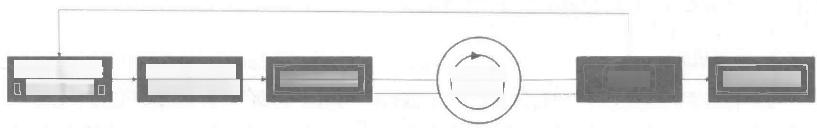
(5)提升质量(实施):组成专题工作小组，数据质量管理部门召集相关业务部门和技术部 门的人员一起分析解决方案信息，由专题工作小组负责落实数据质量改进解决方案，提升数据 质量。

(6)评估改进成果(衡量与监测):业务部门负责评估数据质量的改进成果，数据管理员更 新数据质量问题的追踪状态，结束之后进入下一个循环。

**3.** **数据安全分级和授权流程**

数据安全分级和授权流程是一套相对稳定的工作流程。此处涉及比较通用的数据安全分级 和授权工作的流程。

数据安全分级和授权工作可以分为5大阶段，如图6-4-4所示。

2.制定数据安全

1.制定或更新数 据安全策酯

**数据归口** **管理部门**

**3.审核与发布**

5.使用与反馈

**4.落实**

分级和授权方法

**图6-4-4** **数据安全分级和授权流程**

数据安全分级和授权各工作阶段的具体说明如下。

(1)制定或更新数据安全策略(定义):数据安全管理部门，根据业务需求对应的数据需求， 制定数据安全策略，继而制定与具体数据对应的数据安全分级与用户授权。

(2)制定数据安全分级和授权方法(发现):根据业务部门在数据使用中反馈的相关意见更 新数据安全策略及相关清单，把制定或更新的结果提交给数据归口管理部门审核。

(3)审核与发布(发现):数据归口管理部门审核数据安全策略及相关清单，做出必要调整， 经有关管理层审批之后，签署发布。

(4)落实(实施):技术部门通过系统开发等方式，落实数据安全策略，实现数据安全分级， 制定授权清单。

(5)衡量与监测(使用与反馈):业务部门在日常工作中也需要根据权限落实与之相关的数

**70** 数据治理——工业企业数字化转型之道

据安全分级和授权方法，并且通过使用信息系统等方式，评估数据安全策略及相关方法中数据 安全分级和用户授权的合理性和有效性等，对数据安全策略及相关方法的整体效用进行评估， 并提出必要的反馈意见。

**6.5** **设计机制**

所谓机制，即可以通过建立数据治理的组织、角色、制度、流程，使数据治理这项工作得 到执行和落实。它是数据治理工作得以有效开展的重要保障。数据治理执行的最终效果是衡量 机制落实的标准。

数据管控通常包括以下6种机制。

**1.** **决策机制**

决策机制是其他机制的基础，而且又贯穿于各机制的始终。决策机制在组织中主要体现为： 决策层审批各项规划、裁决重大争议、审批各职能的管理办法，管理层应统一准备决策所需的 材料，执行层应按照管理组织要求提供信息和素材等。

**2.** **监督机制**

在执行数据治理各项工作时，需要对执行过程和结果进行监督。监督层出具年度监督计划， 根据数据资产管理工作进度，监督数据治理工作是否有效执行，出具监督报告，并保存整改报 告或记录文档。

**3.** **保障机制**

数据治理工作的开展需要依赖技术平台的支撑保障，数据治理相关技术平台的管控需要建 立相应的技术规范，技术规范是保障数据技术平台可持续管理的重要基础。技术平台和技术规 范将为数据管控的有效运行提供强有力的保障机制。

此外，资金的保障也是数据治理保障机制的重要一环。由于数据治理工作内容繁多，项目 也相应较多，重要程度有所差异，企业在制订资金计划时应当充分考虑保障重点工作的开展， 并建立资金统筹优化机制，及时根据数据治理工作的开展情况对资金进行再分配。

**4.** **认责机制**

根据关键数据治理对象，确定各数据治理对象的最终负责部门，负责数据资产的标准建立、

第6章 数据管控 **71**

数据质量问题发现和分析、数据资产的日常维护等工作。

数据治理工作覆盖数据生命周期的整个过程，因此数据责任也必然存在于数据生命周期的 各个环节中，应当由参与到数据生命周期中的数据提供者、业务管理方、数据操作方及技术支 撑方等角色分担。

对数据提供者来说，其数据责任主要是负责保障提供数据的高质量、维护数据供应目录、 分配合理的数据权限等。

对业务管理方来说，作为专业领域数据的责任主体，其数据责任主要是明确统一的数据定 义，制定数据标准、安全保障要求和规则，监控业务系统相关数据问题并及时解决。

对数据操作方来说，主要在数据录入、加工、处理等操作过程中负责执行数据管理规则，

生成各项数据并解决相关数据问题。

对技术支撑方来说，主要负责为数据管理提供技术支持，推动数据架构、标准和规则等内 容的落地，对因技术工具缺陷、性能缺陷等问题造成的数据问题负有直接责任。

对于数据认责工作的开展主要有以下通用步骤。

首先，梳理认责数据范围，即明确对哪些数据进行认责管理，企业必须根据自身的数据量 来制定数据认责的范围，尤其对数量巨大的企业来说，这一步不可能一蹴而就，需要分批次进 行。在认责数据项的梳理和筛选上可采用“问题+价值”双驱动的策略，即优先对问题多发且业 务影响较大的数据项开展认责管理，通过责任落实提升数据质量，从而控制和解决问题，支撑 业务发挥价值。

其次，建立数据认责矩阵，即数据各项责任与机构、岗位、人员间的对应关系，将相关数 据责任落实到对应岗位人员的日常工作和数据操作中。

最后，数据责任的落实通常可以与数据质量整治工作结合进行，在明确岗位人员数据责任 的同时，明确责任落实要求。例如数据录入责任与数据项录入规范的同步执行，通过规范录入 行为及纠正录入错误强化员工的责任意识。

**5.** **激励机制**

运用多种激励手段可以使各岗位员工的行为规范化和标准化。激励手段包括精神激励、薪 酬激励等。决策层应该建立相应的激励机制，并发布明确的激励标准和原则。

建立数据治理工作的奖励机制，对在数据治理工作中表现突出、业绩优秀的集体和个人予

**72** 数据治理——工业企业数字化转型之道

以精神及物质层面的表彰。

建立数据治理创新激励机制，鼓励各部门和单位推进数据治理方法创新，推广典型经验， 对产生明显提升效果的创新案例进行奖励。

**6.** **沟通机制**

明确各组织的日常沟通方式、沟通频次、沟通内容，至少需要包括管理层向决策层的汇报 机制和执行层向管理层的汇报机制。沟通内容包括但不限于数据治理整体及部分工作的开展状 况，如数据质量专项提升情况、主数据质量提升情况、数据安全管理达标改进状况等工作内容。

在实际工作开展中，管理层至少应当在组织执行层建立月度例会机制，及时掌握各部门和 单位的工作进展，讨论和解决实际工作中存在的问题。

**6.6** **绩效体系**

数据治理考核用于保障数据治理制度的落实，是一种正式的员工工作评估制度，其通过系 统的方法、原理来评定和测量企业员工在一段时间内的数据治理相关的工作行为和工作效果， 进一步激发员工的积极性和创造性，提高员工的数据治理责任心和基本素质。

数据治理考核的最终目的是引导和激励员工承担数据治理工作责任，使员工的行为符合企 业核心理念的要求，在企业中形成“竞争、激励、淘汰”的良性工作氛围，在实现数据治理目 标的同时，提高员工的满意程度和成就感，从而确保企业的战略目标的有效实现，最终达到企 业和个人发展的“双赢”。

企业应构建明确的数据治理绩效体系，制定相应的考核办法，并把数据治理考核纳入企业 年度考核。通常由数据治理归口管理部门负责制定数据资产管理考核指标，在上报企业决策层 审批后下发执行。

在实施考核之前，需要具备以下前提条件。

(1)企业高层必须对考核工作予以高度重视和支持，否则即使有好的方案也会流于形式。

(2)必须要有清晰明确并且可量化的数据治理目标，通过组织内部自上往下地逐层传递， 使各部门、各岗位的员工目标统一，共同实现数据治理目标。

(3)合理的组织结构、清晰的责权利及流畅的业务流程也是考核体系成功构建的关键因素， 完善的岗位职责体系是衡量各岗位员工绩效的基础。

第6章 数据管控 **73**

结合以上3点前提条件，数据治理绩效体系的构建需注意以下几个方面：

(1)以数据质量提升目标为根本，以结果为导向。

(2)考核指标尽量量化，并且能有客观的数据和技术支撑。

(3)所有与数据治理相关的人员都应积极参与。

(4)结果与过程考核相结合，既要看最终结果，又要看努力程度。

(5)多种考核方法综合运用，例如 KPI考核+关键事件法+360°综合评价。

(6)考核结果要与薪酬挂钩，达到激励作用。

(7)加强沟通和培训，因为沟通是贯穿整个考核管理始终的，只有通过有效沟通，才能引

导员工积极主动地改进数据治理工作。

在明确以上前提条件及注意事项后，需要有具体的实施步骤和措施，可归纳为4个步骤： 制定考核方案，确定考核指标，明确考核标准，开展考核评估。

**1.** **制定考核方案**

考核方案是考核的纲领性文件，不仅要制定好，还要进行广泛的宣传；不仅要让各级管理 者清楚，还要让全体基层员工清楚；不仅要让领导同意，还要让绝大多数员工接受。

年度考核方案一般要在上一年年末完成，主要包括考核的基本原则、考核形式、考核内容、

考核分工、考核程序、考核周期、考核数据来源、数据审核部门。

还需要明确考核周期、考核指标调整原则，以及对各级统计人员、数据人员、考核组成员

的纪律要求。

除以上的内容外，还要同时向各单位、各部门下达考核表，考核表中包括具体的考核指标、 考核标准。还要明确“分级考核”的原则，企业级绩效考核只面对各个单位，分公司考核只面 对部门，部门考核要落实到人。

**2.** **确定考核指标**

考核指标设定的原则是“量化为主、定性为辅”,能量化的考核指标就纳入考核方案，不能 量化但有明确标示且不被误解的考核指标就定性描述，否则就不要纳入考核方案。考核指标主 要包括两个方面： 一是主要控制指标，二是基本工作任务。

**74** 数据治理 工 业 企 业 数字化转型之道

考核指标的设定非常重要：设定过低，考核没有意义；设定过高，执行人员完成不了，反 而没有了工作的动力。所以，考核指标的具体设定需要基于“适当先进”和“够一够”的原则， 对各部门、各单位及各层次人员有区分的效果。

各项考核指标由数据治理归口管理部门制定并提交，由数据治理办公室审核，再经绩效考 核专业组核查，然后上报决策层审定。同时，关于考核指标的准确性问题，由负责审核的职能 部门进行判断。

**3.** **明确考核标准**

考核指标确定后，可确定考核指标的权重和标准。对于考核项目，采取“只扣不加”的原 则，即完成下达的指标和任务是必需的，完不成是不可以的，完不成就要接受处罚。各项规章 制度不能只有原则，而没有具体的处罚标准。

**4.** **开展考核评估**

数据治理的考核评估由数据治理归口管理部门负责，企业考核委员会成员、考核组的人员 参与评估，评估的内容就是对各个考核对象提报的考核结果进行审议，并对考核问题进行研究 决策。

考核评估完成之后，要做两件事： 一是下发考核通报文件，通报的内容有绩效考核、行为 考核、专业考核的结果，明确改进意见，对下一次的重点事项提出要求；二是下发考核通知单， 把具体的奖金数额，以及惩罚事项，分别发给各个被考核单位。

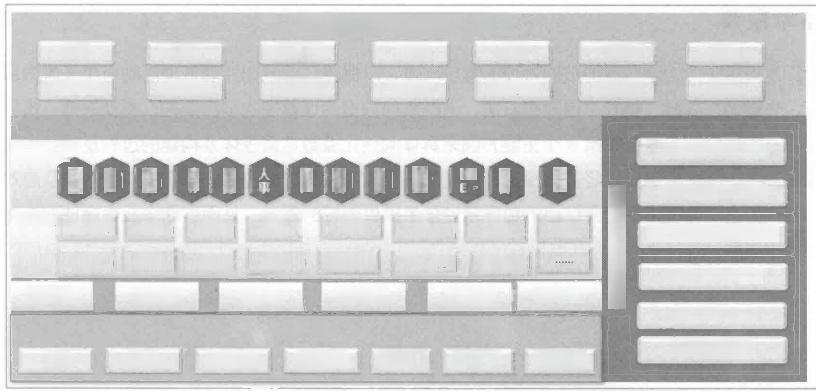
**6.7** **标准体系**

数据标准是数据质量管理的基础与前提。长期以来，大多数工业企业重应用系统的建设， 轻标准规范的制定，严重制约了企业内部数据的打通与共享。

企业可以参考《DAMA 数据管理知识体系指南(原书第2版)》、《数据管理能力成熟度评 估模型》(DCMM)、《大数据标准化白皮书》,以及《数据资产管理实践白皮书(4.0版)》,并 结合多家大型集团公司的数据标准管理实践，设计数据标准体系框架图，如图6-7-1 所示，其 中包含技术标准、数据标准、应用标准与管理标准(管控制度流程体系)。

贸 产

**第6章** **数据管控75**



**应用标准(行业数据模版及模型)**

建材

核工业

数据标准

数据安全与隐私保护

|  |  |
| --- | --- |
| 单位  设备  数据规范 | 别 雾  人事  资产 |

全生命周期管理

技术标准

评价与考核体系

描述模型

数据管理制度与规范

数据质量管理与认责体系

数据资产目录管理

元数据 时序数据 数属交易 数据共享

主数据 通用基础

物料

质量模型 分类方法 技术架构

制造 能源化工 医疗医药

汽车

电力电气

物流

重型装备

医疗

工程建造

钢铁

石油化工

机械制造

航空航天

管 控 制 度 流 程 体 素

生产

采 购

船舶

焊炭

数据指标

数据代码

航空航天

参考架构

项目

合同

财务

质显

总则

市 场

工 程

销 售

设福

项目

投深

HSE

库存

术语

**HS**

**图** **6** **-** **7** **-** **1** **工** **业** **企** **业典** **型** **数** **据** **标** **准** **体** **系** **框** **架** **图**

技术标准，主要包括大数据术语、总则、参考架构、技术架构等基础标准，以及描述大数 据集、进行数据全生命周期操作的技术标准，如描述模型、质量模型、分类方法等。

数据标准，整体包含元数据、数据指标、主数据、数据代码、数据规范、时序数据、数据 交易与数据共享等标准。目前大型工业企业在数据指标与主数据上做得比较完善。数据指标覆 盖了企业经营管理与安全生产的方方面面，包含采购、生产、销售、库存、财务、人事、资产、 设备、投资、市场、HSE(健康安全、环境管理体系的简称)、项目、工程等领域。主数据包括 通用基础、单位、人事、财务、资产、物料、质量、项目、合同等主题领域，以及按行业分的 主数据，如能源化工、航空航天等领域。

应用标准，按工业行业来编制，覆盖工业各部门。 一般而言，技术标准与数据标准可以参 考国际标准和行业标准，或者直接采标，但是具体的应用标准需要企业自身按照业务特点与需 求来编制。

管理标准(管控制度流程体系),是企业数据资产管理体系建设与落地的基本保障，主要包 括数据管理制度与规范、数据安全与隐私保护、数据质量管理与认责体系、数据资产目录管理 及全生命周期管理等。而评价与考核体系是闭环管理中最为重要的一环，特别是集团型企业， 需对各层级的数据治理工作进行评价、考核与激励。

**76** 数据治理——工业企业数字化转型之道

**本章精要**

本章从企业构建有效的数据管控的需求出发，以保障企业数据治理各项活动有序开展为目 标，介绍相关内容、模式、过程和方法，进而围绕数据治理管控的组织架构、制度规范、执行 流程、设计机制、考核体系5个主要方面来具体阐述企业数据管控体系构建的过程及相关的管 理要求，并结合实际案例经验，探讨在实现数据管控建设落地的过程中遇到的实际问题及应对 方法，为企业构建符合自身特点的数据管控体系提供更好的参考，从而帮助企业有序地构建数 据管控体系。

第 7 章

数据战略

**7.1** **数据战略概述**

数据战略已成为企业精细化数据管理不可或缺的基础，只有切实落实好数据战略工作，才 能提升企业数据质量、实现企业数据价值的升华，为企业数字化转型奠定基础。

数据战略是整个数据治理体系的首要任务，是企业开展数据治理工作首先应该考虑的事情。 数据战略应由数据治理组织中的决策层制定，需要指明数据治理的方向，包括数据治理的方针、 政策等。

战略是根据选择和决策的集合绘制出一个高层次的行动方案，以实现高层次的目标。通常， 数据战略是一个数据管理计划的战略，是提高数据质量，保证数据的完整性、安全性和可用性 的计划。然而，数据战略可能还包括利用信息达到竞争优势和支持企业目标的业务计划。数据 战略是企业数据资产管理的总体目标和发展路线图，指导企业在各阶段根据路线图中的工作重 点开展数据治理和运营工作。

**7.2 数据战略规划**

随着企业对数据越来越重视，企业的数据治理发生了根本性的变化，它不再是一个完全在 IT部门中实施的技术规程，而在业务管理方面扮演着日益重要的角色。如今，对于“数据是企 业重要的数据资产”这一理念，企业的各级领导已经基本达成共识，企业也逐渐接受单独编制 数据战略的观点。数据由业务产生，又服务于业务，还能创造新业务，并支撑企业数字化转型。

简单地收集数据，甚至分析数据，并不是数据战略的终极任务。数据战略的核心在于如何

从数据中获取有价值的信息。

要在企业中培育数据文化，最有效的一种方法就是让关键人员参与制定战略和实施战略。 数据战略规划为数据管理工作定义愿景、目标和原则，是所有利益相关者达成的共识。企业要 从宏观及微观两个层面确定开展数据管理及应用的动因，并综合反映数据提供方和消费方的

需求。

**7.2.1** **愿景和日标**

愿景是制定企业战略的起点，愿景的实现是企业的长期战略，而目标是企业短期内要达成 的明确任务，目标的实现是企业的短期战略。企业数据战略目标的规划设计不仅要考虑“诗和

远方”,也要考虑“眼前的苟且”。

如一家零售企业，其数据战略是围绕零售业务开展的，利用零售客户的数据，提升为零售 客户服务的水平。其数据战略目标是注重客户/会员的发展能力和客户服务水平的提升，那么客

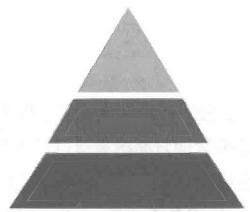
户画像、行为预测、精准营销无疑是企业数据战略关注的重点。

企业首先要建立数据战略规划的目标，维护和遵循数据管理战略；然后针对所有业务领域， 在整个数据治理过程中维护数据管理战略(目标);接着基于数据的业务价值和数据管理目标， 识别利益相关者，分析各项数据管理工作的优先权；最后制订、监控和评估后续计划，用于指

导数据管理规划的实施。

企业数据战略目标大致可以分为三个层次：满足基本的管理决策和业务目标，进行创新 与转型，定义企业在数字化竞争生态中的角色和地位。这三个层次并不是不同企业制定的不同 的数据管理目标，而是企业数据战略在不同阶段、不同成熟度下的三个具体形态，如图7-2-1

所示。



长期目标

中期目标

**短期目标**

**图7-2-1** **数据战略的三个层次**

第7章 数据战略 **79**

**1.** **第一个层次——短期目标**

第一个层次是满足基本的管理决策和业务协同。通过解决企业在数据管理中的各类问题， 可以满足决策分析和业务协同的需要。该层次的战略目标是解决企业最基础、最迫切需要、最 能击中企业痛点的问题。随着多年的信息化建设，企业中建设了多套业务系统，而这些业务系 统是由业务部门驱动建设的，如果缺乏信息化的顶层规划，则各系统各自为政、各成体系，成 为信息孤岛，系统之间的数据不标准、不一致，导致应用集成困难、数据分析不准确。可以说 目前国内绝大部分企业都处于这个状态。而信息技术的发展速度又太快，已逐步形成了技术倒 逼企业数字化转型的趋势。而高质量的数据资产，无疑是企业数字化转型的基石。

**2.** **第二个层次——中期目标**

第二个层次是创新与转型。基于数据实现企业管理的升级和业务的创新，通过数据拓展新 业务、构建新业态、探索新模式是企业数据战略的第二个层次，也是企业数据战略的中期目标。 数据战略不再是企业战略的支撑，而是引导，或者说两者相互作用， “IT 即业务”!对于传统制 造企业，利用数据治理，可以加速管理创新、产品创新、销售模式创新，例如：利用数据治理 加强集团管控、基于数据实现供应链协同和优化、基于市场预测实现创新产品设计与快速上市 等。对于服务行业，利用大数据探索服务的新模式，可以拓宽服务的视野，实现模式领域的横 向拓展、服务精度的纵向延伸。未来的服务业的竞争将更加白热化，而数据资产的价值将愈发 明显。

**3.** **第三个层次——长期目标**

第三个层次是定义企业在数字化竞争生态中的角色和地位。科技的变革将改变企业的业务 形态和竞争模式，在未来的数字化竞争中，数字化将是不可被忽视的核心因素，企业数据战略 的部署和成功实施，将决定企业在未来的竞争和数字化生态中是领导者、挑战者还是被淘汰出 局者。“什么样的愿景，决定了什么样的未来”,企业数据战略愿景的规划一定要有未来的“诗 和远方”。要将数据战略愿景融入企业行动方针和核心价值观中，勾勒出企业未来的蓝图。

**7.2.2** **基本原则**

一般企业的数据战略按照下述基本原则进行规划。

**1.** **数据战略与企业的业务战略保持一致**

企业的业务战略影响数据战略的方向和设计，数据战略目标应与业务目标和更高级别的治

**80** 数据治理——工业企业数字化转型之道

理目标保持一致，以互联网思维提升企业的价值。角色、收入共享、信任和控制是平台治理的 关键职能。数据治理中的角色指的是一种责任明确的数据认责方式，它允许企业保护数据和数 据所有者、使用者的权利。收入共享要求平台所有者应该考虑对数据贡献者的奖励。信任被认 为是成功的先决条件，为了提高信任，数据的高透明度在数据治理中至关重要，可以通过与平 台用户共享决定权来增加信任。否则，必须由平台所有者实施严格的控制机制，并且决策的结 果或过程必须向所有用户开放。

**2.** **企业各级领导高度重视**

明确数据战略规划不仅是企业领导的“一把手”工程，更是各级领导的重点工程。各级领 导应对数据战略高度重视，进而确保数据战略能够顺利推行；要定期召开工作会议，及时了解 项目进展状况，并按实施阶段参与项目审查、评估；同时抽调业务骨干与管理负责人加入数据 战略设计项目组。

**3.** **业务全面配合**

业务管理部门应积极配合项目实施，不应单纯地认为数据战略规划是 IT 部门的技术实现， 而应认为它是一次业务管理上的革新。业务管理部门与信息化部门共同组成项目组，业务管理 部门人员从未来业务开展与部门运营管理角度提出建议，协助实施团队开展业务需求分析。业 务部门要深度参与详细的数据战略流程梳理与优化工作，使优化后的流程满足业务管理部门的 业务执行要求。

**4.** **加强规范管理**

对于数据战略规划应做到统一领导、职责清晰、制度规范、流程优化。企业的数据治理工 作，应严格遵照企业统一制定的数据战略规划开展。在制度建设与流程优化方面，由企业统 一制定管理制度与流程规范，下属单位贯彻执行，企业总部对数据战略执行情况定期进行考 核。

企业只有制定科学的数据战略，才能指导数据治理工作循序渐进、持续优化，达到“数以 致用”的目标。

**7.2.3** **战略举措选择**

对企业来说，数据治理范围和内容该如何选择，是摆在企业面前不得不回答的问题。企业 的数据治理定位应充分考虑以下几点因素：企业数据管理的痛点需求是什么?希望实现的目标

运营 战略

第7章 数据战略 **81**

是什么?实施数据治理能解决哪些痛点问题?数据治理项目的投资计划是什么?期望的投资回 报率是多少?把以上问题都想清楚了，企业的数据战略定位也就清晰了——或选择全域治理，

或选择个别亟待治理的主题。

比如企业先进行主数据治理，通过该项目建立财务类、客商、物料类主数据标准，提升数 据质量。企业通过梳理、建立数据指标，更好地支撑企业数据指标共享与应用，提升数据指标

质量，满足企业内部分析和外部监管要求。

**7.2.4** **模型工具**

数据战略规划的工具有很多，例如：战略地图、差距分析、SWOT 分析、PEST 分析、5W1H 分析、发展驱动力分析、波特五力分析、BCG 矩阵分析等。本书引用业界数据战略规划的两个 经典模型：战略一致性模型和阿姆斯特丹信息模型，这两个模型展示了管理数据的高阶关系图。

**1.** **战略一致性模型**

战略一致性模型将各种数据管理方法的基本驱动因素抽象化。模型的中心是数据和信息之 间的关系。信息通常与业务战略、组织和流程相关。数据通常与信息系统、IT 战略相关。围绕 信息和数据的是战略选择的4个基本领域：业务战略、IT 战略、组织和流程、信息系统，如图

7-2-2所示。



业务

Ⅱ战略

数据

**信息系统**

组织 和流程

**业务战略**

信息

T

**图7-2-2** **战略一致性模型**

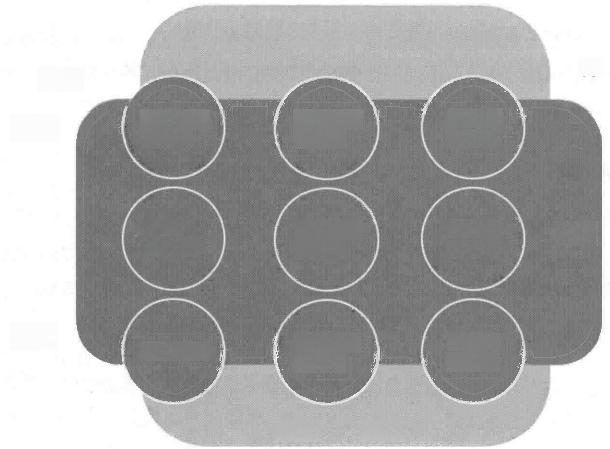
**2.** **阿姆斯特丹信息模型**

阿姆斯特丹信息模型于2010年发布在商业IT 联盟博客中。阿姆斯特丹信息模型从战略角

**82** 数据治理——工业企业数字化转型之道

度考虑业务和 IT 的一致性，强调信息架构规划和中间层，提出信息治理和数据质量的必要性支

撑。如图7-2-3所示，模型的横向为业务、IT; 纵向为业务战略、业务运营。



**信息治理**

IT

Ⅱ战略

和治理

信息架构

和规划

信息管理

和使用

**数据** **质** **量**

业务战略 和治理

信息战略 和治理

**T** **服务** **和开发**

组织

和流程

T 构架 和规划

**业务执行**

**运** **营**

**业** **务**

**构架**

**战略**

图7-2-3 阿姆斯特丹信息模型

**7.3** **数据战略实施**

规划数据战略仅仅是第一步，如何在整个企业中落实和合理执行数据战略是难点。企业高 层管理者应带头在企业内培养数据文化，将数据视为企业最关键的资产之一。数据文化的建立 必须从顶层驱动，并向下逐级贯穿到企业的每个层次。通过加强业务部门信息化能力培养，开 展数据资产管理系列课程培训，可以培养员工的数据管理能力，选拔数据资产管理的核心人才； 提升全员对数据资产的全面认识，充分借助信息化管理手段引领企业数字化转型之路。

数据战略的正确执行，需要从企业高层管理者开始。只有在高层管理者的支持下，才能创

造出自上而下的连锁反应，让数据就是核心资产的意识渗透到企业的每一个角落。

数据战略实施过程是企业完成数据战略规划并逐渐实现数据职能框架的过程。在实施过程 中要评估企业数据管理和数据应用的现状，确定现状与愿景、目标之间的差距；依据数据职能

框架制定阶段性数据任务目标，并确定实施步骤。

第7章 数据战略 **83**

数据治理的实施，应规划项目里程碑，具备可控性，并对阶段性工作做出评估，总结经验， 及时调整并做好下一步工作的准备。为确保项目实施的成功，应使用成熟的实施方法论。

**7.3.1** **实施策略**

实施策略解决了怎么做、由谁做、做的条件、成功原因等问题，是数据战略的核心内容。 数据治理项目涉及的业务范围广、系统范围大、参与人员多，并且是一个需要不断迭代、持续 优化的过程，不能一蹴而就。那么数据治理项目该从何处入手?谁来主导?谁来配合?怎样才 能保证项目成功实施并能够取得效果?这些问题不好回答。根据笔者这些年见到、听到或亲身 经历的数据项目，其成功或失败很大程度上是由这个“制胜逻辑”决定的。我们看大多数失败 的项目都可能会有以下几个特点：目标不明确、范围不清晰、主导人员分量不足、参与人员不 够积极、过分迷信技术和工具、过度依赖外部资源等。做正确的事远比正确地做事更加重要， 事前想清楚数据战略的制胜逻辑，要比事后总结教训的成本低很多。数据治理项目的成功一定 是将以上因素有机整合，忽视任意一个因素都可能会影响数据治理的成效。

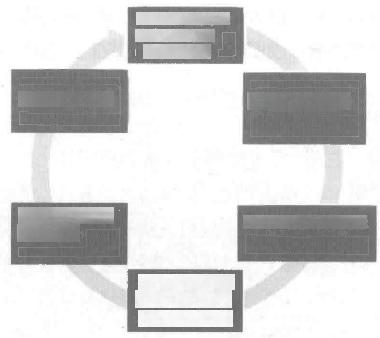
**7.3.2** **实施路径**

实施路径是落实战略目标或指导方针而采取的具有“协调性”的计划安排。行动计划解决 了“谁”“在什么时间”“做什么事”“达成什么目标”的具体问题。行动计划要具备可执行性， 能够量化、能够度量，遵循 PDCA 的闭环管理，要定期进行复盘和检讨。项目建设过程需要企 业高层管理者的高度重视并给予足够的资源支持，需要有经验丰富的顾问团队，需要技术部门 和业务部门的通力协作，只有这样才能提高项目建设的成功率。然而，项目建设阶段的成功并 不代表数据治理的成功，建设阶段的成功不是企业数据治理项目的终点，而是企业数据治理项 目的起点。路漫漫其修远兮，企业数据治理需要的是持续运营。将数据治理形成规则融入企业 文化，是企业数据治理的根本之道。

**7.3.3** **实施步骤**

数据战略实施分为6大步骤，如图7-3-1所示。

**84** 数据治理——工业企业数字化转型之道



1.企业战略环境

的分析和预测

(PEST分析)

6.回顾和考核 2. 识别数据战摩

3.制定数据战略目标

4.编制数据战

略行动实施纲

要和实施计划

5.落实实施战 略的措施

**图7-3-1** **数据战略实施步骤**

**1.第1步：企业战略环境的分析和预测**

企业要分析影响企业数据战略的内外部环境。内部环境包括：企业的业务战略、相关政策， 业务部门的现状和未来的发展方向；企业数据治理的成熟度，以及现行的数据治理对业务的支 撑程度，要找出差距，明确改进和提升方向。外部环境包括：社会、经济、政治、文化、技术 等各个领域现在或将来可能发生的变化情况。数据战略的制定要包括内外部环境的各个相关因 素，使得数据战略成为企业战略不可分割的重要组成部分。

**2.** **第2步：识别数据战略**

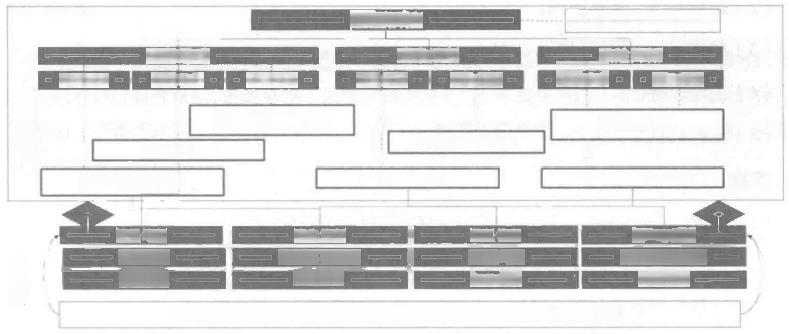
企业还需根据自身发展业务战略、信息化战略的要求来识别本企业的数据战略。数据战略 来源于业务并服务于业务，企业需要结合自身的业务发展要求来制定数据战略。例如： 一家生 产制造企业，其数据战略是紧紧围绕企业的生产开展的，通过数据治理实现“降低成本、提高 效率、质量”的目标。

数据治理的需求始于数据所承载的业务价值，而非由技术或者IT 驱动，如图7-3-2所示。

价值导向

主数据

第7章 数据战略 **85**



规模与效益的协同发展

风险降低

存描区照 信货风验

渠道存货评估

优化产品结构

客户保持/交叉销售

供应商信息

供应商产品信息

分听报告

战略举措的成功执行均有赖于可用、精准、一致的企业全局数据

成本控制

材料成本 物流成本

客户信息

财务(科目》信息

地址信总

警户感知 产品创断 糖准酱销

物料信息 BOM信息

工艺信息

产品信息 组织信息 你格信息

企业价值

客户/供应商信用评估

为客户营销合适产品

业务提升

与物流商战略合作

供应商询价/比价

图7-3-2 价值引导，业务驱动的数据战略识别

**3.** **第3步：制定数据战略目标**

数据是企业各部门共同拥有的资源和资产，数据不能“私有化”,应对数据资产进行集中管 理，统一治理，按需使用，从而使数据资产的效用最大化。制定数据目标要以业务应用为目标， 数据管理为手段，在实现数据标准化管理的同时提升数据的应用效率，并确保数据的合规应用。

现阶段，很多企业的数据资产管理仍是分散模式，由各个部门在自己的业务领域内推进不 同的应用场景，缺乏牵头部门对不同应用场景的整合管理。企业需要根据自身的特点，提出“业 务数据化”“数据业务化”的数据战略的总体目标和阶段目标，并将其拆解为可评估、可衡量、 可操作的目标。

**4.** **第4步：编制数据战略实施纲要和实施计划**

企业要按单位、按部门进行数据战略目标的分解和细化，并制订每个细化目标的实施时点 和详细行动计划，确定每个行动计划的起止时间、负责部门/岗位/角色/人员、明确输入/输出成 果。行动计划的制订要与企业实际相结合，可执行，可量化，可评估。

企业要编制实施纲要和实施计划，列明为实现各子目标应采取的具体行动措施，以及相应

的责任。

数据战略实施纲要主要包括：

(1)实施数据战略纲要的现状和基础；

**86** 数据治理——工业企业数字化转型之道

(2)指导思想、基本原则；

(3)总体目标、阶段目标；

(4)主要任务；

(5)配套机制及保障措施。

实施计划主要包括：

(1)按部门进行细化，并按具体时间段制订详尽的行动计划；

(2)对照数据战略目标的实现日期，确定每个行动步骤明确的起始时间；

(3)以各部门职责分工为基础，确定行动步骤的负责人；

(4)明确分阶段的短期目标。

**5.** **第5步：** **落实实施战略的措施**

实施战略的措施是指为实现数据战略而建立的相关保障措施，主要包括数据管控和技术工

具体系。管控体系包括数据治理组织、数据标准规范体系、数据管理流程、数据管理制度等。

**6.第6步：** **回顾和考核**

数据治理的绩效考核主要是指对各相关部门的数据治理进行定性和定量的衡量、打分，并 公布考核结果。绩效考核一方面是为了促进数据治理工作的有效开展，另一方面也是为了对数 据战略目标进行验证，以发现问题和不足并及时实施改进措施，从而使数据战略目标不断地完 善和优化。

下面总结数据战略实施难点。

(1)业务战略的关联关系不强；

(2)仅仅是“纸面工作”;

(3)目标太高、太大；

(4)缺少配套的资源；

(5)缺少可实施的路线图。

第7章 数据战略 **87**

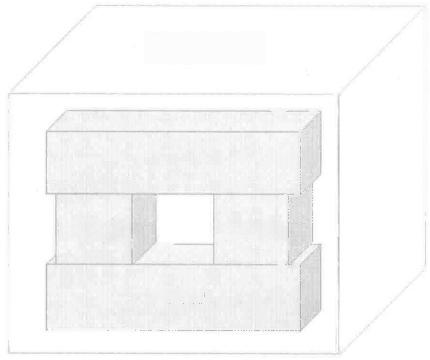
**本章精要**

数据战略不仅是企业领导的“一把手”工程，更是各级领导的重点工程，各级领导应对数 据战略规划项目高度重视，进而确保项目能够顺利推行。本章首先阐述了数据战略规划的目标、 基本原则、主要活动和主要内容；其次，介绍了在数据战略实施中重点介绍实施策略、实施路 径和实施步骤。

第8章

数据架构

数据架构是企业架构(Enterprise Architecture,EA) 的一部分。Zachman认为企业架构是构 成组织的所有关键元素和关系的综合描述，而企业架构框架(EAF) 是一个描述企业架构方法 的蓝图。经过近30年的发展，企业架构理论已经相当成熟，目前，国际上影响力比较大的企业 架构框架有 Zachman 框架、DoDAF 框架、FEAF 框架、TOGAF 框架等。 一个完整的企业架构 通常被划分为业务架构、应用架构、数据架构和技术架构，如图8-1-1 所示。

企业架构

数据架构

**应用架构**

**技术架构**

**图8-1-1** **企业架构划分**

在企业架构中，业务架构描述了企业各业务之间相互作用的关系结构和贯彻企业业务战略 的基本业务运作模式；数据架构将企业业务实体抽象为信息对象，将企业的业务运作模式抽象 为信息对象的属性和方法，建立面向对象的企业数据模型，数据架构实现从业务模式向数据模

第8章 数据架构 **89**

型的转变，业务需求向信息功能的映射，企业基础数据向企业信息的抽象；应用架构以数据架 构为基础，建立支撑企业业务运行的各个业务系统，通过应用系统的集成运行，实现企业信息 自动化流动。

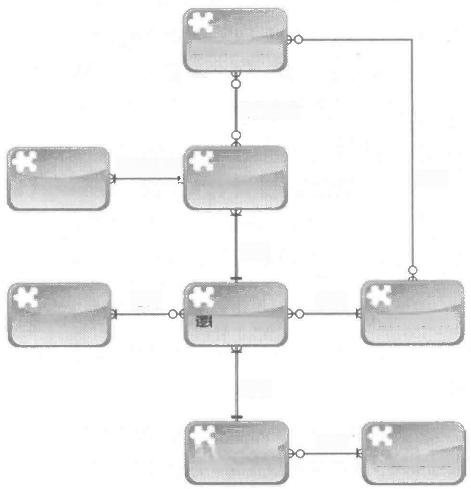
**8.1** **数据架构概述**

依据美国电气及电子工程师协会标准 IEEE/ANSI 1471/2000 和国际化标准组织 ISO/IEC/ IEEE 42010/2011的定义，架构是某些事物的基本组织，体现在其组成部分、它们彼此的关系、 它们和环境的关系，以及它们的设计原则和演进。所以数据架构的基本组织/架构元素是：数据 组件、数据组件之间的关系、数据组件和环境的关系，以及数据组件的设计原则和演进记录。

参考企业架构国际标准 TOGA F(The Open Group Architecture Framework ),数据组件的类 型有数据实体、逻辑数据组件和物理数据组件。数据模型是数据组件、组件之间的连接关系和

关系基数的图形呈现 。

图8-1-2展示了 TOGAF 企业架构内容元模型中定义的数据实体、业务服务/流程和应用等 相关架构元素的关系。



业务服务/流程

消费/提供

供应/消费。

行动者

封装

包含

位置

实现

使用

物理数据组件

逻辑应用组件

物理应用组件

辅数据组件

数据实体

使用

实现

图8- 1-2 企业架构内容元模型中的相关元素

**90** 数据治理——工业企业数字化转型之道

数据实体是数据的封装，可被业务领域专家认知为某种事物。逻辑数据实体可绑定到应用、

存储库和服务上。

逻辑数据组件是将相关数据实体封装到有边界的区域以便保留逻辑位置，例如外部采购信 息。

物理数据组件用于将相关数据实体封装到有边界的区域以便保留物理位置。例如采购订单 对象，由采购订单表头和采购订单项对象节点组成。

**8.2** **框架设计**

**8.2.1** **数据分布**

工业企业在采购、生产、销售到客户服务等过程中，无不伴随着数据的产生、流转和运用。 通过有效的组织、存储、分发和管理可实现在不同业务线之间的数据共享。狭义的数据架构可 以用来特指数据分布，包括数据业务分布与数据应用(系统)分布。数据业务分布指数据在业 务各环节的 CRUD 关系；数据系统分布指单一系统中数据架构与系统各功能模块的引用关系， 以及数据在多个系统间的引用关系。数据业务分布是数据系统分布的基础和驱动。

对于拥有众多分支机构的大型工业集团企业，数据是集中存放还是分散存放是数据分布设 计中重要的考虑内容。从地域角度看，数据分布有数据集中存放和数据分布存放两种模式。这 两种数据分布模式各有其优缺点，需要综合考虑自身需求确定具体的数据分布策略。

一般的数据分布常采用“操作型业务系统数据库+操作型数据存储库(+数据仓库)+数据 集市”的方式。业界领先的实践(如数据中台)一般结合面向服务架构(SOA)、 商业智能(BI) 技术和数据虚拟化技术，利用数据整合平台将数据仓库中的数据转变为能被其他系统所访问的 数据服务，为那些需要满足 BI需求、访问数据仓库数据的系统提供访问路径。

**1.** **数据目录**

数据目录作为数据共享交换的基础数据，对促进企业内部数据共享与交换、对外上报和公 示相关信息都非常重要。

可以自动或者手动从不同信息资源中抽取数据并生成需要的数据目录。信息资源包括任何 数据湖、数据仓库、数据集市、生产系统数据库，以及其他被确定为有价值、可共享的数据资 产实体。数据目录是数据管理、数据管控和数据治理的核心。

**螺**

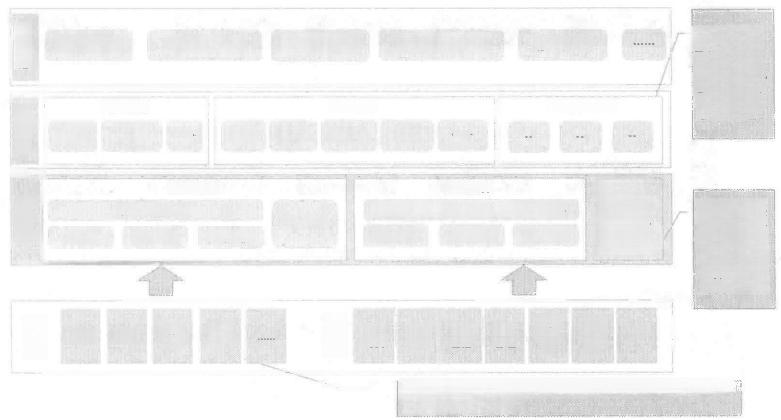
**中** **心**

数据 来源

数据目录可以是一种图形化的数据资产管理工具，它提供了多层次的图形化展现，并具备 各种力度的控制能力，满足业务使用、数据管理、开发运维不同应用场景的图形查询和辅助分 析需求。

**2.** **数据资源全景图**

数据资源全景图是企业全部数据资产的总体视图，既包括数据分布、流向和交互关系，又 包括数据治理、数据服务和数据后期应用的完整视图，是企业资产管理和运营的重要基础。数 据资源全景图如图8-2-1所示。



对策三：建设 企业级数据中 心(包括股份

总部及园区)

逐步完普数据

应用，提离数

据资产价值

生 产 供应链

**数据计算**

大敷据算法库(回归、预测、决策……)

含 主 数 润

批量计算 实时计算 内存计算

一 数 据 标 准

与管理制度

安健

环管

理

对策二：采集并完善缺失的信息；实现数据采集从人工录

入到自动采集的转变

多维分析模型

业务数据 指标数据

对策一：通 过主数据平

台，进行数 据 治 理 ， 统

管理驾驶舱 多维分析报表

集团管控 人力 资源

数据质量、 数据安全)

**数** **据** **治** **理**

**平** **台** **(** **包**

数据应用 数 据 服 喜

大数据挖掘分析

自主查询报表

生产

管控 系统

实验 室管 理

经营 管理 系统

区域园区分析

设备 管理

移动分析

财务 共要

物流 系统

生产 调度

能源 管控

ERP CRM

核心业务

**数据存储**

生产 执行

元数据

数据平台

主数据

研发

销售

财务

………

…

\*

**图** **8** **-** **2** **-** **1** **数** **据** **资** **源** **全** **景** **图** **(** **示** **例** **)**

**3.** **数据地图分布应用**

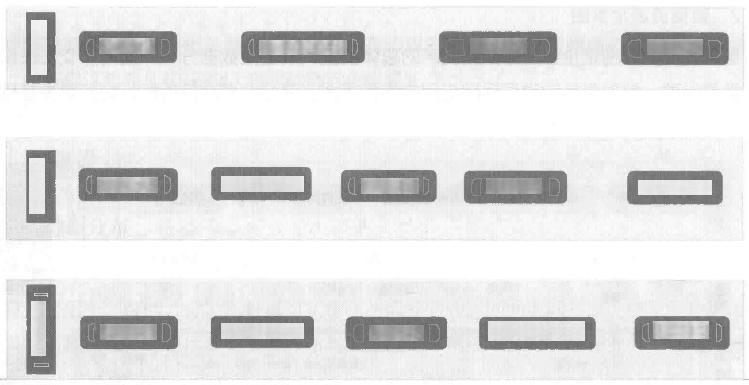
数据地图分布应用是指站在数据资产全景图的视角查看企业各数据域，在每一个数据域下，

可以识别企业各项业务的核心数据主题，明确各个主题间的交互关系，将数据实体分类，形成 企业级数据地图。

构建企业数据地图的意义在于弄清楚企业的数据资产，未来可在此基础上明晰数据在系统 和业务中的分布和流向，保证企业内部信息系统之间共享数据的一致性，亦可在此基础上开展 数据建模、主数据管理、数据标准化等数据管控、治理工作。

**8.2.2** **数据主题域**

数据主题域是最高层级的，以各个主题概念及其之间的关系为基本构成单元的数据主题集 合。企业应划分统一的数据主题域，形成统一的企业数据视图，如图8-2-2所示。



战略规划 计划与预算管理 投资管理 绩效管理

人力资源管理

科技与工艺管理

生产调度管理

内控审计管理

生 执 行

生产管理

**财务管理**

HSE管理

物资管理

项目管理

设备管理

管理支持

战略晨展

**图8-2-2** **工业企业数据分布(示例)**

**1.** **战略发展类主题**

战略发展类主题主要涵盖以下主题。

●“战略规划”主题：包括战略规划编制信息、规划执行信息、股本结构信息、债权人信息、 股权人员信息、高管人员信息。

· “计划与预算管理”主题：包括计划编制维度、计划编制信息、计划执行信息、计划考核 信息、预算科目、预算编制信息、预算目标、预算执行信息、预算考评信息。

●“投资管理”主题：包括投资需求、投资执行信息、投资方案、投资评估、市场兼并收购 信息、投资后评估信息。

●“绩效管理”主题：包括企业绩效目标、考核标准、考核指标、考核结果。

**2.** **管理支持类主题**

管理支持类主题主要涵盖以下主题。

· “财务管理”主题：包括会计凭证、会计科目、金融机构、应付账款、应收账款、固定资 产、内部订单、成本费用信息、财务三大报表。

第8章 数据架构 **93**

· “人力资源管理”主题：包括组织机构信息、岗位信息、员工信息、薪酬福利信息、招聘 信息、员工考勤信息、培训课程信息、职业生涯规划信息、员工绩效考评信息。

“物资管理”主题：包括采购计划、采购方案、寻源文件、采购订单、采购合同、出入库 单、物资领料单、物资调拨单、配送计划信息、物资配送信息、承运商信息、配送监控 信息、合同监造信息、到货接收单。

●“项目管理”主题：包括项目基本信息、项目科研信息、项目计划信息、项目进度信息、 项目施工信息、项目成本信息、项目安全质量信息、项目效益评估信息、项目概预结算 信息。

· “内控审计管理”主题：包括风险评估标准、风险事件库、风险评估报告、审计信息、审 计项目、控制活动、专项检查信息。

**3.** **生产执行类主题**

生产执行类主题主要涵盖以下主题。

●“生产管理”主题：包括生产年度/月度计划、日排产计划、生产工单、作业指令、物料 消耗信息、生产成本信息、产品化验结果等。

“生产调度管理”主题：包括生产调度计划、调度标准指令集、标准问题集、调度指令下

达信息、调度问题、调度日志、应急事件。

●“HSE 管理”主题：包括风险/隐患信息、事故/事件信息、安全检查计划、安全检查记录、 应急预案、应急物资与装备、职业病记录，污染源检测信息、节能减排计划、节能减排

量。

· “科技与工艺管理”主题：包括生产技术标准、工艺专利信息、科技项目信息。

●“设备管理”主题：包括设备台账信息、设备功能位置、设备配套计划、设备采购计划、 设备领用记录、设备维修计划、故障类型、设备维修记录、设备资产折旧信息。

**8.2.3** **数据关联关系**

**1.** **基本概念**

(1)实体。

实体是一个系统内可定义的事物或概念，如人/角色(例如学生)、对象(例如发票)、概念 (例如简介)或事件(例如交易)(注：在实体关系图中，“实体”通常用来代替“表”,但它们 是一样的)。在考虑实体时，可以尝试把它们想成名词。在ER 实体关系模型中，实体显示为圆

**94** 数据治理——工业企业数字化转型之道

角矩形，其名称位于上方，其属性列在实体形状的主体中。图8-2-3为 ERD (实体关系图)中 实体的一个用例。

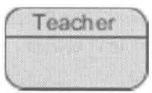


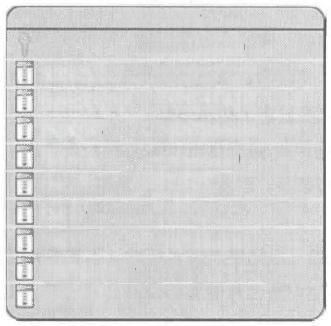
图8-2-3 一个实体用例

(2)属性。

属性也被称为列(Row), 意思是持有它的实体的属性或特性。

一个属性有一个描述属性的名称和一个描述属性种类的类型，例如代表字符串的 varchar, 代表整数的int。当为物理数据库开发绘制 ERD时，应使用目标 RDBMS 支持的类型，以确保 设计和物理数据库一致。

图8-2-4中的 ERD示例显示了一个包含属性的实体。



Customer

**integer(10)**

varchar(255) varchar(255) varchar(255)

integer(10) char(1)

chan(1)

varchar(50) date

date

ID

First\_Name Last\_Name

Address

Telephone

Gender

Actve

Email

Create\_Date Last\_Update

N

N

N

N

N

N

N

N

N

**图8-2-4** **一个包含属性的实体**

(3)主键。

主键(Primary Key,PK),是一种特殊的实体属性，用于界定数据库表中的记录的独特性。 一张表中不能只有一个主键属性值的记录，证明身份的 ID 便是典型的例子。两个人即使姓名 相同， ID 也不会一样，若身份证明是一张表，那么 ID 便是主键了。图8-2-5中的 ERD 示例 显示了拥有主键属性 “ID” 的实体 “Product”, 以及数据库中表记录的预览。第三个记录是 无效的，因为 ID“PDT -0002” 的值已被另一个记录使用。



第8章 数据架构 **95**

 Product ID integer(10)

Name vachar255)N

(4)外键。

外键 (Foreign

**Name**

ID

Tiger T7 Bluetooth Headphones

PDT-0001

PDT-0002

PDT-0002

PDT-0003

**■■■**

DD-027 In-Ear Headphones,Black

SDB-21 HI-FI Stereo Over-ear Earphones

Mr.1022 Deep Bass Earbuds

图8-2-5 主键 (ERD)

Key,FK),是对主键的引用，用于识别实体之间的关系。请注意，有别于

主键，外键不必是唯一的，多个记录可以共享相同的值。图8-2-6中的 ERD 示例包含一系列的 实体，其中一个外键用于引用另一个实体。

 Manufacturer ID **Name Conlact**

ID Inte ger(10)

Name varchar(255) N ■■■ M-DI Hello World Tech. 534-55-7478

M-02 ABC Technologies 283-92-8511

Contact vachar(255) N

1

ID **ManufacturerlD Name**

Product

anufacturerlD ■ PDT-00OI M-01 Tiger T7 Bluefooth Headphones

Name vachar(255)N PDT-0002 M-01 DD-027 In-Ear Headphones,Black

PDT-0003 M-02 Mr.1022 Deep Bass Earbuds

图8-2-6 外键 (ERD)

(5)关系。

两个实体之间的关系表示这两个实体以某种方式相互关联。例如，学生可能参加课程，因

此实体“学生”与“课程”相关，而这一关系在 ERD 中则以连接线表示。

(6)基数。

基数定义了在一个实体与另一个实体的关系里，某方可能出现的次数。例如， 一个团队有

许多球员，若把这一关系呈现于ERD 中，那么团队和球员之间是一对多的关系。

**2.** **三种关系**

在 ERD 中，三种常见的主要关系是一对一、一对多和多对多。

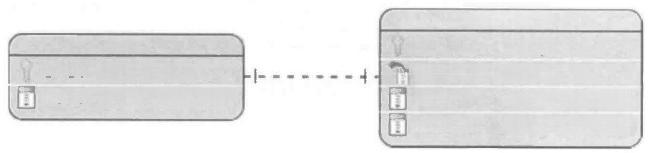
UserProfie

Student

**96** 数据治理——工业企业数字化转型之道

( 1 ) 一 对 一 。

一对一关系主要用于将实体分成两部分，简洁地将信息呈现，使读者更容易理解。图8-2-7 显示了一对一关系的例子。



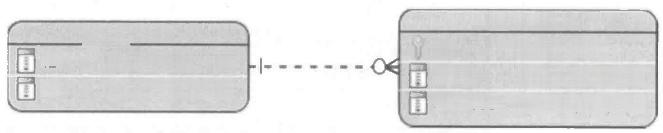
|  |  |  |  |
| --- | --- | --- | --- |
|  | User | ID | integer(10) |
| Email | integor(10) | *UserEmai* | *integor(10)* |
| Name | varchar(255)N | Address  Remarks | varchar(255)N varchar(255)N |

图8-2-7 一对一关系

(2)一对多。

一对多关系是指两个实体X 和 Y 之间的关系，其中X 的一个实例可以连接 Y 的许多实例，

而 Y的一个实例仅可以连接X 的一个实例。图8-2-8显示了一对多关系的一个例子。



Department

ID integer(10) N

Name varchar(255)N

**integer(10)**

varchar(255)N varchar(255)N

ID

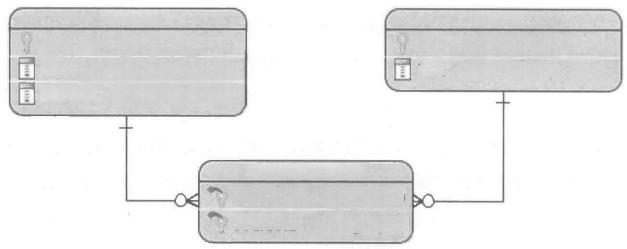
Name

Address

图8-2-8 一对多关系

**(3)多对多。**

多对多关系是指两个实体 X 和 Y 之间的关系，其中X 可以连接Y 的许多实例，反之亦然。 图8- 2-9显示了 一 个多对多关系的例子。请注意，多对多关系在物理 ERD 中被定义为 一 对 一 对 多的关系。



Student

ID integer(10)

Name varchar(255)N

Address varchar(255) N

Student Course

StudentID integer(10) CourselD integer(10)

Course

ID integer(10)

Title varchar(255) N

图8-2-9 多对多关系

第8章 数据架构 **97**

**2.数据血缘关系**

数据血缘关系是指数据从产生、处理、流转到消亡过程中，数据之间形成的一种类似于人 类社会血缘关系的关系。和人类社会血缘关系不同的是，同一个数据可以有多个来源(多个“父 亲”)。 一个数据可以是多个数据经过加工而生成的，而且这种加工过程可以有多个。

(1)表级血缘关系。

针对表结构的情况，最终用户和运维用户最需要关注目标表中每个字段的数据的来源有哪 些。即建立源表、源字段与目标表、目标字段的映射关系。 一个目标表可以对应多个来源表的 字段，比如：姓名字段，可能来自户籍人口表，也可能来自流动人口表，也就意味着这两张表 合并起来的人口，才是这个区域的所有人口。

(2)字段级血缘关系。

从当前记录出发，可以按时间查看该记录所有的变更过程。 一条记录的生成可能对应两个 表的两条记录，这种对应是可跟踪的。

如果再精细跟踪，就可以做到字段级血缘分析，可以与表结构的血缘分析完美呼应。单击 某一个字段，可查看该字段的血缘关系。

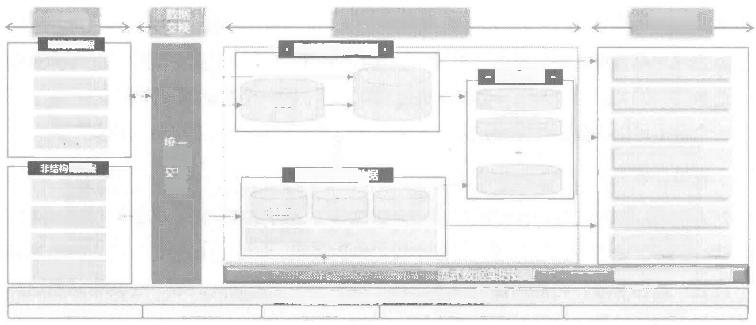
**3.** **数据流转关系**

数据流转线路表现的是数据的流转路径。数据流转线路从数据流入节点向主节点汇聚，又 从主节点流出向数据流出节点扩散。

数据流转线路表现了三个维度的信息，分别是方向、数据更新量级、数据更新频次。方向 默认从左到右流转；数据更新量级通过线条的粗细来表现——线条越粗则表示数据量级越大， 线条越细则表示数据量级越小；数据更新频次用线条中线段的长度来表现，线段越短则表示更 新频次越少，线段越长则表示更新频次越多， 一根实线则表示只流转一次。

数据流转通过系统间的接口进行交换和传输。图8-2-10 是比较典型的企业数据流转架构， 其中业务系统中的数据由统一数据交换平台流转分发给传统关系型数据库和非关系型大数据平 台，数据仓库和大数据平台汇总这些数据后，供各个应用集市分析使用。

**98** 数据治理——工业企业数字化转型之道



政 据 通

结构化数质

传统关系型数据

核心要族

应用集市

倘贷系族

CRM摆练

财离系族

T

tetat

自助查询

化数累

非 关 系 型 大数

OLAP

外部站点

内邮文档

系统日志

第三方鼓据

小 理

**数** **据** **备** **份** **/** **数** **据** **恢** **复** **/** **运** **维** **监** **控** **/** **数** **据** **安** **全**

调廑

决策仪表盘

统计查询

固定报表

数据共享

数据挖避

实时预警

应用主题1

应用主题

分布式文件存储系统(HDFS)

端位用

据

平台

企业级教属平台

应用访问控

**作业与数据**监控

数据 实验室

行业 大数据

应用主题n

数掘仓库

资源管理

高户 大数据

**作业是行**

oDs



→

**图8-2-10** **企业数据流转架构**

**8.3** **数据建模**

数据模型是一种工具，用来描述数据、数据的语义、数据之间的关系，以及数据的约束等。 实体关系图经常被用来描述现实世界中的数据模型。当我们理解了实际问题的需求之后，需要 用一种方法来表示这种需求，数据模型就是用来描述这种需求的。

数据模型依据抽象层次可分类为概念数据模型、逻辑数据模型和物理数据模型。概念数据 模型和逻辑数据模型是需求分析的产出，而物理数据模型是设计活动的产出。数据模型是呈现 数据需求、分析和设计的规格说明书。

数据模型中常见的术语如下。

(1)实体 (Entity): 现实世界中的对象，可以具体到人、事、物。就一般企业而言，实体 是业务专家认知的某种事物，如采购订单、产品、服务、客户等。

( 2 ) 属 性 (Attribute): 实体所具有的特性。属性用来描述实体，是组成实体的数据定义、 格式和值域，如采购订单编号、产品编号、客户电话等。

(3)键属性(Key Attribute): 可唯一识别数据实体实例 (Instance) 和数据库表行记录的属 性，如客户编号可识别不同客户。每个实体一般都有主键(Primary Key)属性，也可能有外键

(Foreign Key)属性。

(4)关系(Relationshi p): 实体和实体之间的关系。可通过连线(C onnection ) 表示，关系 可帮助辨识主键和外键。

第8章 数据架构 **99**

(5)范式 (Normal Form): 规范实体中属性之间的依赖和分解关系，如第一范式(1\*NF) 强调的是属性的原子性，即属性不能够再被分解成其他属性。

图8-3-1 是信息工程数据建模中常用的图形符号。



Connection

Entity

Attribute

Primary key

Foreign key

**图8-3-1** **信息工程数据建模中常用的图形符号**

在设计数据库时， 一般都要求能符合第三范式(3Normal Form,3NF)。第三范式必须是 第 二 范 式 ( 2Normal Form,2NF),所有非主键属性必须能直接依赖于主键，而且其他属性间

不能存在传递依赖关系。

第二范式必须是第一范式，此外，还需满足两个条件：

(1)实体/表必须有一个主键；

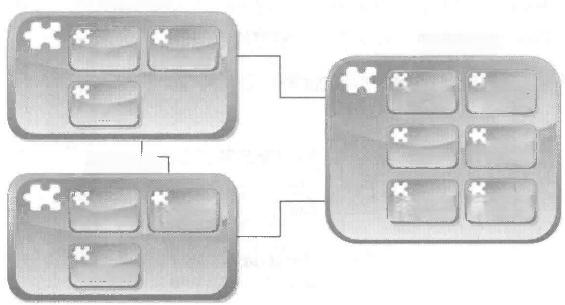
(2)没有包含在主键中的属性必须完全依赖主键，而不能只依赖主键的一部分。

符合第三范式可保证数据的完整性和一致性，这是数据质量的最低要求。在特殊情况下(例 如为了提高系统性能),允许不符合第三范式，但是数据的完整性和一致性需要通过特殊控制来 保证。

**8.3.1** **概念数据模型**

高阶的概念数据模型可以是数据实体和主题领域的目录清单及组成关系。图8-3-2描述了 产品主题域、采购主题域、库存主题域，以及各主题域包含的数据实体。

**100** 数据治理 — — 工业企业数字化转型之道



**产品主题域**

**产品** **产品规格**

|  |  |
| --- | --- |
| **采购订单**  **收货**  供应商绩效 | 供应商计 划交付物  供应商协议  供应商通货 |

**备用零件**

**产品组**

**库存主题域**

**库存和位置** **通输中库存**

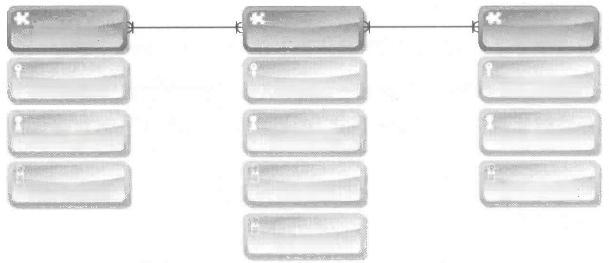
**采购主题域**

**图8-** **3-2** **概** **念** **数** **据** **模** **型**

在概念层次，主要工作是发现核心流程(产生客户价值的流程)使用的数据实体(人、事、 物),将它们以清单的方式列出，并分组到对应的主题域。

**8.3.2** **逻辑数据模型**

从概念数据模型中选取与采购订单实体相关的产品及其库存和位置，分析每个实体的属性、 主键和外键，可以得到简化的逻辑数据模型。图8-3-3 中未列出所有的实体属性，如采购时间、 订单状态、到货时间、交货地点、财务信息等。



**产品**

**产品编号**

**供应商编号**

**产品描述**

**产品单位**

**库存和位置**

**库存位置编号**

**库存产品编号**

**库存产品数量**

**采购订单**

**采购订单编号**

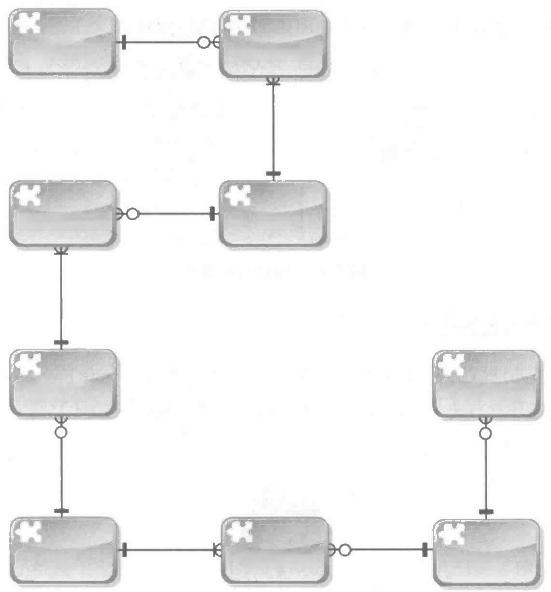
**采购产品编号**

**采购产品数量**

**图8-3-3** **逻** **辑** **数** **据** **模** **型**

逻辑数据模型在转换为物理数据模型时，必须解决实体之间的多对多关系，常用的方法就 是将多对多关系转换成关联型实体(Associative Entity)。例如将图8-3-3中的采购订单和产品之 间的多对多关系转换为“采购订单×产品”实体和两个一对多关系，范例如图8-3-4所示。

第8章 数据架构 **101**



产品x 供应商

采购订单x 产品

产品x需求

客户x需求

采购订单

供应商

产品

需求

客户

图8-3-4 优化的逻辑数据模型

实体“采购订单x产品”是采购订单录入流程的主要数据源，实体“产品×供应商”是产品 供应流程的数据源，实体“产品×需求”是产品开发流程的数据源，而实体“客户×需求”是客

户需求分析流程的数据源。

**8.3.3** **物** **理** **数** **据** **模** **型**

物理数据模型也可以用统 一 建模语言 (UML) 的类图表示，需要将逻辑数据模型中的属性

细化，如库存位置编号可分解为通道(Aisle)、 货 架(Shelf)、 层 (Level)、 容 器 (Bin) 等 。

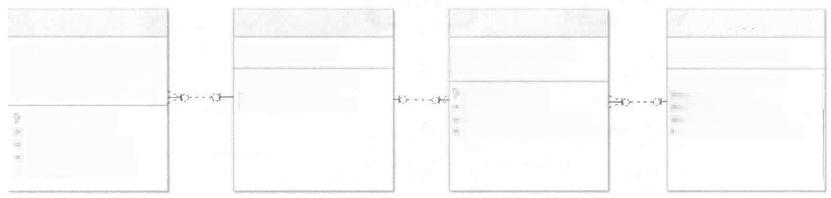
逻辑数据模型的采购订单为了符合第三范式，需要被分解为两个实体 — — 采购订单表头和

采购订单细项。

对于图8-3-5 中的物理数据模型，市场上的许多数据建模工具都可生成这样的模型：例如 用 SQL 的 数 据 定 义 语 言( Data Definition Language,DDL)建立数据库，来支持应用系统的开

发。在产生 SQL 数据定义语言之前，需要先定义属性/字段的详细信息。例如“产品编号”的

数据类型是字符，长度是10,需要将CHARO 修改为 CHAR(10)。



**库存和位置**

库存位置端号-aiste:CHARO

库存位置调号-shelf CHARO

库存位雪编号-levelCHAR0

库存位置端号-binCHAR0

采购产品数里：INT

产品编号：CHARO

库得产品数量INT

库存产品单位：CHARO

入库时间 DATETIME

仓库地址1 VARCHARO

**采购订单**

参采购订单编号：CHARO

采购需求机构：CHARP

采购人：CHARO

采购时间：DATETIME

采购订单状况：CHARO 要求到货时间 DATETIME

采购订单项

采购订单端号：CHARO 采购订单项确号：INT

产品编号：CHARO

产品编号：CHARO

产品描述 VARCHARO

产品单位： CHARO

客户产品编号：CHARO

采购产品到货状况：CHARO

来购产品到货时间：DATETIME



产 品

图8-3-5 物理数据模型

**8.3.4** **数据模型开发方法**

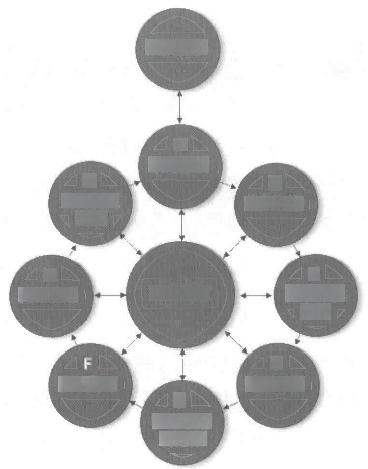
数据模型开发方法取材于 TOGAF(The 结构框架)的架构开发方法( Architecture

如图8-3-6所示。

Open Group

Development

Architecture Framework, 开放组体系 Method,ADM), 它一共有10个阶段，

**预备阶段**

A

**架构愿景**

H

架构变更

管 理

B

业务架构

G

**实施治理**

**需求管理**

C

信息系统

架构

D

**迁移规划**

E

机会及解 决方案

技术架构

图8-3-6 数据模型开发方法

架构开发方法是一个在全球被广泛使用和证明的方法论，它的特色是迭代和循环，专门用

来处理业务需求。

第8章 数据架构 **103**

预备阶段进行准备工作：确定受影响的组织范围，发现架构工作需求，确认治理和支持框 架，建立架构组织、识别架构原则，调整TOGAF 框架，以及制订工具与技术的实施战略和计划。

架构开发阶段包含：

阶段 A——架构愿景；

阶段 B——业务架构：

阶段 C——信息系统架构(数据架构和应用架构);

阶段 D——技术架构，代表了完整的企业架构开发。

阶段 E——依据架构蓝图开发解决方案(架构实施项目);

阶段 F——迁移规划的工作内容；

阶段G——架构开发和解决方案的实施治理；

阶段 H——架构变更管理只覆盖架构变更管理，不包含一般的变更管理。

阶段A 到阶段H 都必须依据需求开发，开发完成后需要验证，当需求能被满足后才能通过 评审和验收。

阶段C 包含了数据架构和应用架构的开发，可从任意一个架构开始。建议先从数据架构的 开发开始，再进行应用架构的开发。利用数据模型的开发方法可以定制阶段C 的数据架构的开

发方法，主要包括如下9大步骤。

(1)选择参考数据模型、数据视图和工具；

(2)梳理当前数据模型，作为基线；

(3)定义未来数据模型，作为目标；

(4)进行基线和目标数据模型的差距分析；

(5)消除差距，定义数据模型发展路线图及需要的数据组件；

(6)解决跨架构景观的影响；

(7)引导正式的利益相关者审查；

(8)定稿数据模型；

(9)创建数据模型文件。

数据架构的输入包括：

**104** 数据治理 工 业 企 业 数字化转型之道

(1)架构工作请求书；

(2)能力评估；

(3)沟通计划；

(4)企业和数据架构组织模型；

(5)已裁剪的企业和数据架构框架；

(6)数据架构原则；

(7)架构工作说明书；

(8)架构愿景；

(9)架构存储库；

(10)架构定义文件初稿(含业务架构);

(11)架构需求规格初稿，其中包括差距分析结果和相关技术需求；

(12)架构路线图的业务架构组件。

数据架构的输出包括：

(1)架构工作说明书( Statement Of Work,SOW);

(2)经验证的数据架构原则或新的数据架构原则；

(3)架构定义文件初稿(含业务架构与数据架构);

(4)架构需求规格初稿；

(5)架构路线图的数据架构组件。

架构定义文件输出物包含数据架构组件，内容如下：

(1)基线数据架构(当前的实际情况)。

(2)目标数据架构，其中包括：

①业务/概念数据模型；

②逻辑数据模型；

③数据管理流程模型；

④数据实体/业务功能矩阵。

第8章 数据架构 **105**

(3)与选定视点对应的数据架构视图(针对关键利益相关者的关注)。

架构需求规格输出物包含数据架构组件，内容如下：

(1)差距分析结果；

(2)数据互操作性需求；

(3)为了与架构变更保持一致，业务架构可能需要改变的地方；

(4)对即将设计的技术架构存在的约束；

(5)最新的业务/应用/数据需求(如果符合实际情况)。

数据架构阶段中常用的模型/视图如下：

(1)数据实体/数据组件目录；

(2)数据实体/业务功能矩阵；

(3)应用/数据矩阵；

(4)概念数据图；

(5)逻辑数据图；

(6)数据传播图；

(7)数据安全图；

(8)数据迁移图；

(9)数据生命周期图。

**本章精要**

数据架构便于企业运营决策者以数据视角分析整个企业的数据分布与业务域之间的关系， 是企业了解自身价值和制定战略决策的最重要依据。数据架构从跨业务、跨级层、跨应用系统 的视角统一对数据进行组织和规划，提高数据集中存储和跨系统间数据共享的效率。

数据架构设计可满足企业信息化各层级要求： 一是经营管理更加集中，实现业务集约化管 理；二是业务更加融合，按照业务主线深度集成所有业务流程，实现企业整体资源共享与业务 协作；三是决策更加智能，实现经营决策智能分析、管理控制智能处理、业务操作智能作业， 深化数据分析利用能力，提高管理决策支撑能力。

第9章

主数据管理

随着互联网和信息技术应用的日益普及，很多企业为了有效解决日趋复杂的业务经营和管 理问题，都在持续大力推进信息化建设工作，陆续开发部署如 OA 、HR、PMS 、PLM 、ERP、 MES 等信息系统，为企业经营效率和管理水平的提升不断注入活力。但在越来越多的信息系统 投入运行后，会发现这些系统在数据互联互通、融合协同方面存在着严重的问题，形成诸多信 息孤岛和数据烟囱，阻碍了企业由信息化向数字化的换代升级。

构建完整的主数据管理体系可以对主数据实施统一、规范、高效的管理，确保分散的系统 间主数据的一致性，改进数据合规性。不仅如此，主数据还是数据标准落地的关键载体，是企 业实施全面数据治理的核心基础，成功实施主数据管理可以很好地推动企业全面建设数据治理 体系。

**9.1** **主数据和主数据管理**

**9.1.1** **主数据的特征**

主数据(Master Data,MD): 指满足跨部门业务协同需要的、反映核心业务实体状态属性 的基础信息。相对而言，主数据属性相对稳定，准确度要求更高，且可唯一识别。

主数据具有以下几个方面的特征。

(1)跨部门。

主数据不是那种局限于某个具体职能部门的数据，而是满足跨部门业务协同需要的、各个 职能部门在开展业务过程中都需要的数据，是所有职能部门及其业务过程的“最大公约数据”。



第9章 主数据管理 **107**

(2)跨流程。

主数据不依赖某个具体的业务流程，但是主要业务流程都需要的。主数据的核心是反映对

象的状态属性，它不随某个具体流程而发生改变，而是整个完整流程中不变的要素。

(3)跨主题。

与信息工程方法论'中通过聚类方法选择主题数据不同，主数据是不依赖于特定业务主题却 又服务于所有业务主题的、有关业务实体的核心信息。

(4)跨系统²。

主数据管理系统是信息系统建设的基础，应该保持相对独立，它服务于但是高于其他业务

信息系统。因此对主数据的管理需要集中化、系统化、规范化。

(5)跨技术。

由于主数据要满足跨部门的业务协同，它必须适应采用不同技术规范的不同业务系统，所 以主数据必须应用一种能够被各类异构系统所兼容的技术。从这个意义上讲，面向微服务架构 为主数据的实施提供了有效的工具。

**9.1.2** **主数据管理的基本概念**

主数据管理( Master Data Management,MDM) 是一系列规则、应用和技术，用以协调和 管理与企业的核心业务实体相关的系统记录数据³。

主数据管理的关键活动包括：理解主数据的整合需求，识别主数据的来源，定义和维护数 据整合架构，实施主数据解决方案，定义和维护数据匹配规则，根据业务规则和数据质量标准 对收集到的主数据进行加工清理，建立主数据创建、变更的流程审批机制，实现各个关联系统 与主数据存储库的数据同步，方便修改、监控、更新关联系统的主数据变化。

主数据管理通过对主数据值进行控制，使得企业可以跨系统地使用一致的和共享的主数据， 提供来自权威数据源的、协调一致的高质量主数据，降低成本和复杂度，从而支撑跨部门、跨 系统的数据融合应用。

1 高复先.信息资源规划——信息化建设基础工程，北京；清华大学出版社，2002

2 王波等.通过标准化主数据实现高效数据交换.管理技术，2008年第12期。

3 《数据资产管理实践白皮书(4.0版)》,中国信息通信研究院。

**108** 数据治理——工业企业数字化转型之道

**9.2** **主数据标准管理**

主数据标准管理的目标是通过统一的标准制定和发布，结合制度约束、系统控制等手段， 为主数据的唯一性、完整性、有效性、 一致性、规范性管理提供支持和保障。

主数据标准管理是主数据管理的重要内容，也是主数据全生命周期管理、主数据质量管理 和主数据应用管理的重要基础。成功实施主数据标准管理，是主数据管理的首要要求。

主数据标准包含业务标准(编码规则、分类规则、描述规则等)、主数据模型标准。主数据 标准体系在建设梳理的过程中， 一般会衍生出一套代码体系表(或称主数据资产目录)。

**1.** **主数据业务标准**

主数据业务标准是对主数据业务含义的统一解释及要求，包括主数据来源、主数据的管理 级次、统一管理的基础数据项、数据项在相关业务环境中产生过程的描述及含义解释、数据之 间的制约关系、数据产生过程中所要遵循的业务规则。

主数据业务规则包含主数据各数据项的编码规范、分类规则、描述规则等。

●编码规则：主数据代码的编码规则。例如，物料代码采用以“1”开头的8位无含义数字 流水码。

●分类规则：依据相关业务环境和管理需求形成分类规则。例如，根据物料的自然属性及 所包括范围的大小，可以将物料分为大、中、小三类。

●描述规则：又被称为命名规范。例如，物料描述规则包括具体物料描述规则的定义，主 要解决物料描述的规范化问题。

**2.** **主数据模型标准**

主数据模型标准包含主数据逻辑模型和主数据物理模型。

●主数据逻辑模型：将高级的业务概念以主数据实体/属性及其关系的形态在逻辑层面上更 详细地表达出来，主要表现形式是 ERD(实体关系图)。

●主数据物理模型：又被称为主数据的存储结构表。业务在应用环境中对数据的统一技术 要求包括数据长度、数据类型、数据格式、数据的缺省值、可否为空的定义、索引、约 束关系等，以保证数据模型中设计的结果能够真正落地到某个具体的数据库中，并提供 了系统初始设计所需要的基础元素，以及相关元素之间的关系。

第9章 主数据管理 **109**

**3.** **主数据代码体系表**

主数据代码体系表(在某些领域内，它又被称为主数据资产目录):是描述企事业单位信息 化建设过程中所使用的主数据代码种类、各类主数据代码名称、代码属性(分类、明细、规则 等)、采(参)标号及代码建设情况的汇总表。它是企业主数据代码查询和应用的依据，同时也 是主数据代码的全局性和指导性文件。主数据代码体系表主要结合了企业的经营管理特点，服 务于企业信息化建设，主要包括两部分内容：第一是企业信息代码体系表的框架结构及分类， 第二是所有分类下的信息代码标准明细及建设情况。

**9.3** **主数据全生命周期管理**

主数据全生命周期管理是指：采用必要的管理工具，依据管理职责，按照规范的流程对主 数据生命周期各环节实施管理行为。它是确保主数据质量的重要手段。主数据全生命周期管理 不但对主数据系统的数据进行管理，而且对所有承担主数据管理的系统进行管理，包括代管主 数据的业务系统(如HR 系统)、主数据采集和预处理系统等，这样才能确保主数据全生命周期 管理的完整性，不留死角。根据企业自身的情况和特点，建立符合企业实际应用情况的主数据 全生命周期管理流程，是保证主数据标准规范得到有效执行，实现主数据的持续性长效治理的 关键。

主数据全生命周期管理通常包括申请、校验、审核、创建、发布、变更、冻结、归档等。 由于不同的主数据具有各自的情况和特点，各类主数据在企业内的全生命周期管理流程都不尽 相同，因此需要对每类主数据按照不同的流程进行全生命周期管理的设计，以实现主数据管理 最合理的分工和协同。通常，主数据全生命周期管理有以下几种模式(如图9-3-1所示)。

**1.** **集中管理模式**

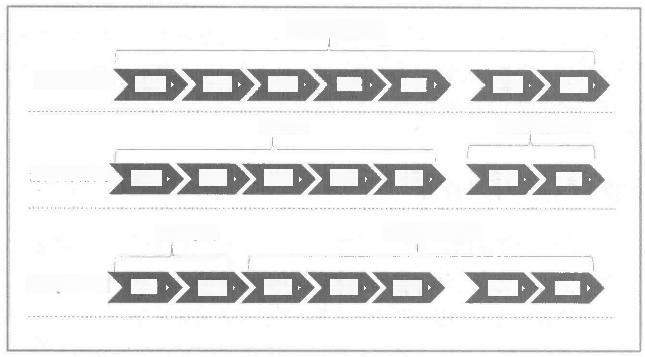
集中管理模式就是主数据全生命周期管理都在主数据管理系统内完成，不依赖其他业务系 统的支持和协同。这种模式一般针对不需要其他业务数据支持的主数据，如物料代码主数据。

**2.** **源头托管模式**

源头托管模式是指将主数据全生命周期管理全部委托给业务系统的管理模式。这种模式适 用于采用业务数据作为参考数据的主数据，如 HR 的员工主数据和内部组织主数据。源头托管 模式又分为单源托管和多源托管，单源托管是指主数据来源于单个系统，多源托管是指主数据 来源于多个系统。单源托管主数据直接接入即可，多源托管在接入时做数据整合。

**3.** **协同共管模式**

协同共管模式是指源头业务系统和主数据系统按照协同规则对主数据全生命周期共同管理 的模式。这种模式适用于源头系统参与部分管理的主数据，如客户主数据。



**主数据管理系统**

申请 校验 审核 创建 变更 发布 归档

HR 系统 **主数据管理系统**

申请 校验 审核 创建 变更 发布 归档

**CRM系统** **主数据管理系统**

**申请** **校验** 审核 创建 **变更** **发布** **归档**

**源头托管模式”**

**集中管理模式”**

**协同共管模式**

**图9-3-1** **主数据全生命周期管理模式**

**9.4** **主数据应用管理**

主数据是被业务系统广泛共享、重复使用的标准化基准数据。主数据管理只有被广泛而准 确地使用，才能解决企业内部各业务系统之间数据不一致的问题，从而体现主数据管理价值。 因此，主数据应用管理是企业主数据管理的重要内容，也是主数据价值管理的核心。

主数据应用管理就是：利用各种技术和管理手段对主数据共享应用进行有效管理，确保主 数据被准确、便捷地使用；在有效满足各业务系统主数据共享需求的同时，确保主数据在应用 过程中的隐私保护和数据安全。

**1.** **统一源头集中共享**

企业实施主数据管理就是要解决数据一致性的问题。能否确保“一致性”贯彻始终，是决 定主数据管理成败的关键。通常，规模较大的企业的内部业务系统都比较多且数据关系复杂， 如果不对主数据共享数据源进行统一管理，则多点引用会带来诸多数据不一致问题，包括数据 代码不一致、数据完整性不一致、数据时效性不一致和数据状态不一致等。采用统一源头集中 共享方式，可以有效避免这些问题，为主数据在应用环节的一致性提供保障。

第9章 主数据管理 **111**

要实现统一源头集中共享，需要建立统一的主数据代码库，将集中管理和分散管理的各类 主数据代码完整汇集到统一主数据代码库中，并以此为基础构建统一主数据共享资源视图，部 署统一的主数据共享平台。所有业务系统都通过统一的共享平台来获取主数据资源，真正实现 主数据同源共享，消除主数据在应用过程中出现不一致的情况，确保主数据的一致性。同时， 通过对业务系统数据集成管理、主数据应用监控等手段，可以加强对主数据统一源头集中共享 的管理，注重实效。

**2.** **主数据应用需求管理**

企业的主数据应用要靠业务需求来驱动，没有需求就没有应用。通过一套标准的制度和流 程对来自不同业务和不同系统的需求进行管理，可以有效识别需求，确定满足需求的主数据类 型、对接方式和数据共享机制等，为主数据应用提供可行的实施方案，以及对主数据应用过程 中的需求变化进行管控，合理驱动主数据的应用方式和策略不断改进。

主数据管理是标准化的，但主数据应用需求往往是个性化的，如何解决标准化和个性化的 矛盾，使主数据在应用中最大限度地满足业务需求，充分发挥主数据价值，是企业主数据应用 需求管理必须解决的问题。通过主数据应用需求管理识别个性化需求的合理性，消除不合理的 差异性需求，对合理需求进行抽象和归并，形成统一标准，可以很好地化解主数据标准化和应 用需求个性化的矛盾，避免主数据管理与主数据应用出现“两张皮”现象。

**3.** **主数据应用服务管理**

不同业务、不同系统对主数据应用的方式会有所不同，比如 CRM 系统需要存储客户主数 据以支持销售管理、信用管理和客户服务管理等业务需求，而订单管理系统一般只是用主数据 核实客户数据，不需要将主数据保留在本地使用。为满足不同应用场景和需求，提高主数据共 享服务的运行效率，主数据应用集成平台一般会部署多个应用服务，主数据应用服务管理可以 保证为每个业务和系统匹配最合适的主数据应用服务。

根据企业主数据应用常见的业务场景和需求，主数据应用服务一般需要提供以下几种方式。

(1)主数据查询服务。

主数据查询服务是为用户提供主数据在线查询的常用方式，用户通过检索条件查询到需要

查看的主数据。

**112** 数据治理——工业企业数字化转型之道

(2)主数据调用服务。

主数据调用是为业务系统实时调用主数据的服务方式， 一般由业务人员直接调用主数据

(3)主数据同步服务。

主数据同步服务是为了确保业务系统的主数据副本与主数据代码库实现数据同步，确保数 据一致性，通常采用订阅/分发的方式。

(4)主数据即时服务。

主数据即时服务是为各个业务系统提供主数据智能检索(模糊、精准等)和即时调用的服 务方式， 一般支持主数据“只用不存”的业务需求，减少业务系统的数据冗余，保证业务数据 的高一致性、高可靠性。

(5)主数据资源服务。

主数据资源服务是利用数据资源提供的数据只读服务，通过数据库视图等方式，满足业务 系统、主数据盘点等应用，也便于BI 等数据分析系统大批量使用主数据，以弥补在线服务难以 支持大批量数据操作的不足。

**4.** **主数据应用质量管理**

实施主数据应用质量管理就是建立一套机制，从数据一致性、有效性、完整性、合规性等 方面进行审计，检查业务系统的主数据应用在质量方面存在的问题，督促对问题的整改，并对 存在的问题和整改情况进行有效的管理和监控，确保主数据被业务系统正确使用。

主数据应用质量管理一般分事前管理、事中管理和事后管理3个阶段：事前管理就是在业 务没有发生之前对业务系统的主数据应用环境、主数据副本等是否存在影响主数据质量的情况 进行管理；事中管理是指在业务操作过程中对业务数据引用的主数据是否存在质量问题进行管 理；事后管理就是在业务操作完成后对业务数据是否存在主数据应用质量问题进行管理。

**9.5** **企业常用的几类主数据**

**9.5.1** **物料主数据**

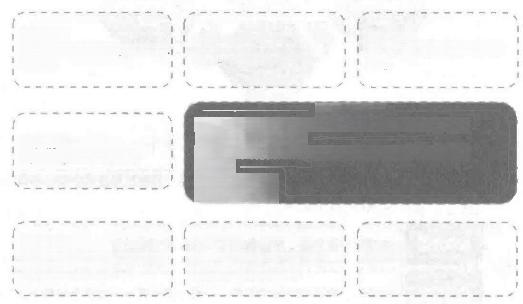
物料数据标准化是加强物料分类管理、提高物资采购管理水平、实现信息共享的基础工 作。物料数据由物料主数据和业务属性数据组成，物料主数据是物料数据核心的属性，包含

第9章 主数据管理 **113**

物料分类编码、物料明细编码等，而业务属性数据体现各业务部门在各业务环节上对物料管理

的需求。

物料是指包括与产品生产有关的所有物品，如原材料、辅料、成品等。对于信息化系统而 言，物料特指企业中的物资、产品和服务的总和。物料按用途分成物资类物料、产品类物料(成 品、中间产品和废次品)和服务类物料。ERP 系统将同一类物料业务的相关参数划分到不同的 视图(界面)中，如物料基本视图、采购视图、销售视图、生产视图和财务视图，物料基本视 图中的信息包括物料编码、物料名称及规格型号、基本计量单位，如图9-5-1所示。



生产视图

生产版本、主产品、关联产品、

权数等

物用基本视图

长描述：热轧普通碳素H型 钢VH200x200x8x12mm Q2358 GB/T 11263

短描述：热轧普通碳素H型钢VH200x200x8x12m m

02358

物料分染：钢材

计量单位：平克

采购视盟

所采购组、MRP 类型、采

购单位、采购单位等

销售视图

可用性检测、利润中心、销 售单位、税分类等

财务视图

评估类、价格控制、ML 作 业等

库存视圈

所处仓库、ABC 分类等

工厂视图。

业务视图的参数信息等

质检视围

图9-5-1 物料基本视图

**9.5.2** **设备主数据**

设备指可供企业在生产中长期使用，并在反复使用中基本保持原有实物形态和功能的劳动

资料和物质资料的总称。

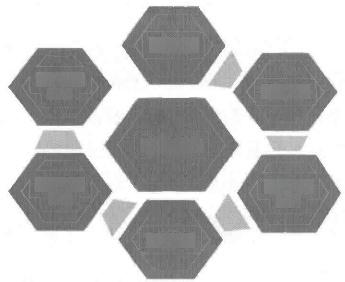
设备主数据包含的信息有基本信息、位置信息、组织机构信息、结构信息、技术参数信息 等相关信息，如图9-5-2所示。

设备主数据是设备管理系统的基础。为确保设备管理系统的有效运行，建设设备主数据管 理系统非常重要。设备主数据管理系统通过企业设备主数据管理人员与外部 EPC (工程总承包) 合作伙伴及供应商的合作请求、企业数据收集人员的现场数据校对、用户在日常使用中对数据 问题的发现及警示，来保障设备管理系统所用的功能位置、设备、备品备件清单、维修任务清 单、测量点各类文档资料等数据收集的及时性、完整性和准确性；并通过将各类设备管理系统 进行集成，在建立和修改各类设备主数据的同时，触发这些系统中各类设备主数据的同步更新，

确保各系统中各类设备主数据的一致性，从而保证业务处理和分析结果的正确性。设备基础数

据是用来定义在企业设备管理中所用到的全部静态数据，主要包含以下8大类数据，如图9-5-3

所示。

基本信息

相关文档

位置信息

资料

**设备主数据**

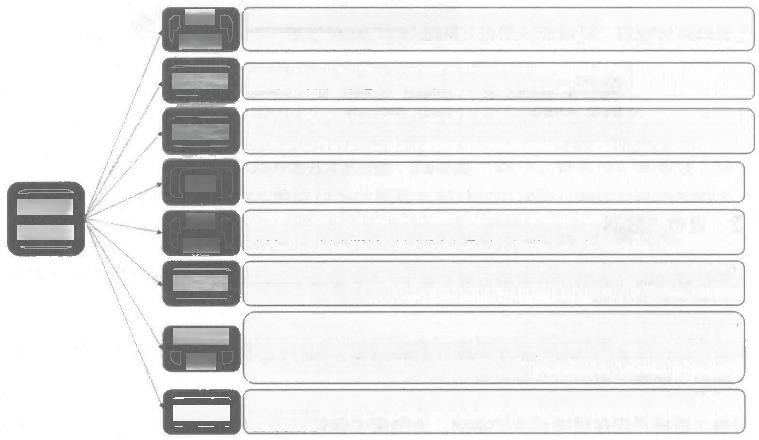
技术参数

组织机构 信息

信息

**结构信息**

**图9-5-2** **设** **备** **主** **数** **据**



设备是指可供企业在生产中长期使用，并在反复使用中保持原有实物形态和功能的劳

动资料和物质资料的总称

**主要指设备分类，是设备主数据的重要组成部分**

**功能位置是设备安装的位置，** **一般来说会有一个或多个设备安装在每个功能位置上**

**设备的技术参数**

**设备BOM** **即设备的备品备件清单，是设备档案的重要内容**

**BOM**

任务清单即标准设备检修作业程序，任务清单可以作为标准检修内容或预防性检

**任务清单**

修工单的参考内容

**按照设备的不同种类、分类、A/B/C** **类，在不同组织层级用树状展示设备数量，并**

**可穿透到设备清单，在设备清单中可以显示功能位置、设备、分类、特性(技术参**

**数)、特性值(参数值)和技术标示等信息**

**设备档案是设备全生命管理的集中体现，包括设备选型、监造、组件更换、**

**设备档案**

**更新报废等**

**设备**

**主数据**

**设备分类**

**功能位置**

设备基础 数据管理

设备专业 台账

**特性**

**设备**

**图9-5-3** **设** **备** **基** **础** **数** **据** **管** **理**

**9.5.3** **资产主数据**

企业固定资产是指企业为生产产品、提供劳务、出租或者经营管理而持有的、使用时间超 过12个月的、价值达到一定标准的非货币性资产，包括建筑物、机器、运输工具，以及其他与

生产经营活动有关的设备、器具、工具等。

第9章 主数据管理 **115**

固定资产主数据是企业资产管理的应用基础，固定资产编码包括固定资产分类编码和固定 资产明细编码。固定资产分类编码适用于固定资产的管理、清查、登记、统计等工作。固定资 产明细编码是对固定资产数据进行系统识别和检索的唯一标示。固定资产分类编码应包括企业 所拥有的全部固定资产，并以资产的自然属性为第一分类原则，兼顾管理要求与实用性。自然 属性是指资产的物理或者化学方面的属性，如经济用途、所有权、使用情况等。

固定资产分类还应适应固定资产全生命周期管理的需求，适应信息系统集成、整合、应用

一体化管理的要求，做到实用、适用、方便。在固定资产分类制定过程中，应遵循以下原则。

(1)执行、参照国家标准、国际标准或行业标准，做到有据可依；

(2)固定资产分类一经制定，不要随意调整；

(3)一个固定资产对应一个分类，避免不同审批人员的审批尺度不统一；

(4)同一大类下的中类或同一中类下的小类的划分维度要一致，便于使用人员理解和查找，

避免分类交叉和混淆；

(5)同一类别下的子类，要将使用频率、价值较高的物资放在较前面，便于查找。

2011年，我国发布了《固定资产分类与代码》(GB/T14885—2010)。 该标准规定了固定资 产的分类、代码及计量单位。企业在规划制定本企业适用的固定资产分类标准时，要根据新《企 业会计准则》和国家相关法规，在现行的《固定资产分类与代码》( GB/T 14885—2010) 的基 础之上，进行细化和补充调整。

**9.5.4** **财务主数据**

财务主数据主要包括会计科目主数据、银行主数据。

会计科目主数据是企业财务核算与财务预算的应用基础。国家一级会计科目包括资产类、 负债类、所有者权益类、成本类和损益类，且国家已对一级会计科目的编码进行了统一编制。 企业在对会计科目主数据进行编码管理时，应在执行国家统一的一级会计科目的基础上，制定 满足本企业经营需要的会计明细科目。会计科目主数据应采用审批制进行管理，其编制标准主 要包括编码规则和属性信息两部分内容。

银行主数据包括国内银行和金融机构，以及在境内开展业务的国外银行和金融机构。银行 主数据编码由编码规则和属性信息两部分组成。

**116** 数据治理——工业企业数字化转型之道

**9.5.5** **组织机构和员工主数据**

组织机构主数据应用于企业的生产经营管理、人事管理、财务管理，以及企业的决策支持 分析。组织机构主数据的统一管理，为不同部门、不同业务、不同信息系统提供了沟通的“相 同语言”,确保企业组织机构主数据的一致性、正确性和共享性，能够提升企业管理水平和信息 系统质量，降低信息传递与使用成本。企业组织机构主数据的管理分为组织机构分类编码管理 和明细编码管理。

企业员工应包括企业所有用工人员，如合同工、临时工，以及代训工和实习生等。员工主 数据包含员工的基本信息，应结合员工行政隶属单位与组织机构主数据使用。

组织机构和员工主数据的统一管理，直接为企业人力资源组织管理、招聘管理、人事管理、 薪酬管理、绩效考核、培训管理和统计分析奠定数据基础。同时，也为企业应用信息系统、信 息安全项目、信息系统集成建设提供统一、标准化的组织机构数据和员工数据。

**本章精要**

主数据是满足跨部门业务协同需要、反映核心业务实体状态属性的基础信息，是企业的核 心数据资产。管好、用好主数据是实施企业数据治理的重要内容。本章针对工业企业主数据管 理特点从4个方面进行了阐述：主数据的概念和基本特征；主数据标准管理是有效实施主数据 管理的基础；主数据全生命周期管理是主数据管理的重要保障；主数据应用管理是体现主数据 管理价值的重要手段。此外，为了便于读者更好地认识和理解主数据，本章还对工业企业中常 见的几类主数据进行了详细讲解。

第10章

元数据管理

随着工业企业的数字化转型，物联网(IoT) 设备与数字化平台的普遍应用，工业企业每年 收集和使用的数据成倍增长，多数工业企业的数据形态多样，标准不统一，多种数据源之间的 采集、传播和共享遇到了瓶颈。元数据管理作为大数据治理的核心基础设施，是有效管理这些 海量数据的基础和前提，在工业企业数字化转型中发挥着重要的作用，并日益得到工业企业的 重点关注。同时，以元数据为核心的大数据治理也被企业广泛认可和实施部署。

**10.1** **元数据的定义**

元数据最常见的定义是“关于数据的数据”,非常简单也容易引起误解。元数据的信息范围 很广，它不仅包括技术和业务流程、数据规则和约束，还包括逻辑数据结构与物理数据结构等。 它描述了数据本身(如数据库、数据元素、数据模型)、数据表示的概念(如业务流程、应用系 统、软件代码、技术基础设施)、数据与概念之间的联系(关系)。

元数据可以帮助企业理解其自身的数据、系统和流程，也可以帮助用户评估数据质量。对 数据库与其他应用程序的管理来说元数据是不可或缺的。元数据有助于处理、维护、集成、保 护、审计和治理其他数据。

**10.2** **元数据分类**

通常我们把元数据分为3种类型：①业务元数据；②技术元数据；③操作元数据。

业务元数据主要是数据中业务语义的描述；技术元数据主要关注技术的实现；操作元数据

**118** 数据治理——工业企业数字化转型之道

主要关注数据处理的过程与关系。这3种元数据的结合能够很好地表述企业信息的全貌。如在 企业数据中，业务含义为“客户名称”的数据，是业务元数据；存储客户信息的数据库表字段 为 “CustomerName”, 类型是 String, 是技术元数据；将该表通过 ETL 过程从客户关系管理系 统中抽取而来，并经过去重处理，这是操作元数据。

**10.2.1** **业务元数据**

业务元数据包括业务名称、业务定义、业务描述等。

业务人员更多关注的是与客户、结算日期、销售金额等相关的内容，这些内容很难从技术 元数据中体现出来。

业务元数据使用业务名称、定义、描述等信息表示企业环境中的各种属性和概念。从一定 程度上讲，所有数据背后的业务上下文都可以被看成是业务元数据。与技术元数据相比，业务 元数据能让用户更好地理解和使用企业环境中的数据，比如用户通过查看业务元数据就可以清 晰地理解各指标的含义、指标的计算方法等信息。

业务元数据广泛地存在于企业环境中，业务元数据的主要来源如下。

**1.ERP 系统**

企业的 ERP系统中存储着大量的业务元数据，比如财务计算公式、过程逻辑、业务规则等。

**2.** **报表**

报表的表头也是一种业务元数据，特别是那些包含合计、平均数等带有总结性质的列，以

及报表中的一些计算公式等。

**3.** **表格**

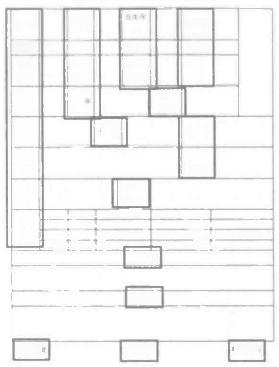
与报表类似， Excel的表头和公式也是很重要的业务元数据。与报表不同的是，大多数表 格中会有单独一列“描述”,有些表格中还会有一列代码和代码描述，这些都是很有用的业务 元数据。

4. 文件

文件中到处都是业务元数据，比如标题、作者、修改时间等，文件内容中的业务元数据的 获取相对比较困难，涉及机器学习等技术。

…

中



申请表名称

m

民

4

应

专

带断



烟在



PB



mm

\*



工作期



费 。

新员工入职中消表



\* 读



\*时间。



城压



身登



月

■

**各种申请表**

第10章 元数据管理 **119**

**5.Bl** **工具**

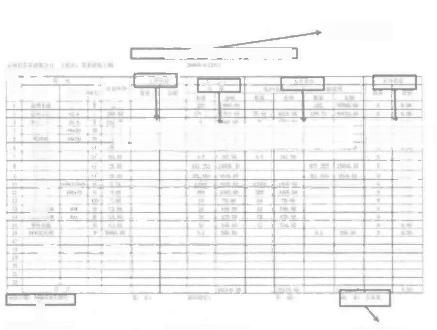
在 BI 工具中经常会用到的操作就是“钻取”,在向上和向下钻取的过程中通常定义了企业 的各种分类结构，例如产品等级和组织结构等级等，这些都是很重要的业务元数据。

**6.** **数据仓库**

数据仓库中也有业务元数据，比如，在构建数据仓库之前通常需要做大量的调研来研究如 何集成多个数据源，这些与数据仓库构建过程相关的文件中存在着大量的业务元数据。

下面举两个具体的例子。

如图10-2-1所示，左图是某建筑公司中的一张报表，我们可以看到，报表中包含了报表名 称、填报时间、制表人和报表表头名称等，这些都是高价值的业务元数据；右图是某公司新员 工入职申请表，和报表类似，申请表中的各栏都是业务元数据。



报表名称

材收发有且招



-  



上月结存本月收入：本月发出



二





T

填报时间 **企业报表** 制表人



本月结存

**图10-2-1** **某建筑公司报表和新员工入职申请表**

与技术元数据相比，业务元数据来源更复杂，分散在企业中的方方面面，为实现业务元数 据的管理，企业需要有效的方法和手段。

**10.2.2** **技术元数据**

技术元数据描述有关数据的技术细节、存储数据的系统，以及在系统内和系统之间数据流

**120** 数据治理——工业企业数字化转型之道

转过程的信息。以下为技术元数据的示例：

(1)物理数据库表名和字段名；

(2)字段属性；

(3)数据库对象属性；

(4)访问权限；

(5)物理数据模型，包括数据表名、键和索引。

**10.2.3** **操作元数据**

操作元数据描述了处理和访问数据的细节，例如以下内容：

(1)批处理程序的作业执行日志；

(2)抽取历史和结果；

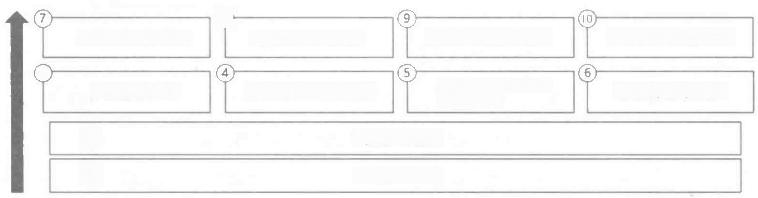
(3)调度异常处理；

(4)审计、平衡、控制测量的结果；

(5)错误日志。

**10.3** **元数据核心能力**

认识元数据架构后，还需要能够使用元数据。元数据管理工具基础能力如图10-3-1所示。



⑧-

元数据变更控制能力

3)

元数据查找能力

②

①

元数据存储能力

元数据采集能力

元数据 对比分析能力 数据生命周期管理能力

基于角色的访问控制 和分层能力”

血缘分析/关系分析能力

与其他系统的集成能力

业务元数据管理能力

**图10-3-1** **元数据管理工具基础能力**

**1.** **元数据采集能力**

元数据采集能力是指从错综复杂的企业环境中自动实时解析和采集各种元数据的能力，为

第10章 元数据管理 **121**

应对各种数据环境，这个环节通常需要使用各种技术和语法来支持大数据平台、关系型数据库、 第三方工具、存储过程、脚本、文本文件、表格文件中的自动化数据采集。

**2.** **元数据存储能力**

元数据存储能力是将采集到的元数据进行统一存储的能力。为支持各种元数据，以及元数 据之间关系的存储，元数据存储需要灵活、可扩展的架构支撑，另外，能够实时更新存储也是 很重要的一点。

**3.** **元数据查找能力**

元数据查找能力是一种提供统一的端口对元数据进行查找的能力。完善的元数据管理工具 应该能支持按照企业的各种分类方法来对元数据进行查找(有一部分分类方式蕴含在元数据本 身中，需要通过对元数据进行分析之后才能获得)。比如你可能会按照系统、表、指标、接口等 不同维度查找信息，甚至会根据自己的查找习惯新建一个完全不同的类别。

**4.** **血缘分析/关系分析能力**

血缘分析/关系分析能力是指分析数据的来源和数据的流向，揭示数据的上下游关系的能力。 在元数据管理工具中可以分析、描述并可视化其中的细节，方便用户对关键信息进行跟踪。完 善的血缘分析需要是横向(当前)和纵向(历史)皆可用的，以方便对同一时期的不同对象进 行分析和不同时期的同一对象的变化进行分析。

**5.** **基于角色的访问权限控制和分层能力**

元数据的增、删、改等权限的控制是元数据管理工具中需要特别注意的地方。元数据管理 工具应该支持访问权限控制，比如，数据管理员具有所有权限，开发人员可能更关注开发环境、 测试环境元数据，而企业管理者可能只关注生产环境的元数据，总经理级别的用户需要访问企 业多种环境下的元数据，而部门负责人级别的用户可能只关注与本部门相关的元数据。

**6.** **业务元数据管理能力**

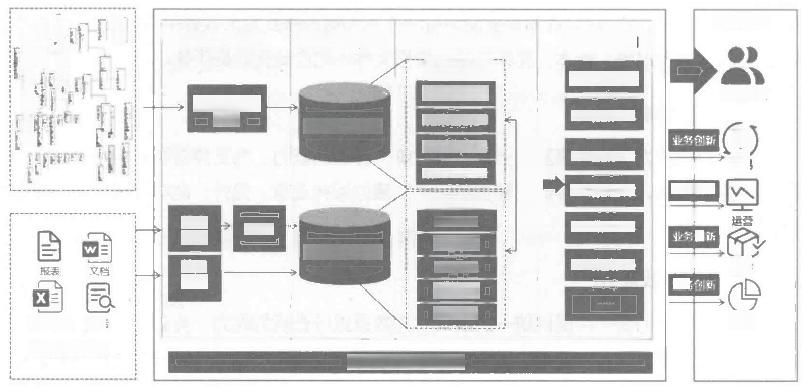
采集企业环境中的业务元数据，并完成业务元数据与技术元数据的映射，为元数据赋予业 务属性，这也是发挥元数据管理工具业务价值的一个关键，如图10-3-2所示。

心

…

**数据生命周期**

**122** 数据治理——工业企业数字化转型之道



面向用户

**业务自助应用**

360数据展示

企业员工

自助数据查战

**技术元数据库**

自助数据整合

数据历史记录

自助数基载取

业务元数据

业务含义

创

业务术语

数据标准

业务

据标含义

应量方法

**非结构化数据**

**自动化的数据生产线**

技术元数据

数据库表结构

数据转换规则

用户访问权限

自助数据共享

自助数据转换

自 黏 猷



技术元数据 采集

业务与技术的映射

**企业元数据存储**

数据资产采集

**业务元数据库**

电子表格 报告

接 口 采集

信息 采集

羊 编

业务创断

**NUP**

分析

生产

!

图10-3-2 技术元数据与业务元数据的映射

**7.** **元数据变更控制能力**

当元数据需要变更时，元数据管理工具提供了变更控制能力，明确元数据版本，保存元数 据的历史状态，在发生任何问题时可以自动恢复到之前的版本。在某个元数据项发生变更时， 可能还需要对该次变更将要产生的影响进行分析和评估。

**8.** **元数据对比分析能力**

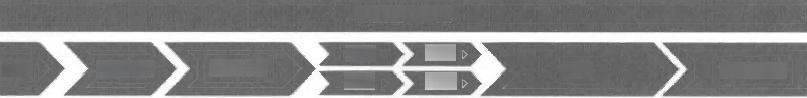
无数据管理工具还可以对不同环境中的元数据进行对比分析，分析其中的异同，必要时还

能根据分析结果产生相应的分析报告。

**9.** **数据生命周期管理能力**

在理想状态下，元数据管理工具应该保留数据“创建→存储→过时”这个过程中被删除/ 备份等各种状态下的元数据，从而管理数据在整个生命周期中的流动，如图10-3-3所示。作为 一 项规则，较新的数据和那些很可能被更加频繁访问的数据，应该存储在容易被访问的位置；

而那些不是很重要的数据，则可以备份存储在比较便宜的、慢 一 些的媒介中。



维护 集成

创建 获取

归档&清除

开发实施

标准规范

使用

规划

图10-3-3 数据生命周期

第10章 元数据管理 **123**

**10.** **与其他系统的集成能力**

要想让元数据管理系统发挥业务价值，还有非常重要的一点就是元数据管理工具与其他系 统的集成能力。

**10.4** **元数据的价值**

为了理解元数据在数据管理中的重要作用，可以试想一个大型图书馆中有成千上万册的图 书和杂志，但是没有目录卡片。如果没有目录卡片，那么读者将不知道如何寻找一本特定的图 书甚至是一个特定的主题。

目录卡片提供了必要的信息(图书馆拥有哪些图书和资料，以及它们被存放在哪里),这使 读者可以使用不同的方式(主题领域、作者或者书名)来查找资料。如果没有目录卡片，那么 寻找一本特定的书将是一件不可能的事情。 一家企业没有元数据与一个图书馆没有目录卡片是 一样的。

那么工业企业的元数据管理究竟能给企业带来哪些好处?下面举3个例子来说明。

**1.** **自动采集企业元数据，全面梳理企业信息资产**

企业数据通常呈现碎片化分布， 一共有多少个系统，各个系统之间有什么关系，系统中都 有哪些表，哪些表是孤立的可以删掉，是很多企业很难直接弄清楚的问题。

合适的元数据管理工具可以通过自动化的方式，帮助企业完成数据信息、服务信息与业务 信息的采集，自动化抽取企业内部所有元数据，为企业展现完整的信息资产视图，从而帮助企 业集中管理所有信息资产，在方便数据交互和共享的同时很好地解决了上述一系列问题。

**2.** **分析数据流向，迅速响应业务数据问题**

在企业中，往往会遇到这样的问题：业务人员发现分析报表中的数据有问题，要求IT 部门 尽快修改。但由于数据加工链路很长，此次修改会涉及多个部门，甚至多个下属公司，其中涉 及的技术手段各式各样，所以很难定位到该问题数据的相关表和字段。

而元数据管理可以帮助企业分析数据流向，具体到字段级的数据解析可以帮助企业分析数 据之间的上下游关系，通过可视化的方式可以展现数据上下游的关系图，快速定位问题字段， 帮助企业降低数据问题定位的难度。

**124** 数据治理——工业企业数字化转型之道

**3.** **通过多场景对比分析，消除系统上线隐患**

通常，企业系统建设会分为开发、测试与生产3种场景。在开发过程中，往往会出现开发 库、测试库已经测试通过，而在上线过程中又出现问题的情况。

若通过元数据管理系统进行上线变更，自动采集并管理这3种场景中的元数据，保证各个 场景中元数据的及时性和准确性，对比上线场景与测试场景的元数据，分析上线系统对其他系 统的影响，就能够避免此类问题发生。

**本章精要**

元数据管理是数据治理的重要抓手，也是数据治理成果呈现的最佳工具，企业若能做好元 数据管理，就可以解决数据查找难、理解难的问题。元数据是数据的说明书，有了完善的元数 据，数据使用者才能了解企业都有什么数据，它们分布在哪里，数据的业务含义是什么,数据 口径及颗粒度是怎样的，若想使用应该向谁提出和如何获取。要达到这样的目标，需要做好元 数据采集、存储、变更控制和版本管理。在此基础上，实现数据血缘分析、关系分析、影响分 析等元数据的高级应用，通过可视化的方式展现数据上下游关系图，快速定位问题字段，可以 帮助企业降低数据问题定位的难度。

第11章

数据指标管理

对企业数据治理来说，实现数据指标标准化、统一数据指标标准，可以规范企业业务统计 分析语言，帮助企业提升分析应用和数据质量，进而提高企业数据质量和数据资产价值。

**11.1** **数据指标管理概述**

**11.1.1** **数据指标应用和管理中的挑战**

企业在数据指标应用方面经常存在数据不一致、不规范等问题。对集团公司而言，还存在 下属公司数据重复填报等问题，具体介绍如下。

**1.** **指标口径不统一**

在企业管理和业务活动中，经常存在数据指标的名称相同，但统计口径、计算方法却有较 大差异的情形；或者反过来，数据指标的计算方式相同，但名称却各异。定义不统一的指标会 带来极大的沟通障碍，让沟通效率降低，甚至“差之毫厘，谬以千里”。

**2.** **指标体系不完整**

企业各部门根据自身业务需求，都有一部分的量化指标，但不够全面，也缺乏方法论指导。 这不利于提升企业整体数据分析和应用能力，且在使用过程中会孤立地强调某些指标的重要性， 而忽略综合分析、长期跟踪与定期比对指标的重要性。缺乏整体考量而设置的指标体系，以及 错误的指标分析方法，会产生错误的分析结果，进而影响运营层面、产品改进方面的决策。

**126** 数据治理——工业企业数字化转型之道

**3.** **指标问题追溯难**

数据指标大多经过多重计算得到，有些指标需要经过很长的加工过程才能得出，如果无法 追溯指标的加工过程，就不知道指标所用的数据来源，也无法快速找出指标出错的原因和对应 的责任部门。指标的一致性、完整性、准确性和可追溯性得不到保证，出现问题时各部门间相 互推诿的情况时有发生，这些都导致了指标问题难以解决。

**11.1.2** **设计目的**

数据指标可以打通信息壁垒，加大信息共享力度。指标体系框架设计需从企业发展战略与 目标出发，要满足各职能部门业务管理需求，构建“横向协同、纵向贯通”的指标体系。

●横向协同：集团相关职能部门结合业务决策、分析、管理工作需要，持续更新、完善、 扩展指标体系框架内容，以满足集团化经营管理。

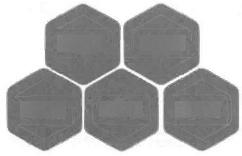
●纵向贯通：下属企业提供集团级经营管理所需的数据指标源，支撑集团级业务决策分析 和管理。同时，要结合自身业务特点，扩展和完善自身业务指标框架内容。

**11.1.3** **设计思路**

企业级的指标体系并不是一些指标的简单堆积和重组，而是基于业务管理职能的原则建立 的。指标体系框架设计得是否科学合理，很大程度上与指标体系框架设计的构建思路相关。在 构建指标体系框架时，应理顺指标体系构建的基本思路。

**1.** **指标体系框架设计原则**

指标体系框架设计是一项系统工程，需要完整构建评价指标体系，其基本设计原则包括5 个方面：系统性、全面性、结构性、差异性和重要性，如图11-1-1所示。



**系统性** **全面性**

**重要性** **差异性** **结构性**

**图11-1-1** **指标体系框架设计原则**

●系统性原则：充分承接整个企业的战略目标，形成以战略目标为核心的指标体系，指标 之间有清晰的关联逻辑，有效促进战略执行。

第11章 数据指标管理 **127**

●全面性原则：全面覆盖企业管理相关因素，含外部环境、核心资源、业务活动、产品服 务与经营业绩等方面，推动整体优化运营。

●结构性原则：以数据和信息为基础，将指标划分为战略层、管理层相互支撑的结构。

●差异性原则：区别于下属企业的指标体系，结合了企业的实际业务特点，可以有针对性

地建立落地的指标体系。

●重要性原则：选取核心业务，明确与之相关的关联业务，关注管理瓶颈与重点，推动集 团总部各部门之间的管理协同。

指标工作的目标和内容见表11-1-1。

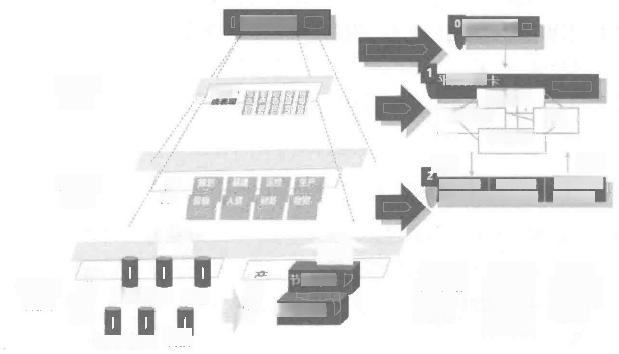
**表11-1-1** **指标工作的目标和内容**

|  |  |  |
| --- | --- | --- |
| **序** **号** | **建设内容** | **说** **明** |
| 1 | 建立指标体系 | 建立指标涉及的业务域划分、数据指标字典元数据、人员/组织/流程的体系 |
| 2 | 定义指标内容 | 整理指标的业务属性、技术属性和管理属性，构建指标关系体系和分析体系 |
| 3 | 建立指标规划 | 持续发展能够支撑横向业务领域划分、纵向加工计算逻辑分层的指标架构体系 |
| 4 | 制定相关规范 | 建立公司级指标搜集、整理、维护流程，为各类数据应用夯实坚实基础 |

**2.** **指标体系框架设计原理**

企业在遵循系统性、全面性、结构性、差异性和重要性原则的基础上，要以企业指标体系 框架设计相关理论为指导，从企业战略及运营管理的各层级考虑，构建支撑企业战略决策的指

标框架体系，以保障企业战略目标的实现为最终目标(指标体系要素框架如图11 - 1 - 2所示)。



战略目标

|  |  |
| --- | --- |
| 平衡计分  财务战略 客户战略  学习成长  *分解*  21 集团 22能 KP 的分解 流探维导 | 流程  承接  231R刺  盐相关者 |

|  |  |  |
| --- | --- | --- |
| “由上到 下”分解 承接战略 层指标 | “由下到 上”反映  各部门及  下属公司  业务指标 | “横向” 反映利益 相关者期 望要求 |

战略层

指标

管理层

指标

接

雷心 人力财务 市场

研发信息 科技

**集团战略**

市

能环保

严业金融

信息与 数据

晶服房

业 旁

**图11-1-2** **指标体系要素框架图**

**128** 数据治理——工业企业数字化转型之道

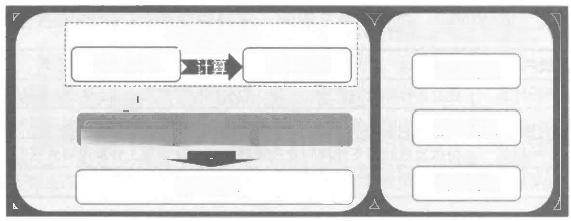
**11.2** **体系框架**

指标体系框架全面反映企业的整体情况，而不是局限于局部或某些具体方面。它是以企业 战略为核心，以各部门的工作职能为出发点，对核心业务逐层分解、分类梳理，并归纳核心和

关键业务环节而进行的数据指标分类选取的体系化设计。

**11.2.1** **典型的数据指标定义框架**

数据指标标准定义框架规范了数据指标标准的分类和属性，如图11-2-1所示。



指标数据标准属性 业务属性

十

技术属性

管理属性

基本类指标

基础指标 复合指标

维度

业务板块。业务阶段等)

十

(时间、组织

指标分类

派生指标

**图11-2-1** **数据指标标准定义框架**

**1.** **数据指标标准分类**

●基础指标是表达业务实体“原子量化”属性的概念集合，是直接对单一变量的明细数据 进行简单计算得到且不可进一步拆解的指标，如调运量、销售量等。

●复合指标是建立在基础指标之上，由若干个基础指标通过一定运算规则计算得到且在业 务角度无法拆解的指标，如签约率、利润率等。

●派生指标是基础指标或复合指标与一个或多个维度值相结合产生的指标，如月计划调运 量、月/日均销售量等。

**2.** **数据指标标准属性**

●业务属性是数据在业务层面的定义，描述了数据和企业业务相关联的特性，是数据业务 含义的统一解释及要求。业务属性包括指标名称、业务定义、业务规则、主题、 一级子 主题、二级子主题、参考标准、计量单位、维度、上报频率、数据层次、映射类型、处 理逻辑、统计时间。

●技术属性是数据在技术层面的定义，描述了数据与信息技术实现相关联的特性。技术属

第11章 数据指标管理 **129**

性包括数据类型、数据格式、数据源系统、数据源表名、数据源字段名。

●管理属性是数据在管理层面的定义，描述了数据标准与数据管理相关联的特性，是数据 管理在数据标准管理领域的统一要求。管理属性包括指标编码、数据提出者、数据使用 者、数据负责人、颁布日期、废止日期。

**11.2.2** **指标选取原则及方法**

指标选取需要遵从科学的选取原则，保证指标的正确性、完整性和有效性，同时应遵循数 据指标梳理的方法论。

**1.** **数据指标选取原则**

数据指标的选取在指标体系框架设计中至关重要。指标选取是否合适，直接影响到指标体 系框架的完整性和有效性。指标过多，事实上是指标重复，会导致干扰；指标过少，则可能选

取的指标缺乏代表性，产生片面性。所以，指标选取应当以如下原则为基准。

(1)正确性，每项指标都必须正确，能够科学地反映评价对象某一个方面的信息。

(2)唯一性，保证指标选取的全面，避免指标重复。

(3)完整性，指标的信息应避免缺项，保证内容完整。

**2.** **数据指标选取方法**

数据指标的选取方法包括以下3个步骤。

(1)借鉴行业最佳管理实践，综合考虑内外部环境因素及上级单位要求，建立企业指标体 系的总体框架，保证指标体系框架的完整性。

(2)通过对企业战略、管理层及下属企业业务的全面梳理，识别企业生产运作中关键和主 要业务环节中的要素，初步形成企业级指标体系框架。

(3)进一步筛选、合并同类指标，去除重复和不重要的指标，通过专家法调整指标，使指 标体系框架更加科学、合理。

**3.** **数据指标选取依据**

指标体系框架设计要结合行业发展趋势和高层关注点，形成基于战略引领、价值创造的指 标体系框架。图11-2-2所示为某大型集团指标体系高阶框架，包括增长引擎、运营卓越、财务 表现、创新驱动、人员保障、宏观环境和绿色发展7大核心主题指标框架体系。

**130** 数据治理——工业企业数字化转型之道



**某大型集团指标体系征架**

增长引摩 运营卓越 财务表现 创新驱动 人员保障 左观环填 绿色发展

管理创解

科技创新

资金支持

资产营断

利消贲献

业务板块运额

行业动态

外部环境

经明现状分析

可揭级发氟

人员现状

图11-2-2 某大型集团指标体系高阶框架

(1)增长引擎，关注内外部环境条件分析与战略目标选择、现有业务与新业务增长路径

的选择等。

(2)运营卓越，关注集团特色，按照价值链进行区分。核心的运营活动包括从产业链协同 层面进行专项分析，同时关注各板块运营的规模、效率、效益和安全，从工程基建到管理和服 务分别展开分析等。

(3)财务表现，关注利润贡献(价值成果一战略实现)、资产营运效率、偿债、资金保障(价

值创造)等。

(4)创新驱动，关注管理及制度创新、科技创新环境、方向、体系、投入及产出等。

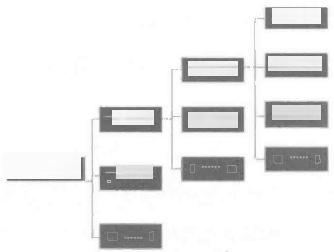
(5)人员保障，关注人员配置、人工效能、培训成长、分配激励等。

(6)宏观环境，关注宏观的政治经济要素、行业动态的分析及竞争策略等。

(7)绿色发展，关注经济增长与发展、环境可持续、能源获取和安全等。

**11.2.3** **指标体系层级设计**

指标体系具有递阶层次结构，通常被划分为三级，如图11-2-3所示。

三级指标

三极指标

二级指标

三级指标

一极指标

二级指标

体系积情际

**图11-2-3** **指标体系层级结构**

第 1 1 章 数据指标管理 **131**

(1)一级指标为战略决策和规划发展类，依据指标框架的第一级框架来确定。每项一级指

标下设多项二级指标。

(2)二级指标为集团管控类，依据指标框架的第二级框架来选取和设计。每项二级指标下 设立多项三级指标。

(3)三级指标为具体业务类，在二级指标框架下设计。

三级指标用于数据采集与测算，二级指标与一级指标为指数型指标。

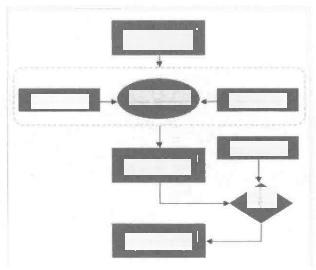
(注：指数型指标是指其数值是由下一级指标根据一定的算法计算而得到的指标，不直接用 于数据采集。)

**11.2.4** **指标体系评价方法**

指标体系评价是基于一定的评价目标，使用评价指标对特定的评价对象进行评价，并结合 专家知识对获得的测算结果进行研判，最终得出科学评价结果的一系列过程。

指标体系评价的目标是评价指标体系实施的正确性和有效性，以帮助其持续改进。

指标体系评价是针对评价对象的特点及其衡量水平，通过对数据指标的度量得到指标值， 形成评价结果，从而实现指标体系的评价目标，如图11-2-4所示。



评价目标

评价对象

专家知识

**测量结果**

研判

评价结果

评价指标

评价算法

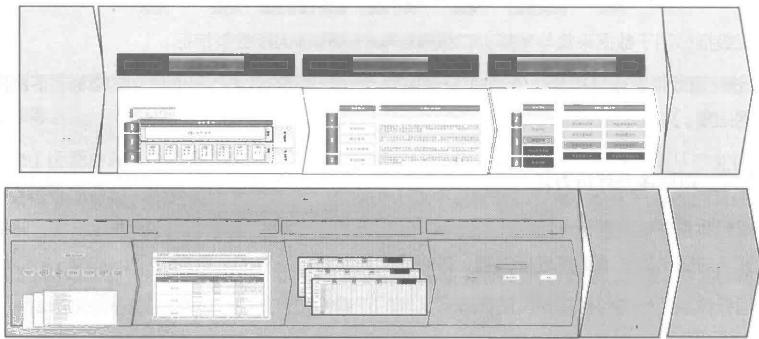
**图11-2-4** **指标体系评价方法**

在梳理指标体系的过程中，应结合企业层面的指标管理范围，明确各层级单位的指标项责 任，统一实现各指标项的定义、口径及取数来源，保证企业各级管理和决策层能够获得一致的 数据理解和数据来源，逐步完善企业指标体系。

**132** 数据治理——工业企业数字化转型之道

**11.3** **找指标**

找指标的工作主要采用自上而下与自下而上相结合的方式。其中指标分类应从业务管理需 求出发，自上而下逐层展开。而具体指标则以业务系统为导向，自下而上逐层筛选，如图11-3-1 所示。



**数据指标体系框架自上而下设计**

**识别业务环节**

数据指标体系自下而上梳理

整理指标维度框架 基本指标与维度映射 数据指标体系分析

如如应四心交

数指指标体 比 配

数 据 指 标 系 理

数整指标体 系 配

数 据 指 标 体 系 柜 架 o p

**定义业务分析主题**

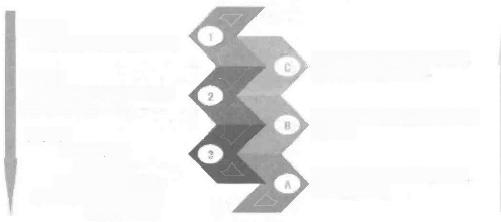
确定指标范围

自下而上废计

**确定业务领域**

图11-3-1 找指标示意图

下面是一个指标梳理过程的示例，如图11-3-2所示。



业务器域

(人力资源、财务、销售……)

业务功能

(销售-市场，销售-客户，销售- 订单，销售-商品……)

指标分类

(一级：销售；

二级：订单；

三级：订单执行绩效；

指标：已执行金额占销售达 成的比例……)

指标名称、接标定义，推标公式 计量单位。数据来源……

功能模块(订单管理、富户箭 理……)

信息系统 (ERP、CRM,PDM.

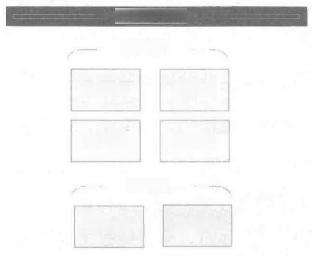
SRM 系统……)

图11-3-2 指标梳理示例

自上而下： 一般先将一级业务域划分为财务、人力资源、销售等，然后划分二级业务域，

如将销售划分为市场营销、客户管理、订单管理等(是否需要三级分类视业务复杂程度而定), 这样我们就得到了一个指标分类的初步框架。

自下而上：从业务系统出发，从系统中识别其功能模块，例如 ERP、CRM、SRM 等系统， 从这些模块包含的业务单据中可以找到相应的数值字段，然后挑选出可以作为指标展示的。如



**指标技术定义**

口径规石

**数据来源** **取数频率**

**取数规则** 加工规则

技术属性

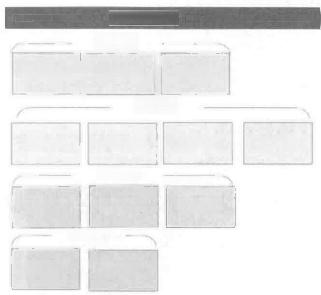
**属性字段** **数值限制**

第11章 数据指标管理 **133**

果这些系统中有单独的统计报表功能，那么我们的工作会相对简单，只需要从报表中找到相应

的指标即可。

此时只是找到了指标的名称，还需要找到指标的各项属性。指标的属性分为业务属性和技 术属性两类：业务属性包含业务人员所关心的指标分类、名称、计算公式、展现方式和查询权 限等；技术属性包含技术运维人员所关心的数据来源、取数字段、取数频率、加工规则等，如 图11-3-3所示。



指标业务定义

指标说明

指标分类 指标名称 计算逻班

分析方法

指标单位 分析凝度 分析维虚

属现方式

属现赖率 属现形式 属现层次

权限设计

所有权限 查询找服

下级指标

图11-3-3 指标属性识别

将找到的指标及其属性形成一张总表，即指标字典，如图11-3-4所示。具体字段的选取工 作最好能让业务人员和技术人员共同参与。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 业务定文 |  |  |  | 出 |  | 破频 | 单位 |  | 单位粒 |  | 超 | 留上  解国 | 时间  粒度 |  | 射  类里 | 计空  夏情 | 整薄  富基 |
| 1101 | 原油采 购途耗 量 | 指报告期内 原油自购入 到炼油企业 量点的实 接收量之 差异量的  月累计量 | 采购 | 板块果 购监控 | 原油 采购 | 炼油事 业部 | 规油论 业 | 炼油事 业部 | 吨/万 吨 | 数字 | 集团公 司/股份 公司/非 上市/地 区/分子 公司 | 2位 | 月 | 月底 | 月完成/年 累计 | 板块统 计量/ 企业上 报量 | 直接  映射 |  | 综合统计 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

图11-3-4 指标字典

**134** 数据治理——工业企业数字化转型之道

**11.4** **理指标**

搭建指标体系是一个系统工程，需要从战略及愿景分析、价值因素分析与 KPI (业绩考核 指标)设计、指标框架的设计、指标分析体系的设计、指标设计展现5个方面系统着手，具体 介绍如下。

(1)访谈高层管理者和各职能部门，理解战略愿景；

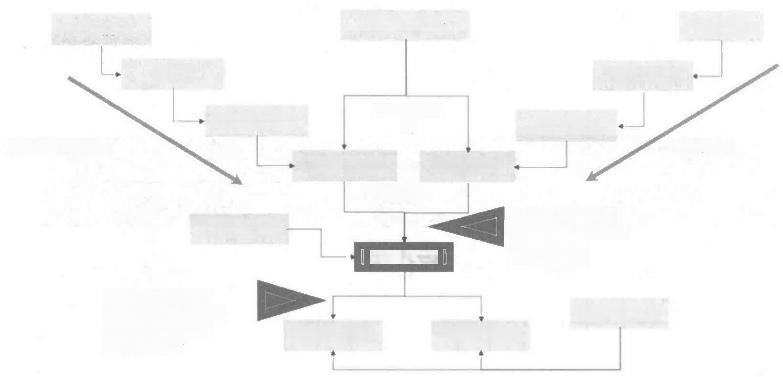
(2)逐步分解战略价值因素，并设计关键KPI 体系；

(3)基于业务能力模型设计指标框架，并实现KPI 指标的归集；

(4)设计符合企业管理决策需求的指标分析体系；

(5)将指标以图形化、多维化、互动性展示(指标设计)。

通过找指标(见图11-4-1),可以获得了企业中的指标列表。但是这些指标并不能被称为体 系，因为没有进行分类，指标没有业务含义，指标和指标之间也没有任何逻辑。此时还需要梳 理指标。在梳理指标时，需要结合企业战略和岗位职责，设定全面的衡量指标，并将指标分解 到具体的业务过程中，如图11-4-2所示。



理解战略

关键成功因素

引进最佳实

现有指标

使战略生效

指标长名单 指标长名单

二者统一

√ 是否可以测量?

√评估频度适当吗? √ 是否重复?

√ 是否平衡?

√层次是否适当?

理解组织结构

分解分析维度 业务主题分类

√是否管理层可控? √ 是否易于理解?

√ 是否平衡?

体现企业价值增长

优先价值动因

确定指标分类

行业指标参考

指标短名单

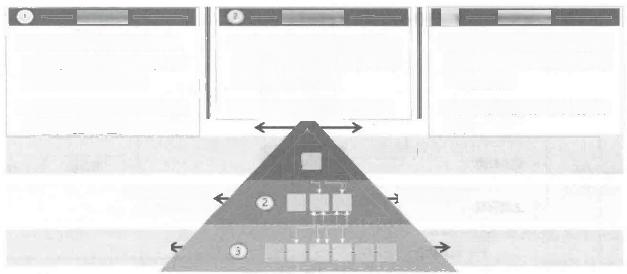
价值模型

价值树

践

图11-4-1 指标体系梳理最佳实践

第 1 1 章 数 据 指 标 管 理 **135**



企业级KPI

·和大的业务流程紧密联系，

通常用于高层管理者对企业

整体运营情况的监控

·关键的结果导向指标，直接

和整个企业的业务目标关联

高层管理者层面 企业织

中高层管理者层面。

-业旁板块

业务执行人员层西

业务人员

业务板块KPI

·和企业业务目标和业务流程 紧密联系，通常用于对企业 运营情况的监控

·可以向上汇总到企业级KPI

·和各作业区的目标紧密联系， 通常用于对生产流程和技术 工艺的监控

·关注运营层面和具体的操作 环节

0 运营层KP

>业务管控KPI

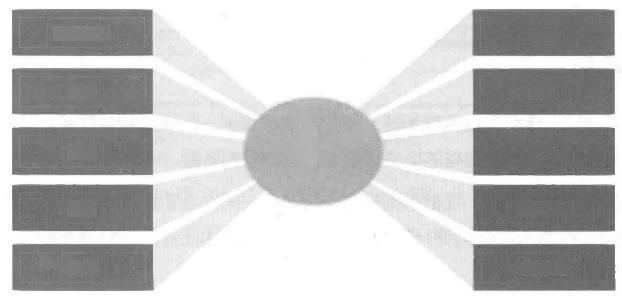
运营层KPI

企业级KPI

**图11** **-** **4** **-** **2** **全** **面** **衡** **量** **指** **标**

还需要将企业中的各业务职能域进行划分，定义业务主题域(如果是集团企业，那么还应

该进行行业板块的划分),如图11-4-3所示。



**产品规划**

**研** **发**

计 划

**采** **购**

**制造**

**主颗域分析**

**综合办公**

**人力资源**

**销** **售**

服 务

财 务

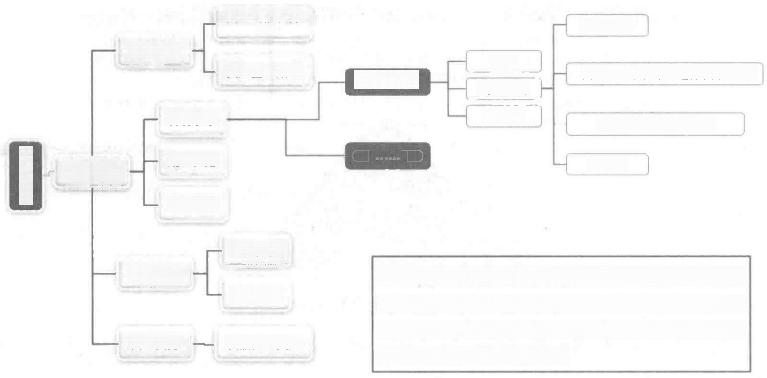
**图11** **-** **4** **-** **3** **定** **义** **业** **务** **主** **题** **域**

上面只是对业务域进行了大的划分，接下来企业需要对业务域的业务流程中的管控目标进 行细化，与指标一起形成功能矩阵，标注出指标的产生、引用关系，从而发现指标字典中存在 的重复统计、统计口径不同等各种情况。在对其进行处理后，将形成含有分类和业务目标的指

标字典。

指标字典可能会非常庞大，有时可以通过指标树的方式进行表示，每棵树表示一个业务域 的指标关系，通过链接跳转到相应的详细指标定义信息。图11-4-4是财务域的局部指标明细

示例。



营业成本

营业税金及附加金额及期间费用

成本费用

其他损益/营业外收支

所得税

**上缴利税**

**短期偿债**

**财务管理以公司盈利能力为核心，以杜邦分析框架为主**

**线，将利润指标逐层分解至资产营运、偿债保障等财务分**

**析层面，并进一步深入分析直至切入到业务运营，实现财**

**发展资金支持**

**务分析到运营提升的融会贯通。**

资产规模及构 成

资产营运效率

**利润水平**

**成本费用**

**利润贡献**

**偿债保障**

**资产营运**

**资金支持**

财 务 管 理

**长期偿债**

销售净利定

净利润

收入

图11-4-4 指 标 明 细 ( 示 例 )

**11.5** **管指标**

管理数据指标工作和其他数据管理工作并无太大差别，即通过推动数据治理体系建设，制

定主数据、业务数据、统计数据的标准和规范，提升数据质量。

类似基础数据标准管理，在数据指标标准管理中，同样需要数据治理归口管理部门来牵头 负责数据指标标准的制定工作，并建立业务归口部门与技术主管部门的协作机制。其中主要涉 及以下3种管理部门。

(1)数据治理归口管理部门：作为牵头者的角色，需要组织数据治理小组，将技术管理与 业务管理的相关人员协同起来，完成数据标准制定工作。为数据标准制定提供资源协调、统筹 安排等便利。

(2)指标标准业务归口管理部门：作为指标所属领域的业务主管部门，对数据指标标准进 行归口管理。对于业务涉及多个板块的指标，以指标产生部门对数据标准进行归口管理；对于 多个部门同时计算的指标，以业务牵头主管部门进行归口管理。其职责主要包括确定数据指标 的使用部门、基础属性、业务含义和业务口径等标准，并对指标的技术口径统计结果进行测试 和确认。

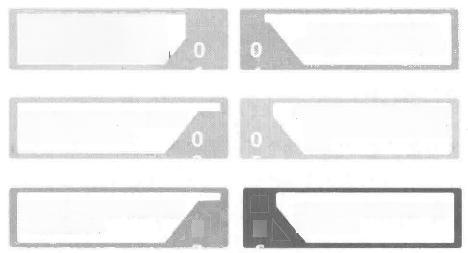
(3)技术主管部门：作为指标标准管理的技术支撑，对指标的取数方式和指标条件进行确 认，并统筹数据指标标准的落地实施工作。

第11章 数据指标管理 **137**

通过制定指标管控制度和流程，可以明确指标责任人、指标用户、指标管控团队在各项管 控活动中应遵循的管理要求和工作流程。在信息系统(尤其是数据中心、智能报表等分析类应 用)的建设中，各方需要遵从企业在指标管控领域提出的细化要求。

**11.6** **用指标**

企业仅仅制定指标体系并不能起到规范数据的作用，只有将指标体系落实到信息系统中才 能发挥其管理作用，因此，构建指标体系往往与应用系统建设同步进行。指标体系包含但不限 于以下应用场景，如图11-6-1所示。



**数据分析(商务智能/数据** **仓库/决策分析/大数据分** **析/移动分析/领导驾驶舱)**

报表或人工填报(统一集团和 子公司的统计格式)

战略和绩效管理(计划/预算 /绩效系统等)

数据服务(通过统一接口定 义，在不同系统中共享指 标 )

对标项目(审计项目、对标 考核指标)

数据管理(主数据管理、 指标管理、数据质量管理)

0

0

图11-6-1 数据指标体系应用场景

通过数据指标体系的应用，可以规范企业内指标的使用，提高数据的准确性、 一致性和可 追溯性。

**本章精要**

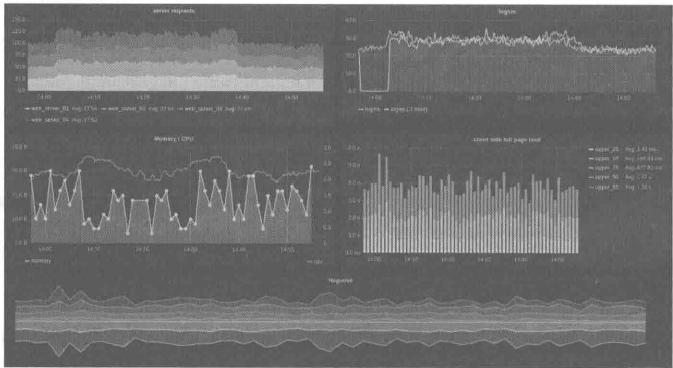
企业要实现经营精细化，需要强化关键指标管控，充分挖掘存量资产的价值潜力。企业要 提升数据价值，需要完善标准体系、统计体系、指标体系。细节决定成败，数据指标为企业精 细化管理提供了强有力的工具支撑。本章首先详细阐述了如何构建数据指标标准定义框架，规 范了数据指标标准的分类和属性；然后系统论述了指标体系的建立和应用步骤，即找指标、理 指标、管指标、用指标4步。

第 1 2 章

时序数据管理

**12.1** **时序数据管理概述**

工业企业为了监测设备、生产线及整个系统的运行状态，在各个关键点都配有传感器，采 集各种数据。这些数据是周期或准周期产生的，有的采集频率高，有的采集频率低。这些采集 来的数据一般被发送至服务器中进行汇总并实时处理，对系统的运行做出实时监测或预警。这 些数据也可以被长期保存下来，用以进行离线数据分析，比如分析故障，即分析主要的故障是 什么;分析产能，即分析如何进一步优化配置来提升生产效率；分析能耗，即分析如何降低生 产成本；分析潜在的安全隐患，以降低故障时长。图12-1-1为典型的时序数据大屏展示。



**图12-1-1** **典型的时序数据大屏展示**

第12章 时序数据管理 **139**

这些通过设备、传感器采集来的数据有一个显著的特点，即数据是按照时间顺序产生的，是 一个按时间分布的序列数据，因此被称之为时序数据(Time-Series Data)。时序数据有别于企业 ERP 、CRM 系统里的数据，有很多显著的特点。其中一点就是数据量大，因为它是机器自动产 生的，只要机器在运行，就会源源不断地产生。在一座数字化工厂中，90%以上的数据都是通 过各种渠道采集的时序数据。

**12.2** **时序数据的特点**

工业领域的时序数据具有鲜明的特点。

(1)数据是时序的， 一定带有时间戳：联网的设备按照设定的周期，或受外部事件的触发， 会源源不断地产生数据。每个数据点是在哪个时间点产生的，这对于数据的计算和分析十分重 要，必须要记录。

(2)数据是结构化的：微博、微信中的海量数据都是非结构化的，其中包括文字、图片、 视频等。但物联网设备产生的数据往往是结构化的，而且是数值型的，比如智能电表采集的电 流、电压数据就可以用4字节的标准的浮点数来表示。

(3)数据极少有更新操作：物联网设备产生的数据是机器日志数据， 一般不允许修改而且 也没有修改的必要。很少有场景需要对采集的原始数据进行修改。但在一个典型的信息化或互 联网应用中，其中记录的数据是一定可以被修改或删除的。

(4)数据源是唯一的： 一个物联网设备采集的数据与另外一个设备采集的数据是完全独立 的。 一台设备中的数据一定是这台设备产生的，不可能是人工或其他设备产生的，即一台设备 中的数据只有一个生产者，数据源是唯一的。

(5)相对互联网应用，写多读少：对于互联网应用， 一条数据记录往往是一次写，多次读。 比如一条微博内容或一篇微信公众号文章，都是一次写，可能会有上百万人读。但物联网设备 产生的数据不一样， 一般是计算、分析程序自动读，而且计算、分析的次数不多，只有在分析 事故等场景中才会有人主动看原始数据。

(6)用户关注的是一段时间的趋势：对于一条银行记录，或者一条微博内容，对其用户而 言，每一条都很重要。但对于物联网数据，每个数据点与数据点的变化并不大， 一般是渐变的， 大家更关心的是一段时间(比如过去5分钟、过去1小时)内数据的变化趋势， 一般对某一特 定时间点的数据值并不关注。

**140** 数据治理——工业企业数字化转型之道

(7)数据是有保留期限的：采集的数据一般都有基于时长的保留策略，比如仅仅保留一天、 一周、 一个月、 一年，甚至更长时间，为节省存储空间，系统最好能自动删除过期数据。

(8)数据的查询和分析往往是基于时间段和某一组设备的：对于物联网数据，在做计算和 分析时， 一定会指定时间范围，不会只针对一个时间点或者整个历史进行。而且往往需要根据 分析的维度，对物联网设备的一个子集采集的数据进行分析，比如某个地理区域的设备，某个 型号、某个批次的设备，某个厂商的设备等。

(9)除存储查询外，往往需要实时分析计算操作：对于大部分互联网的大数据应用，更多 是离线分析，即使有实时分析，要求也不高。比如用户画像，可以在积累一定的用户行为数据 后进行分析。但是对于物联网应用，对数据的实时计算要求往往很高，因为需要根据计算结果 进行实时报警，以避免事故的发生。

(10)流量平稳、可预测：给定了物联网数量、数据采集频次，就可以较为准确地估算出所

需要的带宽和流量、每天新生成的数据大小。

(11)数据处理具有特殊性：与典型的互联网相比，还有不一样的数据处理需求。比如要检 查某个具体时间的设备采集的某个量，但传感器实际采集的时间不是这个时间点，这时往往需 要做插值处理。还有很多场景需要基于采集量进行复杂的数学函数计算。

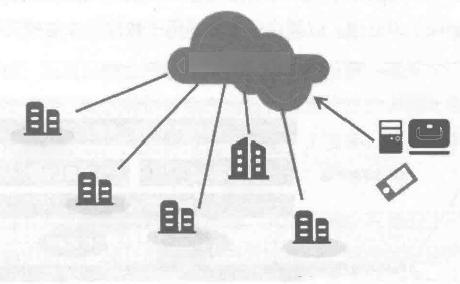
(12)数据量巨大：以智能电表为例， 一台智能电表每隔15分钟采集一次数据，每天自动 生成96条记录。全国有接近5亿台智能电表，每天生成近500亿条记录。 一辆联网的汽车每隔 10~15秒就采集一次数据并发送到云端， 一辆汽车一天就能产生近6000条记录。如果有2亿辆 汽车全部联网，那么每天将产生12000 亿条记录。5年之内，物联网设备产生的数据将占世界 数据总量的90%以上。

时序数据是典型的流式数据，就像视频流，而且单个数据点的价值很低，甚至丢失一小段 时间的数据也不影响分析的结论及系统的正常运行。但看似简单的事情，由于数据记录条数巨 大，导致数据的实时写入成为瓶颈，查询分析极为缓慢，成为新的技术挑战。传统的关系型数 据库、NoSQL 数据库及流式计算引擎由于没有充分利用物联网数据的特点，性能提升极为有限， 只能依靠集群技术，投入更多的计算资源和存储资源来处理，使得系统的运营维护成本急剧上 升。

第12章 时序数据管理 **141**

**12.3** **时序数据的应用**

对时序数据进行实时处理和分析后，可以得到很多有价值的信息。以数字化透明工厂为例， 我们可以按图12-3-1所示建立轻量化的生产管理云平台，运用物联网和大数据技术，采集、分 析生产过程中产生的各类时序数据，实时呈现生产现场的生产进度、目标达成状况，以及人、 机、料的利用状况，让生产现场完全透明，这样可以提高生产效率、降低安全风险、增强决策 的科学性。



**生产管理云平台**

**终端**

**工厂01**

**总部**

**工厂02**

**工厂03**

**工厂04**

**图12-3-1** **生产管理云平台**

**12.3.1** **技术挑战**

数字化工厂产生的时序数据量是巨大的，处理它有相当的技术挑战。以数控机床加工生产 为例，由于工业行业的要求，需要将包括报警在内的各种工况数据存储起来。假设企业每个厂 区具有2000个监测点，5秒为一个采集周期，全国一共200个厂区。这样粗略估算起来每年将 产生惊人的几十万亿个数据点。假设每个点为0.5KB, 数据总量将达 PB 级别(如果每台服务 器的硬盘容量是10TB, 那么总共需要100多台服务器)。这些数据不仅要实时生成，写入存储 器，还要支持快速查询，实现可视化的展示，帮助管理者分析决策；并且也能够用来做大数据 分析，发现深层次的问题，帮助企业节能减排，增加效益。这样看来，需要解决的关键技术问 题如下。

(1)高并发、高吞吐量的写入能力：如何支持每秒上千万数据点的写入，这是最关键的技 术能力。

(2)数据高速聚合：如何支持以秒级的速度对上亿的数据进行分组聚合运算，如何能高效 地在大数据量的基础上将满足条件的原始数据查询出来并聚合，要知道统计的原始值可能因为

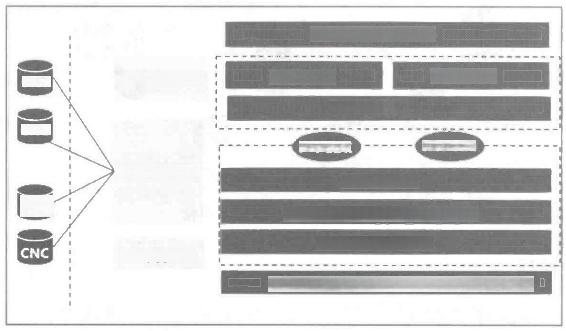
时间比较久远而不在内存中，因此这可能是一个非常耗时的操作。

(3)降低存储成本：要降低海量数据存储的成本，就需要时序数据库提供高压缩率。

(4)多维度的查询能力：时序数据通常会用多个维度的标签来刻画一条数据，而根据几个 维度进行高效查询就是必须要解决的一个问题。

**12.3.2** **典型的技术架构及特点**

传统的做法是使用 Hadoop 体系来处理时序数据，但现在流行的做法是使用专门的时序数 据处理工具(如 TDengine)来处理。以某企业开发的用于数控机床监控云平台为例，其架构如 图12-3-2所示。



**B** **层产品(用户大屏)**

**画面晨产品** 管理层引学

SaaS层后端

**CNC**

IoT SDK HTP API

**PaaS平台产品**

消息处理引肇

**PanS产品引擎(MQTl Server)**

**分布式时序数据库层**

**laaS平台产品**

**公有云|私有云(云主机、云网结、云存铺)**

**SaaS平台产品**

cnc

CNC

图12-3-2 数控机床监控云平台架构示例

此软件架构分为 IaaS 层、PaaS 层和SaaS 层架构，其中分布式时序数据库位于PaaS层的核 心位置，用于实现数据的存储和分析等功能。CNC 数据采集器(SMART NC BOX)采集的数 控机床监控数据被汇集到 PaaS 层的MQTT Server上。随后，数据存储模块实时从MQTT Server 上抽取数据，并将其转换为时序数据库能理解的 SQL 语句并写入引擎。时序数据库可支持接入 各种应用程序，实时查询最新数据。另外，各种配置信息等也会经过内置模块的转换被写入一 个关系型数据库中，方便各个应用查询使用。

**12.3.3** **系统核心功能**

时序数据一般被用于生产线、生产环节等经营过程的实时监控，而且基于历史时序数据的 分析，可以对资源的使用和生产的配置做出科学的决策。

第12章 时序数据管理 **143**

以图12-3-2所示的数控机床监控云平台为例，其具有如下核心功能。

(1)生产线状态监控：监控每台数控机床实时状态，完成故障及瓶颈告警；实时监控指标 执行情况，即时发现、即时处理、减少损失；基于邮件、短信、大屏等多种方式实时主动告知 生产中的异常状况；采集、发布每条生产线的实时生产数据，让用户可以实时了解生产进度、 不良率、人员生产效率等。

(2)生产环节监控：实时采集各岗位的生产数据，由人工录入数据变为扫描或者自动采集 数据，减少人为误差，使得用户可以直观地了解每个生产环节的进度、效率、不良率，及存在 的瓶颈。

(3)智能生产排程：实现订单生产管理导向式的生产配置，根据生产线配置情况，轻松完 成生产计划排程；可按订单、批次等进行生产设置，满足企业的多样化要求；在订单生产过程 中，对订单生成的数据进行监控，实时掌控生产进度、不良率、生产效率等，以及订单完成情 况，实现事后追溯。

(4)生产配置管理：对产品、原料、生产线、人员等建立完整的生产数据档案，以及完整 的物料标示；可根据编码进行产品正反向追溯并界定责任，减少因召回产品所带来的损失，提 高质量的稳定性。

**本章精要**

工业企业在生产经营过程中，会运用物联网技术采集大量的数据并进行实时处理，这些数 据都是时序数据，而且具有显著的特点，比如带有时间戳、结构化、没有更新、数据源唯一等。 这些数据所具有的特点也很明显，比如数据往往只保留一定时长，需要做降采样、插值、实时 计算、聚合等操作，其展示的是一段时间的趋势，而不是某一特定时间的值等。

由于数据量巨大，对时序数据的处理具有相当大的技术挑战，因此一般要使用专业的大数 据平台来进行处理。对时序数据的实时处理能够帮助企业实时监控生产与经营过程。对历史时 序数据的分析有助于对资源的使用和生产的配置做出科学的决策。因此，时序数据的采集和处 理有助于提高企业的生产效率，并降低安全风险。

第13章

数据质量管理

互联网、智能手机、可穿戴设备及智能家居的快速普及，使得每一个人和每一台接入互联 网的设备都在产生数据，这些数据被相关企业或组织通过合法的渠道收集、存储并加以分析， 进而产生价值。“数据即资产”的概念得到了人们的广泛认同，并且对数据的重视程度被提到前 所未有的高度。然而，不是所有的数据都能成为资产，数据的价值与数据质量密切相关。

**13.1** **数据质量需求**

在现阶段，大中型企业已经开始了数据化运营的实践，而其中存在的问题不容被忽视，因 此这就要求企业采用各种措施和方法，加强对数据质量的控制，使数据真正发挥出应有的价值， 也为企业的数字化转型提供可靠的支撑。

近年来，随着人工智能技术的兴起，数据质量管理技术和人工智能技术开始融合，这使得 数据质量管理开始向“智能化”转变：

(1)在检查数据质量时，可以针对少量、核心的检查规则，从大数据中选取训练数据样本， 利用机器学习算法进行深度分析，提取公共特征和模型，定位影响数据质量的因素，预测数据 质量问题，并进一步形成知识库，进而增强数据质量管理能力。

(2)在数据模型的管理过程中，通过机器学习技术可以分析数据库中数据实体的引用热度， 通过聚类算法可以自动识别数据模型间的内在关系，也可以对数据模型质量进行检测和评估。

(3)在数据传输监控中，利用机器学习技术可以对数据历史到位情况进行分析，预测数据 的到位时间，为保证数据处理的及时性和应对数据晚到的影响提供支撑。

第13章 数据质量管理 **145**

(4)在数据问题发现方面，可以应用自然语言处理技术对住址、单位名称等数据进行词性、 句式、语义分析，避免隐私数据泄漏；在出现数据不一致等问题时，提供数据质量管理线索， 增强企业的数据质量和数据安全管理能力。

利用人工智能技术还能提取数据质量评估指标。改善数据质量最理想的模式是从数据源中 剔除脏数据，但是这在现实中并不可行，因为：

(1)数据源众多且难以控制数据源的数据质量。

(2)直接从数据源中改善数据质量付出的成本过大。

因此，根据业务期望，企业应有针对性地提升各个业务线上数据流的数据质量。利用机器 学习可以通过提取有效的数据质量评估指标，最大化地提升该指标下的数据质量。同时，监督 学习、深度学习也可以实现对数据清洗和数据质量的效果评估，进而改善转换规则和数据质量 评估维度，并随着数据量和业务期望的逐渐变化，动态更新数据质量提升方案。

**13.2** **数据质量检查**

通过数据质量检查，用户能够了解各种数据的状况，轻松确定数据与具体业务规则和既定 数据标准的符合程度(主要针对数据的完整性、 一致性、及时性、准确性、有效性、唯一性等)。 通过数据探查和数据剖析可以检查数据质量情况。

在实施数据质量管控及数据治理的过程中，首先需要对数据进行全面、有效的探查。比如 在企业数据中缺失数据的所有权与责任等，这些既有可能导致数据一致性的问题，也有可能因 数据的一致性问题而出现其他问题。假如数据存在一致性问题，则往往在企业管理中会产生数 据定义和标准不同的问题。而数据探查与剖析作为数据质量检查的核心技术，在数据清洗及优 化规则发现方面扮演着重要角色。在大数据时代，透过数据剖析可以发现业务关联关系，寻找 业务创新的关键线索，驱动业务变革及创新。

数据质量检查可以帮助企业全面了解数据质量，并确定这些数据的可用性。数据质量检查

主要包括以下几项。

(1)数据内容及背景分析；

(2)数据结构及路径分析；

(3)数据成分及业务规则合规分析；

**146** 数据治理——工业企业数字化转型之道

(4)数据间的关系及相关资源匹配；

(5)识别数据转化机制；

(6)建立数据有效性及准确性规则；

(7)校验数据间的依赖性。

**13.3** **数据质量分析**

数据质量通常是指数据值的质量，包括准确性、完整性和一致性。数据的准确性是指数据 不包含错误值或异常值，完整性是指数据不包含缺失值， 一致性是指数据在各个数据源中都是 相同的。广义的数据质量还包括数据整体的有效性，例如，数据整体是否可信、数据的取样 是否合理等。本节中介绍的数据质量分析是指对原始数据值的质量进行分析。没有可信的数 据，数据分析将是空中楼阁，因此，数据分析的前提就是要保证数据质量是可信的。

数据质量分析的主要任务是检测原始数据中是否存在脏数据。脏数据一般是指不符合要求， 以及不能直接进行相应分析的数据。脏数据一般包括：

(1)缺失值；

(2)异常值(离群点);

(3)不一致的值；

(4)内容未知的值；

(5)无效值。

通常情况下，原始数据中都会存在数据不完整(有缺失值)、数据不一致、数据异常等问题， 这些脏数据会降低数据的质量，影响数据分析的结果。因此，在进行数据分析之前，需要对数 据进行清洗、集成、转换等处理，以提高数据的质量。其中主要的方法是基于数据检查后的结 果来审核数据质量，发现数据中可能存在的异常和问题，为根本原因分析、所需数据纠错和预 防错误提供优化的基础。可以通过以下6个方面对数据质量进行分析，如图13-3-1所示，从而 发现潜在的数据问题。

(1)准确性；

(2)一致性；

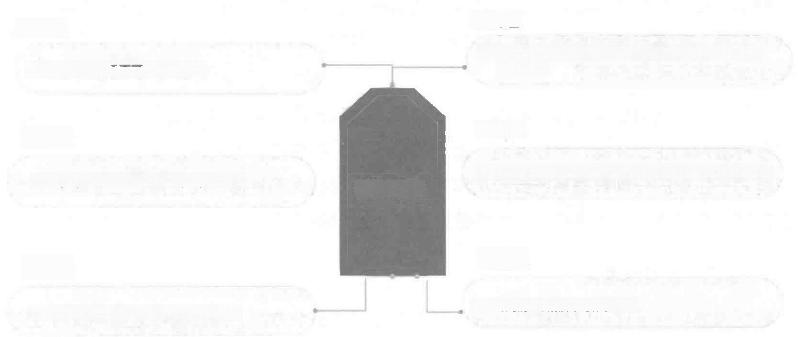
(3)完整性；

第13章 数据质量管理 **147**

(4)相关性；

(5)有效性；

(6)即时性。



**准确性**

■这些数据反映事实吗?所有数据源都传达一样的 信息吗?

**一致性**

■ **所有系统的数据都是一致的吗?有没有数据重复?**

**数据质量**

**完整性**

■ **所需的数据都同时存在吗?**

**相关性**

■ **不同数据源之间的数据存在关联吗?**

**有效性**

■ **所有的数据值都在对应的字段里吗?**

**即时性**

■ **当需要数据是否可以即时拿到**

图13-3-1 数据质量分析指标

**13.4** **数据质量提升**

在企业数字化转型的过程中，数据质量问题成为重要的影响因素之一。对数据进行质量管 理及优化也是企业数据应用工作的重点。当前，工业企业在数据质量管理体系建立过程中遇到 的问题和难点主要有以下几个方面。

(1)企业系统众多，在数据模板、物料编码等方面没有统一的标准，各系统都自成一套体 系，执行工作标准化难度大；

(2)各系统的编码体系存在重复、错误、不一致等现象，且数量众多，数据清洗难度大；

(3)缺乏IT 手段支持数据标准化管理，没有统一的数据管理、优化系统，数据管理平台建 设经验少；

(4)随着企业数字化进程的发展，结构化和非结构化数据越来越多，需进一步建立业务规 则以进行划分；

(5)建立数据管理系统时无响应的数据质量管理组织，系统上线后无自动化的标准化处 理体系；

**148** 数据治理 教字化转型之道

(6)在进行数据质量管理时未能打破旧思维，存在管理制度松散、执行标准化力度不够 等现象；

(7)企业数据质量管理人才缺乏，员工的数据质量管理意识淡薄。

针对企业在数据质量管理、优化过程中遇到的问题，企业需要在内部制定并执行统一、服 务业务需求、科学合理的数据质量优化体系，以提升数据的使用率及信任度，挖掘数据价值以 满足企业数字化转型的需求。

本节所讲的数据质量提升是指帮助企业将原本杂乱无章的数据转化成有价值的信息的过程， 针对各种复杂的企业环境，可以实时、批量改善数据的可靠性、可用性以及业务适用性。而这 个过程对于企业的任何数据系统或应用程序都适用，其强大的兼容性还支持企业全球化的数据 质量管理。可以从以下5方面提升企业数据质量，输出有价值的资产。

**1.** **数据清洗与标准化**

数据清洗与标准化可以帮助企业统一和规范数据的各个方面，制定跨行业通用及行业定制 的数据标准，规范数据的采集、录入、传输、处理等过程，进一步更正、修复企业系统中的错 误数据，并对数据进行归并及整理，使企业在数据应用上更方便，企业内的信息交流更高效， 加快数据变现的速度。

**2.** **数据匹配**

通过使用在线数据匹配工具，可以轻松地把不同来源的企业信息数据匹配到统一的编码下。

数据匹配帮助企业解决以下问题。

(1)在系统集成时进行数据合并与编码统一，保证系统的一致性；

(2)防止企业重复采购外界数据，减少采购成本；

(3)提取合适的数据用于企业营销项目，最大化营销效果；

(4)通过对新导入的数据与现存数据匹配，减少重复数据的产生；

(5)帮助客户发现数据关联组织，最大限度地控制风险。

**3.** **数据校验和补充**

基于企业对数据完整性的需求，可以引用具有时效性的第三方权威数据资源，如人口统计 信息、地理编码、邮政编码等。通过比对第三方数据与企业原始数据，可以确保数据的完整性 与有效性。

第13章 数据质量管理 **149**

**4.** **查找和删除重复数据**

通过清理信息系统中近似重复的企业信息(如客户数据、供应商数据等),可以保证数据的 一致性和正确性，支持信息系统进行正确的决策和应用。为了提高数据质量，必须要查找重复 数据和删除重复数据，确保企业有效使用数据。

**5.** **关联与统一服务**

基于企业数据库的通用内容对各个数据源进行数据关联，可以识别记录之间的相互关系， 进而确定关联规则以符合企业特定需求，并通过数据关联获得最佳数据记录，形成统一视图， 让企业对数据对象有更深的认识。

在企业对数据质量优化完成后，还需要进行持续的数据质量监管，往往通过数据报告与记

分卡来监测以及时管控数据质量。其主要用于以下情况：

(1)监控一段时间内数据质量的合规情况；

(2)展示数据质量优化成果；

(3)及时检测数据质量问题；

(4)为数据质量投入成效提供依据。

**13.5** **数据质量评估**

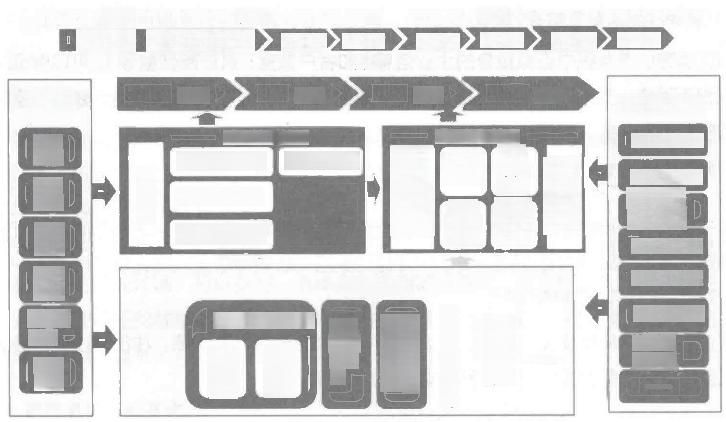
数据质量是分析和利用数据的前提，是获取数据价值的重要保障。业界比较通用的方式是 基于完整性、 一致性、及时性和准确性4个维度来评估数据质量。但如何能真正辅助企业判断 数据价值却是一直以来大家都在探讨的问题。而在当下，为了应对在数据资产整合、数据标准 化管理、数据质量提升等多方面中的挑战，各企业也逐渐提出了数据管理能力成熟度评估模型 以评估数据质量。其从企业数据的采集、存储及应用等环节进行全方位的评估，并根据该企业 数据能力的制度建设、过程监督和管理、组织人员的建设、工具的应用等多个方面进行评分， 然后根据评分的结果汇总成数据能力成熟度等级分布，如图13-5-1所示。

霜露

;疆

》量养墨

**150** 数据治理——工业企业数字化转型之道



**存储**

2.存接》3加工

**管理**

**质量管理工具**

**准确性检查**

**质量检测统计** **质量问题池**

**完整性检查** **相似及重**

**复性检查**

**一致性检查**



**精确度检查**

**合理性检查**

**有效性** **性检查**

…

质量管 理组织

组织 管理

质量 专家

业务 专家

技术

专家 质量管 理员

技术 实施

问题分问题

析报告考核

报告

问题

**数据质量评估**

**数据质量检测**

**采集** **1** **采** **集**》

技 术 释 义 标准 库

质量 评估 规则 库

**应用**

**7.运营**

**技术** **检测**

规则

业务 检测 规则

**质量检测规则库**

主数据标准库

质量检测管理

跟 踪 池

质量问题管理

**8.归档** **9.**

**标准规则库**

调度中心

4开发

6.应用

**5.发布**

警报

**检查**

评估

处理

………

**图13-5-1** **数据管理能力成熟度评估模型——数据质量应用范例**

通过数据能力成熟度的评估，企业可以更加准确地发现自身存在的问题、与相关企业在数 据质量管理和应用方面存在的差异，以及自身的优势，从而明确下一步改进的方向，为数据资 产的价值变现和提升奠定了基础。

本节将在技术基础上，从数据质量管理的技术指标和业务指标两大部分对数据质量评估进

行更深入的分析。

**13.5.1** **数据质量问题的起因**

随着数据呈爆发式地增长，多数传统企业也开始走上了数字化转型的道路。与此同时，数 据中蕴藏的商业价值也逐渐被人们挖掘出来。而大数据类的项目都有一个特点：都以数据为核 心。数据将作为产生业务价值和实现业务目标的基石，因此，数据质量就成为影响这类项目的

一个极其重要的因素。引起数据质量问题的原因有以下几个。

(1)公司IT 建设：系统建设调整，技术平台升级；

(2)业务需求：表单设计不科学，数据描述不一致；

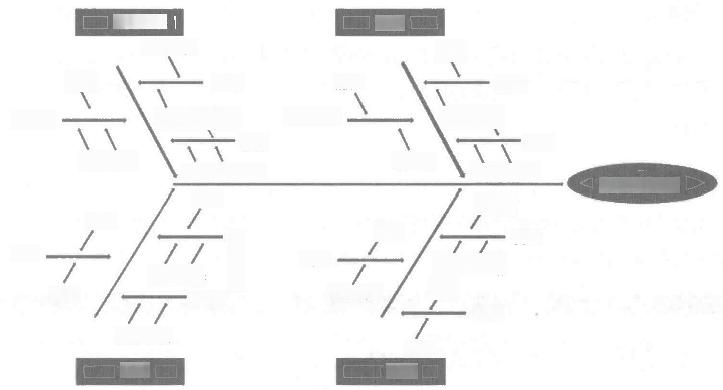
(3)操作水平：企业操作人员技术水平不一；

(4)控制机制：缺少有效的审核纠错机制；

(5)应用程序：数据完整性要求被忽略，缺乏全面校验流程；

(6)数据交换：数据量大，各系统对数据的标准不一致。

在企业的实际经营中，引发数据质量问题的因素广泛、复杂，涉及企业的信息系统、组织 架构、人员、制度流程、企业文化等。因为数据的多样化和复杂化，在连续采集、高速存储、 有效整合、实时分析、多维度呈现等各方面，都不是通过简单的数据统计分析所能解决的，在 此可以使用鱼骨图进行定量问题分析。可以分别在信息系统、流程、技术及人员方面分模块进 行分析，如图13-5-2 所示。



**流程**

设计

漏传 流程

数据传递 优化复章

实施

及时性

测试配置

责任心

|  |  |
| --- | --- |
| 进修  培训  上岗培训 | 管理  目标优先级  奖励  激励 |

**设计工具**

反馈

人员

|  |  |
| --- | --- |
| 定义  元数据  模型加工  缺陷  产品  性能 | 数据  覆盖率有效性  质量  完整性精确性 网络  OS 硬件  架构 |

**数据质量问题**

基础设施

变化频度

**信息系统**

**技术**

图13-5-2 将数据质量问题形成鱼骨图进行分析

**13.5.2** **数据质量管理技术指标**

数据质量管理技术指标是从技术角度对企业数据进行评估，主要包括以下指标，如图13-5-3

所示。

**1.** **唯一性**

唯一性是指存储在不同系统中的同一个数据是一致的。此项主要明确企业所有系统中的数 据是否一致，是否有重复数据。

**2.** **完整性**

完整性是指数据信息不能存在缺失的情况。数据缺失的情况可能是整个数据记载缺失，也

可 能 是 数 据 中 某 个 字 段 信 息 的 记 载 缺 失 。

**152** 数据治理. 工业企业数字化转型之道

**3.** **相关性**

相关性是指数据之间的关联程度。此项指标主要明确不同数据元之间的数据的关联程度。

**4.** **有效性**

有效性是指数据应遵循预定的语法规则的程度，应符合其定义，比如数据的类型、格式、 取值范围等。此项指标主要明确企业系统里所有的数据值是否都在对应的字段里。

**5.** **及时性**

及时性是指数据从产生到可以查看的时间间隔，也叫数据的延时时长。如果数据延时超出 统计的要求，则可能导致分析得出的结论失去了意义。此项指标主要明确当需要数据时是否可 以即时拿到。

**6.** **非重复记录**

非重复记录是指用于度量哪些数据是重复数据或者数据的哪些属性是重复的。此项指标主 要明确企业系统中的数据是否存在多个记录表现同一实体的现象。

|  |  |  |  |
| --- | --- | --- | --- |
| **指标类型** | **说明** | **衡量标准** | **备注** |

**完整性**

(Completeness)

**相关性**

**(Reliability)**

**唯一性**

**(Uniqueness)**

**有效性**

**(Validity)**

**及时性**

**(Timeliness)**

**非重复记录** **(Non-duplicate**

**records)**

不存在或缺失的记录数据百分比 。即实录的空值率

=空值记录总数/总记录数×100%

的每个属性都有明确的值，不存在“空”

或“未知”的属性

满足外键参照完整性记录的百分比，对于 外键无对应主键的记录比=外键无对应主键的记录总数/总 数据库中的某些实体，它们的存在可能 记量数×100

依赖于其他的实体

|  |  |
| --- | --- |
| 满足主键唯一性约束记录的百分比，即一1-主键的重复率 个表中的一组属性的值是唯一的  满足域和数据有效范围定义的记录的百分 1-异常值比率  比，即实体属性的值要在用户定义的有效  范围之内 | =1-主键重复记录总数/总记录 数×100%  =1-超出值域的异常值记录总数/ 总记录数×100% |

满足时间要求的记录的比率=满足时间要求的记录总数/总记录 ×100%

记录是否满足业务应用的时间要求

是否存在多个记录表现在同一个实体的本数据非重复记录的比=1-样本数据重复记录总数/

象 率 总记录数×100%

**图13-5-3** **数据质量管理技术指标**

**13.5.3** **数据质量管理业务指标**

数据质量管理业务指标是从业务角度对企业数据进行评估，主要包括以下指标，如图13-5-4 所示。

第13章 数据质量管理 **153**

指标类型

真实性

(Facility)

精确性

**(Accuracy)**

**一致性**

**(Consistency)**

**可理解性**

**(Understandable)**

**可用性**

**(Availability)**

说 明

真实记录的占比。真实性是指数据库 中实体必须与对应的现实世界中的对象 一致

指数据精度符合业务需要的记录的占比

与其他系统(或者系统内部)一致的记录的 占 比

含义明确和易于理解的记录的占比 即记录本身的含义是否简单、明确

可获得、满足业务使用的记录占比

街量标准

真实记录比率

满足业务对精度需求的 记录的比率

样本记录不同存储的匹 配率

易于理解的记录比率

可获得记录的比率

备注

=1-失真记录总数/

总记录数×100%

=满足业务精度需求的

记录总数/总记录数X100%

=1-不同意义记录总数/总记录

数×100%

=1-费解的记录总数/总记录数

×100%

=可获得的记录总数/总记录数

×100%

**图13-5-4** **数据质量管理业务指标**

**1.** **真实性**

真实性是指数据库中的实体必须与对应的现实世界中的对象一致，以样本数据的真实数据 为衡量标准。

**2.** **精确性**

精确性是指数据精度符合业务需要，以样本数据满足业务对精度需求的比率为衡量标准。

3. 一致性

一致性是指数据与其他系统(或者系统内部)一致，以样本数据不同存储的匹配率为衡量 标准。

4. 可理解性

可理解性是指数据含义明确和易于理解，以样本数据易于理解的记录比率为衡量标准。

5. 可用性

可用性是指数据可获得，可满足业务使用，以样本数据可获得记录的比率为衡量标准。

**154** 数据治理——工业企业数字化转型之道

**本章精要**

数据质量管理贯穿数据生命周期的全过程，其中覆盖数据质量需求、数据探查、数据诊断、 质量评估、数据监控、数据清洗、质量提升等方面。

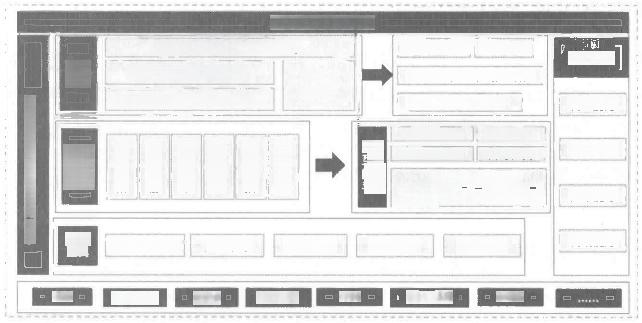
数据源在不断增多，数据量在不断加大，新需求推动的新技术也不断诞生，这些都给工业 企业的数据质量管理带来了困难和挑战。因此，基于数据质量管理框架的数据质量管理方法， 应包含事前计划(数据质量需求)、事中检测(数据质量检查和数据质量分析)、事后评估及处 理提升(数据质量提升和数据质量评估)4个阶段。

第14章

数据安全管理

**14.1** **数据安全管理概述**

构建完整的工业互联网安全体系，是实施数据安全管理的重要保障。根据行业特点和管理 要求，这里将工业互联网安全体系架构分为工业互联网合规保障体系、组织建设、数据安全管 控、工业互联网云平台、人员能力、技术工具，如图14-1-1所示。



**工业互联网安全体系**

**决第层(高管、信息安全官)**

人员能力

**管理层(团队管理人员)**

管理能力

运营能力

工业大数据资源池

技术能力

合规能力

PLC 智能设备 MES SCADA ERP 虚拟仿真 PDM

数据处理

數 据 存 储

数据传输

数据采集

流程、规范、指南、模板

计划、报告、记录、

**数据分级分类** 安全审计 权限管控

留

技术 工具

**执行层(技术团队、安全运营)**

存储安全 多租户隔离

班束库 主 资 库

工业互联网合规保陵体系

数据脱敏 数据加密

权限控制 备份恢复

工业 互联露

监督层 (审计)

方针和总纲

组织 建设

管理制度

數 据 销 毁

數 据 交 换

图14-1-1 工业互联网安全体系架构图

(1)工业互联网合规保障体系：构建工业互联网安全体系框架，就是在以法律法规为基础， 融合业务需求的同时，通过建设组织能力和使用安全技术工具，进行工业互联网云计算平台和 数据安全能力的建设，同时配备相应的安全运营人员，最终实现对工业互联网安全的全生命周 期监管。

**156** 数据治理 工 业 企 业 数字化转型之道

(2)组织建设：建立数据安全组织架构和职责分配，以及沟通协作机制。组织可分为决策 层、管理层和执行层3层。其中，决策层由参与业务发展决策的高管和数据安全官组成，负责 制定数据安全的目标和愿景，在业务发展和数据安全之间做出良好的平衡；管理层由数据安全 核心实体部门及业务部门的管理层组成，负责制定数据安全策略和规划，以及具体管理规范； 执行层由数据安全相关运营、技术和各业务部门接口人组成，负责数据安全工作的推进和落地。

(3)技术工具：与制度流程相配套并保证有效执行的技术和工具，可以是独立的系统平台、 工具、功能或算法技术等。需要综合所有安全域整体规划配套的技术工具，且要与组织的业务 系统和信息系统等进行衔接。其中包括适用于所有安全域的通用技术工具和部分阶段或安全域 试用的技术工具。

(4)人员能力：为实现上述安全组织、制度和技术工具的建设和执行的人员应具备的能力。 其中核心能力包括数据安全管理能力、数据安全运营能力、数据安全技术能力及数据安全合规 能力。企业可以根据不同数据安全能力建设维度匹配不同人员的能力要求。

(5)数据安全管控：数据安全能力支撑数据安全体系框架，为工业互联网云平台中的数据 提供安全保障，其中的安全措施覆盖工业大数据的全生命周期。

**14.2** **数据安全体系框架**

数据安全体系框架通过3个维度构建而成，包括政策法规及标准规范、技术架构层面和安 全组织与人员，如图14-2-1所示。构建数据安全体系框架，就是在符合政策法规及标准规范的 同时，还需要在技术上实现对数据的实时监管，并配备经过规范培训的安全组织与人员。

在整个体系框架中，核心监管技术体现在技术架构层面，包括安全运营中心、数据中心及 安全基础资源。

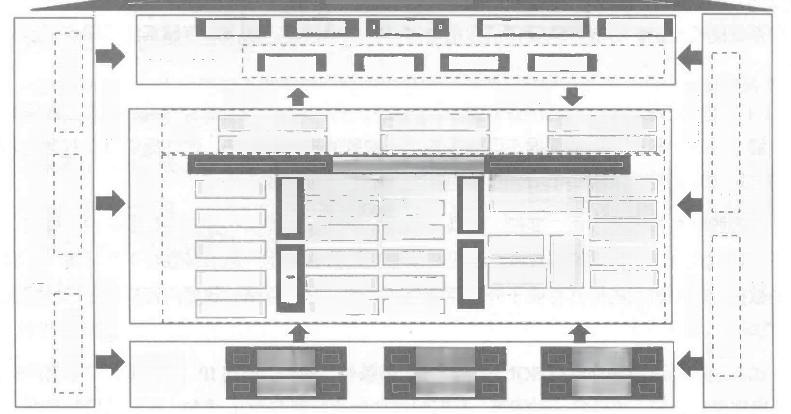
安全基础资源通过提供基础技术工具支撑数据中心，安全运营中心对整个数据中心进行实 时的响应控制。

安全运营中心的功能包括数据的资产管理、合规监管、实时监测、数据安全态势分析及通 报预警。安全运营中心通过采集数据中心中的数据，对数据进行汇聚、分析及治理来实现对整 体数据的实时管控，并以监管的角度分析集成数据，实时处理异常情况，保障数据安全。

数据安全基础资源的功能包括身份管理、认证管理、权限管理、密码管理、审计管理及通 用管理。数据安全基础资源是整体技术框架的支持组件，在提供最基础的技术保障的同时，以

服 挑 数

**工具的形式保障数据安全。**



**数据安全治理整体框架**

|  |  |
| --- | --- |
| **合规监管**  **安全运营中心**  然报汇集  **数据采集**  访问控制  应用攻击防护 | **实时监测** **数据安全态势分析** 通报预警  数探治理 数据服务 数据分析  **响应控制**  应用通信加座  应用脆弱性防护 |

**数据全生命周期与流转监管**

[数退销段处置 文件销毁处置

分级分类

权限控制

DLP

备份恢复

用户认证]

[数据隔离

**身份管理**

**密码管理**

安全组织与人员

组 织 与 人 员 保 障 安 全 运 营 梳 理

人员保麗 运行支挥

权限审批

DLP

数据脱敏 数据加密

数据脱敏 权限管理

数据处置

DLP

泄露防护

接口安全 行为审计

政 策 规 范 标 准 规 范 管 理 制 度

[接口安全

数据水印 监管审计

合法采集

数源认证

采集审计

应用内容保护 应用特权防护

**权限管理**

**通用管理**

**认证管理**

**市计管理**

政策法规及标准规范

数据交施 数摇传喻

安全基 础资源

数 据 中 心

有规可循

数据

资源

有法可依

数据采集

存情加閣

数据在情

**资产管理**

应用

图14- 2- 1 数 据 安 全 体 系 框 架

数据中心分为数据资源和应用两个层面，数据资源在支撑应用的同时，覆盖数据生命周期 的6个阶段，包括数据的采集、传输、存储、处理、交换和共享安全、销毁。

(1)数据采集阶段：要明确数据采集规范，制定数据采集策略，完善数据采集风险评估及 保证数据采集的合规合法性。在数据采集规范中要明确数据采集的目的、用途、方式、范围、 采集渠道等内容，并对数据来源进行鉴别和记录。制定明确的数据采集策略，体现在采集周期 和采集内容的定义，只采集经过授权的数据并进行日志记录。还要对数据采集过程中的风险项 进行定义，形成数据采集风险评估规范，包括评估方式和周期细节等。

(2)数据传输阶段：使用合适的加密算法对数据进行加密传输，其中主要用到的是对称加 密算法和非对称加密算法。对称加密(也叫私钥加密)算法指加密和解密使用相同密钥的加密 算法，有时又叫传统密码算法，其加密密钥可以从解密密钥中推算出来，解密密钥也可以从加 密密钥中推算出来。主要的对称加密算法有 DES 、IDEA 、AES 、SM1 等。非对称加密算法需 要两个密钥，即公开密钥和私有密钥。公开密钥与私有密钥是一对，如果用公开密钥对数据进 行加密，则只有用对应的私有密钥才能解密；如果用私有密钥对数据进行加密，则只有用对应 的公开密钥才能解密。非对称加密算法的加密和解密使用的是两个不同的密钥。常用的非对称

**158** 数据治理 工 业 企 业 数字化转型之道

加密算法有RSA 、ECC 、SM2。

(3)数据存储阶段：制定存储介质标准和存储系统的安全防护标准。存储介质标准需要覆 盖存储介质的定义、质量，存储介质的运输、使用记录及管理，以及存储介质的维修规范。对 存储系统的安全防护标准，需要包括数据备份、归档和恢复，以及对存储系统的弱点识别及维 护。

(4)数据处理阶段：明确需要数据脱敏的业务场景和统一使用适合的脱敏技术是数据处理 的关键。在这个阶段中，要根据不同的场景统一脱敏的规则、方法，评估提供真实数据的必要 性和脱敏技术的使用。脱敏技术主要分为静态脱敏和动态脱敏。

静态脱敏直接通过屏蔽、变形、替换、随机、格式保留加密和强加密等多种脱敏算法，针 对不同数据类型进行数据掩码扰乱，并可将脱敏后的数据按用户需求装载至不同环境中。其导 出的数据是以脱敏后的形式存储于外部存储介质中的，此时存储的数据内容实际上已经发生改 变了。

动态脱敏通过精确地解析 SQL 语句匹配脱敏条件，例如，访问IP、数据库用户、客户端工 具、操作系统用户、主机名、时间等，在匹配成功后改写查询 SQL 语句，将脱敏后的数据返给 应用端，从而实现对敏感数据的脱敏。此时存储于生产数据库中的数据未发生任何变化。

(5)数据交换和共享安全阶段：建立数据交换和共享的审核流程和监管平台。在此阶段要 建立数据导入/导出的流程化规范，统一权限管理、流程审批，以及监控审计，确保对数据共享 的所有操作和行为进行日志记录，并对高危行为进行风险识别和管控。

(6)数据销毁阶段：从整个销毁过程管理和技术保障措施上进行管理。首先数据销毁要符 合工业企业的数据销毁管理制度、办法和机制，对销毁对象、原因和流程需明确；其次是在整 个销毁过程中要进行安全审计，保证信息不可被还原，并验证效果；最后针对物理销毁的介质， 要进行登记、审批和交接工作。 一般在技术上采用删除文件、格式化硬盘、文件覆写和消磁等 方法进行数据销毁，且最后要保证数据无法被复原，防止可能出现数据泄露的风险。

为了实现对数据安全的监控和审计，对数据的分级、分类是必不可少的。在对数据分级、 分类之前，需要通过数据测绘来发现敏感数据，以及数据主要存储的位置。对数据进行结构化 分级、分类，可以实现对数据资产进行安全敏感分级管理，并依据级别分别部署相对应的数据 安全策略，以保障在数据资产全生命周期过程中，数据的保密性、完整性、真实性和可用性。

第14章 数据安全管理 **159**

**14.3** **数据安全防护策略**

工业互联网云平台在实现了数据大集中的同时，也导致了数据风险的大集中。如何识别数 据风险，进而采取有针对性的数据安全防护控制措施来缓解、转移、规避数据安全风险，是数 据安全防护体系建设必须考虑的一环。

从数据安全的全生命周期角度来看，数据的采集、传输、存储、处理、交换、销毁各个阶 段，均面临着不同程度的风险。

①采集和传输阶段。

采集前端存在仿冒、伪造的行为，导致数据交换共享平台存在被入侵的风险；传输链路被 监听、嗅探，导致数据被篡改、窃取。

②存储阶段。

数据库管理员等特权用户越权访问、违规操作、误操作，数据库或文件未加密，导致数据 泄露。

③处理阶段。

终端用户通过 USB、蓝牙等外部设备发送敏感数据，或通过截屏、拍照等方式窃取数据；

内部人员通过应用系统违规窃取或滥用数据； BI 分析人员越权、违规操作数据。

④交换阶段。

传输链路被监听、嗅探，导致数据被篡改、窃取；外部应用系统假冒数据接收对象获取数 据；敏感数据分发给外部单位。

⑤销毁阶段。

重要存储介质在维修/报废前缺乏数据销毁管控，未做到安全删除，存在数据泄露风险。

从工业控制系统到工业信息管理系统，再到互联网应用，其中的数据的流动面临着各种风 险。要以工业大数据为核心，构建覆盖数据全生命周期的安全防护体系，则需要在数据采集、 传输、存储、处理、交换和销毁环节采取相应的安全技术防护策略，保障数据安全。主要防护 策略介绍如下。

(1)数据识别与分类打标策略：根据数据识别和数据分类的结果对数据进行打标操作。首 先通过数据识别分析出每个字段的数据类型，例如姓名字段、身份证字段、邮箱字段等，然后 根据数据分类信息，给每个字段打上对应的标签。

**160** 数据治理——工业企业数字化转型之道

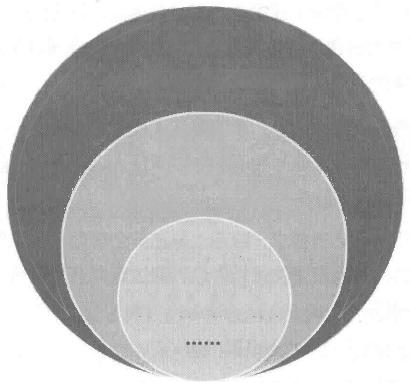
(2)数据访问控制策略：对接 LDAP 等第三方身份管理系统，从第三方身份管理系统里同 步用户身份信息。根据常见的组织关系对用户进行管理，包括多级部门、姓名、ID、级别、电 话、邮件等，为用户认证提供依据。首先通过访问账号信息表获取关系型数据库和大数据组件 的账号信息。然后逐个绑定账号所属的用户信息，将账号与用户关联。再定义不同的数据访问 权限，包括4类状态：非受控应用、受控应用、不可信应用、可信应用。最后根据用户身份认 证结果，结合账号分级分类和数据分级分类，实现基于数据的访问控制功能，避免产生用户越 级访问数据的安全风险。

(3)数据库审计策略：主要对用户的数据库操作行为、网络行为监控、网络传输内容进行 审计。它不仅能够识别谁访问了系统、在什么时候访问，还能指出系统正被如何使用。对于确 定是否有网络攻击的情况，审计信息对于确定问题和攻击源很重要。同时，通过系统事件的记 录能够更迅速和系统地识别问题，并且它是事故处理的重要依据，为网络犯罪行为及泄密行为 提供取证基础。另外，此策略也是对网络潜在威胁者的威慑。

(4)数据脱敏与加密策略：从保护敏感数据机密性的角度出发，此策略包括两类场景。 一 类是在生产环境中敏感数据要在脱敏后展示；另一类是在测试、培训等环境中敏感数据要在脱 敏后使用。在进行数据展示时，应用系统需要对敏感数据进行模糊化处理，特别是对姓名、手 机号码、身份证号码等个人敏感信息。当需要查询原始敏感信息时，需要应用进行二次鉴权。 业务系统或后台管理系统在展示数据时需要具备数据脱敏功能，或嵌入专门的数据脱敏技术工 具。在测试区进行系统测试或数据挖掘算法验证时，需要对生产数据进行批量脱敏并导入测试 环境。针对敏感数据在网络中传输面临中间人攻击、数据窃听、身份伪造等安全威胁的情况， 需要采用专线传输数据或者加密的方式传输数据，如 SSL 的加密方式。

(5)数据防泄密(泄漏)防护策略：主要通过数据梳理服务模型(见图14-3-1)及数据风 险控制框架(见表14-3-1),来实现数据资产分类分级及对数据泄密(泄露)风险的控制。

第14章 数据安全管理 **161**



**知识库**

**数据安全** **业务安全**

**分级分类**

**数据分析**

**数据分类** **数据分级**

数据梳理

数据发现

**分布统计体量统计**

**图14-3-1** **数据梳理服务模型**

**表14-3-1** **数据风险控制框架表**

|  |  |  |  |
| --- | --- | --- | --- |
| 序 号 | 类 别 | 级 别 | **管控措施或思路** |
| 1 | 1类 | 极度敏感级别 | 实施严格的技术和管理措施，保护数据的机密性和完整性，确保数 据访问控制安全，建立严格的数据安全管理规范及数据监控机制。此 类数据严格控制外传 |
| 2 | 2类 | 敏感级别 | 实施较为严格的技术和管理措施，保护数据的机密性和完整性，确 保数据访问控制安全，建立数据安全管理规范及数据标准监控机制； 此类数据在满足相关安全条件下，可以外传 |
|  | N类 | 低敏感级别 | 实施必要的技术和管理措施，确保数据生命周期安全，建立必要的 监控机制；此类数据在满足相关安全条件下，可外传 |

数据防泄密(泄露)防护( Data leakage prevention,DLP)策略主要分为终端数据安全和网 络数据安全。终端 DLP 通常包括敏感数据的识别、威胁监控、日志审计及终端外设的端口管理 等。网络 DLP通常以旁路镜像流量或串联的方式来发现是否有敏感数据在网络中传输，对于网 络中传输的低敏感级别数据采取网络审计策略，对于高敏感级别数据采取阻断和告警的安全防

护策略。

**14.4** **数据安全审计**

数据安全审计包括以下几个方面。

**162** 数据治理——工业企业数字化转型之道

账号审计：根据账号管理要求，在安全管理平台中实现对应用与系统账号的集中管理，实 现“一人一账号”,以及账号创建与销毁符合企业管理制度，并通过定期开展账号管理审计，防

止出现违反账号管理要求的问题。

授权审计：根据账号授权管理要求，系统责任人在安全管理平台中负责对账号权限分类、 实现对账号最小化权限控制、保证账号授权到账号责任人，以及通过定期开展账号授权审计，

防止出现账号权限滥用和责任人缺失等问题。

认证审计：根据账号认证管理要求，维护人员在安全管理平台中登录并访问资源，禁止绕 过安全管理平台直接或跳转访问资源的违规行为，通过定期开展账号认证审计，防止出现违反

账号认证的访问要求等问题。

访问控制审计：根据访问控制管理要求，通过制定基于地址、端口等多维的访问控制策略 加强对重要资源的访问控制，实现对重要资源的多层安全防护；通过定期开展重要资源访问控

制审计，防范重要信息泄漏风险。

重要操作审计：根据安全运维操作记录要求，全面记录运维人员的维护操作记录，通过定 期开展重要操作、违规操作审计，防止出现安全事件无法追溯到责任人的问题，以及对内部违

规行为产生威慑。

敏感信息审计：根据客户信息安全保护管理要求，全面分析业务敏感信息，在前/后台维护 中加强对敏感信息的查询、删除、导出等操作行为的安全管控，以及通过定期对敏感信息操作

审计，防范客户敏感信息泄漏风险。

**14.5** **数据安全风险评估**

数据安全风险评估首先要具有综合漏洞扫描能力，涵盖系统漏洞扫描、Web 漏洞扫描、数 据库漏洞扫描、基线扫描等功能。综合漏洞扫描服务是以 Web、 数据库、基线核查、操作系统、 软件的安全检测为核心，以弱口令、端口与服务探测为辅助的综合漏洞探测技术手段，并且要 采用分布式、集群式漏洞扫描功能，缩短扫描周期，提高长期的安全监控能力。通过 B/S 框架 及完善的权限控制系统，最大限度地满足用户安全协作的要求。具体包括如下功能。

**1.** **系统漏洞扫描**

系统漏洞扫描主要用于分析和指出有关网络的安全漏洞及被测系统的薄弱环节，给出详细

的检测报告，并针对检测到的网络安全隐患给出相应的修补措施和安全建议；全方位检测信息

第14章 数据安全管理 **163**

系统存在的主机和软件的安全漏洞、安全配置问题、弱口令，以及不必要开放的账户、服务、 端口。

**2.Web** **漏洞扫描**

Web 漏洞扫描可以帮助用户充分了解 Web 应用存在的安全隐患，建立安全、可靠的 Web 应用服务，改善并提升应用系统抗各类 Web 应用攻击的能力(如注入攻击、跨站脚本、文件包 含、钓鱼攻击、信息泄漏、恶意编码、表单绕过等)。

**3.** **数据库漏洞扫描**

数据库漏洞扫描可帮助用户充分了解数据库存在的安全隐患，通过定期进行安全检测与评 估，提升各类数据库的抗风险能力，同时可以协助用户完成数据库建设成效评估，协助数据库 安全事故的分析调查与追踪。

**4.** **基线扫描**

基线扫描全面覆盖操作系统、数据库、中间件、防火墙、路由器、交换机等设备类型，支 持Windows系统中的离线检查，无须对每台设备建立任务，即可一键提取系统配置信息，并可 导入远程安全评估系统及出具修复加固建议报告。

另外，工业互联网产生的数据都被存储于数据库中，还要对它进行安全评估服务。服务内 容包括自动完成对几百种不当的数据库配置、潜在弱点、数据库用户弱口令、数据库软件补丁 等的漏洞检测(风险趋势管理、弱点检测与弱点分析、弱口令检测、补丁检测、存储过程检测)。 可以采用数据库漏洞扫描工具帮助用户充分了解数据库存在的安全隐患。通过定期对数据库系 统进行安全检测与评估，可以提升各类数据库的抗风险能力。具体包括以下功能。

(1)数据库监测。

建设工业互联网云平台租户的业务数据库监测能力，为用户提供数据库漏洞扫描、弱口令 检查、扫描策略管理等能力。

(2)数据库漏洞检测。

漏洞检测：根据当前配置的漏洞库，对数据库进行扫描，判断是否存在相应的漏洞。

安全相关信息展示：对数据库系统中与安全相关的信息进行深入分析、提取，通过列表的

形式展现给用户。比如：test1、systest1 是非默认的拥有数据库管理员权限的用户。

(3)弱口令检查。

**164** 数据治理——工业企业数字化转型之道

通过对数据库口令的存在形式(明文、MD5 加 密 、Hash 加密)、可能的存储地址(数据库 表、历史文件、环境变量、配置文件、客户端)、口令的算法(允许的长度、Hash 生成规则) 等进行深入的分析，生成其特有的口令字典。根据已经存在的口令字典完成数据库默认账号的 识别，以及数据库登录账号长度较短及强度不高的弱口令的识别。内置的口令字典支持用户自 定义设置，用户可按需增加、删除或修改口令字典。

(4)扫描策略管理。

策略即数据库检测的依据和标准。通过策略管理可以灵活制定不同的检测标准，可以根据 用户的实际测试目的，定制不同的策略，并可以自行添加策略项扩充策略库用于检测数据库的

安全漏洞。

(5)检测结果输出。

实时结果展现：在扫描过程中，数据库漏洞扫描工具会实时显示详细的扫描和检测结果。

报告格式：支持 PDF 、DOC 、XLS 、XML 等格式的输出。

报告输出模式：风险报告可定制化。在输出报告前，可以对报告的各部分数据进行设置， 并可以定制报告的页眉、页脚、标题等内容。详细的风险扫描报告中包括漏洞名称、描述信息、 风险等级、修补建议等。报告中提供了风险等级统计图表和安全风险类型统计图表，用户可以

方便直观地了解数据库漏洞信息。

(6)自身安全控制。

用户管理功能：提供管理员、操作员、审计员3种不同的角色，分配给具有不同使用需要

的用户，合理管理用户的使用权限，防止对系统的滥用和误用。

屏幕锁定功能：在扫描过程中，如果用户停止操作10分钟，则程序将自动锁定屏幕(并且 锁屏时间可配置)。用户可以离开运行扫描系统的机器而无须担心信息泄露及扫描系统被非法使

用 。

日志管理功能：具有日志管理权限的用户可以查看所有系统日志。其中提供按照日期检索

的功能，并且可以将系统日志导出为 CSV 文件进行备份。

最后，数据安全风险评估会根据用户的各类日志进行综合深度分析，发现潜在的威胁，起

到追根索源的目的，并可作为案件的取证材料。

第14章 数据安全管理 **165**

**14.6** **数据应急保障**

工业互联网数据安全事件的应急保障是指对工业互联网数据在存储、传输或使用过程中被 泄露、篡改或破坏，可能或已经对工业控制系统、工业信息管理系统和工业互联网系统等造成 经济损失或不利影响的事件的应急组织管理、事件先期处理和后续跟踪、事件信息发布等保障。 工业企业应根据自身情况，制定数据安全事件应急保障预案，指定专门的部门或人员负责应急 处理，明确内部应急处理流程、规定各项应急处理措施和管理应急事件的信息发布等内容。

另外，还需要将第三方机构也纳入本机构数据安全事件的应急处理范围，建立与第三方机 构之间的应急处理流程，并监督其落实数据安全管理要求，配合应急预案措施的实施。各成员 机构应结合实际情况，有计划、有重点地组织进行应急处理预案演练，不断完善应急处理流程。 当发生敏感数据泄露时，应立即采取应急处理措施。如暂停涉嫌发生账户信息泄露事件的各类 终端、工业控制系统、ERP 系统等运营和生产信息系统，并逐一排查信息泄露点，必要时应联 系系统开发方协助调查，初步确定事故发生原因；对涉嫌泄露敏感数据的操作人员进行调查， 要求其暂时离岗。如有必要，则应向公安机关报案，并积极配合案件侦破工作，以及安排其他 的应急处理措施等。另外，需尽快落实各项应急措施，协助控制不利影响，防止风险蔓延。在 事后，应提出改进措施及整改方案，再进一步进行数据安全漏洞扫描，消除技术漏洞并更正或 修补相关管理制度存在的问题或缺陷，做好数据安全风险评估的整改工作。

**本章精要**

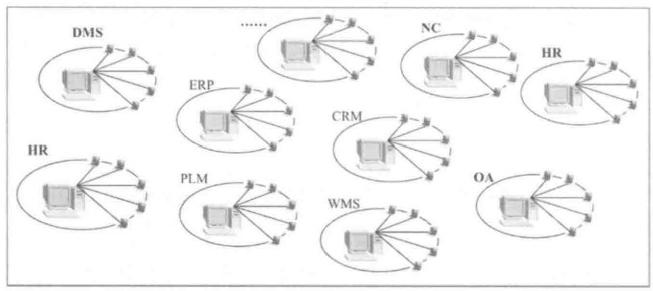
本章介绍了数据安全管理体系框架，可以通过政策法规、技术层面和安全组织人员3个维 度构建数据安全体系框架。

数据安全管理贯穿数据采集、传输、存储、处理、交换、销毁的各个阶段。另外，从数据 安全管理的全局考虑，企业需要引入数据安全风险评估方法论和技术措施，制定数据应急保障 流程和方法，以便在发生数据安全事件时可以很好地进行风险控制。

第15章

数据交换与服务

工业企业经过多年的发展，通常会按照各个阶段及各业务部门的业务需求，分别采用不同 的技术和不同供应商，分期、分批地搭建了一些IT 系统，如 PLM、ERP、MES、WMS、CRM、 SCM 等。 一般小企业中有3~5个系统，大企业中有上百个系统。这些异构的、离散的信息孤岛， 在一定程度上满足了各业务部门内部的协同和信息化管理需求，如图15-0-1所示。



**图15-0-1** **信息孤岛式的业务系统**

随着各个企业业务的快速发展，业务管理协作的瓶颈已经从部门内部迁移到部门与部门之 间，甚至企业与企业之间，从执行层面迁移到管理层面、战略层面，并且出现了大量跨系统、 跨部门、跨组织的业务协同自动化需求。特别是在数据管理方面，企业需要建设一套企业级的、 跨平台的，能够贯穿研发、生产、采购、销售、售后等多个部门业务过程的统一数据视图，提 高数据质量，实现业务数据与流程的同步和逻辑约束，以及数据的可视化和自动化流动，全面 支持公司的收益管理、营销创新及决策支持，实现数据驱动流程、流程驱动业务的智能化企业。

M08

如何整合多个业务系统的数据，以便既快又好地服务于新的业务模式，已成为各个企业亟

待 解 决 的 关 键 问 题 。

**15.1** **数据交换与服务的意义**

由于企业中的各个IT 系统都是独立地维护自己的数据，造成大量的数据冗余和不一致。哪 个系统中的信息是准确、最新、可用的?没有谁能说得清楚。业务数据分别被存放在各业务应 用系统中，需要通过系统集成实现数据交换，进而完成数据整合。只有利用成熟、完善的分析 工具和方法，才能实现以产品、物料、人员、客户为中心的业务数据的统一服务和统一管理， 实时响应信息需求，保证业务信息的完整，实现统一的业务视图。

数据交换与服务，就是连接各业务系统的信息孤岛，将各业务单元中自有的数据共享出来， 供其他业务单元使用，从而将各业务系统的能力进行整合，为企业新的管理模式和业务模式提 供灵活、可靠的数据支撑。

例如各业务系统将其管理的业务数据(如 PLM 系统中的产品研发数据， ERP 系统中的采 购、计划数据， MES 系统中的生产数据等)共享出来，通过数据交换实现工业全要素、全价值 链、全产业链的互联互通；通过数据交换实现数据的感知，通过感知实现业务协同、业务智能，

如图15-1-1所示。



QNS

**服务**

CAD

**研发**

PLM

**DMS**

TDM

**生产**

**FMS**

SCM

**WMS**

**MES**

**销售**

**MDM**

MSM

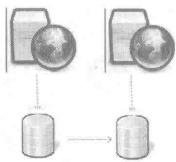
CAPP

CRM

**ERP**



**图15-1-1** **系统业务能力图**



**168** 数据治理——工业企业数字化转型之道

**15.2** **数据交换与服务技术演进**

数据交换与服务技术发展了很多年，其间的技术也层出不穷，但总结下来主要有以下代表

性的技术。

**15.2.1** **文件共享技术**

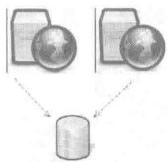
在信息化早期阶段，当多个业务系统需要交换数据时，可以通过文件共享的方式进行数据 传递，如导出或手工录入一个系统中的数据文件，再将其导入另外一个系统中。这种双方约定 好数据格式并按照约定进行功能开发的方式的缺点是：数据传递不及时，需要人工干预，数据

可能出现异常，以至于不安全。

**15.2.2** **数据库中间表技术**

早期的数据库技术是采用中间表的方式实现数据交换与服务的，双方约定一个公用数据库 及表， 一个系统负责向该表写入数据，另外一个系统负责从该表中读取数据，从而实现两个系 统的数据同步。但随着数据库技术的发展，跨系统直接进行数据传输可以通过跨数据库的方式 来实现。通过联邦数据或者 DBLink 等技术，可以实现多个系统之间的数据直接传递，如图15-2-1

.所示。



共享数据库

**图15-2-1**

或

建立DBLink

**数据库中间表技术**

**15.2.3** **点** **对** **点** **接** **口** **技** **术**

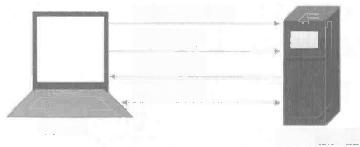
随着互联网的迅猛发展，以前很多 C/S 架构的系统逐步被淘汰，大量的系统都采用了 B/S 架构。随着网络技术、网络运行理念的发展，人们提出一种新的利用网络进行应用集成的解决 方 案 Web Service 。Web Service 是 一 种新的Web 应用程序分支，可以执行从简单的请求到 复杂的逻辑处理的任何功能。 一旦部署了 Web Service,其 他 Web Service 应用程序就可以发现

并调用它部署的服务。因此， Web Service 是构造分布式、模块化应用程序和面向服务应用集成

第15章 数据交换与服务 **169**

的最新技术和发展趋势。在Web Service技术中最典型的莫过于 SOAP(Sipmle Object Access Protocol, 简单对象访问协议)和REST(Representational State Transfer, 表述性状态转移)协 议，这两种协议现在占据着应用层集成的半壁江山，其中 REST 协议随着微服务技术的兴起，

地位已经越来越重要。图15-2-2所示为点对点接口技术示意图。



1.连接到服务器

2.发送请求

3.发送响应

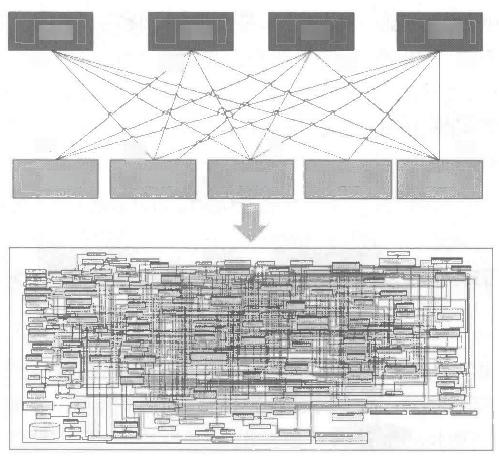
4.关闭连接

**Web** **服务器**

**客户机**

图15-2-2 点对点接口技术

两个系统怎么交换数据?最直接的方法就是写接口，写接口是程序员最擅长的方式。如果 企业只有2~3个系统，接口数量也不多，交换的数据量也不大，就没有问题，写接口也许是最 佳方案。目前接口技术在数据交换与服务领域几乎没有什么缺点，但最大的问题是随着系统越 来越多，接口越来越多，服务接口之间的调用总是存在不透明、不规范的操作，尤其是随着信 息化的不断深入，这种问题会越来越明显，接口之间会成为一个多对多的网状交换关系，难以 管理，如图15-2-3 所示。

PLM

MES

DMS

ERP

137

40

18

32

40

11

2

29

WMS

SCM

OA

CAPP

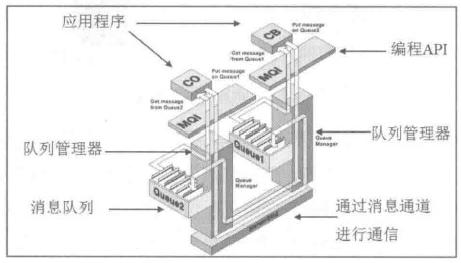
HR

**图15-2-3** **蜘蛛网式的接口调用**

**170** 数据治理——工业企业数字化转型之道

**15.2.4** **消息队列技术**

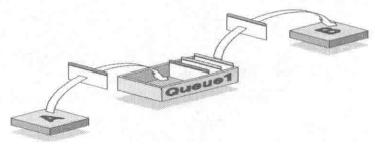
随着接口技术的成熟以及接口越来越多，接口重用及异步传输等需求则慢慢凸显，面对各 种协议、各种技术框架，如何能让一种接口被其他多种协议直接调用，而不用重复开发多种协 议版本的相同服务?我们有时需要将大量数据共享到多个业务系统中，点对点的接口开发则显 得有点无力。这时出现了消息队列技术(见图15-2-4)。我们日常使用的QQ和微信等即时通信 软件，以及邮件都是异步消息传输的典型应用。



**图15-2-4** **消息队列技术**

需要交换的数据被打包成一个个的“消息”,然后被放入消息队列中，等待需要方来取。这 就跟我们日常生活中的快递包裹一样，快递员将包裹放入包裹存储柜中，收件人有时间了就去 取。“消息队列”是在消息的传输过程中保存消息的容器。消息队列管理器在将消息从它的源中 发送到它的目标时充当中间人。队列的主要目的是提供路由并保证消息的传递；如果发送消息 时接收者不可用，那么消息队列会保留消息，直到可以成功地传递它。消息可以被持久化，这 样的机制能够保证消息一定被送达，而且只被送达一次。

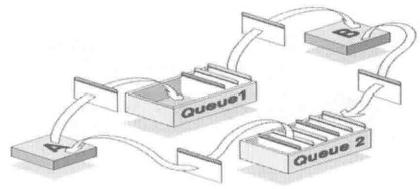
如果把消息队列用公共邮箱的方式来实现，就是将数据打包成一个个的片段(俗称“消息”), 并用信封封好，放到公共邮箱中，收信人根据信件的标签找到源系统传输来的数据(即“消息”), 这就是消息的传输方式。邮件系统就是一个典型的消息传输系统(见图15-2-5)。



**图15-2-5** **消息队列原理图**

第15章 数据交换与服务 **171**

两个系统之间的数据交换虽然可以将数据打包成数据片段异步传输，但是如果发送方希望 及时得到接收方的反馈，则可以设定一个发件箱用于存储发送的数据，再设定一个收件箱用于 收集接收方反馈的数据。这样通过及时获得收件箱的反馈信息可以做到同步的请求响应，便于 满足一些实时性要求高的应用场景(见图15-2-6)。



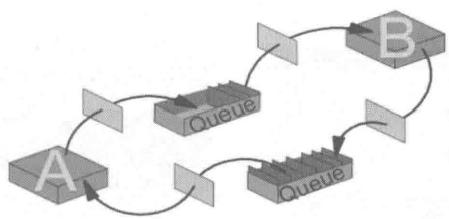
**图15-2-6** **设置收件箱用于存储发送数据**

**15.2.5** **企业服务总线交换技术**

**1.** **从消息队列** **(MQ)** **到服务总线** **(ESB)** **的演变**

消息队列是点对点的传输，要将消息从A 系统传输到B 系统，必须将消息放到 B 系统对应 的接收队列中， B 系统自己去取。B 系统的反馈信息，也必须放到A 系统的接收队列中， A 系

统自己去取反馈信息(见图15-2-7)。



**图15-2-7** **消息队列** **MQ**

此种方式有一个限制条件：A、B 两个系统产生的消息格式一致，接口协议一致(MQAPI)。

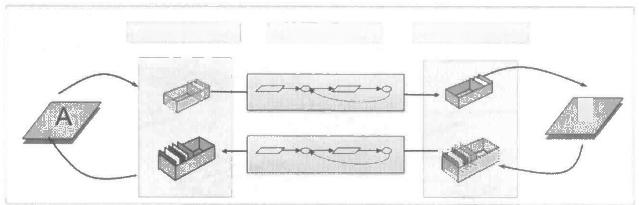
并且会面临以下问题：A、B 两个系统产生的消息格式不一致，接口规范不一致怎么办? A 系 统发出的消息需要在传输过程中增加信息后再传递给B 系统，怎么处理?

**172** 数据治理——工业企业数字化转型之道

**2.** **从消息队列到消息流**

当从 A 系统发送到B 系统的消息需要加工处理时，中间需要增加消息的处理逻辑，这就是

消息流。消息流是企业服务总线技术的核心概念(见图15-2-8)。

队列管理器 消息处理流 队列管理器

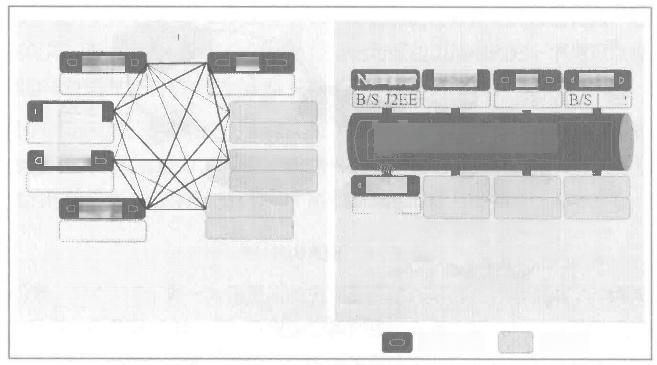
B

**15-2-8** **消息流**

服务总线技术是基于消息队列而产生的，是消息队列技术的扩展。其目标就是解决接口快 速开发和接口泛滥的管理问题。通过服务总线的协议转换、路由分发等技术，可以很好地控制 接口的泛滥，保证接口的重用。同时，服务总线提供了快速开发平台，可以辅助技术人员针对 大量不同的异构系统进行适配。服务总线中有丰富的适配器可以对接不同的协议，其中还有中 间的处理节点，使得接口的开发非常方便、简单。

从图15-2-9可以看出，有了服务总线之后，系统之间的数据交换变得非常清晰明了，系统

之间的接口数量大大减少。通过服务总线可以实现对系统之间的数据交换的管理与监控作用。



总线集成

NC-ERP CAPE OA MES

C/S BS .net

**企业服务总线**

SGRP MDS CRM 新应用

B/S NEW B/S B/S

简化 IT结构，减少接口数量，降低维护 成本松耦合，增加了系统的灵活性和扩 展能力

SOA 基础，为将来的应用提供丰富的可 扩展能力

现有应用 新应用

传统集成 N(N-1)

CAPP

B/S:

NC-ERP

C/S

SCRP

B/S

MDM

B/S

结构复杂，接口繁多，维护成本增加， 紧密集成，系统僵化

可扩展能力脆弱

MES NEW

CRM NEW

EC NEW

OA

B/S:

N

**图15-2-9** **ESB** **技术**

第15章 数据交换与服务 **173**

**15.2.6** **ETL 数据交换技术**

ETL(Extract-Transform-Load, 数据抽取、转换和加载)是基于 SQL 数据库复制批量数据 的技术，底层是数据库的 SQL技术。其在数据仓库项目、数据分析项目中被大量使用，应用场 景主要是从业务系统的数据库到企业数据仓库。它也是一个批量的流式数据加工工具，用于对 数据在转移过程中进行加载、清洗、转换、合并、拆分、补项等操作，以得到更加精确的数据， 便于后续进行建模分析。

ETL 数据交换技术的特点：只依赖于数据库底层，必须对两端的数据库表有精确的认知。 如果业务系统未开放数据库权限，或者对原系统的数据库没有完整的数据库表定义信息，则不 能解析数据，从而不能转换、交换数据。

**15.2.7** **物联网数据采集交换技术**

物联网数据的采集及交换是针对物理设备、自动化系统进行的实时数据采集、加工、交换， 一般被称为 OT 层面的数据交换技术。在该场景下的数据采样频率极高(可以达到毫秒级),数 据量极大。常规IT 层面的数据交换在这个场景中无法应用，需要综合各种数据交换技术来实现。 物理设备数据采集端常见的数据采集传输协议有MQTT和 OPCUA 。MQTT的使用比较广泛， MQTT 与 MQ 消息队列技术是一脉相承的，采用的也是基于消息主题的发布订阅异步传输机制。 物联网数据与 IT 系统中的数据，在数据模型结构上有较大的差别，物联网数据与时间的相关 性较强，所以被称为时序数据，其一般只有3个字段(参量标签、值、时间),常被称为“竖型” 数据结构。而 IT 系统中的数据一般都有多个字段(少则10个，多则上百个),并且都用于管 理，对时间不敏感，常被称为“横型”数据结构。从IoT层面的“竖型”数据结构转换为IT 层 面的“横型”数据结构往往需要进行两次转换(从机器数据转换为技术数据，再从技术数据转 换为业务数据),如图15-2-10所示。

由于物联网数据的采样频率高，所以需要及时处理和转换，不能先存储后处理，因此也就 出现了“边缘计算”的概念。边缘计算也是一个基于消息的流式数据计算处理逻辑，类似 ESB 的 消息流。常见的物联网流式数据处理引擎是NODE-RED, 它是一个基于Node.JS 无阻塞技术的 工具，由 IBM 开源。图15-2-11为一个典型的物联网数据处理边缘计算逻辑结构，物联网数据 处理的典型逻辑是采集、加工、转换、存储、数据服务。我们常见的设备数据展现仪表板并不 是直接从设备上读取并展现信息的，而是从物联网平台的数据库中通过API 读取的。

争

计算 统计 压缩

协议适配器

物职网网关

消息队列

服务网关

**174** 数据治理 工业企业数字化转型之道

设备驱动库

1 **机器数据**

|  |  |  |
| --- | --- | --- |
| 地址编码 | 32位编码值 | 电球时间 |
| D4001 | 00000000000000000110000000000000 | 1:00 |
| D4002 | 000000000001000000000100101100 | 1.00 |
| D4003 | 0000000011000000001111000000000 | 1:00 |
| D4004 | 0001100011001101110101100001100 | 1:00 |
| D4005 | 0001100011001101110101100001100 | 1:00 |
| D4006 | 0001100011001101110101100001100 | 1.00 |
| D4007 | 000110001100110111010110000110 | 1:00 |
| D4008 | 0001100011001101110101100001100 | 1.00 |
| D4009 | 0001100011001101110101100001100 | 1:00 |
| D4010 | 0001100011001101110101100001100 | 1:00 |
| 04011 | 0001100011001101110101100001100 | 1:00 |

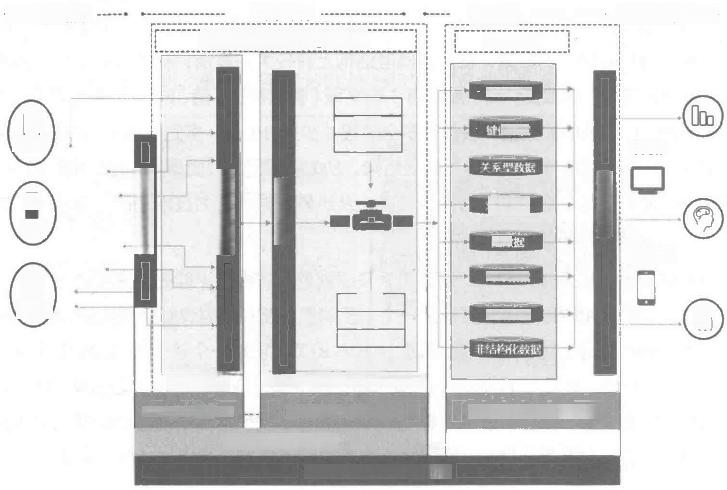
2 **技术数据**

|  |  |  |
| --- | --- | --- |
| 英量参数 | 值 | 时间 |
| 温度 | KXX | 1:00 |
| 压力 | xxx | 1:00 |
| 瞬时流量 | XX | 1:00 |
| 累计流星 | CX | 1:00 |
| 转速 | 8 | 1:00 |
| 功率 | xX | 1:00 |
| 电压 | CX | 1.00 |
| 电流 | xxX | 1:00 |
| 报警1 | XXX | 1:00 |
| 开关状态 | KXK | 1:00 |
|  | XX | 1:00 |

3 **业务状态数据**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 系统标  示 | 组织单位  标示 | 分支机构  标示 | 区域标  示 | 设备标  示 | 参数1(温  度) | 参数 (压力) | 报数 流量) | 参数4 (转速) | 参数5 (功率) | 点 |
| X0QOOX | XXXXX | XNOOOX | XDOOX | 00OOX | XOOCX | XOOX | AOOOX | XXXXX | X0OO\* | 1:00 |
| NO | XXXXX | XXXXX | NOXXX | XXXXX | xXOXX | XXXX | XXXXX | XXXXX | xxXQX | 2:00 |
| XXXXX | XXXXOX | XXXXX | XCOXX | XXXXX | QOOOX | XXXXXX | XXXXX | XXXXX | XXQOX | 3:00 |
| XXXXX | X0OX | XXXXX | CQCXX | OXXXX | XXXCX | XXXXX | XXXXX | xXXXX | XXXXX | 4:00 |
| XOXXX | XQCOXX | XXXXX | XXXXXX | XXXXX | XXXXX | XXXXX | XOCXXX | XXXXX | XCQX | 5:00 |
| XXXX | XXXXX | XXXXxX | XQCxx | XXXXX | xcOOX | XXXxX | OCXX | xOOXX | XCOOX | 6:00 |

**图15-2-10** **物联网数据转换图**

--…-- **数据源**

**边缘计算**

设备磁动(数操测)DEV DRIVER

数据湖 ---- - **数据应用** ……

设备能力化

目

(scnba

日

pLc

目

(传感器

连接

解析

转换

计算

消息流处理引擎

4

规则

活动

事件

时序数据

值数展

对段数规

图数

内存数据

文件数据

Web

**Mobile**

连接控制黑

Data CaptueDiiwetr

感知

计算观则+威知规则+复杂计算消息流

Edge Compute Drtiver

dev Ccnter Cbud Server

kdgetetplte seivar

devBus 一体机(林有云部署)

图15-2-11 物联网数据处理边缘计算逻辑结构

s o s 级 C P S 综 合 集 成 管 理 系 统 建 设

API发布组件建设 P 脂费组性建设

2.3

第15章 数据交换与服务 **175**

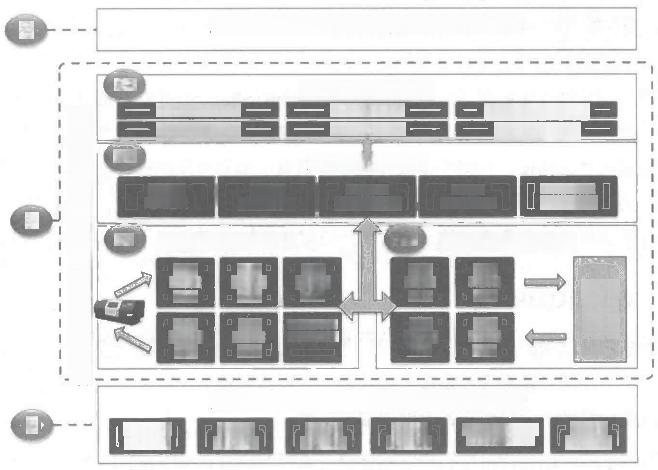
**15.3** **工业企业数据交换与服务标准体系架构**

数据交换与服务技术发展到今天，已经不是某一个单一技术所能覆盖的范畴，而是一个整 体性的解决方案，对于不同层面和不同场景，要有不同的技术作为支撑。IT 层面的数据交换与 服 务 ，IoT 层面的数据交换与服务， IT/IoT 层面融合的数据交换与服务，以及企业外部的数据 交换与服务等，都需要数据交换与服务。如何建立标准的数据交换平台，针对不同的层面和场

景，企业如何建立标准的数据交换与服务体系架构，显得尤为重要。

图15-3-1是企业基于总线的数据交换共享标准体系架构。对工业企业而言，如果需要建立

完善的数据交换与服务体系，则应该从以下几点开展。

3

**工业SoS级CPS** **综合集成管理系统标准建设**

2

1

2.4

**开放互联总线建设**

|  |  |
| --- | --- |
| API运行层建设  AP 管理层建设  **主数据管理建设** | API分析监控组件建设  P 安金银件建设 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 主数据  集成建设 | 志数据 建板原建设 | 主数据质  量管理层建设 | 主数据  标识层建设 | 主数据教据 管理层建没 |

2.1 **设备互联总线建设** 2.2 **应用互联总线建设**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **数据**  采集层  建设  时层  建设 | 信号 处理层 建设  逻组 执行层 建设 | **机理**  **分析层**  建设  工业互联网  平台层重设 | 消息 传输层 建设  监控 管理层  建设 | 协 议  转换展  建设  服务  定义层  建设 | 企业 信息 系统 |

**工业SoS级CPS综合集成管理系统体系建设**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 提供CPS  解决方案 | 标准化  体系建设 | 装成化  体系建设 | 组件化 体系建设 | 工业块义与机模 型支持体系建设 | 工具化 体系建淀 |

图15-3-1 企业基于总线的数据交换共享标准体系架构

一是工业 SoS 级(即系统之系统级，是多个系统级 CPS 的有机组合)CPS (信息物理系统， Cyber-Physical Systems) 综合集成管理系统体系建设。通过系统体系建设，明确各工业企业 CPS 解决方案的具体要求，提供具体的建设内容，使 CPS 解决方案具备标准化、集成化、组件化、

工具化的特性，为企业提供标准化的 CPS 建设方案。

二是 SoS 级 CPS 综合集成管理系统建设。通过管理系统建设，实现企业的设备与信息系统 间的互联互通、企业数据的标准化云集成、企业数据对外部的安全开放，促进企业 IT 与 IoT 的

融合，支持企业智能制造、工业互联网的建设。

**176** 数据治理——工业企业数字化转型之道

三是工业 SoS 级 CPS 综合集成管理系统标准建设。根据项目建设经验，结合国内 CPS 标 准建设基础，制定企业内部 CPS 通用要求类、应用类标准规范。

**15.3.1** **CPS 信息交换模型**

异构服务经过封装“管理壳”的标准定义后，通过路由调度程序，可智能地将服务传递到 任何指定的终端地址。消息通过消息流和队列进行稳定、有效的传输，保证了系统的稳定性和 松耦合。

消息由提供方和消费方进行传输，消费方会将自己的消息传递给集成平台。集成平台通过 标准的“管理壳”给消息进行统一标准的封装。在将其封装成通用的 Administration Shell 消息 后，再通过标准的路由传输组件将消息动态路由到指定的提供方处。在出口处，自动将 Administration Shell 消息的标准报文解析处理以提供方所需要的服务格式进行请求。而提供方 拿到参数进行相关处理后将返回的结果同样以Administration Shell 标准报文进行封装，进入消 息传输层，然后自动返回给消费方。消费方再将其解析成自己认识的业务报文，即可完成一次 标准的数据服务共享传输。在整个传输过程中，消息传输层使用的核心组件为消息流及队列服 务。其中，通过采用通适技术，使得相同协议的服务共用相同的消息流，节省了消息流的资源 空间，提升了消息流的利用率。目前主要应用互联总线设计标准的、通用的消息流模板。

**15.3.2** **设备互联总线**

设备互联总线解决的是 OT(运营技术)层面设备之间，以及设备与控制系统之间的数据连 接性问题。由于现场设备存在数量多、协议杂，现场情况复杂、数据小但数据传输频繁等实际 问题，因此企业需要一个总线平台能处理相关问题。

设备互联总线可以利用标准的物联网协议，诸如MQTT 、MODEL BUS 、OPC等，采用队 列等相关技术(如Kafka) 对大量数据进行相关异步传输，再采用相关的流式开发等技术实现 边缘计算，最终实现设备层的数据交换与服务，如图15-3-2所示。

SCADA PLG.

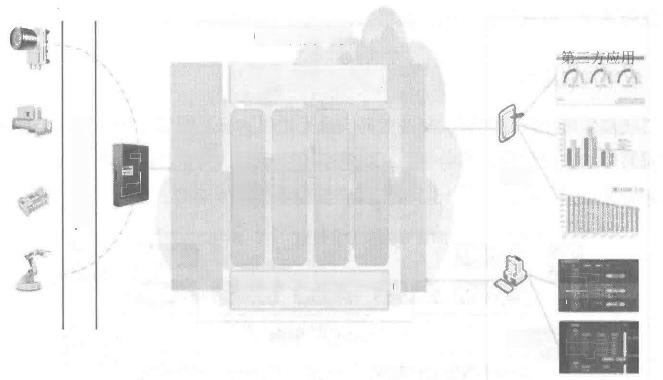
数据服务

数据存储

数据处理

数据采集

第15章 数据交换与服务 **177**

CPS设备互联

虚拟设备定义管理

MQTT

API

Modbus

OPC

APP

系统管理(基础管理)

后台

图15-3-2 设备互联总线

**1.** **设备连接**

首先要针对不同设施、设备的连接方式，采用统一的连接协议，建立连接的标准。这些连 接标准，会使用不同的通信方式(4G/5G、千兆网等)与连接协议(Modbus TCP、OPC DCOM、 MQTT、OPCUA 等),连接不同厂商(西门子、欧姆龙、沃特世、安捷伦、吉尔森等)制造的 设备。

设备互联总线中的设备连接，通过数字资产建模及收集设备连接属性，为设备实体建立虚 拟驱动，完善知识图谱的信息采集，定义两个世界中同一实物的数字孪生体。

**2.** **数据采集**

实现数据治理、共享的前提是必须把已经连接好的设施、设备在生产过程中产生的运行数 据统一保存在物理存储设备中。随着时间的推移，设备连接越来越多、产生的数据也越来越多， 这为以后的大数据分析奠定可塑的环境。在数据采集阶段，业界指定了数据采集格式：使用JSON

数据传输格式作为标准。

**3.** **数据服务**

采集来的数据被存储在物理设备中，想要使其能共享，则必须开放数据服务。数据服务开 放标准通过RESTful 应用接口方式，提供数据下行通道。数据服务可提供监控设施情况、设备

实时监测运行情况、运行状态、能耗值、报警信息等，监测企业中已实现的生产流程、服务运

服务路由 安全控制 缓存管理

日志记录

Sitet 适配

People sof适配

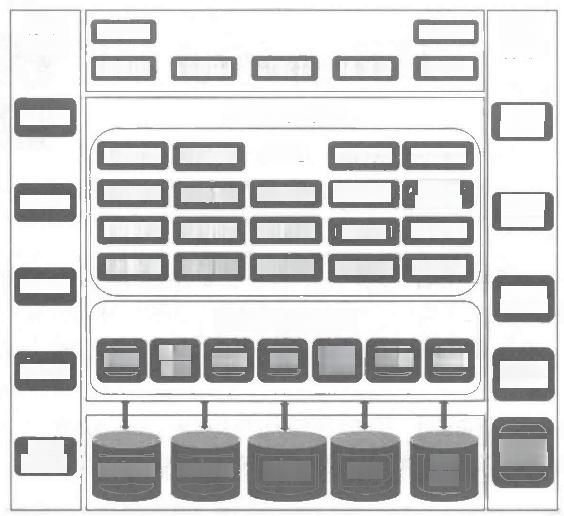
SAP适配

JDE适 配

营流程等业务活动，并辅助产业链实现精细化作业，保障智能化运营可持续进行。

**15.3.3** **应用互联总线**

服务总线是在应用层面为解决某些业务问题而提供的标准服务平台。其目标是解决业务数 据交换与服务中的服务管理、服务开发、服务监控等问题，为企业的业务集成及整合，提供标 准化的服务管理及运行平台。图15-3-3为服务总线平台整体架构。

**标准** **规范**

**编码规范**

**接入规范**

**报文规范**

**服务规范**

统计分析 授权管理 服务管理

服务监控

异常根警

统计管理

日志管理

**ESB运行平台**

|  |  |  |
| --- | --- | --- |
| 发布订阅 格式转换  数据计算  文件处理 | **功能**  异常处理 可靠传输  服务组合 | 协议转换 数加以东 定时器  数报库管理 |

**多协议接入适配**

ODBC

MQ/ JMS

TCP/IP **SMTP**

SOAP FTP

HTIP

JDBC

**治理**

**体系**

服务梳

理流程

服务接

入流程

照考明用

申蔬程

**业务系统**

**注册流程**

DB 设 计规范

订单系统

资金系统 SRM **ERP**

其他

系统

运维异常

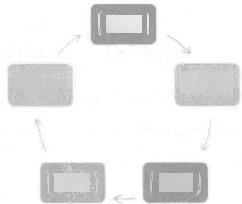
处理流程

**图15-3-3** **服务总线平台整体架构**

**1.** **服务管理**

当企业中存在大量的数据服务时，如果没有一个标准化的平台对其进行管理，则数据服务 将会非常混乱。服务总线提供了一套标准化的管理功能，可以对企业的数据服务或者业务服务 进行有序的管理，还可以通过不同的维度查看服务。服务总线对服务的分类、服务的全生命周 期管理、服务的授权等有一套标准的权限管理机制。图15-3-4所示为服务安全生命周期管理。

第15章 数据交换与服务 **179**



创建

待发布

挂起

发布

退役

**图15-3-4** **服务全生命周期管理**

**2.** **服务开发**

为了快速满足不同业务场景的需要，随时响应市场需求的变化，企业需要一款快速的服务 开发工具，以保证服务的快速开发和重组。通常借助消息流的开发方式，并结合可视化的开发 工具及利用不同的开发组件，可以快速地实现不同异构的服务开发。

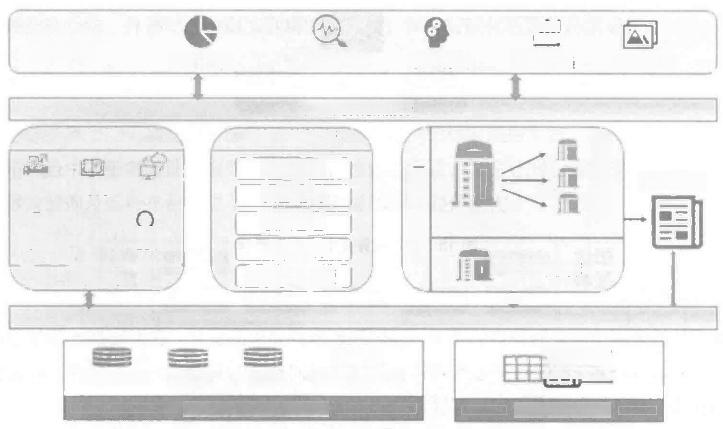
**3.** **服务监控**

服务监控是服务总线的核心。企业中存在大量的异构系统，如果想让服务能适应不同的协 议要求，那么服务总线必须有足够丰富的适配器。同时，服务总线作为企业系统集成的神经中 枢，必须要性能稳定、并发量大，且任何一笔服务交易都要有日志记录，做到服务交易可追溯。

**15.3.4** **数据总线**

数据总线是数据层面的定时或者实时的数据交换和共享平台。其作用是通过数据总线解决 企业中各业务系统数据库层面的数据整合，加强数据管理，提高数据质量，最终为企业的智能 分析和综合决策提供精确的数据支撑。图15-3-5为数据总线平台整体架构。

**180** 数据治理——工业企业数字化转型之道



**数据分析和**

**综合应用**

统一的数据服务

系统运维管理

运行监控

安全控制

权限管理

日志管理

吝灾备份货复

工

**数据交换平台**



SQL Server Orade DB2 xls、ba等文件

**各个系统的应用数据**

国心

**丰结构化数据**

**青洗**

采

存储

**适配器**

同步

查找

山

**数据报表**

**数据集成工具**

元 數 据 管 理

**交换数据库**

**共享数据库**

领 域 主 题 库

信息门户

**应用系统**

综合查询

**决策分析**

**OLAP分析**

**抽取**

缓存



图15-3-5 数据总线平台整体架构

**1.** **数据采集**

数据总线通过数据交换组件中丰富的数据采集节点对各种类型的数据进行采集，如结构化 数据、非结构化数据。采集的数据可以通过相关的数据加工处理形成标准化数据，也可以进行 存储。

**2.** **数据处理**

数据总线提供数据处理节点，对采集的数据可以进行特殊处理，如清洗脏数据、数据计算、 数据合并等，同时还可以结合更复杂的业务处理方法对数据进行转换，直接将原始数据加工成 标准数据。

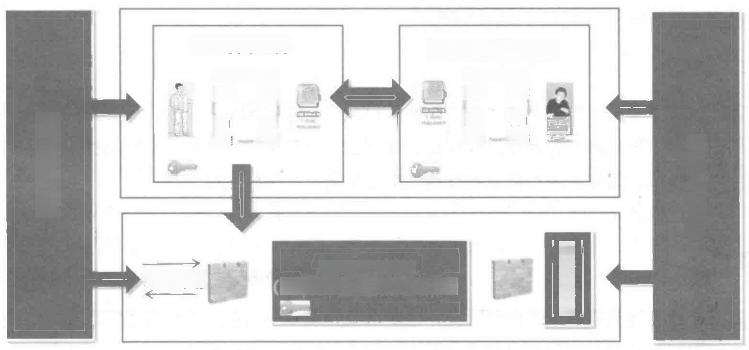
**3.** **数据服务**

数据服务是将数据以业务单元的颗粒度对外提供业务所需的服务，如对于供应商，则需要 提供供应商查询服务，该服务用于对外提供标准化的企业供应商信息查询；对于月度生产报表， 则需要提供月度生产报表查询服务，该服务用于向业务人员提供月度生产报表查询。对于不同 的业务需要，需要提供不同的业务服务。因此，当企业中有不同的业务需要时，会需要大量的 数据服务来供业务消费。数据服务可以被分解不同的颗粒度，小颗粒度的数据服务可以被重组 成颗粒度更大的数据服务。

第15章 数据交换与服务 **181**

**15.3.5** **开放互联API 网关**

API 网关是企业业务服务的管理和运行平台，和服务总线的功能类似，但没有服务总线那 么“重”。随着微服务架构的兴起，大量的 API 不断产生，但都处于无管理状态，因此，企业 需要一款平台对 API 进行有序的管理，同时能对访问的 API 做到安全控制，这就是 API 网关， 如图15-3-6所示。



**存储组件**

API—API

发现 研究

API API

订阅 评估

API网关组件

(安全、节流、服务等级)

|  |  |
| --- | --- |
| 监控  API  组 | 开发  API  发布 |

**API发布组件**

监控分析组件

**API调用**

后端服务

k e y 管 理 组 件

**API**

API—API

图15-3-6 API 网关

在图15-3-6中可以看到，其中涉及专门的安全认证的 Key 管理组件、API 发布和存储组件， 以及 API 网关组件及调用 API 的监控分析组件。这几个组件构成 API 网关的基础功能，保障 API 网关的正常运行，具体介绍如下。

**1.** **管理控制**

管理控制主要包括 API 的发布和管理。在企业中通常由了解 API 、接口、文档、版本等 技术的人来完成 API 开发，而由了解 API 业务的人进行 API 管理。在大多数商业环境中， API 开发者的职责是不同于 API 发布和管理者的。API 在发布之前有一个开发组件，它有一个 结构化的图形用户界面，用于API 的开发、文档维护、测量及版本维护，也便于API 管理等， 例如发布 API、收费、统计分析及升级。

管理控制还包括 API 授权管理。消费组件应用程序为 API 发布者提供协同接口来发布和宣 传他们的API, 以及让API 消费者可以自主注册、发现、评价 API, 以及使用安全的、受保护

的、被认证的API。

**182** 数据治理——工业企业数字化转型之道

管理控制还包括 API调用控制。在API的调用过程中，可以设置API 的收费模式、节流控 制、熔断机制及相关安全措施。

**2.** **容器化**

随着微服务架构的兴起，很多业务系统均以微服务的方式被部署在容器上面，通过容器机 制提高系统的业务安全性及高效性。作为微服务的管理平台，API 网关不但要能部署在容器化 技术之上，还要能与容器化技术很好地结合。

**3.** **安全性**

大部分企业使用API网关用于跨组织的数据交换，尤其是与上下游供应链的数据交换。由 于与供应链或者其他生态的数据交换是在互联网上进行的，因此，其对数据交换与服务的安全 性要求就比较高，如要有黑白名单的支持、熔断机制、节流控制等。

**本章精要**

数据交换与服务是数据的价值所在，也是最低要求。任何企业不管是想实现互联协同，还 是想要实现智慧运营，实现企业各个业务环节的数据交换、共享是基础。从企业内部的纵向连 接、横向连接、端到端之间的连接，再到外部的生态互联等，都需要数据的传递和共享，只有 数据在各个业务域中都能有序地传递和应用，企业才能实现互联协同和智慧运营。

在本章中，介绍了在各个时期使用的相关的数据交换与服务技术。随着技术的不断发展， 数据交换与服务技术将随时变化，但是数据交换与服务的目标是永恒不变的。任何技术的发展 都离不开对应的业务场景和业务需要，不同的层面和不同的场景使用的数据共享技术完全不同。 采用对应的交换技术，遵循相应的标准规范，借鉴最佳的技术实践，才能将数据交换与服务技 术在企业内外使用得更合理、更标准。

第16章

数据共享与开放

数据作为信息的载体，其本身的流动就会带来跨领域信息的传递、融合，有助于原有领域 知识的普及和新知识的产生，进而催生出更多的数据创新应用。与此同时，数据开放本身也会 带来数据交易的机会，从而更好地激发商业模式的升级和发展。综合来看，共享与开放无疑是 数据资源创造价值的关键举措和重要手段。

**16.1** **共享与开放概述**

企业的数据共享与开放在通常意义上分为“数据共享”和“数据开放”两个概念。其中数 据共享主要指的是面向企业内部的数据流动，其中由数据应用单位提出企业内部跨组织、跨部 门的数据获取需求，由对应数据供给单位进行授权，并由信息部门向该数据应用部门开放数据访 问权限。而数据开放则指企业向政府部门、外部企业、组织和个人等外部用户提供数据的行为。

在数据共享与开放的过程中，主要参与的角色可以分为4种：数据拥有者、数据消费者、 数据服务者和数据运营者，如图16-1-1所示。

(1)数据拥有者：通常是指数据的合法拥有方，在数据共享中，则特指信息系统的业务管 理部门及单位。其负责在日常业务活动中，组织人员在信息系统中录入数据，或合法获取外部 数据并提供使用。

(2)数据消费者：在数据共享中，是指发起数据共享需求申请并使用数据用于开展合法、 合规业务的内部部门及单位。在数据开放中，则是指发起数据开放需求申请并使用数据用于开 展合法、合规业务的外部单位，包括政府单位、外部企业或个人。

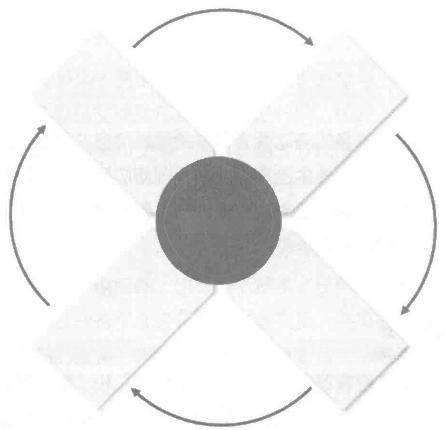
**184** 数据治理——工业企业数字化转型之道

(3)数据服务者：负责在数据拥有者给出的数据资源基础上，根据数据消费者可能的使用 需求，提供各类服务，如将原始数据加工为应用产品、提供数据交易过程中的代理服务、针对

数据真实性或有效性提供验真服务、对数据开放过程的合法、合规性提供审计服务等。

(4)数据运营者：负责提供一个支持数据共享与开放的环境，如统一的服务平台、标准化 的数据产品、数据资源目录查询检索等，以及开展以创造经济价值为导向的运营活动，如客户

管理、订单管理、营销宣传等。

数据运营者

数据服务者

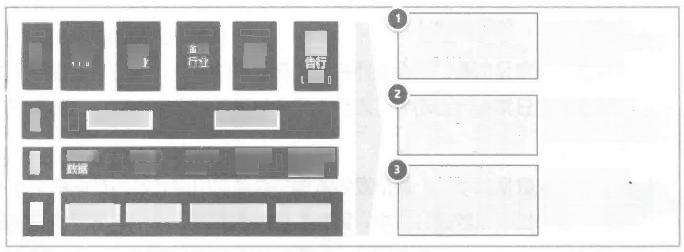
**数据开放**

数据消费者

数据拥有者

图16- 1- 1 数据共享与开放过程中主要参与的角色

从实践案例来看，当前主要存在3种开放业务模式，如图16-1-2所示。



应用商店

行业模型，数据产品

应用开发 应用注册

行型

应用下载

数据服务API

OpenAPI

API 注 册 API 调 用

API 路 由 API 接 入

数据工厂

数据开发PaaS

到数据工厂跟企业数据进行 融合加工提取信息

|  |  |
| --- | --- |
| 位置类  在线使用 | 行为类  数 据  远程调用 |

SaaS: 软件即服务。

利用企业数据能力构建应 用供第三方使用

DaaS: 数据即服务。

适用于需求方通过API 查 询调用企业内部的数据

PaaS; 平台即服考

适用于第三方自有的数据

行 业 角 色 数 机 方 试

开发者门户

数据目录

特定县标 沪群

数据融合 应用开发

互联 同广

业

您 普 铁 数 据

资科类 数 据

批显数据关联

应用开发者

应用使用者

施 薄 行 业

事件协助

签

*政旁*

融

图16-1-2 3种开放业务模式

第16章 数据共享与开放 **185**

( 1 ) 提 供SaaS (软件即服务)开放模式，即通过数据共享与开放，开发并发布数据应用产

品，供企业外部用户在线使用。

●应用的所有权：包括企业内部的应用和合作伙伴的应用。

●使用方式：企业外部用户在线使用应用。

●模式：作为数据产品对外定价打包销售。

(2)提供DaaS (数据即服务)开放模式，即将数据封装为API, 提供给企业内/外部系统或

开发者调用。

●数据来源：企业内部共享的数据、外部开放的数据。

●使用方式：应用系统或开发者调用。

●模式：按次计量。

(3)提供PaaS(平台即服务)开放模式，即第三方将自有数据加入企业提供的开放环境中，

跟企业数据进行融合、加工后提取其中的信息，满足业务应用的分析需求。

●数据来源：企业内部共享的数据、外部开放的数据。

●使用方式：第三方开发者在企业数据中心内访问数据。

●模式：根据议价合同。

**16.2** **数据资源目录**

数据资源目录是依据规范的元数据描述，对企业数据资产进行逻辑集中管理的一种方式。 通过编目形成的数据资源目录中含有各种数据资源的描述信息，便于用户对数据资源的检索、 定位和获取，并提供数据资源显性化的应用入口，真正实现数据的可见、可管、可用。

基于数据资源目录的对外服务，主要是面向企业数据的使用方进行企业数据的访问、获取 等，包括用户对元数据的统一检索，以及对数据的查询服务等。其数据服务形式包括数据使用 者直接登录平台进行数据访问、第三方系统通过接口等方式进行数据获取等。各种访问方式均 受平台统一的权限控制，需要进行访问申请。

编制数据资源目录是启动数据资源共享与开放服务的第一项任务，本阶段的工作成果是后

续各项工作的基础。总体来看，面向共享与开放服务的数据资源目录编制工作包括以下内容。

**186** 数据治理 工 业 企 业 数字化转型之道

**1.** **研究数据资源梳理方法**

对当前企业现有数据资源进行分析和梳理，制定共享与开放数据资源梳理的流程和方法，

包括梳理目标、梳理范围、梳理原则、组织形式、流程步骤、工作要求等。

**2.** **编制数据资源目录**

按照企业制定的相关数据标准，如元数据标准、数据共享与开放管理标准等，开展企业数 据资源的梳理，形成用于共享与开放的数据资源目录。

**3.** **分析数据集的元数据**

针对每一个数据集，分析相关元数据信息，包括但不限于数据集编号、数据集名称、数据 集类型(结构化、非结构化、半结构化)、数据集摘要、数据集关键字、数据领域、主题分类、 数据更新频度、数据提供方单位、数据提供方地址、数据提供方联系方式等。

**4.** **确定数据集的数据逻辑模型**

数据逻辑模型包括数据项英文名称、数据项中文名称、数据项类型、数据项大小、可否为 空、是否主键等。

**5.** **确定数据集的采集方式**

要确定每个数据集通过何种方式进行数据采集，例如，从生产系统采集、从数据中心采集、 人工采集上传数据等。

**16.3** **数据资源准备**

数据资源准备是指将未做处理的原始数据经过加工处理(加工内容包括清洗、比对、脱敏、 分类、打标签等)后形成可开放的数据集并具备共享与开放条件的过程。

数据资源准备主要分为数据采集、数据加工、数据保密、数据装载和数据发布5个过程，

下面介绍具体内容及工作要求。

**16.3.1** **数据采集**

本阶段的任务以共享与开放数据资源目录为目标，采集可以用于共享与开放的数据资源。 除个别数据因条件限制外，数据采集应实现自动化数据抽取、修正或者补录过程，为数据存储

第16章 数据共享与开放 **18 7**

或数据分析提供基础内容。

自动化数据采集包括如下能力。

(1)支持将日、周、月、不定期、实时数据的加载入库。

(2)支持信息安全，如对数据进行加密、脱敏等。

(3)支持结构化数据处理。

(4)支持非结构化数据处理。

**16.3.2** **数据加工**

本阶段的任务是对已采集的数据进行清洗、转换、比对和质量检查等加工操作，从而使得 加工后的数据具备可用性，确保被共享与开放的数据能够满足数据消费方的需求。

对数据加工操作的主要要求介绍如下。

(1)数据清洗：过滤那些不符合要求的数据，主要包括不完整的数据(如身份证字段为空) 或是错误的数据(如字段中存在乱码)。

(2)数据转换：主要对数据进行字段的枚举值转换、空值转换，或是基于规则的计算等。

(3)数据比对：主要对数据进行业务逻辑校验，检查数据的关键数据项是否符合业务规则，

或按照统一标准对不同数据集中业务含义相同的数据进行一致性检查。

(4)质量检查：提供检查数据质量的手段。例如，在数据上线时，对数据进行稽核检查， 保证数据信息的完整性、合理性。

**16.3.3** **数据保密**

由于数据涉及隐私、机密等内容，在共享与开放过程中，需要提供安全防护操作，做到有 组织、有保障、有分级、有步骤的数据共享与开放。

数据安全防护操作通常有以下3种手段。

**1.** **数据权限控制**

数据权限控制是指对用户进行数据资源可见性的控制，通俗的解释就是符合某个条件的用 户只能看到该条件下对应的数据资源。最简单的数据权限控制就是用户只能看到自己的数据。 而在实际的系统环境中，会有很多更复杂的数据权限控制需求场景，如领导需要看到所有员工

**188** 数据治理——工业企业数字化转型之道

的客户数据，而员工只能看自己的客户数据等。

**2.** **数据脱敏处理**

数据脱敏处理是为了防止用户非法获取有价值的数据而加设的数据模糊化处理手段，从而 保证用户根据其业务所需和安全等级，适当地访问敏感数据。

数据脱敏处理方式包括如下7种形式。

(1)数据替换：以虚构数据代替真实值。

(2)截断、加密、隐藏或使之无效：以“无效”或“\*\*\*\*\*”代替真实值。

(3)随机化：以随机数据代替真实值。

(4)偏移：通过随机移位改变数据。

(5)字符子链屏蔽：为特定数据创建定制屏蔽。

(6)限制返回行数：仅提供可用回应的一小部分子集。

(7)基于其他参考信息进行屏蔽：根据预定义规则仅改变部分回应内容。

**3.** **数据加密处理**

数据加密处理是通过技术手段对现有数据进行密码设置，保证数据无法被非授权人员获取、 破解。在数据共享与开放的过程中，需要根据数据敏感度等，支持分级的加密方法：可分别进 行不加密、部分加密、完全加密等不同策略。

常规的数据加密处理方式主要有对称加密和非对称加密。

(1)对称加密。

① 加密和解密使用同一个密钥；

② 加密速度较快；

③缺点是秘钥需要在网络过程中传输，可能会被泄露。

(2)非对称加密。

① 非对称加密有两个密钥，分别为公钥和私钥， 一般用公钥进行加密，用私钥进行解密；

② 非对称加密的速度相比对称加密来说要慢很多。

第16章 数据共享与开放 **189**

**16.3.4** **数据装载**

数据装载也就是数据入库。本阶段的任务是将经加工处理后满足数据共享与开放的质量及 安全要求的数据，存储至指定的数据库或相关存储环境中。

数据加载包括文件加载、流加载、不落地加载等。数据加载功能需要具备将采集、处理后 的数据源文件保存到不同数据库中的能力，具体的功能描述如表16-3-1所示。

**表16-3-1** **数据加载管理要求**

|  |  |
| --- | --- |
| **功** **能** | **功** **能** **描** **述** |
| 支持多种加载模式与策 略定义 | 具备全量、实时、双加载；  允许灵活定义加载策略；  允许对加载事物提交过程进行自定义配置；  支持在加载过程中断点续传 |
| 支持文件落地和不落地 两种存储加载 | 落地加载是将数据源保存在ETL物理服务器中而实现的加载。不落地加载是指将数据 源写入缓冲池中，不在物理服务器中保存而实现的加载 |
| 支持自动和手工两种加 载方式 | 支持数据自动加载的设计与执行；  当数据加载出错时，应提供操作界面以人工干预的方式来重新启动数据的接收和加载 |
| 支持多任务并行加载 | 具备支持数据的并行装载，即支持多个数据库连接同一个装载任务的并发执行 |
| 支持加载对象的参数配 置 | 具备加载对象的参数配置功能，将数据加载过程中需要设置的命令、参数、规则进行配 置，控件会自动生成相应的可执行代码来完成作业 |
| 支持过滤 | 具备基于数据属性值的过滤加载 |
| 支持脚本加载事务处理 | 在加载实现过程中支持提供SQL、HQL、SHELL等不同类别的行为定义脚本，数据加 载执行组件将根据定义行为脚本类型调起相应的脚本执行来加载数据 |
| 支持加载到异构数据库 | 支持加载多种数据库接口 |
| 提供丰富的数据加载作 业执行状态监控管理能力 | 提供丰富的图形化界面设计和监控数据加载过程执行状态 |
| 支持数据加载过程的日 志记录 | 支持在数据加载过程中对数据记录条数、开始时间、完成时间、错误信息等进行记录和 保存 |

**16.3.5** **数据发布**

数据发布阶段的任务是将进入指定存储环境的数据资源，通过门户或数据共享与开放平台 向数据消费者发布。根据不同的服务形式，数据发布的内容可以包括数据集、元数据、数据文 件、数据应用链接、数据开放接口等。

**190** 数据治理 工业企业数字化转型之道

**16.4** **数据服务**

数据共享与开放的实现需要建设数据服务封装能力，通过文件、接口、推送等多种数据服 务形式为数据消费者提供灵活、可靠的数据供给能力，提升数据共享与开放的便捷度和流通效 率，同时也避免将原始数据完全暴露在数据消费者面前，实现数据的“可用不可见”,并支持运 营管理过程中进行的监测、管控及优化处理。

数据服务的主要方式包括以下8种。

(1)数据集：数据的集合，通常以表格形式出现。数据集的服务方式就是通过数据库批量 导出部分数据明细，并提供给数据需求方。

(2)API 接口：预先定义的函数，提供基于软件或硬件得以访问一组例程的能力。API 接 口具备体量轻、使用方式灵活、可管控等优点。众多企业均选择 API 接口为最主要的数据服务 方式。

(3)数据报表：根据规定的业务逻辑，通过简单的统计处理，以数据集合或图形的方式将

结果展现出来。

(4)数据报告：对数据进行深度加工，并基于数据分析，加上文字或图表解释，将数据反 映出的规律和问题展示出来。数据报告提供的是一种知识。

(5)数据标签：对一组数据的基本特性或共同特性的提炼。在数据挖掘或数据分析过程中 可以通过数据标签直接获取符合相应特性的数据集。

(6)数据订阅：通过统一、开放的数据订阅通道，使用户高效获取订阅对象的实时增量数 据。其中包含业务异步解耦、异构数据源的数据实时同步，以及包含复杂 ETL技术的数据实时 同步等多种应用场景。

(7)数据组件：具备特定数据处理逻辑的工具，可以根据需要直接处理数据或作为数据应

用的调用对象。

(8)数据应用：数据服务的高级形式。数据应用将数据通过功能、程序进行处理后，通过 自身的界面展示出来，其可以实现复杂的数据处理和多样化的界面呈现。

**16.5** **共享与开放评价**

数据共享与开放是一项涉及多个部门，涵盖业务、技术和管理多个方面的复杂工作，只有

第16章 数据共享与开放 **191**

建立对于数据共享与开放过程及效果的合理评价体系，才能有助于理顺数据共享与开放过程中 的各种关系，确保数据共享与开放工作高效、有序地开展。为此，企业需要根据系统、科学的 理论，结合数据资源的基本特性，以及数据共享与开放的发展目标来建立相关的评价体系。通 常来说，可以围绕以下4个方面来展开工作。

(1)对数据资源目录的编制过程进行评价，主要从数据资源目录的业务覆盖率、完整性、 规范性等方面来评价是否符合相关规定。

(2)对数据资源目录的内容和应用效果进行评价，主要从数据资源共享和开放工作的落实 状况，评价相关执行方的工作成效，包括共享与开放数据的质量、更新频率等。

(3)对数据共享与开放的组织管理能力进行评价，从管理中最主要的3个方面(制度、流 程、人员)来建立指标体系，并评价管理举措是否落实。

(4)对共享与开放数据资源的应用效果进行评价，主要从数据资源通过共享带来的协同效 果，或是通过开放带来的经济效益等展开综合评价。

**本章精要**

本章从数据共享与开放的需求出发，介绍相关概念、模式和方法，进而围绕数据资源目录、 数据资源准备、数据服务、共享与开放评价等工作阶段来介绍相关职能活动和管理要求，探讨 在数据共享与开放过程中需要解决的关键问题和应对思路。从而帮助企业更好地理解数据开放 与共享的过程，更有效地推动数据资源在企业内、外部的有序流动，释放基于数据跨领域融合 带来的信息红利，推动各类数据应用的创新发展。

**第17章**

**数据管理成熟度评估**

通过对本篇以上章节的学习，相信你已经掌握了数据治理体系中各个职能领域的知识，可 以根据各章节提供的方法论，结合企业实际情况进行落地实施。那么,最终会做得怎样?存在 哪些不足?在同行业处于什么水平?如果你也有这样的疑问，那么本章将为你——解答。通过 本章的学习你可以掌握如何了解企业当前的数据管理现状，识别数据管理能力的不足之处，找 准关键问题和目标差距，提出数据管理能力的改进建议和方向，规划未来数据管理路线图，以 便更好地利用数据提高业务绩效。

数据管理成熟度评估(Data Management Maturity Assessment,DMMA) 是企业在提高数据 和信息质量方面能够用来绘制和评估其进展的一个重要工具。DMMA 是为实现数据资产治理的 预期目标而必须采取的一项重要举措。当企业开始将数据视为最关键的企业资产之一时， DMMA 提供了一种方法来评估企业现状及将数据管理上升到所需的最终状态需要做的工作，它 与成熟的数据治理规范相对应。DMMA 也有助于企业规划近期可行的数据管理方案，尤其是在 企业面临严重的资金压力时。

**17.1** **数据管理成熟度评估模型**

国内外很多机构已经提出了许多数据管理成熟度模型，包括 EW Solution 数据管理成熟度 模型、Gartner EIM成熟度模型、IBM数据管理委员会成熟度模型、CMMI数据管理成熟度模型、 MDM 研究所数据治理成熟度模型、Oracle 数据管理成熟度模型等，每一个模型都有一定优势， 都可以带来一些有价值的建议。DMMA 模型是在借鉴成熟模型和先进经验的基础上，结合最新 的数据管理理念和标杆企业的最佳实践经验设计而来的。

第17章 数据管理成熟度评估 **193**

一般的数据管理包括8大领域，具体介绍如下。

(1)数据管控是数据管理框架的核心职能，是对数据管理行使权利和进行控制的活动集合， 数据管控涉及数据管理的组织、战略等多个方面。

(2)数据架构是用于定义数据需求，指导企业对数据资产的整合和控制，使数据投资与业 务战略相匹配的一套整体规范。

(3)数据质量是指数据的适用性，描述了数据对业务和管理的满足度。

(4)数据安全是指企业中的数据受到保护，没有受到破坏、更改、泄露和非法的访问。

(5)数据生命周期是指数据从采集、传输、存储、处理、交换与共享到销毁的整个过程。

(6)数据价值挖掘是指通过对企业中的数据进行统一管理、加工和应用，支持企业的业务运

营、流程优化、营销推广、风险管理、渠道整合等活动。

(7)数据资产运营包括数据共享与开放、数据服务等活动，可以提升数据在企业运营管理 过程中的支撑辅助作用，同时实现数据价值的变现。

(8)支撑平台是数据管理的IT 技术支撑工具，包括数据管理平台、数据治理工具集及大数

据平台。构建强大的支撑平台可以提高数据管理的效率、效果，避免线上、线下“两张皮”。

数据管理的8大领域，可以被进一步细化为30个评估的核心要素(见表17-1-1),这些核 心要素都是数据管理必不可少的重要组成部分，可以传达评估模型的目标、要求，量化和产出 成果。这8大领域与本篇前几章所讲的内容并不完全相同，这是因为数据管理能力成熟度评估 主要评估企业的数据管理能力，所以包括了企业中所有类型的数据，笔者认为时序数据、指标 数据和结构化数据都可以采用相同的方法论。在表17-1-1所示的在数据架构管理领域中将主数 据和元数据单独提出来，是因为它们所需要的管理能力不同。

**表17-1-1** **数据管理成熟度度评估维度**

|  |  |
| --- | --- |
| **管** **理** **领** **域** | **核** **心** **要** **素** |
| 数据管控 | 数据管理战略 |
| 数据管理组织 |
| 数据管理制度 |
| 数据管理绩效 |

**194** 数据治理——工业企业数字化转型之道

续表

|  |  |
| --- | --- |
| **管理领域** | **核心要素** |
| 数据架构 | 数据标准管理 |
| 数据模型 |
| 主数据管理 |
| 元数据管理 |
| 数据分布 |
| 数据集成与共享 |
| 数据质量 | 数据质量需求 |
| 数据质量检查 |
| 数据质量评估 |
| 数据质量提升 |
| 数据生命周期 | 数据需求 |
| 数据设计和开发 |
| 数据运维 |
| 数据销毁 |
| 数据安全 | 数据安全策略 |
| 数据安全保护 |
| 数据安全审计 |
| 数据价值挖掘 | 数据分析 |
| 数据融合 |
| 数据应用 |
| 数据资产价值管理 |
| 数据资产运营 | 数据服务 |
| 数据共享与开放 |
| 数据资产变现 |
| 支撑平台 | 数据治理工具集 |
| 数据管理平台 |

针对数据管理中的每一个核心要素，根据数据管理成熟度级别不同， DAMA 模型提出了明 确的管理要求，设计了合理的评估指标，并设定了明确的评分标准，最终形成了一套完善的数 据管理成熟度评估打分表，如表17-1-2所示。利用该打分表能够详细了解企业的数据管理能力，

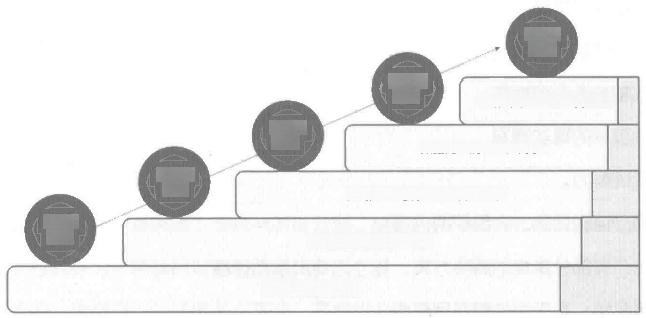
并自动化完成成熟度级别评定。

表17-1-2数据管理成熟度评估打分表(模板及示例)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 管理 领域 | 核心  要素 | 成熟度  等级 | 管理  要求 | 评估  对象 | 可能  出处 | 评估指标 | 符合 度 | 得分 | 支撑  材料 | 总体 评价 | 存在 不足 |
| 支撑 平台 | 数据 管理 平台 | 10~20  分：初  始阶段  级 | 还没有完 整的数据管 理平台 | 数据管 理工具 | 平 台 设 计文档、系 统 演 示 和 运 行 情 况 报告 | 1.已有数据管理 平台建设规划；  2.具有部分数据 管理子系统(有数据 管理的部分功能) |  |  |  |  |  |
| 81~100  分：持  续优化  级 | 不断推动 自身技术创 新 | 平台先 进性 | 平 台 发 版说明、新 技 术 采 用 情况 | 1.数据管理平台 版本每两个月迭代 更新 一 次；  2.平台利用的先 进 技 术 ( 在 3 种 以 上 ) |  |  |  |  |  |
|  | ·· |  |  |  |  |  |  |  |  |  |  |

**17.2** **数据管理成熟度等级定义**

数据管理成熟度分为5级，从低到高分别是初始阶段、基本管理、主动管理、量化管理和 持续优化，如图17-2-1所示。



V持续 优化

**数据成为企业核心竞争力**

**数据完全满足业务需求**

**数据基本能满足业务管理要求**

**数据支持业务基本工作**

**数据尚未发挥价值**

**IV量化** **管理**

I 初始 阶段

皿主动 管理

五基本 曾理

图17-2- 1 数 据 管 理 成 熟 度 等 级

**196** 数据治理 工业企业数字化转型之道

**1.** **初始阶段级：数据尚未发挥价值**

(1)业务能力。

①基本的报表；

②手工作业，依赖于特殊查询；

③信息超载；

④未能反映真实情况；

⑤事后被动发现问题。

(2)系统能力。

①数据：静态结构化的内容；

②集成：无连接、孤立、非集成的解决方案；

③应用系统：孤立模块、依赖特定的应用系统；

④基础架构：复杂、关系混乱，特定平台。

**2.** **基本管理级：数据支持业务基本工作**

(1)业务能力。

①基本的探索、查询和分析功能和基本的报表；

②部分报表自动化；

③完全不同的工作环境；

④有限制的企业可视度；

⑤多种版本的真实情况。

(2)系统能力。

①数据：结构化的、有组织的内容；

②集成：有部分集成的解决方案、孤立的情况依然存在；

③应用系统：基于组件的应用系统；

④基础架构：层级式架构，特定平台。

**第17章** **数据管理成熟度评估** **197**

**3.** **主动管理级：数据基本能满足业务管理要求**

(1)业务能力。

①有脉络的、基于职责的工作环境的导入；

②自动化已提升到一定层级；

③既有的流程和应用系统的增强；

④整合的业务绩效管理；

⑤唯一版本的真实情况；

⑥经由分析的、实时性的洞察力。

(2)系统能力。

①数据：基于标准的、结构化的内容，以及部分非结构化的内容；

②集成：孤立的系统集成、信息的虚拟化；

③应用系统：基于服务的应用系统；

④基础架构：组件式的、面向服务的架构逐步浮现，特定平台。

**4.** **量化管理级：数据完全满足业务需求**

(1)业务能力。

①贯通企业内外的、有弹性的、具有适应力的业务环境；

②战略业务创新的促进能力；

③企业绩效和运营的优化；

④战略洞察力。

(2)系统能力。

①数据：无缝连接并且共享、信息与流程分离、结构化和非结构化内容完全整合；

②集成：信息作为一种随时可用的服务；

③应用系统：流程透过各式服务而集成，有序的业务应用系统；

④基础架构：有随时恢复能力的 SOA, 不限于特定技术。

**198** 数据治理 字化转型之道

**5.** **持续优化级：数据成为企业核心竞争力**

(1)业务能力。

①基于角色的日常工作环境；

②全然融入工作流、流程和系统的能力；

③信息激发的流程创新增强的业务流程和运营管理；

④前瞻性的视野，具有预测性的分析。

(2)系统能力。

①数据：所有相关的内部及外部信息无缝连接并且共享、新增的信息很容易加入；

②集成：虚拟化的信息服务；

③应用系统：动态的应用系统组合；

④基础架构：动态的、可重新配置的侦测和回应。

经过数据管理成熟度评估，在由低级别向更高成熟度级别提升的过程中，为企业所带来的 变化可以描述为以下内容。

●从被动到主动的管理；

●从点解决方案到综合解决方案；

●从“孤立”数据到同步数据(即一致、高质量数据);

●从数据分类和安全级别不一致的本地化系统到一致的数据分类和基于标准的安全管理；

●从短视的传统数据管理到企业范围的数据资产全景图。

**17.3** **开展数据管理成熟度评估**

开展数据管理成熟度评估包括以下两个步骤。

第1步是梳理利益相关者并确定内部支持者，即提倡数据管理的人。

第2步是为上述每个利益相关者创建一套详细的调查问卷。对于每个问题，应找到有关当 前和未来状态的答案。

其中最重要的是找到弱点或不成熟之处，并制订切实的计划来解决它们。要了解企业当前 所处的位置，并根据现状规划未来发展路线图。

第17章 数据管理成熟度评估 **199**

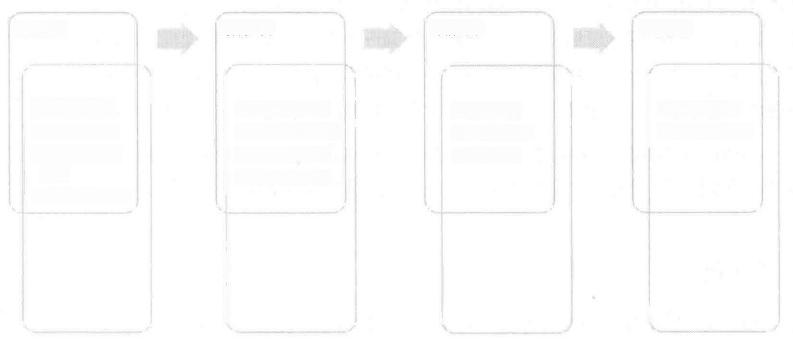
数据管理的本质是获得战略和战术上应对业务挑战的能力，在紧急情况下立即做出响应，

并确保通过信息共享协调组织响应。

**17.4** **数据管理成熟度评估实施**

数据管理成熟度评估工作分为项目启动、培训宣贯、评估执行和总结分析4个阶段，如图

17-4-1 所示。

**项目启动** **培训宣贯** **评估执行** **总结分析**

·建立评估团队

·评估标准介绍

·评估方法论介绍 ·收集相关资料

·下发调研问卷

· 成熟度评定

·完成评估报告

·现场分析

·面对面访谈 ·总结问题

·制定评估计划

·明确项目目标、

范围

·召开项目启动会

**图17-4-1** **数据管理成熟度评估步骤**

**1.** **项目启动阶段**

项目启动阶段的主要工作是了解企业自身的发展情况，建立评估团队，制订评估计划，并 召开项目启动会。项目启动阶段是明确项目目标、范围的阶段，对推动整体评估工作的顺利开

展具有重要意义。

**2.** **培训宣贯阶段**

培训宣贯阶段主要的工作是进行标准介绍，帮助评估人员了解标准的组成、评估的方法和

过程、各方面评估的重点等，并且可以指导相关人员开展自评估。

**3.** **评估执行阶段**

评估执行阶段的主要工作是根据自评的情况，在了解相关资料之后，评估人员在现场对数 据管理能力评估模型中的各方面进行评分，主要方式包括现场分析、面对面访谈等。

**200** 数据治理——工业企业数字化转型之道

**4.** **总结分析阶段**

总结分析阶段的主要工作是根据对企业数据管理现状的了解，制定整体的数据管理成熟度 等级分析及评估报告。

**本章精要**

数据是企业未来生存的生命线。打通数据孤岛是消除数据冗余和实现自动化业务流程的先 决条件。DMMA模型是评估企业当前状态所需的工具，它是根据许多世界领先组织的实践经验 和其他来之不易的经验总结创建的，也是业界第一个以数据资产的管理视角设计的评估体系。 它综合了传统数据治理与数据资产管理相关的最佳实践，将数据管理相关的支撑工具纳入数据 管理成熟度评估范围，也是提升数据管理水平，实现数据治理成功的关键。

DMMA 模型包括8大领域，这些领域又进一步被细分为30个核心要素。它们详细定义了 数据管理的目标，并建立了可持续进行数据管理所需的操作要求。

**参考资料**

[1]IEEE/ANSI1471 软件密集系统架构描述的建议实践，IEEE,2000

[2]ISO 42010 系统与软件工程：软件密集系统架构描述的建议实践， ISO,2011

[3]THE TOGAF STANDARD,VERSION 9.2 企业架构框架，Open Group,2018

[4]D AMA 中国分会翻译组.DAMA 数据管理知识体系指南(原书第2版).北京：机械工业

出版社，2020

[5]James Martin,Information Engineering,Book I:Introduction,Prentice Hall,1989

[6]APQC,Cross Industry Process Classification Framework,Version 7.2.1,Sep.2019

[7]The IT4IT Reference Architecture,Version 2.1,Open Group,2017

[8]蔡莉，梁宇，朱杨勇，何婧.数据质量的历史沿革和发展趋势.计算机科学，2018,45卷(第 4期):第1页

[9]GB/T 36073-2018《数据管理能力成熟度评估模型》

[10]GB/T34960.5-20 18《信息技术服务治理第5部分：数据治理规范》

第17章 数据管理成熟度评估 **201**

[11]李广乾.政府数据整合政策研究.北京：中国发展出版社，2019

[12]《数据资产管理实践白皮书(4.0版)》,中国信息通信研究院

[13]《主数据管理实践白皮书(1.0版)》,中国信息通信研究院



**第** **3** **篇** **工具篇**

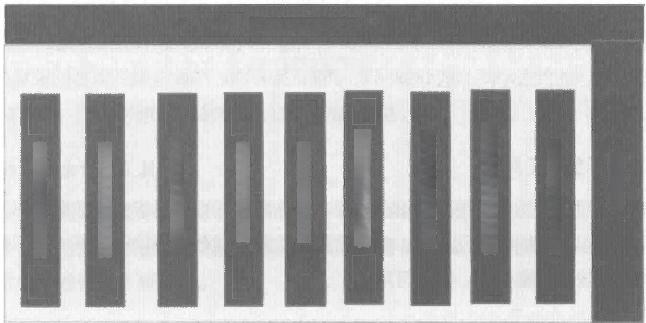
第18章

数据治理工具概述

工业数据治理需要多种数据治理工具和软件的支撑，包括以数据资产目录为核心的数据资 源管理工具、以元数据和数据模型为核心的数据中台等。这些工具互有侧重，需要根据实际需 求予以选择。目前，国内外还没有一家公司能提供覆盖工业数据治理全部领域的成熟套装软件， 主要原因在于：

工业领域的数据治理滞后于互联网、金融、电信行业，国内能够提供工业领域高水平数据 治理解决方案的供应商不多。工业领域的数据治理需要具备数据和工业数据领域的技术和能力， 属于数据治理的新兴市场，进入行业的技术门槛比较高。

本篇对工业数据治理工具平台的定义如图18-0-1所示。



数据中台

数据安全管理工具

期 交 按 与 留 务 工

数据质量管理工具

时月督期盛理工口

元数据管理工具

主数据管理工具

数据指标管理工具

数据衡型管理工具

数据资产运曹工具

**工业大数据治理门户**

**工业大数据治理工具**

**图18-0-1** **工业数据治理工具平台框架图**

**204** 数据治理 工 业 企 业 数字化转型之道

**1.** **工业数据治理门户**

工业数据治理门户是工业企业数据治理组织的工作平台，用于定义数据治理组织的工作流 程和工作标准，包括组织架构、制度规范；发布各类数据标准，包括主数据、数据指标等；评 估数据治理组织绩效和数据质量；查询基于知识工程的知识库。

**2.** **数据资产运营工具**

数据资产运营工具包含数据资产目录工具和数据资产价值评估工具。数据资产目录，也被 称为数据资产地图，支持用户按照自己的业务需求和企业标准，构建企业级数据资产目录，实 现对海量数据进行梳理和归类和数据资产的全面盘点，为用户提供完整的数据资产视图，并提 供数据资产展示、交换和共享。数据价值管理主要通过对数据内在价值的评估、数据成本和收 益的管理来实现。

在工业领域，通过数据资产目录工具，可以让数据拥有者直观、清晰地掌握所拥有的信息 资源；数据使用者也可以通过数据资产目录发现自己所需要的数据现状，并发出需求申请。通 过对数据资源的梳理与编目，为数据的挖掘分析和开发运用提供了准确、全面的数据支撑。数 据资产目录在数据治理解决方案中属于非必选项，但在数据资产管理解决方案中属于必选项。

**3.** **数据模型管理工具**

数据模型管理工具是企业数据模型的管理、比对、分析、展示的技术支撑，用于提供统一、 多系统、基于多团队并行协作的数据模型管理，解决企业数据模型管理分散、无统一的企业数 据模型视图、数据模型无有效的管控过程、数据模型标准设计无法有效落地、数据模型设计与 系统实现出现偏差等多种问题。在工业领域，企业架构与数据架构往往在一起构建，核心业务 流程、业务对象、业务活动定义先在数据架构中逐步落地，然后映射到数据的逻辑模型，之后 落地到物理模型。

**4.** **数据指标管理工具**

数据指标管理工具用于管理数据指标标准，包含数据指标信息维护、数据指标治理及指标 应用等功能，用以打通指标、元数据、数据质量、数据标准各子模块的关系，确保数据接入规 范、标准统一，数据质量可控、数据可用。

**5.** **主数据管理工具**

主数据管理工具用于定义、管理和共享企业的主数据信息，可通过数据整合工具或专门的

第18章 数据治理工具概述 **205**

主数据管理工具来实施主数据管理。主数据管理工具具备企业级主数据存储、整合、清洗、监 管及分发5大功能，并保证这些主数据在各个信息系统中的准确性、 一致性、完整性。

**6.** **元数据管理工具**

通过元数据管理工具可以了解数据分布及产生过程。该工具是针对元数据管理职能而开发 的。元数据管理已经深入数据的物理模型。工业数据领域的元数据管理组件往往根据具体项目 来进行定义。

**7.** **时序数据管理工具**

时序数据管理工具用于物联网、车联网、工业互联网领域中的过程数据采集、过程控制， 并与过程管理建立一个数据链路，属于工业数据治理的新兴领域。从工具维度看，时序数据管 理工具与传统时序数据库差异很大，后者局限于车间级的可编程逻辑控制器，而非企业级。

**8.** **数据质量管理工具**

数据质量管理工具从数据使用角度监控、管理数据资产的质量，是针对数据质量管理职能 而开发的，以实现数据全生命周期的质量管理。其能根据标准规则配置数据质量检查策略，通 过调度中心实现数据质量的检查，发现问题数据，以及将问题数据分派给相关人员修正，并能 根据需要形成数据质量评估报告和问题处理报告等。

**9.** **数据交换与服务工具**

数据集成交换工具是一系列数据技术工具的集合，涉及采集、汇聚、加工、共享等多个环 节，是支撑其他数据治理工具的中间件，被广泛应用于源系统的数据采集、系统之间的数据交 换、多源头数据的汇聚与加工等，属于 PaaS 层中的一系列组件。数据交换与服务工具的搭建， 除要考虑技术外，还要考虑到工业数据治理的长远战略。

**10.** **数据安全管理工具**

数据安全管理工具是结合信息安全的技术手段，保证数据资产在使用和交换共享过程中的 安全。其中包括数据采集管理、数据传输管理、数据存储管理、数据处理管理、数据交换和共 享管理、数据销毁管理6类工具。

**11.** **数据中台**

数据中台通过企业内、外部多源异构的数据采集、治理、建模、分析、应用，使数据在企

**206** 数据治理——工业企业数字化转型之道

业内部可以优化管理，在企业外部可以释放数据合作价值，是企业数据资产管理和服务的中枢。

数据中台技术起源于B2C 互联网平台的应用。在数字化转型的大趋势下，数据中台在工业 领域的应用迅猛发展，本篇把数据中台解读为工业数据的应用平台，是数据治理平台服务的对

象，因为数据从采集到处理的整个过程都会产生成本，只有被应用才会创造出价值。

数据治理工具是为了更好地确保数据的质量和安全，是数据中台的支撑平台。为了确保读 者更好地理解数据治理工具，本篇特意补充数据中台的相关章节。

第 1 9 章

数据资产运营工具

数据资产运营是指把数据作为资产，通过对数据的采集、清洗、加工、分析、挖掘，在合 规化的条件下进行共享与开放。数据资产运营的目的在于获得收益。通过数据资产运营，不仅 可以赋能传统产业，而且可以对现有业务模式进行颠覆式创新。数据资产运营是企业数字化转 型的核心能力，只有将数据资产在全社会流通，实现数据融合，才能使数据的商业价值最大化。

企业应该加强对数据资产的盘点、筛选、甄别、编译、存储、清洗加工及分析挖掘，让数 据资产满足用户的需求，以及将数据资产及时汇聚、整合，并定期更新、注册入库，最终通过 数据资产目录将标准化的数据资产对外发布，统一对外提供数据服务，让数据资产方便用户搜 索和浏览，易于理解和解读。

数据资产运营的核心是促进数据的流通，让数据资产的价值最大化。这就必然需要进行数 据资产的价值评估，以便识别核心数据资产，并分析影响数据资产价值发挥的因素，从而制定 相应的数据资产运营策略，最终实现数据资产收益的最大化。

高质量、可信任的数据是数据资产运营的基础。数据资产运营的对象是数据资产，目的是 数据的资产化，主要手段是数据资产的共享与开放。而数据治理的对象是数据，目的是通过数 据的标准化来保障数据的高质量和可信任。因此，数据治理是数据资产运营的支撑，二者既有 关联又各有侧重。

**19.1** **数据资产目录**

数据资产目录使得数据资产易于查找、便于理解、值得信任，让数据资产更好地满足数据 应用需求。企业若将数据视为资产，就需要建立数据资产目录和数据资产清单。工业企业普遍

**208** 数据治理——工业企业数字化转型之道

存在数据分散(甚至可能不知道所有这些数据位于何处)、数据来源途径多、数据不一致等问题， 往往会花费了很多时间去寻找有意义的、可信赖的数据。

**19.1.1** **总体概述**

数据治理倾向于“组织”和“标准”:包括成立数据治理委员会，设置数据所有者和数据管 理员的角色，制定主数据标准和主数据维护的流程制度。这些活动的重点是实现数据全生命周 期管理。

数据资产管理还关心更高阶的目标，例如：

(1)扩大数据的可用性和简化数据的可访问性，使更多的数据消费者能够查找、访问和 共享数据资产；

(2)标准化数据语义，让数据使用者对共享数据资产有一致性的理解；

(3)实现数据的质量可测量，让用户高度信任数据。

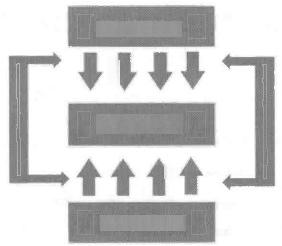
这些数据资产管理目标有一个共同的驱动因素：数据消费或信息使用。如果无法实现这些 目标，则会降低企业有效共享和利用数据的能力。

定义企业的数据战略，对数据资产的获取、转换、资产化、共享和隐私保护实施监督是大 趋势。数据治理趋向于将传统的管理优先事项(如理解数据资产的使用和统一业务术语),与新 兴的评估企业数据前景的关键需求(包括盘点、分类和记录整个企业的数据资产)相结合，以 指导数据的可知、可信和可用。

**19.1.2** **数据资产目录系统构建**

企业拥有大量的、各种类型的、分散在各处的数据资源，如果没有数据资产目录和数据资 产清单，那么许多数据资产实际上是被隐藏的。这意味着需要进行数据盘点来识别企业的数据 资产，并将其分类。数据资产目录的设计采用“业务驱动自顶向下”和“盘点驱动自底向上” 相结合的工作思路。“业务驱动自顶向下”是指按照业务视角全面梳理企业业务价值链、各种业 务场景、端到端的业务流程，包括业务流程中涉及的表单、术语、业务数据项等。“盘点驱动自 底向上”是指通过现状调研，盘点并提取源业务系统中的数据项，将其作为“果子”挂接到相 应的“目录树”上，如图19-1-1所示。

第19章 数据资产运营工具 **209**

**业务数据梳理**



**数据资产目录**



**数据资产盘点**

图19-1-1 数据资产目录构建思路

数据资产目录需要从多个视角构建。 一种是从业务人员角度，通过业务分类呈现，比如人、 财、物、产、供、销。另一种是从技术人员角度，通过主题域呈现，如客户、交易、事件等。 此外，数据资产目录具备一个很重要的功能是可以利用标签实现更便利的数据搜索和查询， 如按部门、系统等。

数据资产目录管理需要一个强有力的系统平台支撑，需要利用先进的技术最大化提升数据 资产的自动化、智能化管理水平。

随着涉及保护个人隐私数据的全球性法规越来越多，识别敏感数据正在迅速成为一个刚需。 其中至关重要的是能够确定哪些数据包含有关个人的信息。建立数量在千亿级的海量样本库、 数量在百万级的敏感词库，以及基于HMM、CRF、Word2Vec 等自然语言处理技术，并结合传 统正规表达式构建的复杂模型体系，可以高效识别各类场景中的敏感信息。结合有关数据资产 业务敏感性和数据血缘关系，数据管理员可以确定谁是数据资产用户，以及他们对具有敏感数 据的数据资产的访问权限，并通过控制来增加数据处理工作流，以防止未经授权的访问。

在数据资产目录中，通过收集数据资产清单，可以使数据使用者能够通过数据资产目录和

标签快捷搜索到最能满足其需求的数据集。数据资产目录可用于共享不同类型的元数据，包括：

(1)描述源系统结构的物理元数据，如表和字段。

(2)描述语义信息的逻辑元数据，如数据库描述、数据质量评估和相关的数据管理策略。

(3)描述如何在各种业务场景中使用数据资产的行为元数据。行为元数据可能是最重要 的，因为它可以自动洞察系统中的每个对象的活性或者热度。

数据资产目录将传统的元数据管理功能(如业务术语表、结构化元数据管理、对象元数据 和数据血缘)与人工智能算法结合，通过简化数据发现，自动推断元数据，提高这些推断的准 确性，以及提供业务术语表、数据元素定义、数据血缘、数据质量、数据安全等信息的可见性，

**210** 数据治理——工业企业数字化转型之道

有助于向用户展示正确的数据资产。如图19-1-2所示，此数据资产目录系统、简单地展示了数 据资产目录系统的大体效果。



我的关注

我的数据源

检索条件：自有数据及目录： X& 标答：AAAAX

后台管理

发票 金条

源表盘点

数据表 文件 音频 视模 图 片

系统盘点

众翻业务

传理业务

标签管理

供应销金题

保险科技部

保险(CSU)

值用卡代运营

农村金融

-其他

库名：fms invoic

AA

业务描述：业务描述XXXXX

**Invoice table** [ 色入在

发票/值息表

数据目录：

来源系统：发票

是否入仓：否

来源系统

数据应用

自有散据

全部

数据库盘点

日录管理

**数据标签**

散据库

AAAA

**图19-1-2** **数据资产目录系统**

此数据资产目录系统实现了全域数据资产的统一登记和展现，消除因为信息不对称而导致 的不同部门之间数据的重复存储和抓取，共同完善底层数据资产的扩充和共享，有力地支持了 数据资产的探寻、应用和流通等价值的发挥。

**19.1.3** **数据资产目录能力评估模型**

数据资产目录能力评估模型有助于全面了解企业数据资产目录的管理现状，指明了达到 目标状态的方向和实现路径。数据资产目录能力评估是开展数据资产目录建设的第一步。

数据资产目录能力评估模型定义了企业提高数据资产利用率所需的5种能力。

**1.** **发现能力**

发现能力描述了在搜索数据资产时浏览数据资产目录的过程。支持数据资产发现能力是数 据资产目录最重要的功能。对所有可用数据资产进行清晰而全面的概述是提高数据资产利用率 和货币化的关键。支持搜索、推荐和查询功能进一步促进了数据资产发现过程，使用户能够快 速识别相关数据资产，并查看样本数据条目及摘要统计信息。

**2.** **理解能力**

理解能力是向数据资产目录添加新条目并使用元数据丰富它们的过程，例如谁拥有数据?

第19章 数据资产运营工具 **211**

为什么收集它?测量单位是什么?以及如何评估价值?数据资产目录应支持两种元数据的生成 方式：手动和自动，例如应该可以自动捕获某些类型的元数据，如数据存储空间占用的信息；

以及支持手动输入业务定义和数据收集目的描述。

**3.** **信任能力**

信任能力支持用户评估数据的可靠性和质量。

有助于让用户对数据集产生信任的3种信息是关于数据质量和覆盖范围的信息、数据血缘 的信息及负责人的信息(例如数据管理者或数据所有者)。利用该信息，可以让用户使用合适的 数据源，并且可以避免使用可能会产生不良结果的不适当的数据源。

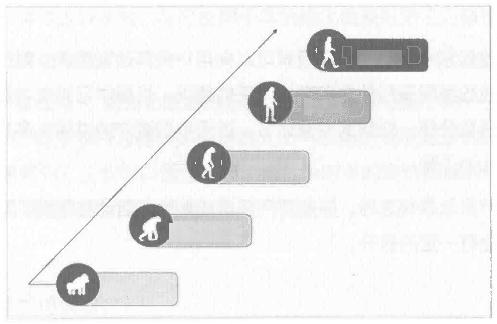
**4.** **协作能力**

协作能力支持其用户进行任务共享并奖励用户创建和修订描述。例如，如果用户需要有关 数据资产的更多信息，那么用户应该能够请求数据管理员添加相关信息。

**5.** **治理能力**

治理能力体现了支持和满足法律及企业标准的能力。例如对访问权限的管理，数据资产目 录中应列出所有数据资产并显示示例条目，但也应符合数据隐私法规。因此，除非用户具有相 应的访问权限，否则不得显示个人数据的示例条目。

数据资产目录能力成熟度分为5个级别，从低到高分别是初始级、主动管理级、基于工具 级、优化级和自动化级，如图19-1-3 所示。



自动化级

优先缀

基于工具级

主动管理级级

初始级

**图19-1-3** **数据资产目录能力成熟度等级图**

**212** 数据治理——工业企业数字化转型之道

**1.** **初始级**

在第1个成熟度级别中，企业没有努力集中编制数据资产目录，数据科学家无法集中获得 大部分有价值的信息。即使数据分析师知道特定数据资产的存在，但理解其业务环境和数据血 缘的唯一方法是联系数据资产所在的业务系统的产品经理或数据工程师。

**2.** **主动管理级**

在第2个成熟度级别中，存在数据资产列表和文档。企业使用标准化模板记录和维护基本 元数据。用户可以从列表中获取有关数据资产的信息，并且可以使用基本的搜索功能。目录条 目更新相对较慢。

**3.** **基于工具级**

在第3个成熟度级别中，数据资产目录工具由专门的团队使用和管理。该数据资产目录工 具有助于自动捕获元数据，例如，它可以自动检测和标记账号或姓名。应用分类和预定义标签 有助于在数据资产的业务环境中手动提供元数据。通过高级搜索功能可以更轻松地查找可能相 关的数据。

**4.** **优化级**

在第4个成熟度级别中，大多数数据资产目录都提供了样本数据条目和统计信息。目录中 还包含有关自动捕获的数据血缘信息。数据血缘信息有助于让用户信任数据资产，因为它们能 够快速确定数据的来源及所有先前的处理步骤。

**5.** **自动化级**

在第5个成熟度级别中，数据资产目录可以向用户推荐数据资产。对于目录中的所有数据 资产，可以自动生成数据质量和数据资产价值评估结果。机器学习算法为那些尚未由用户描述 的数据资产建立标签和分类。在数据准备环节，访问数据资产的过程大多是自动化的，只有在 特殊情况下才需要人为干预。

在实施数据资产目录系统之后，数据资产目录成熟度在自动化程度、激励的有效性和工作

流程效率等方面都会有一定的提升。

第19章 数据资产运营工具 **213**

**19.2** **数据资产价值评估**

数据资产价值评估是实现数据资产化的基础。要针对不同的评估目的和资产类别设计适用 的数据资产价值评估模型，使用户既能了解数据资产的总体价值，也能了解数据资产价值的各 部分构成，从而分析出影响数据资产价值的因素，以便制订相应的数据资产增值计划。

**19.2.1** **总体概述**

数据资产价值评估模型总体上可以分为两类。第一类是非财务方法，即非金融或非经济模 型，它不给数据贴上实际的价格标签，因为一些企业只想创建数据质量特征的聚合，以了解它 的相对或内在的价值。第二类是财务方法，即从会计实务领域中借用一套金融模型。

**1.** **非财务方法**

(1)数据的内在价值模型。

这个模型根本不考虑数据的业务价值，而是考虑数据的内在价值。该模型将数据质量分解 为精度、可访问性和完整性等特性。然后对每一个特性评分，按最终分数进行统计(包括数据 的稀缺性)。对于企业，更独特的数据更有潜力为企业提供更多的价值。这个模型可以针对企业 进行量身定做，例如，它可以为每个特性分配权重因素。

(2)数据的商业价值模型。

这个模型用于衡量关于一个或多个业务流程的数据特征。例如准确性和完整性被评估为“及 时性”,因为即使数据与业务流程相关，但如果不是及时的，那么无法衡量它到底值多少钱。该 模型可以根据组织的需要进行定制，甚至适用于非结构化数据或第三方数据等特定的数据。

(3)数据的绩效价值模型。

这个模型更具“实证性”,因为它衡量的是数据随着时间的推移，对一个或多个关键绩效指 标(KPI) 的影响。以销售部门为例，如果销售人员可以访问竞争对手的定价数据，那么他们 可以更快地完成销售额吗?企业可以通过比较获得竞争对手的定价数据的实验组与未获得竞争 对手的定价数据的对照组来进行实验。

**2.** **财务方法**

(1)数据的成本价值模型。

该模型用于衡量“获取或替换丢失的数据”的价值的成本。在2001年发生“9·11”恐怖

**214** 数据治理——工业企业数字化转型之道

袭击事件之后， 一家企业想知道如何评价丢失数据的价值，因此其开发了一种方法来量化丢失 数据的价值，也即“更换成本”:测量丢失数据对收入的影响及获取数据所需的成本。这也是估 值专家评估大多数无形资产的方式，这些资产没有明显的市场价值或正在产生市场价值。

(2)数据的经济价值模型。

该模型用于衡量信息资产对企业收入的贡献。这也是 KPI 模型，但这里不考虑任何给定的 KPI, 而是考虑收入。为了更好地说明这个问题，再次回到前面那个销售的例子。实验组可以 获得竞争对手的定价数据，而对照组则不能。在一段时间内，我们不再关注销售时间，而是关 注特定销售人员的收入，这将使我们对这些数据的价值有更好的认识。企业应该考虑购买、管 理和将数据上传到销售人员正在使用的系统中所花费的成本，同时，还应该考虑数据的寿命， 例如，竞争对手的定价数据具有保质期，应将其纳入价值评估体系中。

(3)数据的市场价值模型。

该模型用于衡量出售、出租或交换企业数据所产生的收入，这是评估数据资产的最佳方式 之一。但现在面临的问题是大多数数据资产没有 “开放式公平市场”,或者没有在公开市场上 的价格。解决这个问题的方法是弄清楚来自数据交易市场或竞争对手中类似数据的价格，在此 基础上考虑设定一定的折扣。当我们出售数据时，我们并不是真的在卖它，而是授权消费者使 用它。设定的折扣将根据企业销售数据的次数和其他因素而有所不同。

**19.2.2** **数据资产价值评估模型**

数据资产受控于企业主体，依附于有形资产。数据资产的价值受到众多变量因素的影响，

例如数据内在价值、数据在不同场景下的应用价值。

(1)数据内在价值。

数据只有应用在具体的场景中，才会体现其价值。因此，伴随着不同的场景，同样的数据 会表现出不同的价值。数据的价值是需要不断被挖掘的。随着我们对不同行业的认知逐渐深入， 数据的价值不断丰富。我们可能难以穷尽数据在所有行业的价值，但我们可以就数据在某一行 业、在当前阶段的价值予以评估。数据的价值受两个主要因素的影响：数据质量和数据应用。

(2)数据应用价值。

数据价格围绕数据价值上下波动。对数据价格的评估可以基于两个主要因素：数据成本和 数据收益。数据成本主要是从数据拥有方角度来考虑的，是数据拥有方制定数据价格的主要出 发点。出于财务的考虑，数据价格应该高于数据成本。数据收益主要是从数据需求方角度来考

第19章 数据资产运营工具 **215**

虑的，是数据需求方购买数据时愿意付出的最高价格。出于盈利的考虑，数据价格应该低于数

据收益。

**1.** **数据资产价值评估的理论方法**

(1)层次分析法( AHP)。

层次分析法是美国在20世纪70年代提出的一种将定性和定量分析相结合的多准则决策方 法。层次分析模型把复杂的问题分成若干个组成因素，并按照支配关系分组形成层次结构，然 后通过两两比较的方式，并综合专家的判断，确定分析的结果，具体步骤如下。

①分析各因素的关系，建立递阶层次结构；

②对同一层次的多个元素，关于上一层次中某一准则的重要性进行两两判断，构建比较判 断矩阵；

③由判断矩阵计算比较元素对于该准则的相对权重；

④计算合成权重。

(2)专家打分法(德尔菲法)。

德尔菲法由兰德公司首次使用，后来该方法被广泛采用。这种方法先将所需解决的问题单 独发送到各个专家手中征询意见，然后回收并汇总全部专家的意见，整理出综合意见。随后将 该综合意见和预测问题再分别反馈给专家，再次征询意见，各专家依据综合意见修改自己原有 的意见，再汇总。这样多次反复，逐步取得比较一致的预测结果，具体步骤如下。

①组成专家小组。按照课题所需要的知识范围，确定专家；

②向所有专家提出所要预测的问题及有关要求，由专家做出书面答复；

③各个专家根据他们所收到的材料，提出自己的预测意见；

④将各位专家的第一次判断意见汇总，列出图表，进行对比，再分发给各位专家，让专家 比较自己同他人的不同意见，修改自己的意见和判断；

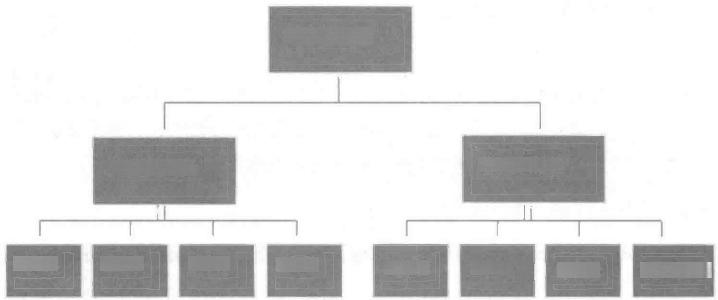
⑤将所有专家的修改意见回收并汇总，再次分发给各位专家，以便做第二次修改。将这一 过程重复进行，直到每一个专家不再改变自己的意见为止；

⑥对专家的意见进行综合处理。

**216** 数据治理- 工 业 企 业 数字化转型之道

**2.** **数据资产价值评估指标体系**

数据资产价值评估指标体系如图19-2-1所示。



**数据资产价值**

**数据内在价值**

**完整性** 正确性 一致性 重复性

多堆性 场景经济性

稀缺性 时效性

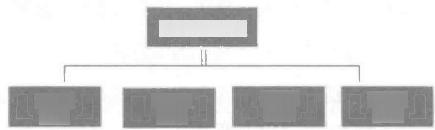
**数据应用价值**

图19-2-1 数据资产价值评估指标体系

(1)数据内在价值评估。

数据内在价值是数据质量的价值体现，是保证数据应用的基础，是数据资产价值得以实现 的前提。随着企业拥有的数据量的急剧扩大，数据质量问题变得日益突出。数据质量将严重影 响企业数据资产的价值，是决定数据价值高低的重要因素。

数据内在价值评估的维度包括数据的完整性、正确性、 一致性、重复性，如图19-2-2所示。 数据内在价值评估能够对整体数据或部分数据的质量状况给出一个合理的评估，帮助数据用户 了解数据的质量水平，进而对数据应用水平予以预测，评估企业数据资产的真实价值。



数据内在价值

完整性

(Q1)

正确性 (Q2)

重复性 (Q4)

一致性 (Q3)

图19-2-2 数据内在价值评估指标体系

①完整性 (Q1)

指标解读：描述数据是否存在缺失记录或缺失字段。数据缺失的情况可能是整个数据记录 缺失，也可能是数据中某个字段的记录缺失。

指标评分方法：完整性=数据集中所有满足条件的数据记录数/数据记录总数×100%

第19章 数据资产运营工具 **217**

②正确性(Q2)

**指标解读：**描述数据是否与其对应的客观实体的特征相一致。任何字段都应该符合特定的 数据格式与值域范围。例如身份证号码应该为15位或18位，移动电话号码应该为11位，人的 年龄应该为0~120岁等。

指标评分方法：正确性=数据集中所有正确的数据记录数/数据记录总数×100%

③一致性(Q3)

指标解读：描述同一个实体的同一个属性的值在不同数据集中是否一致。在各个独立的业 务系统内部，数据不一致的现象大量存在。例如，“客户”和“用户”的意义相同，在相关的多 个数据表中，这两个术语同时存在。

指标评分方法： 一致性=1—数据集中所有不一致的数据记录数/数据记录总数×100%

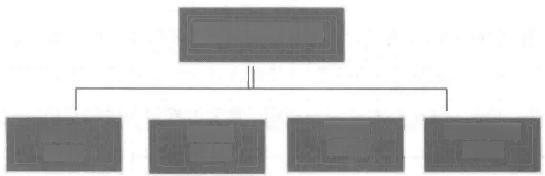
④重复性 (Q4)

**指标解读：**描述数据是否存在重复记录。现实世界中的同一个主体，在不同的数据源中常 常有多个表达形式，在语法上相同或相似的不同记录可能会代表现实世界中的同一个主体，因 而会对同一个主体造成重复性记录。

指标评分方法：重复性=数据集中所有重复的数据记录数/数据记录总数×100%

(2)数据应用价值评估。

数据的价值只有在应用时才得以体现，数据应用价值是数据资产的核心价值。数据应用价 值评估的维度包括稀缺性、时效性、多维性、场景经济性，如图19-2-3所示。数据应用价值在 不同的行业、不同的应用场景中有所不同。在市场环境中，数据是否被垄断也是决定数据价值 高低的重要因素。在不同应用场景中，对数据的时效性要求也不同，有些场景需要实时性数据， 而有些场景需要具有较长时间周期的历史性数据。交叉性的多维数据可以为分析者带来更深刻 的洞察，因而价值也更高。



数据应用价值(A

场景经济性 (A4)

稀缺性 (A1)

时效性 (A2)

多维性 (A3)

图19-2-3 数据应用价值评估体系

**218** 数据治理——工业企业数字化转型之道

由于数据应用价值评估的各个维度缺乏具体的数据，也缺乏衡量的标准，更多的是利用行 业专家的经验，采用专家打分的方法。因此我们采用将本维度的经验估值与相对应的最大估值 进行比较的方法，得到相对数值，来评估数据应用价值。该数值的最大值为10。

更为重要的是，数据伴随着应用场景的不同，具有不同的应用价值。同样的数据集，在 A 场景的应用价值为10,可能在B 场景的应用价值为5,因此，必须在具体的应用场景中进行数 据应用价值评估，如表19-2-1所示。

**表19-2-1** **数据应用价值评估标准表**

|  |  |
| --- | --- |
| **标** **度** | **含** **义** |
| 10 | 表示与最大值相比，具有同样价值 |
| 8 | 表示与最大值相比，该维度的价值低一些 |
| 6 | 表示与最大值相比，该维度的价值明显低于最大价值 |
| 4 | 表示与最大值相比，该维度的价值强烈低于最大价值 |
| 2 | 表示与最大值相比，该维度的价值极端低于最大价值 |
| 0 | 表示与最大值相比，该维度的价值几乎没有 |
| 1,3,5,7,9 | 为上述相邻判断的中值 |

①稀缺性

**指标解读：**描述供给方数量的多寡及数据的供给丰富程度。在数据源市场上，当数据供给 方数量很少，或者市场上该类型的数据量稀少时，相应的数据价值就会较高。

**指标评分方法：** 将供给方数量或数据供给的丰富程度与最大供给方数量或数据供给的丰富 程度相比较。利用行业内的大数据专家经验，采用专家打分法。

**示例：** 例如个人位置数据，目前只有三大运营商及一些 APP 厂商拥有个人位置数据，数据 源少；相对于数据源丰富的个人互联网数据，个人位置数据相对稀少，即稀缺性强，采用专家 打分法，稀缺性数值为9。

②时效性

**指标解读：** 描述数据的时间特性对应用的满足程度。不同类型的应用对数据的时间特性有 不同的要求。通常，实时性应用中的数据时效性较短，而预测性应用中的数据时效性较长。

**指标评分方法：**将数据集的有效时间与应用所需要的期望有效时间相对比，利用行业内的 专家经验，采用专家打分法。

**示例：**例如个人位置数据，同样的数据集对于不同的应用具有不同的价值。如果将10分钟

第19章 数据资产运营工具 **219**

更新一次位置数据的数据集用于旅游商店的实时位置营销，则不能完全满足实时营销的要求， 数据时效性数值为3。如果将同样的数据集用于景区人流监控，则可以满足应用的需要，数据 时效性数值为10。

③多维性

**指标解读：** 描述数据集维度的多寡程度。企业业务的丰富和数据数量的增加，会促使数据 维度不断增加。企业数据维度与数据资产的价值呈正相关关系，更为复杂的多维度数据通常蕴 含着更大的价值。

指标评分方法：将数据集的维度与应用所需要的期望数据维度相对比，利用行业内的专家 经验，采用专家打分法。

**示例：**例如个人交通违章处理数据，对车辆保险应用而言，这个维度的数据基本满足需要， 其多维性数值为8。但是同样的数据集，对个人信用评估而言，则远远不够，其多维性数值很 可能为2。

④场景经济性

**指标解读：**描述在具体场景中数据集的经济价值。由于不同行业的规模、数据应用程度等 具有差异性，因而不同场景中的数据集的价值会相差很大。

指标评分方法：将数据集在某场景中的经济价值与在所有场景中的最大经济价值相对比， 利用行业内的专家经验，采用专家打分法。

**示例：**例如，大学生的个人社交数据的经济价值相对较低，其指标评估值为2;“上班族” 的个人社交数据的经济价值相对较高，其指标评估值为4;而“上班族”的财务数据的经济价 值最高，其指标评估值为8。

**2.** **数据资产价值评估指标的权重计算**

数据资产价值评估指标的权重计算步骤介绍如下。

(1)构建指标层次结构模型，如表19-2-2所示。

**表19-2-2** **构建指标层次结构模型**

|  |  |
| --- | --- |
| **一级指标** | **二** **级** **指** **标** |
| 数据内在价值(Q) | 完整性(Q1) |
| 正确性(Q2) |
| 一致性(Q3) |
| 重复性(Q4) |

**220** 数据治理——工业企业数字化转型之道

续表

|  |  |
| --- | --- |
| **一级指标** | **二级指标** |
| 数据应用价值(A) | 稀缺性(A1) |
| 时效性(A2) |
| 多维性(A3) |
| 场景经济性(A4 |

(2)计算判断矩阵及各级权重结果。

①确定定量的标度及含义

对于各个指标的权重，采取两两因素比较的专家打分法。在比较两个因素时，需要有定量

的标度。这里采用如表19-2-3所示的标度方法。

**表19-2-3** **各标度含义**

|  |  |
| --- | --- |
| **标** **度** | **含** **义** |
| 10 | 表示两个因素相比，具有同样重要性 |
| 30 | 表示两个因素相比， 一个因素比另一个因素稍微重要 |
| 50 | 表示两个因素相比， 一个因素比另一个因素明显重要 |
| 70 | 表示两个因素相比， 一个因素比另一个因素强烈重要 |
| 90 | 表示两个因素相比， 一个因素比另一个因素极端重要 |
| 20,40,60,80 | 为上述相邻判断的中值 |

②数据内在价值判断矩阵及权重

●构造比较判断矩阵

根据企业数据专家的意见，按照重要程度对指标进行两两比较，构造数据内在价值判断矩 阵，如表19-2-4所示。

**表19-2-4** **构造数据内在价值判断矩阵**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 完整性(Q1) | 正确性(Q2) | 一致性(Q3) | 重复性(Q4) |
| 完整性(Q1) | 1 | 1 | 3 | 5 |
| 正确性(Q2) | 1 | 1 | 3 | 5 |
| 一致性(Q3) | 1/3 | 1/3 | 1 | 3 |
| 重复性(Q4) | 1/5 | 1/5 | 1/3 | 1 |

第19章 数据资产运营工具 **221**

●计算权重

完成判断矩阵一致性检验，然后计算数据内在价值评估指标的权重，如表19-2-5所示。

**表19-2-5计算权重**

|  |  |  |
| --- | --- | --- |
| **一级指标** | **二级指标** | **权** **重** |
| 数据质量 | 完整性 | W₁=0.32 |
| 正确性 | W₂=0.32 |
| 一致性 | W₃=0.21 |
| 重复性 | W₄=0.15 |

③数据应用价值判断矩阵及权重

●构造判断矩阵

根据企业中数据专家的意见，按照重要程度对指标进行两两比较，构造数据应用价值判断 矩阵，如表19-2-6所示。

**表19-2-6** **构造数据应用价值判断矩阵**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 稀缺性(A1) | 时效性(A2) | 多维性(A3 | 场景经济性(A4) |
| 稀缺性(A1) |  | 1/3 | 1/5 | 1/9 |
| 时效性(A2) | 3 | 1 | 1/3 | 1/3 |
| 多维性(A3) | 5 | 3 | 1 | 1/3 |
| 场景经济性(A4) | 9 | 3 | 3 | 1 |

●计算权重

**完成判断矩阵一致性检验，然后计算数据应用价值评估指标的权重，如表19-2-7所示。**

**表19-2-7** **计算权重**

|  |  |  |
| --- | --- | --- |
| **一级指标** | **二级指标** | **权** **重** |
| 数据质量 | 稀缺性 | W₁=0.11 |
| 时效性 | W₂=0.14 |
| 多维性 | W₃=0.31 |
| 场景经济性 | W₄=0.44 |

●数据资产价值评估指标的评分

数据内在价值得分计算

构建数据内在价值评估二级指标的评分标准，如表19-2-8所示。

**222** 数据治理——工业企业数字化转型之道

**表19-2-8** **构建数据内在价值评估二级指标的评分标准**

|  |  |  |
| --- | --- | --- |
|  | **指标评分标准** | 得分计算 |
| 完整性(Q1) | Q1=(数据集中所有满足条件的数据记录数/数据记录总数)×100% | S₁=Q1×100 |
| 正确性(Q2) | Q2=(数据集中所有正确的数据记录数/数据记录总数)×100% | S₂=Q2×10 |
| 一致性(Q3) | Q3=(数据集中所有不一致的数据记录数/数据记录总数)×100% | S₃=(1-Q3)×100 |
| 重复性(Q4) | Q4=(数据集中所有重复的数据记录数/数据记录总数)×100% | S₄=(1-Q4)×100 |

计算二级指标得分。以某数据集为例，假设其数据内在价值评估的二级指标得分如表19-2-9

所示。

**表19-2-9数据内在价值二级指标得分**

|  |  |  |  |
| --- | --- | --- | --- |
| **一级指标** | **二级指标** | **质量合格率** | **得** **分** |
| 数据质量 | 完整性 | 94.5% | S₁=94.5 |
| 正确性 | 92.05% | S₂=92.05 |
| 一致性 | 7.34% | S₃=92.66 |
| 重复性 | 12.1% | S₄=88.9 |

计算一级指标得分。通过加权计算公式得到数据质量得分：

S₀=W₁S₁+W₂S₂+W₃S₃+W₄S₄

=0.32×94.5+0.32×92.05+0.21×92.66+0.15×88.9

=92.49

**数据应用价值得分计算**

下面以某一具体的应用场景为例：将个人位置数据应用于旅游景区商家营销的应用中。

构建数据应用价值二级指标的评分标准，如表19-2-10所示。

**表19-2-10构建数据应用价值二级指标的评分标准**

|  |  |  |
| --- | --- | --- |
|  | **指标评分标准** | **得分计算** |
| 稀缺性(A1) | 供给方数量或数据供给丰富程度，与最大供给方数量或数据供给丰富程 度相比较 | S₁=[1,100] |
| 时效性(A2) | 将该数据集的有效时间与应用所需要的期望有效时间相比较 | S₂=[1,100] |
| 多维性(A3) | 将该数据集的维度数量与应用所需要的期望数据维度相比较 | S₃=[1,100] |
| 场景经济性(A4) | 将该数据集在某场景下的经济价值与所有场景中的最大经济价值相比较 | S₄=[1,100] |

第19章 数据资产运营工具 223

计算二级指标得分。如将某移动运营商个人位置数据集应用于旅游景区商家营销应用的二 级指标得分，如表19-2-11所示。

**表19-2-11** **数据应用价值二级指标得分**

|  |  |  |  |
| --- | --- | --- | --- |
| **一级指标** | **二级指标** | **指标评估** | **得分(满分100分)** |
| 数据应用价值 | 稀缺性 | 移动运营商是主要来源，还会有景区Wi-Fi数据， 但很少。总体稀缺性强 | S₁=80 |
| 时效性 | 该数据集只能30分钟提供一次。时效性很差 | S₂=20 |
| 多维性 | 只有位置数据，缺乏用户消费能力数据及个人爱 好数据等。多维性较差 | S₃=50 |
| 场景经济性 | 景区商家多为一次性消费，且消费金额较小，因 而场景经济性一般 | S₄=5 |

计算一级指标得分。通过加权计算公式得到数据质量得分(满分为100分):

*S=W₁S₁+W₂S₂+W₃S₃+W₄S₄*

=0.11×80+0.14×20+0.31×50+0.44×50

=49.1

数据资产价值评估得分

数据资产价值(S) 由数据内在价值(S) 和数据应用价值(Sa) 共同决定，数据资产价值

评估得分由二者相乘得到。

S=(Sq×S₄)/100

在上面示例中，数据资产价值 S=(92.49×49.1/100=45.41。

19.2.3 数据资产价值评估工具

在数据资产价值评估模型设计完成后，企业需要反复在实际中进行检验，证明其科学有效， 并利用工具将其落地，辅助企业进行数据资产价值评估，并提高数据资产价值评估效率。数据 资产价值评估平台架构如图19-2-4 所示。

**224** 数 据 治 理 — —工 业 企 业数 字 化 转 型 之 道



统一登录入口

上传数据 数据资产确权

接口管理 公式管理

数据水印

任务调度

数据加密

传输监控

数据上链

确权认证

权限管理

对外宣传商口

数据质量

评估

评估&

析

评估报告

指标库 指标管理 评估模型 样本管理 联盟管理

用户注册

角色管理

用户管理

数据价值 评估

自动评估

数据资产 评估报告

电子报告

正式报告

档案管理

数据资产 评估

行业交流空间

人工评估

支撑功能

资产认证

工作流

图 1 9 - 2 - 4 数 据 资 产 价 值 评 估 平 台 架 构

其中包括数据资产价值评估流程(用户注册、上传数据、数据资产确权、数据质量评估、 数据价值评估、数据资产评估和数据资产评估报告),以及底层支撑功能、工作流、权限管理、 指标库、评估模型等，最后通过门户统一对外提供服务。

**本章精要**

从数据规划开始，到数据采集、存储和共享、维护、应用、注销，在数据的全生命周期中， 数据的价值只有在数据的应用环节中才得以实现。数据为什么必须成为企业的资产?因为在当 今的技术和市场环境下，获取数据的能力和运用数据的能力更能体现出企业的核心能力。

运营数据资产首先要有一个数据资产目录。数据资产价值的评估是对数据资产化的过程进 行量化评估，用于设定数据资产运营计划的目标，并对结果进行评估。

第 2 0 章

数据模型管理工具

**20.1** **数据模型管理工具概述**

针对企业在不同业务发展阶段建设的一个个“竖井式”系统，其中最大的挑战莫过于在系 统集成过程中数据模型的不一致。解决这个问题的唯一方法就是从全局入手，设计标准化数据 模型，构建统一的数据模型管控体系。数据模型管理工具负责为企业的数据模型的管理、比对、 分析、展示提供技术支撑，解决数据模型管理分散、无统一的数据模型视图，数据模型无有效 的管控过程，数据模型标准设计无法有效落地等多种问题。数据模型管理工具需要具备以下基 础功能。

(1)数据模型设计：支持对于新建系统的正向建模能力，还应支持对原有系统的逆向工程 能力，通过对数据模型进行标准化设计，能够使数据模型与整个企业架构保持一致，从源头上 提高企业数据的一致性。

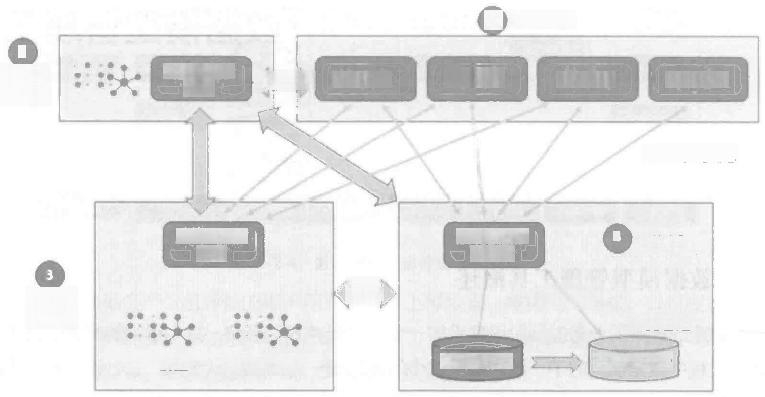
(2)模型差异稽核：提供数据模型与应用数据库之间自动的数据模型审核、稽核对比能力， 解决数据模型设计与实现不一致而产生的“两张皮”现象；针对数据库表结构、关系等差别形 成差异报告，辅助数据模型管理人员监控数据模型质量问题，提升数据模型设计和实施的质量。

(3)数据模型变更管控：支持对数据模型的变更管控(提供数据模型从设计、提交、评审、 发布、实施到消亡的在线、全过程、流程化的变更管理);同时，实现对各系统数据模型的版本 化管理，自动生成版本号、版本变更明细信息，辅助数据模型管理人员管理不同版本的数据模 型；通过工具可以简单回溯任意时间点的数据模型设计状态，实现对各系统数据模型的有效管 控和管治，强化用户对其数据模型的掌控能力。

**226** 数据治理——工业企业数字化转型之道

**20.2** **企业级数据模型管控**

企业级数据模型管控包含数据架构管控、标准管控、事前模型管控和事后元数据管控4 个部分，这4个部分既协同工作又相互影响，如图20-2-1所示。



2)**标准管控**

1

标准代码 业务术语

**落实标准规范**

**事后元**

**系统级开发**

模 型

**事前模**

**型管控**

**OLAP模型** **数据仓库模型** OUTP教据库 大数据平台

**系统级生产**

元教据

企业级遥辑 模型

制定标准 数据标准

**数据架** **构管控**

目录/标签

**质量管控**

**数据管控**

**自动获取**

(

\*

4

**图20-2-1** **企业级数据模型管控**

数据架构管控基于企业级逻辑模型，从业务角度进行企业级数据模型的整体架构。在梳理 企业级逻辑模型的过程中会涉及数据标准的定义和审批发布。而数据标准体系的建设会为企业 级逻辑模型提供有力的支撑。

标准管控主要包括数据标准体系的建设、标准代码管理，以及对业务术语和目录/标签的梳 理，将企业需要共同遵循的数据标准整理并发布。数据标准中包括了数据的业务和技术属性， 所以可以映射到相应的逻辑模型和物理模型。

系统级开发模型是基于企业级逻辑模型进一步细化扩展而来的，根据系统的不同又被分为 应用系统开发模型(OLAP 模型)和数据仓库模型。系统级开发模型通过应用数据标准或者继 承企业级逻辑模型中的数据标准信息实现数据标准的落地。应用系统开发模型作为基线模型， 可以对业务系统的元数据进行比对及监控。

对于已有的业务系统，可以通过提取数据库结构模型，参考数据标准，对原有的数据开发 模型进行完善，以通过数据模型基线比对发现元数据的差异信息并进行后续处理。

数据模型管理工具需要具备以下功能。

第20章 数据模型管理工具 **227**

**1.** **模型抽取和版本管理**

一是自动收集业务系统数据结构，支持主流的数据库，包括 Oracle、SQL Server、Hive、 MySQL 等。并且可以对数据模型进行版本化管理，自动比对数据模型差异，形成模型差异分析 报告。二是实现数据标准的统计分析，快速分析信息系统数据标准的落地情况。

**2.** **模型差异报告**

对于已经发布的模型，在业务系统升级的情况下，其库表结构也会发生变化。为了保证基 线模型与业务系统的一致，数据模型管理工具可以对模型自动检测，发现不一致的情况并预警。

**3.** **模型质量报告**

模型开发基本完成后，需在测试阶段进行模型的评审。

模型评审前需要创建模型基线。评审包含以下内容。

(1)标准的落标引用。

模型工具应该自动提供报告，重点检查标准的引用和落地，发现落地标准(简称“落标”)

的潜在问题。

(2)自定义标准与词典的评审和转化。

模型工具具备自定义数据标准和词典等能力，从而可以提高自定义标准的转化率，完善标

准库。

(3)元数据的充足率。

模型工具应该自动提供报告，列出没有填写中文名称的字段。

(4)其他模型质量。

比如检查模型主题覆盖率等。

**4.** **模型上线流程**

模型的上线需要提交设计文档、测试报告、使用手册等内容，模型的核准环节包含以下工 作。

(1)模型生产库基线与封板。

根据评审时建立的模型分支，建立模型的生产库基线，并进行封板操作。

**228** 数据治理 工 业 企 业 数字化转型之道

(2)模型基线报告。

提供模型标准数据字典、标准落标报告、模型质量报告。

**5.** **新增和变更流程**

在实际落标的过程中，需要新增或修改标准，应设立模型新增和变更流程。

**20.3** **数据标准管控**

**20.3.1** **标准的发布和工具访问**

数据标准是数据模型定义的基础，在定义数据模型的过程中，数据标准的制定需要一些前 提条件。

**1.** **标准的技术规范已经准备好**

数据标准已经具有详细的技术规范，包括数据元属性及取值范围定义，支持物理数据模型 设计，可以直接应用在物理层，并已经建立逻辑数据类型到不同数据库的映射。

**2.** **标准的主题已经准备好**

标准的主题其实是标准的应用范围和检索目录，具备条件的企业应该设计好逻辑模型，对

数据标准进行业务组织。

**3.** **标准已经权威发布**

标准已经经过讨论并进行公开发布，具有流程上的正式性和权威性，已经在组织内部充分

贯彻。

**20.3.2** **模型设计中的应用数据标准**

数据模型是一个很好的数据字典，其向上承接业务语义，向下实现物理数据，不但包含数 据字典，更包含业务主题、业务对象、数据关系，以及数据标准的映射。所以，模型工具的运 用不但是企业数据管理是否成熟的重要标志，也是数据标准落标的重要依托。通过创建一个模 型工具，可以在开发阶段自动管理数据字典和模型，实现以下3个标准落地操作。

第20章 数据模型管理工具 **22 9**

**1.** **建立标准和数据的映射**

(1)标准落地的属性继承。

一般情况下，数据字段标准在落地时要引用数据标准中的内容，还要包含数据的标准代码。

(2)物理字段的落地衍生。

对于一个标准落地的物理字段，如果语义和业务规则没有变化，但是为了满足系统环境需 要而加上了特定/限定环境，比如“电话”在供应商的表里叫“供应商电话”,则被称为物理字 段的落地衍生。在此种情况下并不需要创建一个新的数据标准。

**2.** **建立代码的标准引用**

对字段中数据类型的引用代码进行标准化，坚决杜绝手工写代码的情况。

**3.** **标准化命名**

对字段的命名进行标准化。

**20.3.3** **数据标准应用情况的自动检核**

模型设计工具记录了当前设计的模型库对数据标准的引用情况，并根据每个模型对数据标 准的引用记录统计模型遵从度，以及给出详细的报告，具体包括以下内容。

(1)全域模型的遵从分析。

(2)哪些字段使用了数据标准。

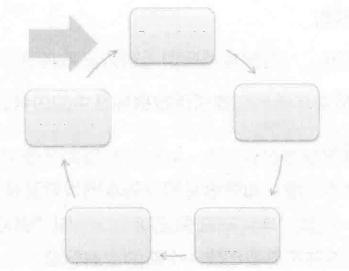
(3)统计使用数据标准的数据项。

**20.3.4** **自定义标准的发布管理**

数据标准通常是在事后制定的，因此， 一定会有新数据没有对应数据标准的情况出现。在 实际解决方案中，数据标准可以采用先创建后验证的模式。

通过建模工具，开发人员可以提交自定义的数据标准到数据管理平台。模型评审组在数据 管理平台对其进行评估，通过验证后可以发布到企业级公共数据标准库中，并不断补充及完善 企业数据标准库，如图20-3-1所示。

**230** 数据治理——工业企业数字化转型之道



创建自定义

标准

治理组评估

打已验证

标签

标准补全 信息

**多人使用**

图20-3-1 创建自定义标准

**20.4** **数据字典的质量检核**

通过模型质量报告可以对模型的数据规范和完整性进行检测，包括元数据是否填写、数据

标准覆盖率、表和字段的重复率、命名的规范性等。在模型的评审环节，此报告是重要的依据。

**本章精要**

数据模型管理工具负责对企业数据模型的管理、比对、分析、展示提供技术支撑，包括数 据模型设计、模型差异稽核、数据模型管控等功能。数据模型管控包含数据架构管控、标准管 控、事前模型管控和事后元数据管控4个部分。数据模型是一个更好的数据字典，其向上承接 业务语义，向下实现物理数据，不但包含了数据字典，更包含了业务主题、业务对象、数据关 系，以及数据标准的映射。数据模型管理工具最重要的功能是可以自动收集业务系统的数据结 构。

第21章

数据指标管理工具

本章介绍如何通过构建数据指标管理工具，为企业的数据指标标准化工作提供落地支撑。 通过数据指标工具可以将企业系统梳理的统一数据指标标准落地，可以规范企业业务统计分析 语言，帮助企业提升数据质量和分析能力，进而提高企业数据质量和数据资产价值。

在应用数据指标工具时，需要先系统梳理数据指标，形成完整的数据指标体系框架。数据 指标体系框架主要用于企业数据指标的体系化应用，数据指标梳理的完善与否决定了数据指标 工具的实施效果。在梳理数据指标的过程中，应结合企业指标管理范围，明确各层级数据指标 项的责任人，统一实现各数据指标项的定义、口径及取数来源，保证企业各级管理层和决策层 能够获得一致的数据指标理解和数据来源，逐步推进企业数据指标体系标准化工作的完善。

数据指标工具可以快速、准确地分析数据指标，提升企业的数据资产价值。因此，在数 据治理过程中，企业需要高度重视数据指标标准的制定和数据指标工具的落地。

**21.1** **指标库管理**

指标库管理并不是一些指标的简单堆积和组织，而是基于业务管理职能的原则建立的。指 标库管理关注指标的统一入库管理，包括指标的批量导入与导出、指标录入、指标信息维护、 指标申报管理、指标查询、指标发布管理等功能。

(1)指标的批量导入与导出。

指标的批量导入与导出可实现按需选择指标，并导出为指标卡片文件。在指标库中根据指标名 称、指标体系或指标的版本范围等条件进行指标过滤，用户将指标查询结果以指标卡片格式导出。

**232** 数据治理——工业企业数字化转型之道

(2)指标录入。

指标录入主要是通过手工方式进行指标的添加、提交、审核等操作。

(3)指标信息维护。

指标信息维护分为3个区域，包括指标工作区、指标候选库、指标库。

指标工作区用于收集各单位与专业部门对指标的需求，形成需求状态的指标卡片后，经过 审核提交到各单位的指标候选库中。

指标候选库根据各单位与专业部门对需求状态的指标卡片进行审批后逐级上报，经过定期

梳理后形成最终生效的候选指标。

指标库是指企业通过计划部门审批的所有指标的集合。所有在指标候选库中为生效状态的 指标通过审批后才能进入企业指标库，并被分配唯一编码和版本号。

(4)指标申报管理。

指标申报管理可实现指标库中指标的生效和失效管理功能。

(5)指标查询。

指标查询可实现快速查询、自定义查询功能。快速查询可以在系统查询界面输入关键字(指 标名或其他指标属性),系统根据查询条件模糊匹配出相关指标并显示，可以对查询结果进行排 序，进一步查询。自定义查询可自行选择指标的属性进行组合，支持先选择指标的属性，并设 置组合条件连接符，再选择或输入查询值。

(6)指标发布管理

指标发布管理主要是指手工录入指标并通过审核，成为正式指标后将其发布使用。

**21.2** **指标体系管理**

指标体系管理包括指标分类库管理、指标体系维护、指标维度管理、指标标准管理、指标 分类查询。

(1)指标分类库管理。

指标分类库管理用于定义指标的不同分类、分类之间的层级关系，方便各层级部门从不同 视角管理、使用指标。

第21章 数据指标管理工具 **233**

(2)指标体系维护。

指标体系维护用于维护指标与分类之间的关系，达到对企业指标库中的指标进行多样化分 类管理的目的；通过指标体系，可以实现指标价值链的多视角分析与查询。同时，指标体系维 护是指标地图、指标血缘分析等高阶应用功能的基础。

(3)指标维度管理。

指标维度管理用于定义指标与不同维度的层级关系、具体维度与库表字段的对应关系，方 便业务人员理解指标的特征，技术人员计算指标不同维度的具体值。指标维度管理包括维度的 新增、修改、停用、审核等。

(4)指标标准管理。

指标标准管理主要是对指标相关标准的管理操作，包括标准的录入、查询、分发及指标标 准的映射。

(5)指标分类查询。

指标分类查询用于快速定位任何一类指标的具体情况，方便查找和使用指标。

**21.3** **指标评价管理**

指标评价管理通过指标评价、评价报告、问题报告来支撑整个工作。

(1)指标评价。

企业各部门根据评价方法，开展指标应用评价，提出改进建议和意见。系统为每个用户提 供了对指标打分的功能，每个用户都能为本部门的应用指标打分，并提出改进建议和意见。系 统根据企业计划部门制定的规则进行统计，最后得出指标的评分，作为计划部门改进指标的依 据。

(2)评价报告。

评价报告是依据指标评价结果出具的报告，支持统计指标被不同报表的引用次数。

(3)问题报告。

问题报告是依据指标问题反馈与解决情况出具的报告，也统计指标被不同报表的引用次数。

**23 4** 数据治理——工业企业数字化转型之道

**21.4** **指标应用管理**

按照企业组织管理维度，指标应用管理分为指标统计报告、指标地图、指标版本管理等。

(1)指标统计报告。

基于企业指标库和指标分类体系(系统会根据指标定义的应用部门进行初始化分配),各单 位可按照各自考核要求，申请指标库中生效的指标进行应用。在指标应用管理中存储的是企业 指标库中指标的引用关系。

指标统计报告主要是从指标应用部门和指标两种维度出发，统计不同类别指标或不同指标 被各应用部门引用的次数。

(2)指标地图。

指标地图通过调用元数据地图展现功能来实现，可以提供企业指标的全景显示、指标的健 康状况。指标地图可以根据指标体系，实现多种视角、多层级的拓扑展现。

(3)指标版本管理。

指标版本管理是按照不同时间、不同分支记录指标版本状态，支持对不同版本指标的引用、 查询、统计。

(4)指标订阅管理。

指标订阅管理可以实现按需订阅指标的功能。

(5)指标查询管理。

指标查询管理可以实现多维度、多条件的查询。

(6)指标分析。

指标分析是通过指标的应用情况对指标问题、指标质量进行分析。分析结果可以通过导出 的方式形成分析报告。

(7)指标血缘分析。

指标血缘分析主要用于分析指标的上下游组织关系、映射关系等，让企业能够清楚地掌握 指标的血缘关系。

(8)指标质量管理。

指标质量管理主要包括指标质量的稽核、指标问题的处理。通过指标质量管理解决指标应

第21章 数据指标管理工具 **235**

用过程中出现的问题，提升指标的质量。

(9)指标应用追溯。

指标应用追溯可以让企业详细了解指标的应用情况，并对指标应用中出现的问题进行追溯、

处理。

**本章精要**

数据指标直接反映企业的生产运营状况，为企业的决策提供数据支撑。随着企业数字化转 型的驱动，数据指标的作用越来越重要，数据指标管理中的诸多痛点使得统一指标管理成为企 业的普遍需求。通过数据指标管理工具可以实现对数据指标标准的集中、统一、规范管理，形 成指标的维护、分发、应用、检查的持续更新的管理闭环，促进数据指标的共享使用和质量改 进，为企业业务运营和领导决策提供完整、 一致、规范的数据指标服务。

第22章

主数据管理工具

在企业信息化建设过程中，主数据建设已经越来越受到管理者的重视。对主数据的集中管 理为企业整合及共享系统中的数据提供了关键的基础支持。因此，构建主数据标准化体系、建 立主数据交互和共享标准、实现主数据全生命周期管理，已经成为提高企业信息化建设效益、 改善业务数据质量、在高端决策上为企业提供强有力支持的重要途径。

主数据全生命周期管理的理念和应用全面解决了原有主数据管理流程不规范、平台不统一、 依靠人工校验的问题，实现了从分散到集成、从局部到全面、从手工(非专业)到专业自动化 流程管理的转变，大幅度提高了数据处理的效率，以及主数据应用的唯一性、准确性和规范性。

主数据管理工具的核心功能是主数据提取与整合、主数据模型管理、主数据清洗管理、主 数据全生命周期管理、主数据质量管理、主数据发布与共享，其还能够提供主数据的建立、审 批、查询、修改等功能，以及主数据系统与其他系统的“接口通道”,将主数据同步传递到数据 消费系统。

**22.1** **主数据提取与整合**

主数据提取与整合是指从主数据的来源系统中提取、汇总主数据信息，并针对来自不同系 统的主数据对象，根据主数据的数据模型，对相同或相似的对象进行辨别，然后将其整合到主 数据管理系统的主数据资源库中，为企业提供全局、统一的核心业务实体的主数据视图。

第22章 主数据管理工具 **237**

**22.2** **主数据模型管理**

主数据模型管理是从模块化、功能化的角度管理主数据模型和主数据的数据结构，实现对 主数据属性数据元、约束条件、校验规则、编码规则等方面的定义与管理。主数据模型管理是 主数据管理工具的核心功能，主要包括主数据模型的创建申请、审批和变更申请、审批过程的 管理，属性数据元的定义与管理，主数据编码生成方式的定义与管理，以及各种主数据属性数 据元的校验规则和约束条件等。

(1)主数据模型管理。

主数据模型管理能够支持定义主数据分类规则和编码规则，主数据属性数据元的显示规则、 校验规则、输入规则、取值规则等，通过属性模板定义、属性特征值取值范围集合、数据多视 图设置，实现数据模型配置的高度灵活性。

主数据模型的版本控制记录了主数据模型在应用过程中的动态变化。主数据管理系统支持 主数据模型的版本控制和变更管理流程，保证在模型转换过程中主数据代码的连贯性。主数据 模型定义了主数据代码的数据结构，描述了属性数据元定义、取值范围约束等。

主数据模型分类遵从主数据分类标准，每一类主数据的数据模型，都遵从该类主数据的主 数据标准。每一类主数据都需要在主数据系统中建立对应的数据模型。

(2)主数据编码管理。

主数据对象即主数据管理的对象，如某个供应商。主数据代码包括定义主数据对象的唯一 ID, 和该 ID 所代表的信息，即是定义主数据对象的一组属性数据元组合。在主数据代码数据 库中，每一行都以主数据 ID 为指针，每一列都有固定的格式，该格式由主数据标准和根据主数 据标准而制定的数据模型所决定。每个主数据对象的主数据代码都是唯一的。

主数据编码是根据主数据分类及编码规则，给每个主数据对象设定一个代码。主数据编码 是主数据系统的重要功能。主数据编码管理的范围除了包括直接参与业务交易的业务对象数据， 还应该包括国家代码、币种、计量单位、地区划分等国家通用信息代码。

主数据代码在各个业务系统中需要保持一致，以利于各个业务系统之间有关联的数据或信 息可以建立关联，避免形成数据孤岛。

主数据编码规则和配置功能应支持多种编码属性标示，如流水码、随机码、特征组合码、 对照码等，还要能够灵活定义主数据编码生成方式，以符合多系统集成的要求。

**238** 数据治理——工业企业数字化转型之道

(3)校验规则管理。

校验规则是指创建主数据实体时的数据校验方法，例如唯一性校验、相似性校验等。具体 来说，校验可以引用字段和属性，执行算术运算、字符串和逻辑运算，调用内置函数，引用事 先定义的校验规则。

主数据管理系统需要支持多种校验规则，并可自定义校验规则。例如属性值组合校验、属 性值表达式校验、数据唯一性校验、多属性组合关联校验、取值范围校验、相关附属表校验， 以及直接利用主数据管理系统定义的数据内容进行取值校验、同名库校验、重复性校验、正则 表达式校验、关联性校验等。这些校验规则实现了对数据之间的精确查重和模糊查重，在校验 结束后会自动提示输入错误，有效地保证主数据的唯一性和规范性，最大限度地降低和避免人 为因素导致的信息错误。

**22.3** **主数据清洗管理**

主数据清洗是指发现并改正不完整、不正确和不一致的主数据属性数据元描述，从而提高 主数据质量。主数据清洗是发现并纠正代码数据中可识别错误的最后一道环节，包括检查主数 据的一致性，处理无效值和缺失值等。由于主数据代码标准的制定往往滞后于信息化建设，主 数据系统通常要从多个业务系统中抽取主数据，这就避免不了有的主数据是错误的、有的主数 据相互冲突，所以要按照数据模型定义的规则，把零散、重复、不完整的主数据清洗干净，得 到准确、完整、 一致、有效、唯一的新的主数据，这就是主数据清洗。

一般来说，通过主数据清洗可以保证主数据的唯一性、准确性、完整性、一致性和有效性。

(1)唯一性：是指主数据不存在重复记录。

(2)准确性：是指主数据与其对应的客观实体的特征相一致。

(3)完整性：是指主数据不存在缺失记录或缺失字段。

(4)一致性：是指同一实体的同一属性的值在不同的系统中是一致的。

(5)有效性：是指主数据的属性数据元满足事先定义的条件或在一定的取值范围内。

如果主数据不能符合以上特征，则会影响到各个应用系统中的数据质量、数据的汇总及分 析，以及基于数据的决策，给企业造成不可估量的损失。例如，主数据的不一致会带来业务交 易数据的不一致，在不同业务系统中的数据不一致会使系统之间的数据共享变得困难，无法对 企业的整体运营情况进行统一的分析和决策。

第22章 主数据管理工具 **239**

**22.3.1** **主数据清洗的内容**

‘在以下几种情况下，需要清洗主数据。

(1)主数据缺失：主数据缺失主要有两种情况：①一些记录或一条记录里缺失一些值(或 者两种情况都有),原因可能有很多种， 一般为由系统导致的或由人为因素导致的。当有空值时， 为了不影响分析的准确性，要么不将空值纳入分析范围，要么进行补值。将空值纳入分析范围 会减少分析的样本量，补值需要根据分析的计算逻辑，选用平均数、零，或者等比例随机数等 来填补。②缺失一些记录，若业务系统中还有这些记录，则可以通过系统再次导入，若业务系 统中也没有这些记录了，则只能进行手工补录或者放弃。

(2)主数据重复：主数据重复主要有两种情况。①相同的记录出现多条，这种情况比较好 处理，只要去掉重复记录即可。②记录中的某些值重复，比如有两条会员记录，其余值都一样， 就是住址不一样，这种情况相对复杂。如果记录有时间属性，则还能判断以新值为准，如果记 录没有时间属性，就无从下手了，只能人工判断。

(3)主数据错误：主数据错误主要是因为没有严格按照规范记录。比如：异常值，价格区 间为[0,100],却有价格为180的记录；格式错误，日期格式被录成字符串格式；数据不统一， 有的记录叫北京，有的叫 BJ或 beijing。对于异常值，可以通过区间限定来发现并排除；对于格 式错误，需要从系统层面找原因；对于数据不统一的情况，系统无能为力，因为它并不是真正 的“错误”,系统并不知道 BJ和 beijing 是同一事物，只能人工干预：通过清洗规则表，给出匹 配关系，第一列是原始值，第二列是清洗值，用规则表去关联原始表，用清洗值进行分析。更 高级的办法是通过近似值算法自动发现可能不统一的数据。

(4)主数据不可用：主数据虽然正确，但不可用。比如地址写成“北京海淀中关村”,当想 分析“区”级别的信息时，还要把“海淀”拆出来才能用。这种问题最好从数据源头解决，事 后补救只能通过关键词匹配，且不一定能全部解决。

**22.3.2** **主数据清洗的一般过程**

在对主数据进行清洗之前，要先对主数据进行预处理。主数据预处理一般分为两个步骤， 第一步是将数据导入处理工具，比如数据库；第二步为分析属性数据元，包括字段解释、数据 来源、代码表等一切描述数据的信息，抽取一部分主数据作为样本数据，通过人工查看，先对 主数据有直观的了解，为之后的清洗做准备。

下面进入主数据清洗的核心内容，包括缺失值清洗、格式内容清洗、逻辑错误清洗、非需

**240** 数据治理——工业企业数字化转型之道

求数据清洗、关联性验证、干净数据回流。

**1.** **缺失值清洗**

缺失值是最常见的数据问题，处理缺失值可以按照以下4个步骤进行。

(1)确定缺失值范围：计算每个字段的缺失值比例，然后按照缺失值比例和字段的重要性，

分别制定策略。

(2)去除不需要的字段：不需要的字段直接删掉即可，但建议每进行一个动作都备份一下 数据，或者在小规模数据中试验成功后再处理全部数据，避免删错了数据，导致数据无法恢复。

(3)填充缺失值：对缺失值进行填充，方法有以下3种。

①以业务知识或经验推测填充缺失值；

②以同一指标的计算结果(均值、中位数、众数等)填充缺失值；

③以不同指标的计算结果填充缺失值。

(4)重新取数：如果某些指标非常重要但缺失率又高，就需要向业务人员了解，以重新获 取相关数据。

**2.** **格式内容清洗**

有些数据是由人工收集或用户填写而来的，很有可能在格式和内容上存在一些问题， 一般 来说，格式内容存在的问题主要有以下几类。

(1)时间、日期、数值、全半角等显示格式不一致。

这种问题通常与输入端有关，在整合多种数据源中的数据时也有可能会遇到，将其处理成 统一的某种格式即可。

(2)数据中有不该存在的字符。

某些数据中可能有不该存在的字符。最典型的就是数据的头、尾、中间出现空格，也可能 出现姓名中存在数字、身份证号中出现汉字等问题。在这种情况下，需要以半自动校验、半人 工的方式来找出可能存在的问题，并去除不需要的字符。

(3)内容与该字段应有内容不符。

将姓名写成了性别、身份证号写成了手机号等，均属于这种问题。但该问题的特殊性在于： 并不能简单地通过删除来处理，因为成因有可能是人工填写错误，也有可能是前端没有校验，

第22章 主数据管理工具 **241**

还有可能是导入数据时部分或全部存在列没有对齐，因此要详细识别问题类型。

格式内容问题是比较细节的问题，但很多分析结果错误都是由此问题引起的，比如跨表关 联失败、统计值不全、模型输出失败。因此，务必要注意这部分与主数据的清洗工作。

**3.** **逻辑错误清洗**

逻辑错误清洗是修正逻辑推理有问题的数据，防止由于数据错误而导致分析结果错误。逻 辑错误清洗主要包含以下几种情况。

(1)去重。

要去除数据表中的重复数据。比如物料代码中经常存在一物多码的情况，因此在这一步就 要标记出重复的数据，以便建立映射关系，以进行数据去重的工作。

(2)修正不合理值。

要修正数据中的不合理值。比如有人在填表时随意填写，不注重检查，将年龄填成580岁， 这时就要将数据修正准确，如果不能修正，则要么删掉，要么按缺失值处理。

(3)修正矛盾内容。

有些数据内容是可以互相验证的。比如：身份证号是1329321990××××××××,年龄是 18岁，在这种时候，需要根据字段的数据来源，判定哪个字段提供的信息更为可靠，去除或重 构不可靠的字段。

**4.** **非需求数据清洗**

非需求数据清洗就是把不要的字段删除。在实际操作中，要具体问题具体分析，在非需求 数据清洗中，经常会遇到一些问题。例如，把看上去不需要但实际上对业务很重要的字段删了； 某个字段觉得有用，但又没想好怎么用，不知道是否该删；操作失误，删错字段了。对于前两 种情况，如果数据量没有大到不删字段就没办法处理的程度，那么能不删的字段就尽量不删， 对于第3种情况，就需要建立数据备份机制，保证数据能恢复。

**5.** **关联性验证**

如果数据有多个来源，则有必要进行关联性验证。关联数据变动在数据库模型中就应该涉 及。多个来源的数据整合清洗是非常复杂的工作， 一定要注意数据之间的关联性，不要因为数 据之间互相矛盾而造成下游系统无法使用的情况。

**242** 数据治理——工业企业数字化转型之道

**6.** **干净数据回流**

当主数据被清洗后，应该将干净的主数据替换数据源中原来错误的主数据。这样不仅可以 提高原系统的数据质量，还可避免将来再次抽取主数据时进行重复的清洗工作。

**22.4** **主数据全周期管理**

主数据全生命周期管理是主数据管理工具的核心功能。在创建主数据模型实体后，业务管 理流程生成相应的实例化业务功能。主数据管理工具(主数据管理系统)应提供数据申请、数 据审核、数据变更、数据查询、数据归档等功能。

**1.** **数据申请**

在创建主数据模型后，各业务系统可以根据需要向主数据管理系统提交主数据申请。在申 请过程中，提前定义的数据约束规则能够自动进行初始数据的校验，提供相似的主数据供用户 参考，并提示申请进度。

数据申请要能够实现以下业务要求。

(1)支持实时创建数据编码；

(2)上传数据文件时支持通用格式；

(3)支持外部系统通过接口服务方式提交数据申请；

(4)在提交数据申请时，根据元数据定义规则，支持查重、检查、搜索功能；

(5)提供附件上传功能，例如 PDF、JPG、DWF 等格式文件的上传；

(6)支持按照模板批量导入编码初始文件，按细分类描述规则自动生成特征描述；

(7)支持现有编码条目的复制、添加等功能。

**2.** **数据审核**

数据审核主要包括自动校验和人工审批两个方面。

在提交数据申请后，主数据管理系统将首先进行自动校验。自动校验实现对数据之间的精 确查重和模糊查重，并通过提前配置的校验规则对数据进行检查，保证主数据的唯一性和规范 性。

经过主数据管理系统自动校验之后的数据申请需要按审批流程进行人工审批。在人工审批

第22章 主数据管理工具 **243**

过程中系统会记录审批意见，生成审批日志。审核的内容及功能如下。

(1)支持自定义审批流程，以及利用图形化工作流实现数据审核过程，并支持工作流的版 本控制；

(2)支持在线查重、检查与搜索，支持相关日志自动创建和下载；

(3)支持根据提供的审批任务列表，实时对审批任务列表中的申请进行数据合法性、业务 合规性的快速审批；

(4)支持对审批流程中的各个环节进行有效管理，跟踪、监控各个审批环节的信息；

(5)工作流支持会签、互斥、汇总、分支等多种审批方式，支持按条件跳转审批，支持通 过文字和图形显示审批进程和审批状态，以及支持通过电子邮件、手机短信或者内部消息等发 送提示通知：

(6)当主数据审核人员发现数据错误时，可以对错误主数据进行审核并退回，以及填写审 核意见供数据申请人查看等。

**3.** **数据变更**

对于已审核的主数据、模板、属性值等内容，可以进行调整、维护和变更。例如，要对物 料编码的设置信息进行调整，首先需要提交变更申请，主数据管理系统会根据已定义的数据约 束规则自动进行数据校验，并对维护内容进行审核，审核通过后变更内容才能生效，同时保留 数据的历史版本，实现可追溯的版本管理。

主数据管理系统支持实时的变更单条数据及批量数据的申请，并生成数据变更日志。和数 据申请过程类似，主数据管理系统也可以提供自动提示功能，辅助用户完成主数据的变更申请。 变更生效后，主数据管理系统会以邮件或其他方式发消息，通知数据申请者和所有使用该数据 的最终用户。

当主数据记录失效时，主数据管理系统不允许将其直接删除，只能进行核销，以便后续有 需要时可以调用该条数据的历史数据。数据核销也是一种数据变更，即将一条可用数据的状态 更改为“已核销”,过程与其他数据变更的过程相同。

**4.** **数据查询**

主数据管理系统能够提供标准查询和模糊查询功能。用户可以实时查询主数据的所有信息， 包括申请、审批、明细属性、变更历史、分发历史、数据分发接口日志等，并按照不同的需求

**244** 数据治理 工业企业数字化转型之道

下载和打印查询结果。同时，主数据管理系统还提供便捷查询和高级查询功能，支持将查询条 件保存作为共用检索，以实现个性化查询。

除了简单的查询功能，用户还可以按照申请单列表项进行过滤、排序、查询和统计，根据 用户需求，生成各种统计报表，供用户查询分析。同时，主数据管理系统还提供了开放功能供 用户自定义报表，用于个性化的查询分析。根据统计方式，可以将报表分为主数据信息统计报 表、主数据提报审核统计报表和主数据分发情况统计报表。

(1)主数据信息统计报表。

主数据信息统计报表主要统计主数据管理系统中维护的主数据情况，其统计内容包括主数

据的类型、每种类型的主数据数量及对应的明细信息。

(2)主数据提报审核统计报表。

主数据提报审核统计报表主要统计主数据管理系统中主数据的提报审核情况，其统计内 容包括根据在一定时间段内查询数据的提报数量、审核完成数量、待审核数量、退回数量，以 及每种类型主数据对应的明细信息。

(3)主数据分发情况统计报表。

主数据分发情况统计报表主要统计主数据管理系统中的主数据分发给其他应用系统的情况， 其统计内容包括分发的主数据类型、每种类型的主数据分发的数量，以及分发成功的数量、分 发失败的数量。

**5.** **数据归档**

数据归档指以物理方式将主数据系统中具有较低业务价值的主数据迁移到更适合、更经济、 更高效的历史库中。因此，在主数据管理系统中，对于不再使用或无法满足业务需求的主数据 可以进行归档及核销处理。归档后的主数据不能被更改，但能被查询及调用。同时，主数据管 理系统也支持定期对日志信息进行归档及多种归档信息的查询。

**22.5** **主数据质量管理**

主数据质量管理是指对主数据全生命周期内的每个阶段里可能引发的各类数据质量问题进 行识别、度量、监控和预警等一系列的管理活动，并通过改善和提高企业的管理水平，进一步 提高数据质量。主数据质量管理是循环的管理过程，其终极目标是通过可靠的主数据，提升主

第22章 主数据管理工具 **245**

数据在使用中的价值，并最终为企业的数字化转型、管理模式优化、经营决策提供重要支撑。

(1)主数据质量管理目标和任务。

主数据质量管理不仅包含对数据质量的改善和管理，还包含对组织的改善和管理。针对数 据质量的改善和管理，主要包括改正和完善数据分析、数据评估、数据清洗、数据监控、错误 预警等功能；针对组织的改善和管理，主要包括确立组织数据质量改进目标、评估组织流程、 制订组织流程改善计划、制定组织监督审核机制、实施改进策略和评估改善效果等。主数据质 量管理可以被划分为5个阶段，即定义、测量、分析、改进和控制。

(2)主数据质量管理评估维度。

主数据质量管理评估维度就是数据质量评估标准，它们衡量了数据的完整性、规范性、 一 致性、准确性、唯一性、关联性、及时性和可用性等。对主数据来说， 一致性和准确性至关重 要。

下面分别针对数据质量评估和管理质量评估通常涉及的几个维度进行介绍。

**1.** **数据质量评估维度**

(1)完整性：主要度量哪些数据是缺失的或者哪些数据不可用，例如超出阈值。该维度是 对数据质量的基本度量，例如完整性/填充率、有效性、范围、最大值和最小值。在主数据管理 系统中，可以通过自定义校验规则来保证主数据的完整性。

(2)规范性：主要对数据标准、数据模型、业务规则、元数据和参考数据进行完整性、 质量及归档的度量。规范性度量了哪些数据未按统一格式存储，为数据质量评估结果提供对比 标准。

(3)一致性：主要度量哪些数据的值在信息含义上是冲突的，是对各种不同的数据仓库、 应用和系统中所存储或使用的信息的等价程度的度量，也是数据等价处理流程的度量标准。满 足一致性是不同系统间进行数据集成的基础。

(4)准确性：主要对数据内容的正确性进行度量。度量数据的准确性需要将数据与其所描 述的实际对象进行比较，并且评估流程通常需要人工介入，因此效率较低。

(5)唯一性：主要度量哪些数据是重复数据或者数据的哪些属性是重复的。保证数据的唯 一性能避免数据冗余，同时也减少出现数据不一致的潜在可能性。

(6)关联性：主要度量哪些关联的数据缺失或者未建立索引。

**246** 数据治理 数字化转型之道

(7)及时性和可用性：主要度量预期时段内的数据对特定应用的及时程度和可用程度。由 于数据的数值随时间而不断变化，因此，对该维度的评估、检查需要贯穿主数据的全生命周期， 评估数据是否是最新且即时可用的，以满足业务需求。

**2.** **管理质量评估维度**

管理质量评估主要评估企业的数据管理流程、资源配置等方面是否符合规范，以及是否严 格按照计划执行。具体包括以下维度。

(1)管控流程：此维度用于度量数据在其全生命周期内的一切资源是否得到了控制和规范， 即在数据的计划、产生、变更直至消亡的过程中，与数据相关的计划、规范、描述是否受到控 制。评估指标包括配置项的细化颗粒度、基线准确度和频度，以及变更流程是否合理完善等。

(2)培训：此维度用于度量数据的生产者和使用者在数据的全生命周期内的一切活动中是 否经过了知识和技能的培训，培训效果是否满足岗位需要；受训的知识和技能是否经过审核和 确认，受训的内容是否与企业的文化和价值观一致；培训流程是否合理、完善，等等。

(3)验证和确认：此维度用于度量数据在其全生命周期内是否得到验证和确认。评估内容 包括是否通过验证流程确保数据满足指定的要求，是否通过“确认”流程保证数据在计划的环 境中满足使用的要求，“验证”和“确认”的流程是否完善。

(4)监督和监控：此维度用于度量产生和使用数据的流程在数据的全生命周期内是否真正 受控(若信息、技术、计划、流程、制度等脱离监督和监控，则会导致数据质量低下),以及监 督和监控的流程是否完善。

**22.6** **主数据发布与共享**

主数据发布与共享是通过数据中间件，根据预定义的分发服务、参数、服务描述、分发频 率，向目标业务系统分发主数据，同时自动创建数据分发同步日志。通过抽取规则的配置实现 从业务系统采集主数据。

应用系统与主数据管理系统间的交互包括两个方面： 一是数据接收，即其他业务系统作为 数据源，主数据管理系统接收业务系统发送的数据；二是数据分发，即主数据管理系统作为数 据源，业务系统接收主数据管理系统发送的数据。

主数据发布与共享是实现主数据同步和主数据一致性应用集成的关键过程。数据分发需要

第22章 主数据管理工具 **247**

支持分发目标系统、分发频率、分发数据范围、数据同步规则等的自定义功能；能够实现分发 日志的自动跟踪和记录；能够支持多种分发方式和分发数据协议，并支持异常处理，例如设定 分发失败后重发数据的处理机制，实现全面的数据监控管理，保证主数据在多系统间的完整性 和一致性。

**1.** **数据传输方式**

从数据传输方式来划分，数据集成有同步传输和异步传输两种方式。

(1)同步传输。

源系统向目标系统发送一条或者一批主数据后，源系统客户端处于等待状态，当目标系统 接收数据成功后，会同时返回处理结果，源系统接收目标系统返回的处理结果后，再开始下一 条或下一批数据的发送，其数据的交互是实时的。

(2)异步传输。

源系统向目标系统按顺序发送所有待处理的主数据，不需要等待目标系统返回处理结果， 目标系统对接收到的主数据进行处理后，另行调用服务返回处理结果，源系统接收目标系统返 回的处理结果，其数据的交互不是实时的。

**2.** **数据获取方式**

从数据获取方式来划分，数据集成有主动分发和被动分发两种方式。

(1)主动分发方式。

主动分发方式由主数据管理系统制定分发策略，包括分发目标系统、分发数据对象范围、 分发频次和时间，提供有针对性的主数据内容。在主动分发方式下，目标系统需要收标准 Web Service,并具备日志管理、版本管理等功能。

(2)被动分发方式。

被动分发方式由目标系统制定数据获取策略。在被动分发方式下，服务的调用时间、调用 频次均由目标系统自行设定。

**3.** **数据集成协议**

要实现数据集成和分发策略的标准化，在分散的业务信息系统间最大限度地保证主数据的 完整性、 一致性，就需要保证业务系统与主数据管理系统集成接口的统一和规范。现在通用的

**248** 数据治理——工业企业数字化转型之道

集成方式采用松耦合方式，以SOA 架构为基础，以 Web Service 为传输协议。

**本章精要**

主数据管理工具是主数据全生命周期管理平台，也是主数据标准、运维体系落地的重要保 障。本章从主数据提取整合、主数据模型管理、主数据清洗管理、主数据生命周期管理、主数 据质量管理、主数据发布与共享6个部分讲述。主数据工具是企业主数据项目建设成功的重要 保障，建议企业选择应用成熟、经验丰富的厂商提供的主数据管理工具。

第23章

元数据管理工具

**23.1** **元数据管理工具概述**

元数据管理工具主要包含元数据采集、元数据识别、元数据分类、元数据展示、元数据应 用和元数据搜索等功能。

**1.** **元数据采集**

元数据采集功能能够适应异构环境，支持从传统关系型数据库、大数据平台、数据产生系 统、数据加工处理系统、数据应用报表系统中采集全量元数据。采集内容包括过程中的数据实 体(系统、库、表、字段的描述)及数据实体加工处理过程中的逻辑。元数据管理系统根据数 据源的连接信息、同步周期及开始时间，定时自动解析、获取，并更新元数据信息，保证系统 中元数据信息的及时、有效。

**2.** **元数据识别**

元数据识别功能能够从本身不包含元数据信息的数据(比如非结构化数据)中提取特征，

并以此识别元数据。

**3.** **元数据分类**

元数据分类功能能够根据业务特点和管理的需要，动态分类元数据(包括技术元数据、业 务元数据和管理元数据等)。

**4.** **元数据展示**

元数据展示功能能够根据类别、类型等信息，展示各个数据实体的信息及其分布情况，数

**250** 数据治理 工业企业数字化转型之道

据实体间的组合、依赖关系，以及数据实体加工处理上下游的逻辑关系。

**5.** **元数据应用**

元数据应用功能能够利用元数据发现数据之间的关联性， 一般包括数据地图、数据血缘分 析、影响分析、全链分析、热度分析等。

**6.** **元数据搜索**

元数据搜索功能可根据数据源的库、类型等搜索元数据信息。

**23.2** **元数据在数据架构管理中的应用**

元数据通常分为三类：业务元数据、技术元数据和操作元数据。业务元数据(Business Metadata) 主要关注数据的内容和条件，还包括与数据治理相关的详细信息。在规划和设计数

据架构时，必须涵盖的元数据包括：

(1)业务域、业务过程、业务活动定义及描述；

(2)数据集、表、字段的定义和描述；

(3)业务规则、转换规则、计算公式和推导公式；

(4)数据模型、属性数据元定义及属性数据元有效值约束；

(5)数据质量规则和核检结果；

(6)数据的CRUD 规则；

(7)数据溯源和数据血缘；

(8)信息代码的数据结构和编码标准；

(9)特定的数据元记录系统；

(10)利益相关者的联系信息(例如，数据所有者、使用者和管理者);

(11)数据的安全/隐私级别；

(12)已知的数据问题；

(13)数据使用说明。

第23章 元数据管理工具 **251**

**23.3** **元数据在数据资产目录中的应用**

在数据资产目录中，元数据管理组件用于解析数据来源系统(简称源系统)的数据结构， 并将源系统的数据结构转换为目标系统的数据结构。数据资产目录中的数据资产清单包括：

(1)描述源系统结构的物理元数据，如表和字段。

(2)描述语义信息的逻辑元数据，如数据库描述、数据质量评估和相关的数据管理策略。

(3)描述如何在各种业务场景中使用数据资产的行为元数据。行为元数据可能是最重要的，

因为它可以自动洞察系统中的每个对象活性或者热度。

(4)源系统数据结构和目标系统数据结构的转换规则。

**23.4** **元数据在主数据管理中的应用**

元数据可以用来描述主数据的数据结构，定义主数据的数据模型。 一条主数据往往由一条 以上的属性数据项组成。每条属性数据项包括属性名称、同名词库、取值范围、计量单位、校 验方式、页面显示控制、前/后置符号、附表取值、版本等，它们都属于定义和描述主数据的元 数据。从主数据管理维度看，数据模型决定了数据质量，因而元数据管理也被延展到数据质量 管理的范畴。

**23.5** **元数据在数据交换和共享中的应用**

数据交换和共享是基于技术元数据 (Technical Metadata) 的。技术元数据提供了关于业务 数据的技术细节，包括在业务系统中的物理存储格式，以及在系统内和系统之间的数据流转过 程中的信息。技术元数据包括：

(1)物理数据库表名和字段名；

(2)字段属性；

(3)数据库对象的属性；

(4)访问权限；

(5)数据的CRUD (增、删、改、查)规则；

(6)物理数据模型，包括数据表名、键和索引；

**252** 数据治理——工业企业数字化转型之道

(7)记录数据模型与实物资产之间的关系；

(8)ETL 作业详细信息；

(9)文件格式模式定义；

(10)从源系统到目标系统的映射文档；

(11)数据血缘文档，包括上游和下游变更影响的信息；

(12)程序和应用的名称和描述；

(13)周期作业(内容更新)的调度计划和依赖；

(14)恢复和备份规则；

(15)数据访问的权限、组和角色。

技术元数据用于ETL 设计和程序开发，与数据交换的应用高度耦合，因此，技术元数据管 理组件往往是数据交换平台 ETL 组件中的一部分，或者与ETL 组件联合使用。

**23.6** **元数据在大数据平台中的应用**

在OLAP (联机分析处理)平台或大数据平台中，操作元数据(Operational Metadata)描述

了处理和访问数据的细节。例如：

(1)批处理程序的作业执行日志；

(2)抽取历史和结果；

(3)调度异常处理；

(4)审计、平衡、控制度量的结果；

(5)错误日志；

(6)报表和查询的访问模式、频率和执行时间。

操作元数据与数据采集、数据交换、数据共享、数据查询、数据存储计算、数据挖掘等关 联密切。

第23章 元数据管理工具 **253**

**本章精要**

元数据管理贯穿数据的全生命周期。从工具层面看，它与各类数据治理工具都有交集；从 应用层面看，它属于数据治理的高阶应用，有成熟的理论体系、工具和案例。元数据管理更是 支撑数据治理体系的组件和技术工具。

元数据可以用来定义主数据的数据结构和数据模型，定义数据元的数据格式和取值范围， 定义数据指标的计算公式，以及定义交易数据的库表结构。

第24章

时序数据处理工具

**24.1** **通用大数据处理工具的不足**

在物联网、车联网、工业互联网兴起之后，大家都想用通用的大数据平台来处理其中的数 据。现在市场上流行的物联网、车联网等大数据平台几乎无一例外都是这类架构，但这套通用 处理工具的效果如何?可以说有很多不足，主要表现在以下几个方面。

(1)开发效率低：因为不是单一的软件，需要集成4个以上模块，很多模块都不是标准的 POSIX 或 SQL 接口，都有自己的开发工具、开发语言、配置等，需要一定的学习成本。而且由 于数据会从一个模块流动到另外一个模块，数据的一致性容易受到破坏。同时，这些模块基本 上都是开源软件，总会有各种漏洞， 一旦被一个技术问题卡住，总要耗费工程师不少时间。总 的来讲，企业需要搭建一支优秀的团队才能将这些模块顺利地组装起来，因此需要耗费较多的 人力资源。

(2)运行效率低：现有的这些开源软件主要用来处理互联网上的非结构化数据，但是通过 物联网采集来的数据都是时序的、结构化的。用非结构化数据处理技术来处理结构化数据，无 论是存储还是计算，消费的资源都大很多。

(3)运维成本高：每个模块，无论是 Kafka、HBase、HDFS 还是 Redis,都有自己的管理 后台，都需要单独管理。在传统的信息系统中，数据库管理员只要学会管理 MySQL 或是 Oracle 就可以了，但现在数据库管理员需要学会管理、配置、优化很多模块，工作量大了很多。由于 模块数过多，定位一个问题就变得更为复杂。比如，用户发现有一条采集的数据丢失了，至于 是 Kafka、HBase、Spark 丢失的，还是应用程序丢失的，则无法迅速定位，往往需要花很长时

第24章 时序数据处理工具 **255**

间，只有将各模块的日志关联起来才能找到原因。而且模块越多，系统整体的稳定性就越低。

(4)产品推出慢、利润低：由于源软件研发效率低，运维成本高，导致将产品推向市场的 时间变长， 让企业丧失商机。而且这些开源软件都在演化中，要同步使用最新的版本也需要耗 费一定的人力。除互联网头部公司外，中小型公司在通用大数据平台上花费的人力资源成本一 般都远超过专业公司的产品或服务费用。

(5)对于小数据量场景，私有化部署太重：在物联网、车联网场景中，因为涉及生产经营 数据的安全，很多还是采取私有化部署。而每个私有化部署，其处理的数据量有很大的区别， 从几百台联网设备到数千万台联网设备不等。对于数据量小的场景，通用的大数据解决方案就 显得过于臃肿，投入与产出不成正比。因此，有的平台提供商往往有两套方案， 一套针对大数 据场景，使用通用的大数据平台， 一套针对小数据场景，使用 MySQL 或其他数据库。但这样 会导致研发、维护成本提高。

**24.2** **时序数据处理工具应具备的功能和特点**

时序数据处理工具(系统)需要具备哪些功能?与通用的大数据处理工具相比，它具备什 么样的特点呢?下面仔细分析一下。

(1)必须是高效的分布式系统。工业互联网产生的数据量巨大，比如，全国有5亿多台智 能电表，每台智能电表每隔15分钟采集一次数据，全国的智能电表一天就会产生500多亿条记 录。这么大的数据量，任何一台服务器都无法处理，因此时序数据处理系统必须是分布式的、 水平扩展的。为降低成本， 一个节点的处理性能必须是高效的，需要支持数据的快速写入和快 速查询功能。

(2)必须是实时的处理系统。对于互联网大数据的应用场景，大家所熟悉的都是用户画像、 推荐系统、舆情分析等，这些场景并不需要数据计算具有实时性，批处理即可。但是对于工业 互联网大数据的应用场景，则需要基于采集的数据做实时预警、决策，延时要控制在秒级以内。 如果没有实时计算，则其商业价值就大打折扣。

(3)需要运营商级别的高可靠服务。工业互联网系统对接的往往是生产、经营系统，如果 数据处理系统宕机，则会直接导致停产，无法对终端消费者正常提供服务。因此，时序数据处 理系统必须是高可靠的，必须支持数据实时备份，必须支持异地容灾，必须支持软件、硬件在 线升级，必须支持在线 IDC 机房迁移，否则服务一定有被中断的可能。

**256** 数据治理 数字化转型之道

(4)需要高效的缓存功能。在绝大部分场景中，都需要能快速获取设备当前状态或其他信 息，用以报警、大屏展示等。时序数据处理系统需要提供高效机制，让用户可以获取全部或符 合过滤条件的部分设备的最新状态。

(5)需要实时流式计算。各种实时预警或预测已经不是简单地基于某一个阈值进行的，而 是需要通过将一个或多个设备产生的数据流进行实时聚合计算(并且不只是基于一个时间点， 而是基于一个时间窗口进行计算)。不仅如此，计算的需求也相当复杂，因场景而异，应容许用 户自定义函数进行计算。

(6)需要支持数据订阅。时序数据处理系统与通用大数据平台比较一致的地方是，同一组 数据往往有很多应用都需要，因此，时序数据处理系统应该提供订阅功能：只要有新的数据更 新，就应该实时提醒应用。而且这个订阅也应该是个性化的，容许应用设置过滤条件，比如只 订阅某个物理量5分钟的平均值。

(7)实时数据和历史数据的处理要合二为一。实时数据被存储在缓存里，历史数据被存储 在持久化存储介质里，而且可能依据时长，被存储在不同的存储介质里。时序数据处理系统应 该隐藏背后的存储介质，给用户和应用呈现的是同一个接口和界面。无论是访问新采集的数据 还是10年前的老数据，除输入的时间参数不同外，其余都应该是一样的。

(8)需要保证数据能持续、稳定地写入。对于物联网系统，数据流量往往是平稳的，因此 数据写入所需要的资源往往是可以估算的。其中变化的是查询、分析，特别是即席查询，有可 能耗费很多的系统资源，不可控。因此，时序数据处理系统必须保证分配足够的资源以确保数 据能够写入系统而不被丢失。准确地说，时序数据处理系统必须是一个写优先系统。

(9)需要支持灵活的多维度数据分析。对于联网设备产生的数据，需要进行各种维度的统 计分析，比如根据设备所处的地域进行分析，根据设备的型号、供应商进行分析，根据设备所 使用的人员进行分析等。这些维度的分析是无法事先设计好的，而是在实际运营过程中，根据 业务发展需求定下来的。因此，工业互联网大数据平台需要一个灵活的机制来增加某个维度的 分析。

(10)需要支持数据降频、插值、特殊函数计算等操作。原始数据的采集可能频次较高，但 在具体分析时，往往不需要对原始数据进行分析，而是需要对数据进行降频。时序数据处理系 统需要提供高效的数据降频操作。不同设备采集数据的时间点是很难一致的，因此，分析一个 特定时间点的值，往往需要插值才能解决，系统需要提供线性插值、设置固定值等多种插值 策略。

第24章 时序数据处理工具 **257**

(11)需要支持即席分析和查询。为提高数据分析师的工作效率，时序数据处理系统应该提 供命令行工具或容许用户通过其他工具，执行 SQL 查询，而不是非要通过编程接口。并且查询 分析结果可以很方便地被导出，以及被制作成各种图表。

(12)需要提供灵活的数据管理策略。 一个大的系统，其中采集的数据种类繁多，而且除采 集的原始数据外，还有大量的衍生数据。这些数据各自有不同的特点，有的采集频次高，有的 要求保留时间长，有的需要保存多个副本以保证更高的安全性，有的需要能快速访问。因此， 工业互联网大数据平台必须提供多种策略，让用户可以根据特点进行选择和配置，而且各种策 略并存。

(13)必须是开放的。时序数据处理系统需要支持业界流行的标准，提供各种语言开发接口， 包括 C/C++、Java 、Go 、Python 、RESTful 等，也需要支持 Spark、R、MATLAB等，方便集成 各种机器学习、人工智能算法或其他应用，让大数据处理平台能够不断扩展，而不是成为一个 数据孤岛。

(14)必须支持异构环境。大数据平台的搭建是一个长期工作，每个批次采购的服务器和存 储设备都会不一样，时序数据处理系统必须支持各种档次、各种不同配置的服务器和存储设备 并存。

(15)需要支持边云协同。时序数据处理系统要有一套灵活的机制将边缘计算节点的数据上 传到云端，根据具体需要，可以将原始数据、加工计算后的数据，或仅仅符合过滤条件的数据 同步到云端，并且同步可以随时取消，同步策略可以随时修改。

(16)需要单一的后台管理系统。单一的后台管理系统便于查看系统运行状态、管理集群、 管理用户、管理各种系统资源等，而且能让系统与第三方IT运维监测平台无缝集成，便于统一 管理和维护。

(17)便于私有化部署。因为很多企业出于安全及各种因素的考虑，希望时序数据处理系统 采用私有化部署。而传统的企业往往没有很强的 IT 运维团队，因此在时序数据处理系统安装、 部署上需要做到简单、快捷，可维护性强。

**24.3** **时序数据的采集**

时序数据的采集一般都是通过传感器自动进行的，包括光电、热敏、气敏、力敏、磁敏、 声敏、湿敏、电量等不同类别的工业传感器。就某一个具体的物理量而言，数据采集是很容易

**258** 数据治理 字化转型之道

的。但就整个系统而言，数据采集是相当复杂的，具体表现在以下几个方面。

(1)工业数据的协议不标准：在现实场景中，往往会出现 ModBus、OPC、CAN、ControlNet、 Profibus 、MQTT等各种类型的工业协议，而且各个自动化设备生产及集成商还会自己开发各种 私有的工业协议，导致在实现工业协议的互联互通时出现极大的难度。很多开发人员在工业 现场实施综合自动化等项目时，遇到的最大问题即是面对众多的工业协议，无法有效地进行 解析和采集数据。

(2)通信方式不统一：由于历史原因，采集的数据往往会通过局域网、蓝牙、Wi-Fi、2.5G、 3G、4G 等各种传输方式被传送到服务器中，导致各种通信方式并行存在，连接管理变得复杂。

(3)安全性考虑不足：传统的工业系统都运行在局域网中，安全问题不是考虑的重点。若 需要通过云端(特别是公有云)调度工业行业中核心的生产数据，又没有充分考虑安全问题， 则很有可能造成难以弥补的损失。

根据上述原因，企业在实际采集数据时，往往配有工业互联网网关盒子，该盒子支持各种 物理接口、通信协议和工业标准协议，将不同协议进行转换，对数据进行安全加密，统一以

MQTT(Message Queuing Telemetry Transport,ISO/IEC PRF 20922)协议或其他协议发往云端。

对于数据采集部分，因为标准性不够，就不对具体工具做介绍了。

**24.4** **时序数据处理工具**

采集后的数据一般通过网络被送往服务器或云端进行处理。相对数据采集工具而言，数据 处理工具比较统一，下面对几个流行的工具进行介绍。

(1)以PI为代表的实时数据库：从20世纪80年代起，就涌现一批实时数据库(时序数据 库的一种),专门用于处理工业自动控制或流程制造行业的实时数据。其中美国OSIsoft 公司的 PI(Plant Information) 实时数据库最典型，它提供成套的工具，包括实时写入、实时计算、存 储、分析、可视化、报警等系列功能，GE、 Simens 、Honeywell 都有类似产品。国内有庚顿、 朗坤、麦杰、力控等产品。这些产品在一定程度上满足了工业数据处理的需求，但在测点数量

暴涨、数据采集频率不断提高的大数据时代，传统实时数据库暴露出以下问题。

①没有水平扩展能力，数据量增加，只能依靠硬件的纵向扩展解决；

②技术架构老旧，很多还是运行于Windows 系统中的；

③数据分析能力偏弱，不支持现在流行的各种数据分析接口；

第24章 时序数据处理工具 **259**

④不支持云端部署，更不支持 SaaS。

在传统的实时监控场景，由于对各种工业协议的支持比较完善，实时数据库还占有较牢固 的市场地位，但是在工业大数据处理上，因为上述几个原因，几乎没有任何大数据平台采用它们。

(2)InfluxDB: 它是美国 InfluxData 公司开发的产品，提供了数据存储、查询、分析、流 式计算等系列功能。其单机版本采用 MIT 许可证，是开源且免费的(用 GO 语言开发),但其 集群版本收费。目前，该款产品在全球时序数据库榜单上排名第一。在IT 运维监测领域，该产 品由于能与多个数据采集工具，以及可视化工具无缝对接，能够方便用户快速搭建一个监测系 统，因此拥有相当大的市场份额。但在工业大数据领域，其优势不够明显，用户量还不大。

InfluxDB存储采用Key-Value 和 LSM技术，支持多列数据写入，而且是 Schemaless 模式， 无须预先定义数据表结构。同时，每条记录可以带有一组标签，便于数据流之间的聚合计算。 对于小数据量，其性能表现不错。但对于历史数据查询，其性能欠佳，而且消耗的系统资源过 多。相对其他NoSQL 数据库产品而言，InfluxDB 的数据压缩做得很好，能节省不少存储空间。

同时， InfluxDB是一个独立的软件，不仅是一个数据库，还有流式计算、报警等功能，不 依赖第三方，因此其安装部署、维护也相对简单。

(3)OpenTSDB: 这是一个 Apache 开源软件，是在 HBase 的基础上开发的，底层存储是 HBase,但其依据时序数据的特点做了一些优化。其最大的好处就是建立在Hadoop 体系上，各 种工具链成熟，但这也是它最大的缺点，因为 Hadoop 不是为时序数据打造的，导致其性能很 一般，而且需要依赖很多组件，安装部署相当复杂。

OpenTSDB采用 Schemaless 模式，不用预先定义数据结构，因此写入灵活，但每个时间序 列只能写入一个采集量，不支持多列写入。每个序列可以被打上多个标签，以方便聚合操作。 总的来讲，OpenTSDB 的性能无论是写入还是查询，都很一般，而且系统的稳定性欠佳。但吸 引人的是，它支持集群部署和支持水平扩展。

OpenTSDB 只是单一的时序数据库，因此，要完整地处理时序数据，还需要搭配缓存、消 息队列、流式计算等系列软件，使整个架构的设计和维护变得困难。

(4)TDengine: 这是来自中国的开源软件，由涛思数据研发推出。它不仅是一款时序数据 库，而且还提供缓存、消息队列、数据订阅、流式计算等系列功能，是时序数据的全栈技术解 决方案。而且它不依赖任何第三方软件，安装包只有1.5MB, 使系统设计、安装、部署和维护 都变得极为简单。

**260** 数据治理——工业企业数字化转型之道

TDengine充分利用了时序数据的特点，因此具有很强的优势，具体表现在5个方面：

①无论是插入，还是查询，性能都高出许多；

②因为性能超强，其所需要的计算资源不及其他软件的1/5;

③采用列式存储，对不同数据类型采取不同的压缩算法，所需要的存储资源不到其他软件 的1/10;

④无须分库、分表，无实时数据与历史数据之分，管理成本为零；

⑤采用标准 SQL 语法，应用可以通过标准的JDBC 、ODBC接口插入或查询数据，学习成 本几乎为零。

(5)TimeScale: 这也是一家美国公司开发的产品，开源，采用的是 Apache 2.0许可证。它 是在流行的关系型数据库Postgres 基础上开发的，因此接口与 Postgres 完全兼容，而且支持各 种复杂的 SQL 查询。因为底层存储没有充分利用时序数据的特点，所以其性能一般。它的社区 版完全开源，不支持集群，但提供企业版和云服务版。

该软件很受 Postgres用户的欢迎，但其目前在IT运维监测、工业大数据处理这些领域，市 场占有率还很有限。

**本章精要**

时序数据处理应用于物联网、车联网、工业互联网领域的过程数据采集、过程控制，并与 过程管理建立一个数据链路，属于工业数据治理的新兴领域。从工具维度看，时序数据处理工 具与传统时序数据库的差异很大。后者局限于车间级的可编程逻辑控制器，而非企业级。

企业级的时序数据处理，首先是基于数据架构和数据模型的。数据架构决定了哪些时序数 据需要采集，如何处理，用于哪些业务场景，用于时序数据采集的规划与设计开发；数据模型 用于解析时序数据的数据结构。

驱动时序数据处理的是工业知识体系，需要业务部门更多地参与，涉及车间层面的专业技 术人员、企业层面的跨领域技术专家和运营管理人员。

第25章

数据质量管理工具

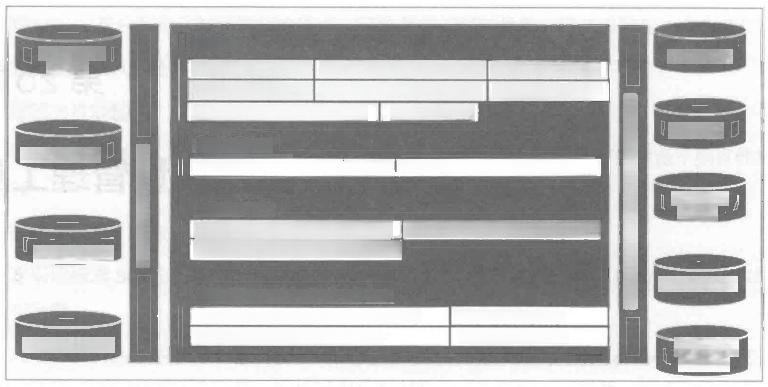
在大数据时代，数据质量问题成为影响数据分析和利用效能的“最后一公里”。目前，数据 治理、ETL、数据分析、数据集成、数据可视化等领域都已经取得飞速发展，而由数据质量问 题引发的数据质量衡量、数据质量检验、数据质量监控、数据质量提升等技术实现手段仍有 较大的发展空间。随着数据价值落地诉求的日益增长，数据质量相关技术重新被业界所关注 并逐步加大投入，以期获得革命性的改进，提升数据质量管理的效率，降低数据质量优化的 成本。

本章介绍数据质量管理理论、工具、行业实践案例等内容，让读者建立从理论到微观技术 实操的完整的数据质量管理认知体系。

**25.1** **数据质量管理工具概述**

数据质量管理平台提出的数据质量策略，从4个方面(建立数据质量评估体系、落实数据 质量优化流程、监控方案部署、建立具有持续改进机制的系统管理)进行多方位优化和改进， 最终形成一套高度灵活的数据解决方案，可以根据企业日益变化的数据条件做出相应的调整， 为企业的业务决策提供高质量的数据支持。数据质量管理平台架构如图25-1-1所示。

**262** 数据治理 — —工业企业数字化转型之道



|  |  |
| --- | --- |
| **数据剖析**  数据内容及背景分析 数据成分业务规则合规分析 数据结构及路径分析 数据间关系及相关资源匹配 数据有效性及准确性规则建立 其他数据剖析 | 识别数据转换机制  校验数据间依赖性 |

**购平台性**

**数据质量诊断**

**统一管理流程**

**完整性及时性** **合法性一致性** **准确性安全性唯** **一** **性租关性**

**数据业务规则**

数据属性业务规则ABR (如字段) 企业特定业务规则(如行业规范) 数据实体业务规则EBR (如库、表)

**多用户多角色**

|  |  |
| --- | --- |
| **数据质量记分卡&质量报告仪表盘**  整体/实体/属性业务规则数据质量分数 数据通过业务规则情况统计分析 | 实体业务规则统计 其他数据质量报告 |

站式可视化数据质量评估监管

将的政重 质量监管

实时处理/ 批处理

数据质量管理平台

**可视化分析报告**

**统一评估标准**

企业级业务

**工具自动化**

规库

**图25** **-** **1** **-** **1** **数** **据** **质** **量** **管** **理** **平** **台** **架** **构** **图**

数据质量管理平台的主要功能介绍如下。

(1)数据业务规则及规则库设计：包括字段、表、业务3个级别的业务规则处理；独立个 性化业务规则库设计；元数据管理，字段级修改历史追踪；协同业务规则库的建设及管理；行 业现成业务规则分享。

(2)数据剖析及异常数据处理：包括自动快速分层扫描检测；内置数据管理快速处理机制； 字段、记录、表、表间关联、键值分析；异常数据分析；灵活的业务规则匹配及数据诊断方法。

(3)集成工具及自动化处理：与业界主要的软件、数据库、业务系统接口。

**25.2** **数据质量稽核规则设置**

数据质量管理平台内置了业务规则中心，它是一个基于网页浏览器的应用，用户可以通过 流程化的设计来搜索、添加、复制、编辑并删除业务规则，帮助管理并组织业务规则以更有效 地监控数据质量状况。通过平台提供的针对具体国家的成熟业务规则，可验证全局、多域数据 集，并识别记录其中的复杂关系，具体包括以下优势。

(1)可以针对多种数据源运行业务规则，以同一个数据标准比对数据并追踪一段时间内的 数据合规趋势。

(2)可以基于业务需求自定义创建简单或复杂的规则，并能在控制中心实施并管理相关规

第25章 数据质量管理工具 **263**

则。

(3)可以通过向下挖掘数据来回顾数据合规分析结果，若显示有不合规的数据，则意味着

这些数据并不符合业务规则的表达标准及阈值水平。

(4)可以在多个项目或数据库中重复实施同一个业务规则，以确保数据集的一致性。

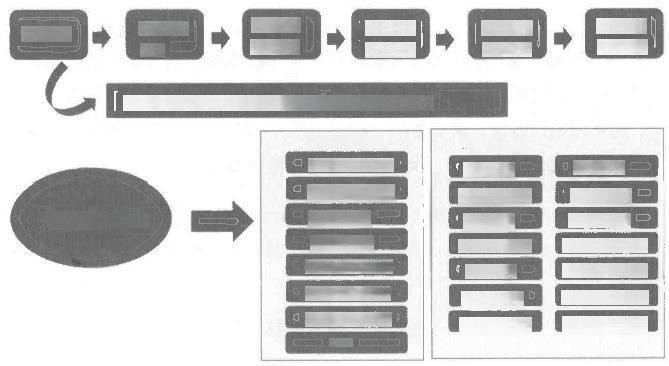
(5)可以配置唯一的优先级，以定义规则的重要性及影响。

(6)可以关联用户自定义的分类及细分进行规则调整，以在一个项目生命周期中配合组织

并管理业务规则。

(7)可以重复使用既有的基础业务规则，不需要针对每个数据源重复创建规则，便可直接 用于评估和审查数据。

任何质量问题的改善都是建立在评估的基础上，知道问题在哪里才能实施改进。通过数据 质量管理工具，可以对企业各个系统中的数据进行稽核，发现问题并通过发布数据质量报告促 进数据质量的提高，其中包括获取数据、定义质量规则、创建质量稽核实例、质量稽核任务调 度、返回质量稽核结果、获取数据质量报告7个步骤，并通过15个质量评估规则以进行全面的 稽核流程，如图25-2-1所示。



定义质量

规则

**各业务系统的数据汇集到ODS, 成为待查数据源**

数据质量评估

数据重复向题率

指标数概准确率

非空问慰率

接口及时率

共享实体一致率

指标数值一致率

接口格式出错率

……

空表检查

垃级表检查

记录表检查

数据重复性检查

数据时效性检查

处理正确性检查

内容一致性检查

技术检查规则

主媒检查

字段类型检查

外键检查

字段值域检查

非空检查

空值率检查

数据规范检查

创建质量 稽核实例

质量稽核 任务调度

获取数据 质量报告

返回质量 榨核结果

**15个质量评估规则**

获取数据

**图25-2-1** **数据质量稽核管理流程**

**25.3** **数据质量任务管理**

数据质量任务管理平台能将可疑数据记录发送给企业内部的责任方，以迅速解决可能影响

**264** 数据治理——工业企业数字化转型之道

企业绩效的数据质量问题。平台的协同门户可以帮助庞大而分散的团队管理问题解决方案、异 常处理机制，并变更数据质量工作管理方式；自动化工作流程帮助业务团队和 IT 团队基于现 有数据共同设计质量流程，并针对需要人工干预的数据质量问题确定解决方案；任务管理使各 企业能够通过建立透明、可审计的工作流程来加强其条例的合规性。

**25.4** **数据质量报告**

数据质量报告面向业务人员和 IT 人员提供协作型环境，对大批量数据进行剖析、诊断， 以评估其质量水平，从而输出优化数据。通过此报告，用户可理解数据内部存在的数据域、格 式、模式和关系，包括数据中的整体情况及完整度、准确度、遵从度、合规性、 一致性、唯一 性的详细情况。在此报告中，对数据内容进行详细分析，输出数据表中的空值、数据值中的空 值、数据值与设计不符的内容、存在异常范围的数据、数据格式不正确的内容；对数据关系进 行详细分析，输出表内字段关系分析结果、表间字段链接分析结果；对数据遵从度进行详细分 析，根据业务规则输出量化指标，并在报告的最后输出总体优化内容。

**本章精要**

本章中介绍的数据质量管理工具，可面向所有类型的数据，实现全局性的数据质量管理及 处理海量数据，并可轻松集成于任何平台和应用程序，支持多种复杂的处理，全面解析、整理、 标准、匹配、强化和监控数据，使业务部门和IT 部门能够发现、开发、部署、管理业务规则和 数据质量流程，从而优化企业数据，发挥最大的业务影响力。

第 2 6 章

数据交换与服务工具

**26.1** **数据交换与服务工具概述**

数据交换指从一个或多个信息系统(源系统)读取(抽取)数据，并基于一定的业务规则 (解析、转换),将数据写入(加载)一个或多个目标系统、企业级数据中心(EDW) 或数据中 台。提供数据交换服务的工具组合被称为数据交换与服务工具。数据交换与服务工具可以将若 干信息系统(企业内部的及外部的)中的数据/信息进行整合，完成数据的抽取、集中、加载(ETL), 使各个异构应用系统之间进行数据/信息的传输、共享、互联互通。

数据交换与服务工具具有集成协议转换、加密、压缩、交换过程监控等多种功能，保证各 系统之间的数据有效交换。在交换过程中涉及的功能调整均通过调整交换平台的应用得以实现， 减少了由于功能调整带来的对源系统和目标系统的影响。数据交换与服务工具由一系列负责结 构化数据交换的中间件、服务、Web Service接口，以及大数据(结构化、半结构化、非结构化 数据)交换引擎组成，具备数据采集、转换、脱敏、加密、压缩、安全、日志、系统权限管理 功能。

从应用角度看，数据交换与服务工具需具备以下功能。

(1)支持不同地域的服务器进行服务器端连接，能够进行连接策略和加密的配置；支持跨 越广域网的服务器间进行透明的请求转发和内容依赖路由机制；提供服务器间的连接保持机制， 保证连接不会因为短暂的空闲被断开；提供网络传输压缩功能，以减少应用在网络上的传输量，

压缩功能必须是可配置的，对应用透明。

(2)支持laaS 架构，要提供多种高可用性，能够在服务节点间进行请求转移；在单机模式

**266** 数据治理——工业企业数字化转型之道

和多机模式下，提供多种不同的负载均衡算法，包括轮询、权值等；支持有计划地停机，提供

工具将业务组件迁移到其他节点，对节点维修后，可以将原有组件迁回并正常运行。

(3)具备事件触发、定时任务、数据触发(当某些敏感数据发生变化时主动触发数据交换) 等发起数据交换请求的技术性能。数据触发服务，即在数据交换过程中，可根据特定情况进行 特殊处理的服务。如在业务高峰期间，外单位向数据交换平台发送大批数据处理请求，为了保 证内部核心业务系统的正常运行，数据交换平台在接到请求后并不及时处理，而是在业务空闲 期再响应外单位的请求。

(4)具备完善的日志管理机制；对交换任务全过程进行监控和日志管理；具备直观、详细

的运行与状态日志，同时还具有日志分析功能，帮助用户调整、优化平台的运行性能。

(5)支持实时数据交换，对于在线交易系统，数据交换平台能在1秒内完成交易(从接受 请求到处理完成);对于实时或批量的数据交换业务请求予以响应，支持大量并发处理机制，支 持实时与批量数据的均衡调度。对于工业数据，能在1秒内完成数据采集和跨网段的上传，以 保证工艺过程的实时可控。

**26.2** **数据采集**

数据采集就是从源系统中提取数据。工业数据的来源有3个途径，分别是物联网、Web 系

统和传统信息系统。

物联网数据占据了工业数据中90%以上的数据量，其中大部分是非结构化数据和半结构化 数据，采集的方式通常有两种， 一种是报文，另一种是文件。在采集物联网数据时往往需要制 定一个采集的策略，其中重点有两方面， 一个是采集的频率(时间),另一个是采集的维度(参 数)。

Web 系统是另一个重要的数据采集途径。随着Web 2.0的发展，整个 Web 系统涵盖了大量 的价值化数据，而且这些数据与物联网数据不同，Web 系统中有结构化、半结构化和非结构化 数据，数据的价值密度比较高。

传统信息系统中的数据占比较小，但由于传统信息系统中的数据结构清晰，同时具有较高 的可靠性，所以价值密度是最高的。传统信息系统中的数据采集与业务流程关联紧密，需要通 过编写专门的系统接口程序来读取各信息系统中的库表，按一定的业务规则生成中间表，传送 给目标系统。

第26章 数据交换与服务工具 **267**

数据采集的主要技术路线包括以下5种。

**1.ETL**

将数据从源系统通过 ETL(包括数据的提取、转换和加载),按一定的周期(如每天、每小 时)传输到目标系统。在转换过程中，需要针对具体的业务场景对数据进行治理，例如进行数 据质量监测与过滤、格式转换与数据规范化、数据替换，保证数据完整性等。

**2.** **实时数据采集**

实时数据采集工具包括 Flume 和 Kafka, 主要用在需要流处理的业务场景，比如，用于记 录对数据源执行的各种操作活动，网络监控中的流量管理、金融应用的股票记账和 Web 服务 器记录的用户访问行为。在流处理场景，数据采集系统会成为 Kafka 的消费者， Kafka 就像一 个水坝一般将上游源源不断的数据拦截住，然后根据业务场景做对应的处理(例如去重、去噪、 中间计算等),之后写入对应的数据存储中。这个过程类似传统的 ETL, 但它是流式的处理方式， 而非定时的批处理，这些工具均采用分布式架构，能满足每秒采集和传输数百兆的日志数据的 需求。

**3.** **互联网数据采集**

Scribe 是 Facebook 开发的数据(日志)收集系统，又被称为网页蜘蛛、网络机器人，用于 按照一定规则，自动抓取互联网信息的程序或者脚本，支持图片、音频、视频等文件或附件的 采集。除了网络中包含的内容，还可以使用DPI 或 DFI等带宽管理技术采集网络流量。

(注意，要在合法的情况下进行。)

**4.** **数据采集接口**

对于企业内部各信息系统中的数据，可以通过与软件技术服务商合作，使用特定系统接口 等相关方式采集数据。

(1)数据库视图：可以从指定数据库的视图中提取数据。

(2)数据表：可以从指定数据库的表中提取数据。

(3)自定义 SQL: 可以用自定义 SQL 从指定数据库中提取数据。

(4)Web Service: 可以从Web Service提供的方法中提取数据。

(5)文件：可以从 XML 、Excel 、Access 、DBF 文件中提取数据；可以把任意类型的文件

**268** 数据治理 工 业 企 业 数字化转型之道

从一个地方交换到另一个地方。

**5.** **工业数据采集**

狭义的工业数据指企业在生产活动中产生的物联网数据。工业数据可通过各种工业传感器 进行采集，通过SCADA(数据采集与监视控制系统)、PLC (工业控制系统)、SFC (车间控制 系统)进行实时处理，并存储在部署于工业互联网环境下的时序数据库中。工业数据的数据量 大，价值密度低，数据解析需要专门的业务知识。在采集工业数据时需要解决以下问题。

(1)明确数据采集的目的，如设备运行监控、工艺过程控制、在制品质量控制、生产进度

控制、生产环境监控等。

(2)了解时序数据库中的数据格式，设计数据采集的策略，定义数据采集标准和数据分析

维度。

(3)部署数据采集前置服务器，采集分布于工业互联网网段内的数据，并传输到企业内网

中。

**26.3** **数据交换**

数据交换系统通常包括前置交换、交换传输、交换管理3部分功能。在数据交换系统的管 理权限约束下，通过数据传输子系统和前置交换子系统，把需要交换的数据定向传输到接收系 统中。

**26.3.1** **前置交换子系统**

前置交换子系统是在工业互联网环境的前端部署的数据采集与数据交换代理服务的节点， 通过内置的Kafka中间件实现数据采集节点与企业内网之间的数据传输。其主要功能如下。

(1)通过内置的服务适配器，实现服务的注册、发现、适配、管理功能。

(2)通过内置丰富的适配器，自定义各种资源交换方式及抽取方式，包括数据库抽取、文

件读取、Web Service 接入、JMS 传输等。

(3)通过自带适配器的表输入、表输出，以及 SQL 执行器，可以实现数据的双向同步。

(4)通过 ESB 管理监控平台，能够对接入系统及流程进行查看和管理。

(5)具备交换数据缓存功能，能够制定缓存规则。

第26章 数据交换与服务工具 **269**

**26.3.2** **交换传输子系统**

交换传输子系统被部署于企业数据中心的中心节点中，与前置数据采集节点、数据交换代

理服务节点之间，通过 Kafka、MQ 消息中间件进行传输。其主要功能如下。

(1)以Kafka、MQ为底层传输中间件，保障安全可靠的数据传输。提供各种事务类型，可 对消息进行差错处理，实现消息同步及异步传输。

(2)通过ESB 内置的服务组件，提供 Web Service 、FTP 、JDBC 、MQ 、Kafka 等多种传输 方式，实现数据库数据、电子表格数据等格式化数据的传输，以及文本、图片、音频、视频等 非格式化数据的传输。

(3)可对数据进行分段、分组传输，具有压缩传输数据的功能。

(4)能够实现交换信息的打包、转换、传递、路由、解包等功能，进行不重、不漏、不错、 不丢的高效传输。

**26.3.3** **交换管理子系统**

交换管理子系统用于实现对整个数据交换过程的权限配置管理、日志管理、交换统计分析、 交换服务管理等。其主要功能如下。

(1)通过设计器中的域管理工具，实现对交换服务器、交换节点的管理。

(2)通过管理菜单，对路由信息进行创建、启动、停用等操作。

(3)提供资源订阅/发布管理，实现为各业务部门提供请求/应答、订阅/发布、消息广播等 多种数据交换方式的配置管理。

(4)提供多种接入方式，包括 Web Service、FTP 、JDBC、MQ 等，满足不同业务部门不同 的业务需求。

(5)提供交换信息统计分析功能，能够对交换当前状态及历史记录进行查询。

**26.4** **数据加工服务**

数据加工服务是将经过数据采集、汇聚的数据，进行清洗、转换、加载，以保障采集的数 据能被正确、完整、规范地加载到目的地。数据加工服务还可以实现数据整合过程中的异常处 理，如处理传输异常、数据加载异常、数据结构与质量异常等。

**270** 数据治理 工业企业数字化转型之道

数据加工服务是企业数据中台的核心功能，数据清洗、整合、应用就是由多个数据加工服 务组成的，其中包括一系列预定义的基础数据处理工作，这些基础数据处理包括以下几项内容。

**1.** **数据抽取**

数据抽取就是从源业务系统或数据中心中获取业务数据的过程。其中数据来源为业务系统 或数据中心，抽取方式为根据具体业务进行全量抽取或增量抽取，根据具体业务制定抽取的时 间、频率，这些参数都是可配置的。

数据抽取方式分为两种： 一种是全量抽取， 一般在数据初始化时使用，将数据源中的数据

原封不动地从数据库中抽取出来；另一种是基于时间戳的增量抽取，主要实现过程如下。

(1)建立数据库连接。

(2)定义一张数据字典表，定义需要进行处理的任务，其中主要包括业务数据库和目标数 据库的表名、字段列表及条件等。

(3)对比源表和目标表的最晚时间，设置抽取数据的时间段为“目标表的最晚时间”—“源 表的最晚时间”。

(4)根据设置的抽取频率循环抽取。

**2.** **数据清洗**

数据清洗指对抽取的数据进行清洗处理，包括数据过滤、数据去重、类型转换、编码映射、

文件拆分与合并、维度转换等功能。

数据清洗的主要任务是进行不一致的数据转换、数据粒度转换和转换规则计算。其中不一 致的数据转换是数据整合，其侧重于将来源于不同业务系统的相同类型的数据进行统一处理； 数据粒度转换是对数据进行统一归整；转换规则计算是按照设计的计算规则对数据进行重新计 算。数据清洗支持批量清洗和实时清洗，针对批量离线数据进行分布式并行清洗转换，针对实 时数据进行不落地清洗转换。

数据清洗主要是针对源系统库中出现的二义性、重复、不完整、违反业务或逻辑规则等问 题的数据进行统一处理， 一般包括空值处理、日期格式转换、数据类型转换等。在清洗之前需 要进行数据质量分析，以找出存在问题的数据。

**3.** **数据加载**

数据加载主要是指将经过清洗转换后的数据，准确、及时地存储到不同的基础数据库中。

第26章 数据交换与服务工具 **271**

依据数据的加载方式，数据加载可分为文件加载、流加载、压缩加载、不落地加载等。依据数 据加载技术的特点，数据加载可分为全量数据加载、流式(实时)数据加载、文件落地双加载、 内存不落地加载。

不同的数据库、不同的加载方式，在数据加载过程中的工作原理基本相同。数据加载包含

以下功能。

(1)在默认情况下提供基础通用的加载控件，支持将数据源加载到不同的数据库中，加载

的数据支持接口、文件加载策略或流式策略。

(2)加载方式支持全量或实时方式：全量加载方式采用落地加载策略，并且需要结合使用 不同平台的加载工具；实时加载与实时采集必须配套使用，二者之间共享内存，实现同步数据 交换，通过引入插件机制来屏蔽不同数据源的差异性。

(3)支持在加载时提交参数配置，允许设定数据文件相关输入路径与加载文件匹配规则等 信息，由数据装载在发现文件、文件获取、加载数据、数据校验等操作流程后进行数据入库操 作。

(4)在加载实现过程中支持提供 SQL、HQL、SHELL 等不同类别的行为定义脚本，数据

加载执行组件将根据定义行为脚本类型，调用相应的脚本执行来加载数据。

(5)在数据加载结束或失败时，都需要记录操作日志，为后续数据稽核与问题排查提供详

细信息。

(6)加载触发模式支持自动加载与手工执行两大类型。对于数据自动加载，当数据加载出 错时，应提供操作界面以人工干预的方式来重新启动数据的接收和加载。

**26.5** **数据共享服务**

数据共享服务应包括数据汇总、分发、查询等基本功能。在大数据中心，数据经过加工后， 按用途分区存储(通常会被划分为共享数据区和主题数据区)。共享数据区可以供数据需求方访 问。

**1.** **数据汇总**

将源系统数据通过数据采集、数据交换和数据加工，汇总到数据中心，并根据数据共享需 求，建立一个面向数据共享服务的主题库。

**272** 数据治理——工业企业数字化转型之道

**2.** **数据分发**

数据分发是以数据中心的角度，主动向各数据使用方提供数据的过程。通过数据共享服务， 可以依照数据使用权限的规则，从数据中心将数据分发到各个数据使用部门，实现数据共享、 信息联动。

**3.** **数据查询**

数据查询是通过统一标准的数据接口，以 XML 为标准数据格式，依据数据存取、访问权 限，通过标准的Web 服务，对各数据需求部门按需提供数据访问及查询服务。

**4.** **维度转换**

维度转换是按数据使用方需求(或视图),对共享数据区的数据进行维度转换。

**5.** **服务管理**

服务管理用于定义和管理数据共享服务的流程及策略，对数据服务进行监控管理、用户权 限管理、运行日志查看、性能统计。通过数据服务日志可以记录、跟踪数据交换的细节，对数 据交换节点进行管理，提供安全策略指南、服务器安全管理配置。

**26.6** **工业大数据技术平台**

工业大数据有3个典型的应用方向：第一个方向是设备优化，即提高单台设备的可靠性、 识别设备故障、优化设备运行等；第二个方向是针对产品线、车间、工厂，提高其运作效率， 包括能耗优化、供应链管理、质量管理等；第三个方向是产业跨界，实现产业互联。

**26.6.1** **工业大数据的采集**

工业大数据具备“多模态、高通量、强关联”的特性。多模态是指在工业领域中约有上百 种不同类型的数据，而且数据模态多样，结构关系复杂。高通量是指数据持续不断地产生，采 集频率高，通量大。强关联是指工业场景下的数据有非常强的工业知识支撑，不同学科之间的 数据是在机理层面的关联，而不是在数据字段上的关联。

对于工业大数据采集，不仅要涵盖已有信息系统中的数据，还将逐步包括半结构化的用户 行为数据、网状的社交关系数据、文本或音频类型的用户意见和反馈数据、设备和传感器采集 的周期性数据、通过网络爬虫合法获取的互联网数据，以及未来越来越多有潜在意义的各类数

第26章 数据交换与服务工具 **273**

据。主要包括以下几种。

**1.** **海量的** **Key-Value 数据**

在传感器技术飞速发展的今天，包括光电、热敏、气敏、力敏、磁敏、声敏、湿敏等不同 类别的工业传感器得到了大量的应用，而且很多时候关于设备的数据大概要到毫秒的精度才能 分析海量的工业数据，这部分数据的特点是每条数据内容很少，但是频率极高。

**2.** **文档数据**

文档数据包括工程图纸、仿真数据、设计的 CAD 图纸等，还有大量的传统工程文档。其

中涉及的数据有以下几种类型。

(1)信息化数据。工业信息系统产生的数据， 一般是以数据库形式存储的，因此这部分数 据是最好采集的。

(2)接口数据。由已经建成的工业自动化或信息系统提供的接口类型的数据，包括 TXT、 JSON、XML 等格式数据。

(3)视频数据。工业企业中会有大量的视频监控设备，这些设备会产生大量的视频数据。

(4)图像数据。例如工业企业中各类图像设备拍摄的图片(例如巡检人员用手持设备拍摄 的设备、环境信息图片)。

(5)音频数据。例如语音及声音信息(例如操作人员的通话、设备运转的音量等)。

(6)其他数据。例如遥感/遥测信息、三维高程信息等。

工业数据采集的技术难点主要包括以下几个方面。

(1)数据量巨大。任何系统，在不同的数据量面前，需要的技术难度是完全不同的。大量 的工业数据是“脏”数据，将其直接存储后无法用于分析，因此在存储数据之前，必须要进行 处理。对海量的“脏”数据进行处理，从技术上来说又提高了难度。

(2)工业数据的协议不标准。互联网数据采集一般都是遵从常见的 HTTP 等协议，但在工 业领域，会出现 ModBus 、OPC 、CAN 、ControlNet 、DeviceNet 、Profibus 、Zigbee 等各类型的 工业协议，而且各个自动化设备的生产及集成商还会自己开发各种私有的工业协议，导致工业 协议的互联互通难以实现，从而无法有效地进行解析和数据采集。

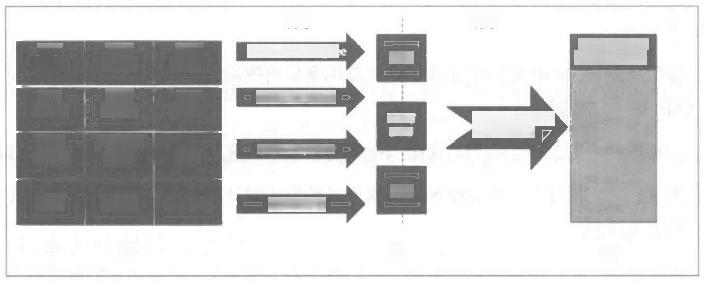
(3)对已有系统的数据采集难度大。数据采集往往不是针对传感器或者 PLC, 而是采集自 动化系统上的位机的数据。这些自动化系统的厂商水平参差不齐，大部分系统是没有数据接口

**274** 数据治理——工业企业数字化转型之道

的，文档也大量缺失，大量的现场系统没有点表等基础设置数据，使得这部分数据采集的难度 极大。

**26.6.2** **工业大数据的交换**

工业大数据平台通常是由“云一管一端”三部分组成的，如图26-6-1所示。与x86 服务器 主板中的南桥和北桥相似，工业物联网平台也有类似的概念。在“管”的左边，与外设传感器 等相关的部分称之为“南向”,在“管”的右边，与数据处理相关的部分称之为“北向”。南向 由各种传感器及 SCA DA/PLC/HMI 组成，负责数据的采集。这些数据经过 RTU/DTU、 网关或 路由被传输到大数据中心，由北向的工业大数据处理平台进行处理，从而完成监控、告警、预 测性维护及分析等功能。



**南向**

NBIOT/Lcxa/Zghes

R⁵232/R5485

监控

4-20mA/0-5v

PLC

WIFI/网线

鳙 管 云

压 力

传感器

霍尔 传感器

SCADA 系统

变频器

温度 传感础

液位 传感器

红外 传感器

害警

预测性维护

**数据分析**

位移 传感器

可燃 传感器

工业互联网大 数据处理平

2G/3G/4G/5G Wi-Fi/网线

仪器仪表

RTU/ DTU

**路由**

北向

网关

HMI

**图26-6-1** **工业大数据平台框架**

工业大数据平台南向传感器采集的数据具有时间属性，并且自带标签与数值。每条数据代 表一个监测指标并且反映数值的变化，同时这些数据又随时间延续而无限增长，因此被称之为 流数据或时序数据。

南向采集数据的设备虽然多种多样，但本质上它们的数据格式都是一样的，都由“timestamp、 tags、metrics”这三部分组成。数据格式看上去很简单，但是对于数据处理系统来说却有着如下 的复杂性。

(1)数据自带时间戳，这意味着数据处理具有实时性。

(2)都是小数据，这意味着数据存储系统需要对此进行专门的设计。

(3)数据随时间延续而无限增长，这意味着数据具有无限性。

(4)数据到达的速度有快有慢、负载有高有低，这意味着需要灵活、细粒度的资源。

第26章 数据交换与服务工具 **275**

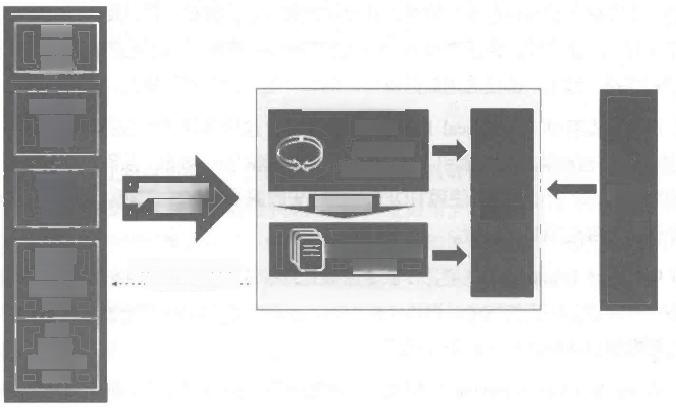
(5)数据可能是有序或无序的，会有持久化需求，以及数据传输的环境可能是复杂的，在

这些约束条件下要保证数据处理结果的正确性。

**26.6.3** **工业大数据的处理**

工业大数据平台北向负责大数据的处理，不同的应用场景意味着不同的数据处理要求与技 术复杂度。要把这些不同的，甚至矛盾的数据处理要求都很好地综合在一个大数据处理系统中， 对现有的大数据处理技术来说是一个非常大的挑战。例如，无人车的数据处理要求具有毫秒级 甚至纳秒级的数据处理实时性，而对于有些工业设备的数据处理，则只需要分析历史数据即可。

在物联网大潮下，智能手机、PC、 智能硬件设备的计算能力越来越强，而企业对数据处理 实时响应能力的需求也越来越强，过去传统的中心化、非实时化的数据处理思路已经不适应现 在的大数据分析需求。新一代的大数据架构IoTA 可以解决上述问题，其整体思路是设定标准数 据模型，通过边缘计算技术把所有的计算过程分散在数据产生、计算和查询的过程中，以统一 的数据模型贯穿始终，从而提高整体的预算效率，同时满足即时计算的需要，以及可以使用各 种即席查询来查询底层数据，如图26-6-2所示。



Web

Edge

SDK

Ad

SDK

Common

Data Model

小程序

Edge

SDK

摄像头

Edge

Server&

SDK

loT

Edge

Server&

SDK

Dumper

**Common Data Model**

**Historical Data Storage**

**Real Time Data**

**Cache**

**RealTime Model -Feedback** …

**Standard Format**

query Engine

QueTV User

图26-6-2 IoTA 架 构

IoTA 架构的整体技术结构分为以下几部分。

(1)通用数据模型( Common Data Model): 贯穿整个业务始终的数据模型。这个模型是整 个业务的核心，可以使 SDK、Cache、 历史数据、查询引擎保持一致。通用数据模型可以将数

**276** 数据治理 工业企业数字化转型之道

据定义为“主一谓一宾”或者“对象一事件”这样的抽象模型来满足各种各样的查询。以大家 熟悉的 APP用户模型为例，用“主一谓一宾”模型描述就是“×用户一事件1-A 页面(2018/4/1 20:00)”。当然，根据业务需求的不同，也可以使用“产品一事件”“地点一时间”模型等。模 型本身也可以根据协议(例如 QrotoBuf) 来实现在 SDK 端定义，而在中央集中存储的方式。 此处的核心是，从 SDK 到存储再处理是一个统一的标准数据模型。

(2)边缘服务器和边缘 SDK(Edge Server&Edge SDK): 数据的采集端，它不再是过去的 简单的 SDK, 它在复杂的计算情况下，可以赋予 SDK 更复杂的计算，在设备端就将数据转化 为统一的数据模型来进行传送。例如，对于智能 Wi-Fi 采集的数据，从 AC 端就变为“×用户 的 MAC 地址一出现一A 楼层(2018/4/1118:00)”这种“主一谓一宾”模型。对于摄像头采集 的数据，会通过Edge AI Server,被转换成为“×的面部特征一进入一A 火车站(2018/4/1120:00)” 模型，也可以是上面提到的简单的 APP 或者页面级别的“×用户一事件1—A 页面(2018/4/11 20:00)”模型。对于APP 和H5 页面采集的数据，没有计算工作量，只要求是埋点格式即可。

(3)实时数据 (Real Time Data) 缓存区：这部分是为了达到实时计算的目的，在接收海量 数据时不可能实时进入历史数据库，那样会出现建立索引延迟、历史数据碎片文件等问题。因 此，要有一个实时数据缓存区来存储最近几分钟或者几秒的数据，可以使用Kudu 或者 Hbase 等组件来实现。这部分数据会通过 Dumper被合并到历史数据中。此处的数据模型和 SDK 端的 数据模型是保持一致的，都是通用数据模型，例如“主一谓一宾”模型。

(4)历史数据存储 (Historical Data Storage) 区：这里保存了大量的历史数据，为了实现 Ad-Hoc 查询，将自动建立相关索引，提高整体历史数据的查询效率，从而实现秒级反馈数百亿 条数据的复杂查询。例如，可以使用HDFS 存储历史数据，此处的数据模型依然是 SDK 端数据 模型，是保持一致通用数据模型。

(5)Dumper:Dumper 的主要工作就是将最近几秒或者几分钟的实时数据，根据汇聚规则 建立索引，并存储到历史结构中，可以使用MapReduce、C、Scala 等语言来撰写，把相关的数 据从实时数据缓存区写入历史数据存储区。

(6)查询引擎(Query Engine):提供统一的对外查询接口和协议(例如SQLJDBC), 把实 时数据和历史数据合并到一起查询，从而实现对数据进行实时的 Ad-Hoc 查询。例如常见的计 算引擎可以使用 Presto 、Impala 、ClickHouse 等数据库。

(7)实时模型反馈(RealTime Model Feedback):通过边缘计算技术，在边缘端有更多的交 互可以做，可以通过在实时数据缓存区设定规则来对边缘 SDK 端进行控制。例如，数据上传

第26章 数据交换与服务工具 **277**

频次的降低、语音控制的迅速反馈，某些条件和规则的触发等。简单的事件处理，将通过本地 的IoT 端完成，例如，现在已经有很多摄像头本身带有犯罪嫌疑人的识别功能。

IoTA 大数据架构主要有如下特点。

(1)去 ETL 化：ETL 和相关开发一直是大数据处理的痛点，IoTA 架构通过通用数据模型 的设计，专注在某一个具体领域的数据计算，从而可以从SDK 端开始计算，中央端只用于采集、 建立索引和查询，提高整体数据分析的效率。

(2)Ad-Hoc 即时查询：鉴于整体的计算流程机制，在手机端，智能物联网事件发生之时就 可以被直接传送到云端进入实时数据缓存区，可以被前端的查询引擎查询。此时用户可以使用 各种各样的查询，直接查询到前几秒发生的事件，而不用等待ETL 或者流的数据研发和处理。

(3)边缘计算(Edge-Computing): 将过去统一到中央进行整体计算的数据分散到数据产生、 存储和查询端，数据产生即符合通用数据模型。同时，也给予实时模型反馈，让客户端在传送 数据的同时马上进行反馈，而不需要在所有事件都要到中央端处理之后再进行下发。

采用IoTA 架构搭建工业大数据平台的难点不是如何封装各种大数据工具和中间件，而是贯 穿业务始终的数据模型。企业必须遵从数据架构的总体定义，从数据采集端开始系统性地建立 贯穿业务始终的数据模型，并体系化开展数据治理的各项工作。

**本章精要**

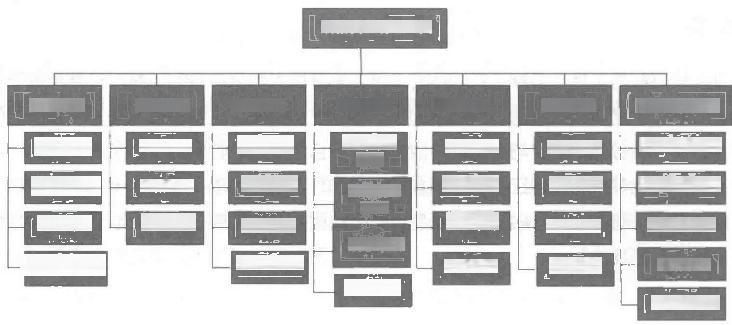
数据交换与服务工具是多种工具的组合，可以将其划入数据中台的范畴。它是服务于数据 治理平台和各类数据应用系统的一系列中间件，被广泛应用于源系统的数据采集、系统之间的 数据交换、多源头数据的汇聚与加工、将数据加载到目标系统、数据共享应用等方面。对数据 的用户而言，数据交换与服务工具是“黑盒”模式的；对企业的信息技术人员而言，它是一个 保障企业信息系统有序运营、融合数据孤岛、推进数据共享的重要工具，可以将其看作是企业 信息系统的“神经系统”。

第27章

数据安全管理工具

**27.1** **数据安全管理工具概述**

工业互联网数据安全管理需要依托具体的工具来实现。目前，针对数据安全管理，业界很 多厂商都开发了相关的工具，其中包括安全采集、安全传输、数据存储、安全处理、安全交换、 安全销毁等管理工具。这些工具有的是单独呈现，有的是相互组合在一起形成包括多种功能的 平台，其具体意义和主要功能如图27-1-1所示。



数据安全管理工具

统一身份认证

用户身份管理

统一身份策略

强密码策略

脱敬联动

应用准入

脱做日志

强认证管理

安全处理

数据静态 脱敬

数据动态 脱敏

安全采集

分类分级

识别与记录

安全存储

介质净化

安全扫描

安全交换

审计平台

流量监测

安全传输

身份鉴别

数据加密

安全销毁

履写工具

消碱工具

质量管理 密钢管理 日志审计

备份与恢复

一采集筑略管理

焚级工具

实体捣碎

日志市计

加密交换

**图27-1-1** **数据安全管理工具**

第27章 数据安全管理工具 **279**

**27.2** **数据采集安全管理工具**

数据采集安全管理工具用来确保数据源的真实性、有效性，并规范数据采集的渠道、数据 格式及相关流程和方式，从而保证数据采集的合规性、正当性和执行上的一致性。工业互联网 数据采集主要通过传感器与感知技术，采集物联网设备及生产经营过程中的业务数据、外部互 联网数据等，并且具有很高的实时性要求。

**27.2.1** **数据分类与分级工具**

在大数据时代，企业需要对数据进行分类与分级：首先是通过对数据的有效理解和分析， 对数据进行不同类别和保密级别的划分；然后根据数据的类别和保密级别制定不同的管理和使 用原则，尽可能对数据做到有差别和针对性地防护，实现在适当安全防护下的数据自由流动。

**1.** **数据分类与分级的依据**

(1)根据数据的重要性和对组织可能产生的影响，对数据进行分类与分级，并制定组织机 构层面的分类与分级原则和要求。

(2)根据数据泄露、丢失、损坏等对组织可能造成的影响，对数据进行分类与分级。

**2.** **数据分类与分级的技术工具参考**

(1)标签库：根据数据分类与分级规则，建立标签库；可以单独构建一个静态库，也可以 直接在打标工具/系统后台进行配置。

(2)结构化数据打标：用户在建表时直接对字段标签进行设置，基于数据库的权限模型对 底层数据表的列进行权限控制。

(3)非结构化数据打标：引入自然语言处理、数据挖掘和机器学习等技术，对内容进行识 别，并实现数据分类与分级。

(4)分类：将生成的分类器应用在有待分类的文档集合中，获取文档的分类结果。由于机 器学习技术在文本分类领域中有着良好的表现，其已经成为该领域的主流技术。

**27.2.2** **采集内容及策略**

在采集数据时要设置统一的数据采集策略(如采集周期、频率、采集内容等),保证数据采 集流程的一致性。并且在采集过程中，对被采集方授权同意采集的过程和信息进行日志记录。

**280** 数据治理——工业企业数字化转型之道

采集内容包括数据类型/类别、采集内容的来源、采集目的、采集用途、采集方式、采集 范围。

采集周期和频率可根据场景进行制定。

**27.2.3** **数据采集人员管理工具**

数据采集人员管理工具用于管理数据采集人员，主要包括识别采集人员、制定采集人员的

操作规范，以及记录采集过程的工具。

(1)采集的数据信息，包括但不限于数据、文本、文件、图片、音频和视频等。

(2)采集数据的传输方式，包括但不限于有线通信传输、无线通信传输和数字通信传输等。

(3)数据采集者在采集前，应进行身份识别，使用人脸识别、指纹识别等技术可以防止数

据采集点的仿冒与伪造。

(4)数据采集者应设置专人负责信息生产或提供者的数据审核和采集工作。

(5)数据采集者应明确数据来源、采集方式、采集范围等内容，并记录存档。

(6)数据采集者应制定标准的数据采集模板，以及数据采集方法、策略和规范。采集策略 参数配置应包括采集周期、有效性、检测时间、入口地址和采集深度等。

**27.2.4** **数据源鉴别及记录**

数据源鉴别即对产生的数据源进行身份鉴别和记录，防止数据的仿冒和伪造。数据源鉴别 是对收集或产生数据的来源进行身份识别的一种安全机制，防止采集到其他不被认可的或非法 数据源(如机器人信息注册等)产生的数据，即避免采集到错误的或失真的数据。

数据源记录是指需要对采集的数据进行数据来源的标示，以便在必要时对数据源进行追踪 和溯源。

利用元数据管理工具提供的数据血缘关系可以对采集的数据来源进行标示。

**27.3** **数据传输安全管理工具**

数据在通过不可信或者具有较低安全性的网络进行传输时，容易面临被窃取、伪造和篡改 等安全风险，因此，需要建立相关的安全防护措施，保障数据在传输过程中的安全性。数据传

第27章 数据安全管理工具 **281**

输安全管理工具就是为了建立相关的加密措施，来保障数据在传输过程中的机密性、完整性和

可信任性。

数据传输安全管理工具具体需具备以下基础功能。

(1)身份鉴别和认证。在建立传输加密通道前，要对两端的主体身份进行鉴别和认证，确 保数据传输双方是可信任的。

(2)数据加密。针对需要加密的场景确定加密的方案，通过加密产品或工具落实制度规范 所约定的加密算法要求和密钥管理要求，保护数据在传输过程中的机密性和完整性。同时，加 密算法的配置、变更、密钥的管理等操作过程应具有审核机制和监控手段。

(3)密钥管理系统。密钥管理系统用于实现对密钥生命周期的安全管理。

**27.3.1** **加密算法**

组织应在定义需要加密的场景后，选择合适的算法针对数据进行加密。

加密系统通常包含明文、密文、加密/解密算法及用于加密/解密的钥匙(密钥)。加密算法 是一些公式和法则，它规定了明文和密文之间的变换方法。密钥是控制加密/解密算法的关键信 息，它的产生、传输、存储等工作是十分重要的。

数据加密的基本过程包括对明文(即可读信息)进行翻译，译成密文或密码的代码形式。 该过程的逆过程为解密，即将该编码信息转换为其原来的形式。

**27.3.2** **对称加密**

对称加密(也叫私钥加密)指加密和解密使用相同密钥的加密算法，有时又叫作传统密码 算法，即加密密钥能够从解密密钥中推算出来，同时解密密钥也可以从加密密钥中推算出来。 而在大多数对称算法中，加密密钥和解密密钥是相同的，所以，也称这种加密算法为秘密密钥 算法或单密钥算法。它要求发送方和接收方在安全通信之前商定一个密钥。对称算法的安全性 依赖于密钥，泄漏密钥就意味着任何人都可以对其发送或接收的消息解密，所以密钥的保密 性对通信安全至关重要。常见的对称加密算法包括：SM1 和 SM4 算法、DES 算法和 AES 算法 等。

SM 算法是国密算法。国密算法即国家密码局认定的国产密码算法。SM 算法主要有SM1、 SM2 、SM3 、SM4,其中 SM1 算法为对称加密，其加密强度与 AES 算法相当，但该算法不公 开，在调用该算法时，需要通过特定的加密芯片的接口进行调用； SM4 为对称加密算法； SM3

**282** 数据治理——工业企业数字化转型之道

为信息摘要算法；SM4 的特点是密钥长度和分组长度均为128位。虽然密钥长度与DES 算法类 似，但SM 算法的加密强度能与AES 算法媲美。在国家推出信息安全战略的大背景下，国密算 法得到了越来越多的支持。

**27.3.3** **非对称加密**

非对称加密算法需要两个密钥，即公开密钥(public key,简称公钥)和私有密钥(private key, 简称私钥)。公开密钥与私有密钥是一对，如果用公开密钥对数据进行加密，那么只有用对应的 私有密钥才能解密；如果用私有密钥对数据进行加密，那么只有用对应的公开密钥才能解密。 因为加密和解密使用的是两个不同的密钥，所以这种算法被称为非对称加密算法。对称加密算 法在加密和解密时使用的是同一个密钥；非对称加密算法需要使用两个密钥进行加密和解密， 这两个密钥是公开密钥和私有密钥。

RSA 公钥加密算法是目前最有影响力的公钥加密算法之一，它能够抵抗到目前为止已知的 绝大多数密码攻击，已被ISO 推荐为公钥数据加密标准。它基于一个十分简单的数论事实：将 两个大素数相乘十分容易，想要对其乘积进行因式分解却极其困难，因此可以将乘积公开作为 加密密钥。

非对称加密与对称加密相比，其安全性更好。对称加密的通信双方使用相同的密钥，如果 一方的密钥遭泄露，那么整个通信就会被破解。非对称加密使用一对密钥， 一个用来加密， 一 个用来解密，而且公钥是公开的，密钥是用户自己保存的，不需要像对称加密那样在通信之前 先同步密钥。其缺点是加密和解密花费的时间长、速度慢，并且只适合对少量数据进行加密。

**27.4 数据存储安全管理工具**

数据存储安全管理工具是为了解决将数据存储在介质上时(比如物理实体介质(磁盘/硬盘)、 虚拟存储介质(容器/虚拟盘)等),因对介质的不当使用而引发的数据泄露问题。

数据存储安全管理工具需具备以下基础功能。

(1)介质净化。主要是针对物理实体介质(磁盘/硬盘)进行净化。

(2)安全扫描。定期对重要的数据存储系统的安全配置进行扫描，以保证符合安全基线

要求。

(3)日志审计。采集存储系统的操作日志，识别访问账号和鉴别权限，监测数据使用的规

第27章 数据安全管理工具 **283**

范性和合理性。

(4)数据备份和恢复。数据备份和恢复的技术工具要统一，并做到自动化执行。

(5)备份管理。对已备份的数据要有安全管理技术手段，包括但不限于对备份数据的访问 控制、压缩或加密管理、完整性和可用性管理。

**27.4.1** **数据存储介质管理**

存储介质是数据存储的载体和基础。存储介质并不是越贵越好、越先进越好，要根据不同 的应用环境，合理选择存储介质。常见的数据存储介质有移动硬盘、可记录光盘、MP3 、MP4 、 U 盘、闪存卡等。

**27.4.2** **数据存储安全**

在过去10年中，存储设备已演变为多个系统共享的一种资源。非常多的案例都表明，只保 护存储设备所在的系统的安全已不能满足需要了。目前存储设备会被连接到非常多的系统中， 因此，必须保护各个系统中有价值的数据，防止其他系统未经授权访问数据或破坏数据。相应 地，存储设备必须要防止未被授权的设置改动，对其所有的更改都要做审计跟踪。

在实践中，建立数据存储安全需要专业的知识，要留意细节，不断检查，确保存储解决方 案能持续满足业务不断改动的需要。必须减少诸如伪造回复地址这样的威胁。最重要的是，安 全的本质是三个方面达到平衡，即采取安全措施的成本、安全缺口带来的影响、入侵者要突破 安全措施所需要的资源的平衡。

**27.4.3** **数据备份和恢复**

数据备份是容灾的基础，是指为防止系统出现操作失误或系统故障导致数据丢失，而将全 部或部分数据集合从应用主机的硬盘或阵列复制到其他存储介质的过程。传统的数据备份主要 采用内置或外置的磁带机进行冷备份。但是这种方式只能防止操作失误等人为故障，而且其恢 复时间也很长。随着技术的不断发展，数据的海量增加，不少企业开始采用网络备份。网络备 份一般通过专业的数据存储管理软件结合相应的硬件和存储设备来实现。常见的备份方式如下。

**1.** **备份磁带/光盘**

(1)远程磁带库、光盘库备份。即将数据传送到远程备份中心制作完整的备份磁带或光盘。

(2)远程关键数据+磁带备份。采用磁带备份数据，生产机实时向备份机发送关键数据。

**284** 数据治理——工业企业数字化转型之道

**2.** **数据库**

在与主数据库所在生产机相分离的备份机上建立主数据库的备份。

**3.** **网络数据**

对生产系统中的数据库数据和所需跟踪的重要目标文件的更新进行监控与跟踪，并将更新 日志通过网络实时传送到备份系统，备份系统则根据日志对磁盘进行更新。

**4.** **远程镜像**

通过高速光纤通道线路和磁盘控制技术，将镜像磁盘延伸到远离生产机的地方，镜像磁盘 数据与主磁盘数据完全一致，更新方式为同步或异步。

造成数据丢失和毁坏的原因主要有如下几个方面。

(1)数据处理和访问软件平台故障。

(2)操作系统的设计漏洞或设计者出于不可告人的目的而人为预置的“黑洞”。

(3)系统的硬件故障。

(4)人为的操作失误。

(5)网络内非法访问者的恶意破坏。

(6)网络供电系统故障等。

为了保障生产、销售、开发的正常运行，企业用户应当采取先进、有效的措施，对数据进 行备份，防患于未然。

**27.4.4** **等级划分**

灾难恢复等级的确定是信息系统灾备建设的重要考虑因素。《信息系统灾难恢复规范》 (GB/T20 988—2007) 将灾难恢复能力划分为六级。

(1)等级一。

基本支持。要求数据备份系统能够保证每周至少进行一次数据备份，备份介质能够提供场 外存放。对于备用数据处理系统和备用网络系统，没有具体要求。

(2)等级二。

备用场地支持。在满足等级一条件的基础上，要求配备灾难恢复所需的部分数据处理设备，

第27章 数据安全管理工具 **285**

或灾难发生后能在预定时间内调配所需的数据处理设备到备用场地；要求配备部分通信线路和 相应的网络设备，或在灾难发生后能在预定时间内调配所需的通信线路和网络设备到备用场地。

(3)等级三。

电子传输和设备支持。要求每天至少进行一次完全数据备份，备份介质场外存放，同时每 天多次利用通信网络将关键数据定时、批量传送至备用场地。配备灾难恢复所需的部分数据处 理设备、通信线路和相应的网络设备。

(4)等级四。

电子传输及完整设备支持。在等级三的基础上，要求配置灾难恢复所需的所有数据处理设 备、通信线路和相应的网络设备，并且处于就绪或运行状态。

(5)等级五。

实时数据传输及完整设备支持。除要求每天至少进行一次完全数据备份、备份介质场外存 放外，还要求采用远程数据复制技术，利用通信网络将关键数据实时复制到备用场地。

(6)等级六。

数据零丢失和远程集群支持。要求实现远程实时备份，数据零丢失；备用数据处理系统具

备与生产数据处理系统一致的处理能力，应用软件是“集群的”,可实时切换。

由此可见，灾难恢复能力等级越高，对于信息系统的保护效果越好。需要对关键数据进行 同城和异地的实时备份，保证业务应用能够实现实时切换。同时要制订灾难恢复计划，并定期 进行测试，确保各个恢复程序的正确性和计划整体的有效性。

**27.5** **数据处理安全管理工具**

工业互联网处理的数据主要是来自工业软件和工业互联网经营管理中的数据，以及外部数 据等，此类数据信噪比低，要求数据具有真实性、完整性和可靠性，更加关注处理后的数据质 量。数据处理安全管理工具用于降低上述各类数据在分析与挖掘过程中的敏感数据被泄露的风 险。

数据处理安全管理工具需具备以下基础功能。

(1)具备数据脱敏工具。

组织要具备统一的数据脱敏工具，数据脱敏工具应具备静态脱敏和动态脱敏的功能。而且，

**286** 数据治理 工 业 企 业 数字化转型之道

脱敏工具应与组织的数据权限管理平台实现联动，可以根据使用者的职责权限或者业务处理活 动，动态化地调整脱敏的规则。职责权限一般用来决定可以访问哪些敏感数据，业务处理活动 则主要决定采用哪些脱敏方式。例如，用户展现的敏感数据可以通过遮蔽部分数据等方式实现， 用户开发测试的数据则通过同义替换的方式实现。

(2)留存脱敏日志。

脱敏工具对数据的脱敏操作过程都应该留存日志记录，以审核违规使用和恶意行为，防止 敏感数据泄露。

(3)静态数据脱敏(S DM)。

静态数据脱敏一般用于非生产环境，将敏感数据从生产环境抽取并脱敏后，用于非生产环 境中，如常用于培训、分析、测试、开发等非生产系统的数据库。

(4)动态数据脱敏( DDM)。

动态数据脱敏常用于生产环境，在访问敏感数据时即时进行脱敏， 一般用来解决在生产环 境中，需要根据不同情况对同一敏感数据读取时，进行不同级别脱敏的场景。对于需要共享的 生产数据或时效性很高的数据测试和培训场景，基于网络代理模式的动态脱敏技术，可以达到 实时模糊敏感数据的效果。动态脱敏可实现对业务系统数据库中的敏感数据进行透明、实时的 脱敏。

所以，数据脱敏能够更好地保证数据的隐私性。简单来说，数据脱敏在此只是对数据(如 姓名、电话号码等)进行简单、粗暴的脱敏处理，即改变数据，并且通常无法返回原始数据。 因此，数据脱敏在保护数据安全层面只适用于测试、开发环境。在不用提供真实数据的情况下， 可以将脱敏后的数据传输给第三方用于测试、开发等，以保障真实数据的安全。

**27.6** **数据交换安全管理工具**

数据交换安全管理工具用来解决因非法人员通过非法技术手段导出非授权数据、敏感数据 等带来的数据篡改和数据泄漏的重大事故，控制数据交换中存在的安全风险。

数据交换安全管理工具需具备以下基础功能。

(1)审核平台。

通过建立数据交换审核流程的在线平台，组织机构内部可对数据交换进行审核并详细记录，

第27章 数据安全管理工具 **287**

确保没有超出数据服务提供者的数据授权使用范围。

(2)日志审计。

针对数据交换日志，建立相应的管理和审计方案，有效记录在交换过程中的相关日志信息， 并通过定期开展审计工作发现其中存在的安全风险。

(3)流量监测。

通过工业互联网云平台的服务提供商，把流量数据引到数据汇集资源处，通过威胁情报、 预制规则、用户行为分析，发现工业互联网云平台各区域的实时安全状况，并进行告警，为工 业互联网云平台的安全运维提供关键预警信息。

(4)加密。

利用数据加密、安全通道等措施，保护数据交换过程中的个人信息、重要数据等敏感信息。

**27.6.1** **数据导入/导出的安全保障**

建立数据导入/导出安全保障包括以下内容。

(1)建立数据导入/导出的安全制度规范，对各业务中的导入/导出场景进行充分合理的安全 需求分析，能够依据不同的场景，并基于数据分类与分级要求定义数据导入/导出的安全策略， 例如访问控制策略、不一致处理策略、流程控制策略、审计策略、日志管理策略等。

(2)建立规范的数据导入/导出的权限审批和授权流程，流程中包括但不限于数据导入/导出 的业务方、数据在组织机构内部的管理方、相应的安全管理团队，以及根据组织机构数据导/ 导出的规范要求所需参与具体风险判定的相关方，如法律团队、对外公关团队、财务数据对外 管理团队等其他重要的与数据价值保护相关的团队。

(3)建立针对导出数据介质的安全技术标准，明确介质的命名规则、标示属性等重要信息， 定期验证导出数据的完整性和可用性。

**27.6.2** **数据交换安全**

在数据交换环节中，业务系统将数据共享给外部组织机构，或者以合作方式与第三方合作 伙伴交换数据。数据在共享后会释放更大价值，并支撑数据业务的深入开展。在数据共享过程 中面临着巨大的安全风险，数据本身存在敏感性，若共享保护措施不当，则将带来敏感数据和 重要数据的泄漏。因此，在本过程中，需要采取相应的安全保护措施，以保障共享后数据的完

**288** 数据治理——工业企业数字化转型之道

整性、安全性和可用性，防止数据被丢失、篡改、假冒和泄露。

组织机构要制定数据共享原则及数据保护措施，从国家安全、组织机构的核心价值保护、 个人信息保护等方面对数据共享中的风险控制提出要求，明确数据共享涉及的机构或部门的相 关职责和权限，明确共享数据相关的使用者的数据保护责任，确保数据使用的相关方对共享数 据具有足够的保护能力，从而保障数据共享安全策略的有效性。

组织机构在原则要求的基础上，要根据组织机构对数据共享涉及的数据类型、数据内容、 数据格式，以及对数据共享的常见场景制定细化的规范要求，以满足数据共享业务场景需求， 提高数据共享效率，指导其对具体数据共享场景的风险把控。

组织机构要建立规范的数据共享的审核流程，在审核流程中包括但不限于数据共享的业务 方、在组织机构内部的共享数据管理方、数据共享的安全管理团队，以及根据组织机构数据共 享规范要求所需参与具体风险判定的相关方，如法律团队、对外公关团队等其他重要的与数据 价值保护相关的团队，确保共享的数据未超出授权范围。

组织机构要制定数据共享审计策略和审计日志管理规范，明确审计记录要求，为数据共享 安全事件的处置、应急响应和事后调查提供帮助。

针对数据交换过程中涉及第三方数据交平台的场景，组织机构要制定明确的安全评估的要 求和流程，以保证该数据交换平台已符合组织机构对数据交换过程中的数据安全要求。

**27.6.3** **数据销毁安全管理**

在弃置工业计算机或设备前，必须将其中所有的数据彻底删除，并使其无法复原，以免造 成信息泄露，尤其是国家涉密数据。

下面列出几种企业经常采用的数据销毁方式，这些方式不见得最完美，却是企业在衡量所 采用方式的时间、金钱及效益上的最佳平衡点。

(1)覆写法。

由于磁带是可以重复使用的，当前面的数据被后面一笔数据覆写时，就算可以透过软件进 行数据还原，随着被覆写次数的增多，非结构性数据在被复原时需要解读的时间也越久。企业 可以通过评估数据被复原的风险是否能够承担而选择合适的覆写方式。其中，低程度的覆写就 是将磁带或磁盘完全覆写；高程度的覆写则需符合美国的 DoD 5220.22-M 标准，结合数种清除 与覆写程序，让硬盘的每一个空间都被重复清除与覆写。

第27章 数据安全管理工具 **289**

(2)消磁法。

磁盘或是磁带等存储介质使用的都是磁性技术，如果能破坏其磁性结构，则既有的数据便 不复存在。 一般企业可以购买小型消磁机做单卷消磁(但消磁机的磁波高),如果是大量消磁， 则委托专门的公司负责会比较迅速、安全。

(3)捣碎法/剪碎法。

破坏存储介质实体，让数据无法被系统读出，也是保护数据机密性与安全性的方法之一。

虽然有人质疑用这种数据销毁方式并不安全，但采用此种数据销毁方式的企业，通常会衡 量其能承担的最大风险，而且还可以搭配焚毁法进行数据销毁。

(4)焚毁法。

需要淘汰的存储介质还可以借由焚毁让数据真正化为灰烬。在整个过程中，信息主管需在 现场监督旧数据焚毁状况与进度，落实数据安全的最后一步。

若有人从弃置的存储介质中取得数据，对企业来说则会存在风险。引进合乎企业最大效益 的数据销毁方式，是多数企业必须要慎重考虑的。

**27.7 统一的身份认证系统**

统一的身份认证系统是面向工业互联网的多个应用系统，提供集中、统一的安全认证服务， 形成统一的、高安全的身份验证中心。其支持用户名/口令、PKI/CA 数字证书等多种不同强度 的用户身份认证方式，提供集中的数字身份管理、认证、授权、审计的模式和平台，从而实现 对企业信息资产统一的身份认证、授权和身份数据的集中管理与审计。

统一的身份认证系统需具备以下基础功能。

(1)建立基于实名制的统一、权威的用户身份数据源，实现用户全生命周期管理，消除账 号分散管理，避免因没有统一身份管理策略和强密码策略所面临的风险。

(2)建立集中、高强度的安全认证中心，以统一的安全认证策略和技术保障用户认证安全。 (3)建立应用接入规范和标准，支持当前主流的认证协议和认证技术，支持异构应用集成。

(4)建立或者接入现有审计平台，提供事后追溯甚至事中监测、报警乃至阻断的能力。

(5)除此之外，身份认证系统应具备一定的灵活的配置和扩展能力，能够和第三方身份认 证、审计平台进行整合，便于降低实施成本。

**290** 数据治理——工业企业数字化转型之道

身份认证系统应对用户、应用、应用账号的登录、单点及管理等操作进行集中的日志记录， 提供基本的安全审计和常用报表功能。根据用户需要，可以将日志转发给第三方处理或者进行 定制化开发，实现更多功能。

当前，工业互联网企业采用云部署、容器部署已经成为趋势，所以该系统需采用云技术实 现系统模块的部署，通过云架构为用户和应用提供更健壮的不间断的服务能力。

**本章精要**

本章介绍了工业互联网数据安全治理中所使用的具体工具，包括数据采集管理、数据传输 管理、数据存储管理、数据处理管理、数据交换和共享管理、数据销毁管理6类工具，这些工 具有的是单独执行的小程序，比如数据分级打标工具等，有的是相互组合在一起形成包括多种 工具集成功能的平台，为数据安全治理过程提供在线的安全服务，比如统一的身份认证系统等。

第28章

数据中台

在信息化领域中，过去把业务运行和数据分析分开，分别叫作在线交易系统(OLTP) 和在 线分析系统 (OLAP); 而现在，随着大数据、人工智能等技术的成熟应用，过去的划分方式不 再适用于当前的环境，业务系统的功能不再局限于“录单据、跑流程和出报表”;每个业务人员 都需要基于数据分析做判断，决定下一步如何处理；每个业务环节以前都是业务人员根据经验 来进行审批、处理、控制的，现在结合人工智能技术，有些业务环节基于数据模型可以实现自 动处理，这在很多行业中都已经应用。这种系统架构和应用也已经成为一种不可阻挡的趋势。

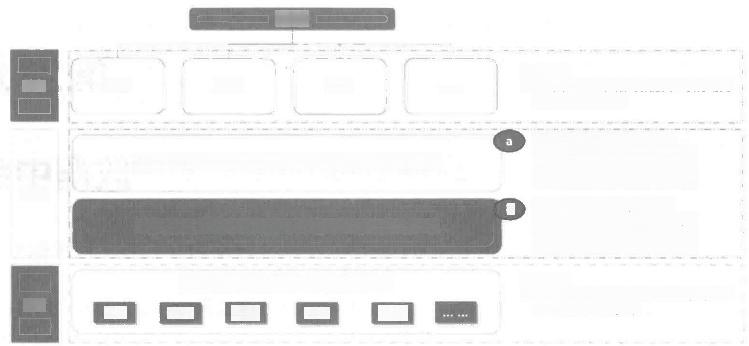
从另一个角度来看，企业内部的生产经营管理要规则清晰、流程明确、稳定固化，而全球 化的市场又对企业提出了灵活多样、敏捷高效的要求，这就要求企业在内部(后台)的稳态和 市场(前台)的敏态之间有柔性过渡的方案，因而“中台”的概念应运而生。数据中台可以实 现业务数据资产化和数据资产价值化。

本章将从数据中台的概念和定位，以及所具备的技术能力等方面来进行介绍。

**28.1** **数据中台的概念和定位**

数据中台通过对企业内、外部多源异构的数据进行采集、治理、建模、分析、应用，对内 优化管理、赋能业务，对外开展数据合作，释放数据价值。它是企业数据资产管理和服务中枢。 数据中台是前台和后台之间的数据桥梁，是创新性应用与传统数据之间的纽带，是解决应用开 发和数据开发不协同问题的工具，如图28-1-1所示。

**292** 数据治理 工业企业数字化转型之道



企业

敏捷前台

√一线作战单元，强调敏捕交互及稳定交 付的组织能力建设

业务中台：能力固化与赋能

√固化通用能力，赋能前线团队 √提升配置效率，加快前线响应 √产品化业务化，开辟全断生态

**中台**

数据中台：资产整合与共事

√整合多维数据，统一资产管理 √连通数据孤岛，共事数据资源 √深入挖掘数据，盘活资产价值

√以共享中心建设为核心，为前中台提供 专业的内部服务支撑

业务中台——企业核心竞争能力的固化与赋能

**数据中台——企业核心数据资产的整合与共享**

后台支撑职能的共享化、服务化建设

ERP CRM SRM MES LIMS

稳定后台

伙伴

客户

渠道

**前台**

**后台**

b

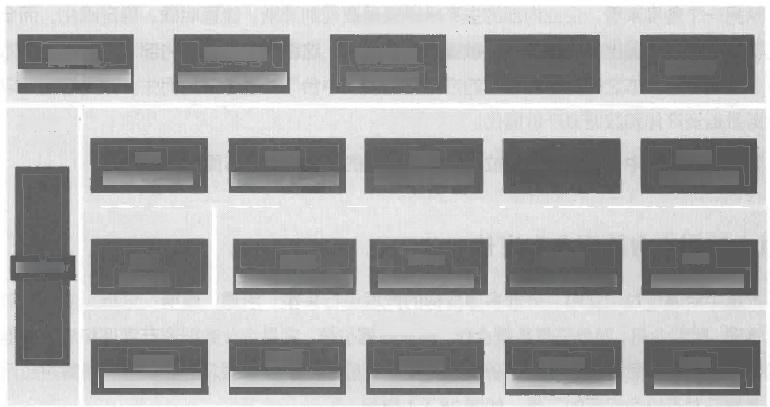
mm\*\*

图28-1-1 数据中台的定位

设计一个平台或了解并应用一个平台要从其技术架构入手。通过研究架构中每个组件的位 置、功能和用途，可以了解不同组件的范围和边界，以及技术能帮我们实现什么、如何实现得 更好。

数据中台的核心组件通常有：数据采集及传输、数据存储、数据计算及查询、数据展示及

分析，如图28-1-2所示。



**数据展示及分析**

可视化软件

Tableaw/Qli

任务调度

管理

Hive

使用MR/ 类SQL

**资源管理**

Yarn

集群管理

**数据采集及传输**

logstas

ELK 技术栈采集引擎

**数据计算及查询**

kyin

多维/Cube 预计算

**数据存储**

HBase

列式存储/NoSQL

可视化框架

Superset/Metabae

Spark

快速/内存计算/批量

Fulme

分布式采集日志引肇

MysQL

应用广/单机/关系型

Filebeat

轻量级日志传输工程

图标控件

HighCharts/Echarts

HDFS

分布式/Hadoop 底层

TiDB

单机+分布混合式

Sqoop

RDBMS 导入工具

Kafka

分布式消息队列

Druid

实时分析/列式

Presto

内存计算/列式

机器学习

数据挖掘

Zookeepur

**图28-1-2** **数据中台架构图**

第28章 数据中台 **293**

**28.2** **数据采集**

数据采集又被称为数据获取，其利用一种装置从系统外部采集数据并输入到系统内部。它 是把在不同数据源中的数据经过收集、整理、清洗、转换后加载到一个新的数据源中，为数据 使用者提供统一数据视图的数据集成方式。

**28.2.1** **数据采集方式**

数据采集分为通用数据采集和流式数据采集，其中通用数据采集包括结构化数据采集和非

结构化数据采集两类。

**28.2.2** **通用数据采集**

通用数据采集包括以下功能。

(1)实现对数据源到目录数据平台之间的结构化数据的抽取、转换和加载功能；

(2)结构化数据采集：支持HANA、SAPBW、HDFS、Hive、Kafka、Flume、Hbase 等文

件类型数据源；

(3)非结构化数据采集：支持HDFS 、Kafka 、Flume 、Hbase 等文件类型数据源；

(4)支持分布式 ETL 任务执行，任务处理能力可以横向扩展；

(5)支持过滤、数学运算、字符运算等数据转换功能。

**28.2.3** **流式数据采集**

流式数据采集可以实现对流数据的抽取、转换和加载功能，支持对应用日志、系统日志 及数据库日志文件的增量采集，具体包括以下功能。

(1)提供高并发、高吞吐量、低延迟、可容错、可持久化的实时消息系统(如 Kafka、 RabbitMQ等);

(2)提供可扩展的分布式流数据ETL 处理功能；

(3)支持任务失败断点恢复；

(4)支持分级的并发任务调度功能；

(5)具备基于Web 的图形化流定义工具界面，提供任务管理、任务监控界面及功能；

**294** 数据治理 工业企业数字化转型之道

(6)具备完善的用户及权限管理功能。

**28.3** **数据存储**

数据存储组件具有横向扩展的分布式数据存储、NoSQL 数据存储等能力。

**28.3.1** **分布式数据存储**

分布式数据存储能力具备以下功能：

(1)具备可扩展的分布式数据存储功能；

(2)支持结构化数据、非结构化数据、半结构化数据的分布式存储；

(3)支持1:4以上的数据压缩比；

(4)支持高可用和故障自动切换部署；

(5)具备数据安全策略(如多副本策略),避免数据丢失或损坏；

(6)支持结构化数据的ANSI SQL 2003语法访问。

**28.3.2** **NoSQL 数据存储**

NoSQL 数据存储能力具备以下功能：

(1)可扩展的分布式Key-Value数据库功能；

(2)支持对象存储；

(3)可以用于图数据库；

(4)支持高可用和故障自动切换部署；

(5)具备数据安全策略(如多副本策略),避免数据丢失或损坏；

(6)具备基于Web 的图形化管理工具。

**28.4** **数据计算**

数据计算组件提供了数据处理分析的计算能力，包括分布式查询、分布式计算、数据建模 和数据分析等。

第28章 数据中台 **295**

**28.4.1** **分布式查询**

分布式查询能力具备以下功能：

(1)支持联机分析处理(R OLAP);

(2)支持标准ANSISQL2003 语法访问，支持复杂SQL 查询，支持UDF;

(3)要求有良好的水平扩展性；

(4)支持高可用和故障自动切换部署；

(5)支持基于角色的数据访问权限控制；

(6)提供基于Web 的图形化管理工具。

**28.4.2** **分布式计算**

分布式计算能力具备以下功能：

(1)支持结构化、非结构化数据的分布式批处理功能；

(2)能够支持大规模水平扩展；

(3)支持高可用部署， 一个节点“挂”了，计算任务会自动转移到其他节点；

(4)支持任务失败恢复能力；

(5)具有高效的分布式内存计算引擎(如 Spark、Flink 等);

(6)支持Java 、Python 等多种开发语言。

**28.4.3** **数据建模**

模型搭建是数据分析计算中的关键环节，也是高难度环节。数据建模能力具备以下功能：

(1)提供数据库建模工具，可自动生成物理实体；

(2)提供统一的基于Web 方式的集成开发环境，支持数据库开发、数据处理、数据集成等 功能；

(3)支持数据中台模型开发，可定义维度、度量、ODS、Cube 等大数据模型；

(4)提供基于Web 的图形化模型设计、训练、评估；

(5)支持特征工程创建；

(6)支持主流统计分析算法及图形。

**28.4.4** **数据分析**

数据分析能力是直接考验数据平台能力及可用性的核心指标，其具备如下功能。

(1)支持对多维模型的钻取(Drill-down) 操作，在维度的不同层次之间的变化，从上一层 降到下一层，或者说是将汇总数据拆分细粒度数据；

(2)支持对多维模型的上卷(Roll-up) 操作，钻取的逆操作，即从细粒度数据向更高层 次的聚合；

(3)支持对多维模型的切片 (Slice) 操作，即选择维中特定的值进行分析；

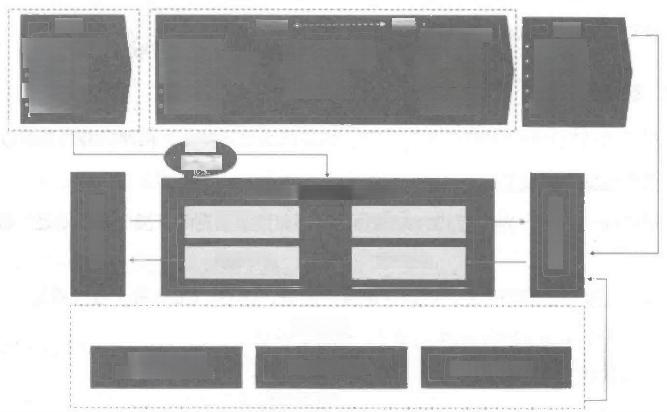
(4)支持用户从两个维度内选择所关心的值进行展示和分析；

(5)支持对多维模型的切块(Dice) 操作，即选择维中特定区间的数据或者某批特定值进 行分析；

(6)支持对多维模型的旋转(Pivot) 操作，即维的位置互换，就像是二维表的行列转换。

**28.5** **数据服务**

数据中台的作用就是将处理好的数据以服务的方式提供给相应的用户，满足他们的分析需 求。而这些服务通常都以API 的方式提供给各类应用去调用，如图28-5-1所示，具体介绍如下。



0ps

·预制准生产

·生产环境

·实例管理 ·持续交付



Push

config

**API同关：调度、分发、治理**

方法请求 集成请求

方法响应 集成响应

**服务提供**

API Manager

Rackend Service

**设计**

·API 在线定义

·测试设计

·安全定义

·访问控制

SLA 设置

·版本管理

·框架与样例代码 ·预制开发环境

在线编译

持续集成

·预制测试环境 ·静态代码检直 ·自动测时

·持续集成

管理 服务审批 服务目录 服务市场 服务消费 ·运营分析

互联网开放API

遗留/已建系统

客 户 端 调 用

服 勢 端 提 供

Dev

**图28-5-** **1** **数** **据** **计** **算** **—** **—** **数** **据** **服** **务** **图**

第28章 数据中台 **297**

●提供API 网关。API 网关可以完成配置调用路由、调用权限验证、调用限制等功能。

●提供 API 生成、API 登记发布、API 调用申请审核功能。调用者可浏览 API、 申请调用 API, 在通过审核之后，可调用API。服务的发布者负责审核申请，提供API 生成、审核 的工作流。API 网关支持 RESTful 、Web Service 协议转发。

**28.5.1** **API** **网关**

API 网关为平台业务系统提供高性能、高可用、高安全的 API 托管服务，以及防攻击、请 求加密、身份认证、权限管理、流量控制等多重手段，保证API调用安全，降低因API开放带 来的风险。

通过API 网关，开发者可以封装后端各种服务，以API 的形式提供给企业内部或合作伙伴 使用。同时，API 网关协助开发者完成 API 文档管理、API 测试，并提供各语言客户端代码， 从而提高API 调用者的工作效率。

无论是服务调用者希望使用API 网关服务，还是服务提供者希望通过服务目录将自己的能 力开放，都可以通过API 网关来实现。服务提供者将自己的业务部署在 API网关上供服务消费 者调用，最终用户(无论是移动客户端、Web 客户端、物联网或其他应用)都可以直接通过 API 网关调用API服务。

API网关为企业业务系统带来以下便利：

(1)规范化接口，企业内部多套业务系统对外提供统一、规范化的 API。

(2)将调用API 的权限鉴别功能下沉到平台层处理，简化服务代码，减少冗余。

(3)通过API 网关将企业内部服务能力以标准API 的形式开放给合作伙伴。与合作伙伴共 享服务、能力和数据，达成深度合作，共建新生态。

API 网关有如下特性。

(1)简单易用：基于云平台，用户可直接在管理控制平台中对 API 网关进行图形化配置， 无须自行搭建 API 网关部署所需要的设备。

(2)易于维护：提供对API 网关的可视化监控，以及资源管理、租户隔离、权限控制等一 系列的运维能力，将用户从繁重的运维中解放出来。

(3)高可靠性： API 网关是分布式集群部署，当网关节点出现故障时，可及时迁移，以实 现高可靠性。

**298** 数据治理 工业企业数字化转型之道

(4)高性能：可承载大规模、大流量的API 调用访问。

(5)安全：通过接入多种认证方式，确保API 的访问安全性；通过严格的流量控制，避免 用户服务的过载；通过全面的监控告警，保证用户服务的可用性。

**28.5.2** **API 生成**

数据管理模块支持API 在线生成功能。在管理页面中，对于权限范围内的数据源，操作员 可以通过拖放等操作定义数据发布内容，同时填写必要的 API 定义。在提交这些配置数据后， 系统可根据配置生成对应的 API 服务。

**28.5.3** **API 发布**

发布者可一键发布经过测试的数据共享 API, 完成 API 基础信息、API 路由信息的创建， API调用限制定义，以及将API 登记到用户指定的运行环境 API 网关上。

API 基础信息包括 API 的版本、调用方法、调用路径、请求参数定义和描述、响应参数定 义和描述等。

API 路由信息包括服务名、调用重试次数、超时时间定义等。

API调用限制定义包括API 发布范围，以及多组API 调用的限制方式和配套的参数设定。

在完成API 发布后，通过服务目录功能，调用者可以看到相关的服务信息、API 信息。

**28.5.4** **API 调用申请**

当服务发布完成后，调用者可以查看并申请调用 API。虽然在服务发布时设置了调用者范 围，但是调用者仍然需要经过审批，通过之后才能正常调用API。 进行 API 调用申请时，需要 调用者提供自身的 API-key,系统将 API-key 作为令牌在调用 API 时验证调用权限和匹配调用 规则。

**28.5.5** **API 调用审核**

执行API 调用申请最终审核的是服务发布者。服务发布者通过审核API 调用申请，可获得 调用者期望调用的 API, 在评估后通过或拒绝审核。

第28章 数据中台 **299**

**28.5.6** **API 信息支持**

数据共享模块为服务调用者提供了页面方式的 API 测试工具，调用者可在页面中完成 API 的测试，从而获得高度还原的仿真结果。

**28.5.7** **API 服务监控**

-API 服务监控信息可以统计API 服务的调用次数、调用者分布、错误码分布。这些数据可 以帮助服务发布者全面了解 API 服务的运行情况。

**28.6** **从** **ETL向** **ELT 转变**

如今，传统的数据仓库已不能满足企业的数据分析需求。企业在数据分析应用方面呈现“五 大转变”(从统计分析向预测分析转变、从单领域分析向跨领域分析转变、从被动分析向主动分 析转变、从非实时分析向实时分析转变、从结构化数据分析向多元化数据分析转变),并且对统 一的数据中台诉求强烈，对数据中台的运算能力、核心算法及数据的全面性也提出了更高的 要求。

如今，数据中台的处理架构发生了变化。 一是实现以 Hadoop、Spark 等分布式技术和组件 为核心的“计算与存储混搭”的数据处理架构，能够支持批量和实时的数据加载及灵活的业务 需求。二是数据的预处理流程正在从传统的 ETL(Extract-Transform-Load, 用来描述将数据从 来源端经过抽取、转换、加载至目的端的过程)向 ELT(Extract-Load -Transform) 转变。传统 的数据仓库集成处理架构是 ETL结构的，这是构建数据仓库的重要一环，即用户从数据源抽取 出所需的数据，经过清洗后被加载到数据仓库中。而大数据背景下的架构体系是 ELT 结构的， 其根据上层的应用需求，随时从数据中台中抽取想要的原始数据进行建模分析。

数据中台整体采用云计算架构模式，将数据资源、计算资源、存储资源充分云化，并通过 多租户技术进行资源的打包整合与开放，为用户提供“一站式”的数据服务。数据中台利用大 数据技术，对海量数据进行统一采集、计算、存储，并使用统一的数据规范进行管理，将企业 内部所有的数据统一处理成标准化数据。数据中台通过挖掘出对企业最有价值的数据，构建企 业数据资产库，向企业提供一致、高可用的大数据服务。图28-6-1展示了传统数据仓库与数据 中台的对比。

**300** 数据治理——工业企业数字化转型之道

|  |  |  |  |
| --- | --- | --- | --- |
|  | **大数据(中台)架构** | **混合架构** | **传统数据仓库** |
| **部署架构** | 水平扩展 | 水平扩展 | 大部分水平扩展 |
| **数据类型** | 文件存储，没有数据类型 | 简单数据类型 | 丰富的数据类型 |
| **数据模型** | 非常简陋的数据模型 | 简单灵活数据模型 | 完善丰富的数据模型 |
| **数据关系** | 没有数据关系描述 | 非常简单的数据关系描述 | 数据关系完善 |
| **数据一致** | 无一致性 | 弱一致性 | 强一致性 |
| **数据安全** | 安全性很弱 | 安全性很弱 | 安全性很高 |
| **计算类型** | 离线批量处理，只读，低 并发 | 实时CRUD操作，海量并发 | 离线批量处理，只读，低 并发 |
| **适用场景** | 低密度数据海量存储，数 据预处理，预计算 | 高并发实时 | 高价值数据统一存储和计 算平台 |
| **人员要求和成本** | 高 | 较高 | 适中 |

**图28-6-1** **传统数据仓库与数据中台的对比**

**本章精要**

数据中台不只是一个技术概念或平台产品，它还是一个体系、 一个生态，以及一个实现数 据服务和供给数据服务的平台。数据中台同时也是业务中台的支撑。数据中台以 API的方式把 各种数据服务直接嵌入业务交易系统中。其多应用于产品研发、控制优化、运营风险管控、物 流规划、生产调度等诸多业务应用场景中。数据中台的建设，应以应用场景设计为先，与具体 业务应用场景结合在一起，才能为企业的经营和生产的运行、分析、决策提供完整、高效的支 持，为企业的数字化转型提供强大的助力。

**参考资料**

[1]William H,Inmon.Building the Data Warehouse.1992

[2]Danette McGilvray. 数据质量工程实践.刁兴春，等译.北京：电子工业出版社，2011

[3]April Reeve. 数据集成的技术、方法与最佳实践.余水清，潘黎萍译.北京：机械工业出版

社，2014

[4] 《数据安全治理白皮书》,中国网络安全与信息化产业联盟数据安全治理委员会

[5] 《数据资产管理实践白皮书(4.0版)》,中国信息通信研究院

[6]GBT36073-2018《 数据管理能力成熟度评估模型》

第28章 数据中台 **301**

[7]《工业互联网数据安全监管技术方案》,杭州安恒信息技术股份有限公司

[8]Hoberman,Steve,Burbank,et al.Data Modeling For The Business[J].2009

[9]W.H.Inmon,Daniel Linstedt.数据架构：大数据、数据仓库以及 Data Vault.唐富年译.北京：

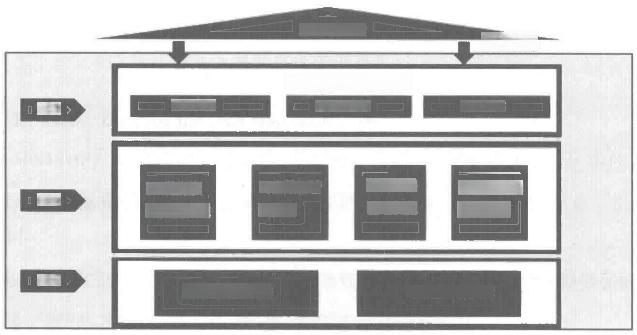
人民邮电出版社，2016

[10]Steve Hoberman. 数据建模经典教程.丁永军，译.北京：人民邮电出版社，2017



**第** **4** **篇** **实施篇**

数据治理的实施是一个系统工程，成功的数据治理首先应该支持企业的发展战略。下图为 数据治理实施框架。数据治理的首要工作就是确定数据治理的目标、内容、方法等，即要进行 数据治理的顶层架构规划与设计；然后在顶层架构规划与设计的基础上，将数据治理的核心举 措落地，包括数据资产运营实施、主数据管理实施、元数据管理实施、数据指标管理实施、数 据质量实施及数据安全实施等；最后，数据治理是否能够成功实施，保障措施也是非常重要的。 因此，企业在实施数据治理时一定要有组织、制度和工具的保障。



企业发展战暗

要求、捐尊

1.顶层架构规划与设计

步骤与方法

3.主数据管

理实施

保障

7.数据安全实施

5.数据指标 管理实施

2. 数据资 产运营实施

4.元数据 管理实施

6.教据质量实能

实施内容

价值评估

促进

支撑

实现

目标

举措

数据治理实施框架

第 2 9 章

数据治理实施策略和路径选择

**29.1** **实施内容**

当下，很多企业的信息化系统还处于大规模的集中建设期，各项信息系统的建设和优化升 级工作正在如火如荼地开展。但是在很多企业中，老旧系统中的存量数据的质量问题非常突出， 很多业务创新和市场营销的数据需求远远得不到满足，而新业务系统上线后，由于数据质量问 题而造成的生产问题也时有发生。摆在企业面前的数据治理工作是尽快解决影响信息系统推广 应用效果和业务用户体验的各种数据质量问题。根据数据治理的不同情况，这里归纳并总结出 以下实施策略，作为企业实施数据治理的参考建议。

**1.** **坚持支持企业战略，突出数据治理绩效**

数据治理是实现企业 IT 规划目标架构的重要举措，务必要在继承企业 IT 规划和以往的信 息化建设成果的基础上推行，这是从企业的长远利益来考虑的。所以，数据治理项目的实施一 定要与IT 规划和当前开展的信息化建设项目相向而行、无缝融合，不能只从数据资源加工处理 的角度来实施，并且需要实施厂商对企业IT 规划和当前信息化建设现状有非常深入的了解和高 度承接。数据质量问题与企业的绩效呈现强相关关系。数据就像汽车中的汽油，汽油质量不好， 再好的车也跑不快。

**2.** **制订数据治理长期计划，兼顾数据短期成效**

从众多企业的数据治理实践来看，数据治理是一个长期和逐步见效的过程，是战略性的举 措，期待一次性就能把数据治理工作做好是不现实的，必须以长期建设的心态来制定实施路线。

**304** 数据治理 工业企业数字化转型之道

而在工业企业中，数据质量相关问题很突出，容不得犹豫和拖延，所以在实施操作上，小步快 跑、以点带面的短线操作更符合各企业现状。从策略上看，数据治理既要兼顾长远目标又要兼 顾短期效果，建议选择与有长期服务及数据治理能力的厂商持续合作，因为其能从产品、技术、 咨询、实施等各角度提供全面的长期支持。

**3.** **严选数据治理工具，聚焦数据实施能力**

“工欲善其事，必先利其器”。数据治理是一项复杂的系统性工程，如果没有有效的工具支 撑就开展数据治理工作，则结果是不可想象的。因此，企业在开展数据治理项目时要引入业界 优秀的数据治理工具。数据治理工具的选用和部署只是万里长征的第一步，工具的选用并不是 决定成功的关键要素。因为数据治理工作的落地要依靠对各相关系统的改造和业务人员大规模 进行业务数据采集、清洗、更新录入等工作。所以数据治理成功的关键在于有效协调企业各相 关方的资源，将数据治理工作和业务经营各信息系统的需求、建设和运维工作各环节有效衔接， 这样才能制定出最优的解决方案和系统性的实施方案。因此，在合作厂商的选择上，建议不要 过于聚焦现有工具的对比，要把重点聚焦在合作厂商在未来1~3年的服务软实力上。企业需要 的是端对端的软服务，而不是生搬硬套拿来让企业削足适履的硬产品。

**4.** **坚持以解决数据问题为导向，重视数据治理的全局把控**

“冰冻三尺，非一日之寒”,很多企业的数据质量问题积重难返，要彻底解决需要投入大量 的时间和资源。数据治理工作先治标后治本是对企业现实要求的折中，首先要解决具体的数据 质量问题，在开展数据治理工作时必须坚持以数据问题为导向，每一步工作都要明确具体解决 什么问题，能体现怎样的业务效益。另外，企业也不能只低头解决一些具体的数据质量问题， 而忽略了数据治理体系的长期建设路径。产生数据质量问题的原因有很多，如果我们只看后果 而忽略了原因，那么数据质量问题并不可能得到彻底的根治，因此，从治本的角度来看，企业 必须要站在IT 战略和总体架构管控的角度，从总体架构把握数据治理的各项决策，确保不偏离 企业的战略方向要求。从这个层面来说，做好充分的顶层设计才是确保成功的关键。

**29.2** **路径选择**

企业数据治理工作的路径选择：根据组织方式，可选择自上而下的顶层设计模式或自下而 上的各个击破模式；根据建设策略，可选择从生产系统切入和从数据系统切入。

第29章 数据治理实施策略和路径选择 **305**

**1.** **根据组织方式，分为自上而下和自下而上两种模式**

如果企业将数据治理纳入战略规划，且企业高层领导拥有较大的决策权，则可以采用自上 而下的顶层设计模式，结合企业业务发展，制定长远的数据治理规划。如果企业的数据管理部门 具有一定的独立性，并且具备专业技能和相关经验，则可以采用自下而上的各个击破模式，以 探索数据治理需求为驱动力，通过以解决数据问题为导向，推动企业数据治理的逐步完善。这两种 建设路径的解释详细见表29-2-1。

**表29-2-1** **数据治理的两种建设路径**

|  |  |  |
| --- | --- | --- |
| **建设路径** | **建设要点** | **优** **缺** **点** |
| 自上而下模式 | 规划先行，组织体系先行，随后是分阶段、分 步骤地建设实施 | 有体系和节奏，规范性好，适合有分支机构的大型 企业；时间和成本投入很大，见效慢 |
| 由下而上模式 | 从具体某一业务需求开始，由点及面，逐渐扩 展到组织的其他业务 | 需求驱动，快速行动，见效快；统一整合比较困 难，适合机构和业务不多的中小型企业 |

(1)数据治理自上而下模式的显著特点是“规范、标准先行”,通常由数据治理咨询项目开 头。这种模式通常包括以下几项重要内容。

①调研数据治理现状：在数据资产盘点、收集调研问卷、现场访谈等手段的基础上展开数 据治理现状调研。结合业务场景，充分了解企业当前的数据资产分布情况，有助于企业在展开

自上而下的数据资产管理前掌握业务人员的数据需求。

②评估数据治理水平：通过自评估或者专业机构进行数据资产管理评估，帮助企业在实施 自上而下模式之前了解当前自身数据资产管理的现状，明确存在的问题和潜在的挑战，规划适 当的数据资产管理蓝图。

③建设数据治理体系：数据管理工作是一项跨业务、跨部门的系统工程，自上而下模式的 实施高度依赖高层管理人员的支持和职能集中化的数据资产管理组织。数据资产管理体系通过 明确管理战略、制定管理制度、搭建组织架构等一系列活动，以企业级的全局视角推进数据资 产管理的实施。

(2)数据治理自下而上模式的显著特点是“问题导向、系统建设先行、快速见效”,以解 决各业务部门和业务系统数据管理中的问题为出发点，通过使用成熟的数据治理工具，快速搭 建数据治理平台，实现对问题的逐个击破，并逐渐探索出全面的解决方案。自下而上模式一般 以解决企业面临的主数据管理、数据质量管理两项核心数据管理任务作为切入点，然后逐步扩 展到数据模型管理、数据标准管理、数据安全管理等其他数据管理职能。

**306** 数据治理——工业企业数字化转型之道

**2.** **根据建设策略，分为从生产系统切入和从数据系统切入**

从生产系统切入的建设模式包括大型生产系统开发建设模式、企业数据模型建设模式和主 数据建设模式。从数据系统切入的建设模式包括统一数据平台模式和数据集市模式。

企业在选择不同的数据治理建设路径时，需要考虑数据对于企业的重要性，以及企业目前 的数据管理水平。如果数据是企业重要的业务资源，同时企业已经具备了一定的数据治理专业 水平和经验，则可以通过从数据系统切入的实施模式，修复数据管理中的漏洞、提升数据服务 应用水平，推进数据管理能力建设。

如果企业的数据管理水平并不成熟，那么选择直接从数据系统切入有些冒险，而从生产系 统切入则较为稳妥，也易见成效。各模式的解释详细见表29-2-2。

**表29-2-2** **数据治理的切入方式**

|  |  |  |
| --- | --- | --- |
| **建设路径** | **常用建设模式** | **建设要点** |
| 从生产系 统切入 | 大型生产系统开发建设模式 | 从大型生产系统的开发入手，借助项目建设契机，建立该应用和业务 领域数据的企业级标准和质量管控 |
| 企业数据模型建设模式 | 从企业数据模型出发，在建模的同时建立标准，规范生产环节的数据 录入，保证数据质量 |
| 主数据建设模式 | 从解决主数据的质量和业务协同入手，推动生产环节中的统一编码 |
| 从数据系 统切入 | 统一数据平台模式 | 以数据仓库、大数据平台等统一数据整合平台为切入点，统一接入各 业务、各分公司的数据，统一语义和标准，提升数据质量 |
| 数据集市模式 | 各业务单独建立自己的数据仓库，满足自己的数据分析需求；或者以 某个特定的分析主题为切入点进行建设，后续统一对每个业务的数据仓 库进行语义和标准方面的规范，实现物理分离，逻辑统一 |

**本章精要**

数据治理是一个长期和逐步见效的过程，是战略性的举措，必须以长期建设的心态来制定 实施路线。本章主要介绍了数据治理实施策略和路径选择，作为企业实施数据治理的参考。

第30章

数据治理顶层架构规划与设计

数据被称为“未来的新石油”,如今已经渗透到企业的战略决策、生产经营和各业务职能领 域中，成为企业数字化转型的基础。本章介绍了数据治理顶层架构规划与设计的3个阶段：调 研与需求分析、顶层架构总体设计、应用场景验证，以及如何加强数据治理管控，强化组织和 制度保障，进一步明确数据全生命周期过程的相关权责，实施标准化、规范化、体系化的管理， 确保数据生产、使用的全过程受控。

**30.1** **实施内容**

企业需要持续深化数据治理的管理工作，围绕数据价值挖掘、数据质量提升及在管理落实 过程中的难点及痛点，建立完整的数据治理体系。通过数据治理体系可以实现以下目标：

梳理各业务部门在数据治理中的职责，对各类数据资产建立标准，统一数据资产管理流程；

搭建数据资产管理平台，编制数据资产管理相关制度，建立数据资产管理组织，为提升企 业数据质量奠定基础；

逐步建立良好的数据治理文化氛围，提高员工对数据及数据质量工作的重视程度，加强业 务部门在数据管理过程中的参与程度；

明确各部门在数据治理中的管理范围、管理职责，把数据治理工作融入日常的工作过程中。

数据治理顶层架构规划与设计应遵循但不限于以下指导原则。

(1)统一性原则。企业数据治理顶层架构规划与设计应遵循统一性原则，基于企业整体业 务、信息化战略目标，统筹考虑企业信息化建设的其他部分内容，并进行统一规划设计。

(2)实用性原则。企业数据治理顶层架构规划与设计应以业务运营和经营管理的实际情况 为贴合点，以解决实际需求为导向，进行数据治理体系的规划、设计和建设，确保管理体系规 划实用可行，为上层创新应用提供有力支撑。

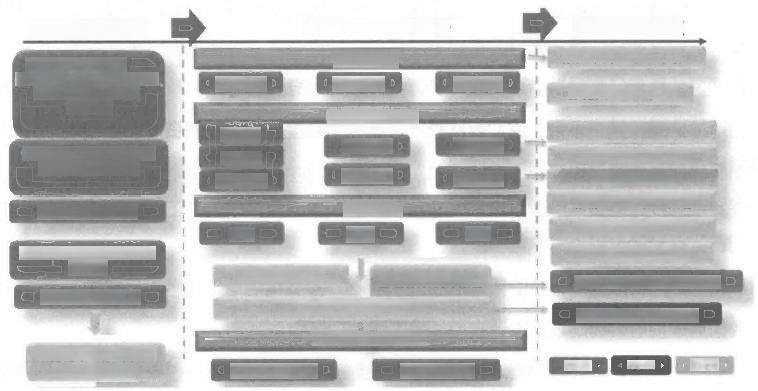
(3)标准化原则。企业数据治理顶层架构规划与设计应遵循标准化原则，根据企业的实际 应用需求制定相关的体系建设标准、规范和指导文件；明确总体规划目标，统一规划思路，提 供组织间的交互公约，稳步推进体系建设分批、分步实施。

(4)开放性原则。企业数据治理顶层架构规划与设计应遵循业界流行的开放标准，能够充 分考虑未来应用发展的需要，使得数据治理体系在公司业务发展的过程中不需要重新进行规划 与设计，能够顺利、平稳地向更新的技术过渡。

(5)易推广性原则。企业数据治理顶层架构规划与设计应充分考虑企业现有的信息化基础， 在保证先进性、实用性原则的基础上，尽量降低推广难度，并提供切实可行的推广策略和推广 方法。

(6)安全性原则。企业数据治理顶层架构规划与设计应遵循安全性原则，符合企业信息化 建设安全和等级保护的相关规范，通过技术手段保障数据信息的安全，提高数据安全风险识别 和处理能力。

数据治理顶层架构规划与设计主要分为调研与需求分析、顶层架构总体设计、应用场景验 证3个阶段，如图30-1-1所示。



**顶** **层** **架** **构** **总** **体** **设** **计**

及展设计

组织构建

致擦治理核心城

数据标准

教据模型

保障措施

流程

管理标准设计报告

顶层设计及总体实施方案

数据治延其他相关域

数据全生命周期 胶据需求管理

**调研与需求分析**

**理论基础：DAMA**

**《数据海产白皮书(40版)》、**

**工业大数据白皮书、**

**DCMM**

外部调研，对标领先企业 数据治理框架体系

**企业参考资料研读**

企业”十四五”信息化发 展规划

**企业主要信息系统**

数据冶理现状评估与 需求分析报告

**应用场景验证**

数据架构与集成方案

数据资产关键模型

数据资产目录

平台技术选型标准

平台建设规划

数据治理实施路径图

运营组织设计报告

数据系统开发标准

年度垂需发布的制度和标准

典型工具主要功能验证

平白吴

战暗规划

元数据 主数据 数据据标

数据质量

数据安全

技术

机制设计报告

架构设计

制度

交

联

图30 - 1 - 1 数 据 治 理 顶 层 架 构 规 划 与 设 计 方 法 论

第30章 数据治理顶层架构规划与设计 **309**

**1.** **调研与需求分析**

(1)理论基础：《数据资产管理实践白皮书(4.0版)》《工业大数据白皮书》等。

(2)外部调研、对标：领先实践数据治理框架体系的企业。

(3)企业参考资料研读：包括企业信息化发展规划、企业主要信息系统等。

**2.** **顶层架构总体设计**

(1)顶层设计：包括战略规划、组织构建、架构设计。

(2)数据治理核心域：由数据标准、数据模型、数据质量、数据安全、数据指标、主数据、 元数据7部分组成。

(3)保障措施：包括制度、流程、技术，共3部分。

(4)数据治理其他相关域：数据全生命周期、数据需求管理。

**3.** **应用场景验证**

(1)数据治理工具验证：例如物料主数据、设备主数据系统验证。

(2)数据资产工具验证：实现数据资产目录、数据资产地图。

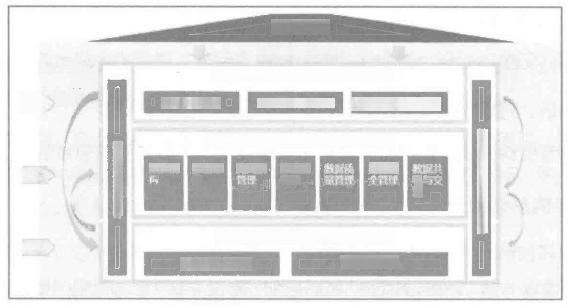
**30.2** **步骤和方法**

**30.2.1** **顶层设计总体思路**

**1.** **引入** **DAMA 数据管理知识体系设计理念**

通过引入DAMA 数据管理知识体系设计理念，企业可以建立数据战略体系(见图30-2-1), 将数据治理、管理制度、管理服务统一在数据管理平台中，对数据架构体系集中统一管理。通 过技术基础和系统集成，能够与业务建立整体的逻辑映射关系，并通过规范、可视、易于理解 的形式表达出来，对于后续的业务需求沟通和系统建设过程可以起到指导作用，指引企业数据 资产管理的落地实施。数据资产管理的顶层设计过程是核心输出成果之一。

**310** 数据治理——工业企业数字化转型之道



**数据战略**

**要求、指导**

数据治理(Data Governance)

支撑

教据战路规划 治理组织设计机制 民据性理形度与流程

**数据管理(Data** **Management)**

主数区

首理

促进

实现

**数据应用(DataApplication)**

目标

数取应用与器务 教时置专普

数 据 治 理 技 术 支 彰

数配标 避

数 报 治 理 审 计

教据架

元裁服

教抵安

保障

举措

支撑



**图30-2-1** **数据战略体系**

**2.** **调研与需求分析**

通过对企业数据治理文档、数据、制度、流程、信息化系统运行记录等现状资料的收集， 可以充分理解企业的战略、管理模式和各主要业务方向，明晰企业发展战略对数据资产管理、 数据治理、数据质量、数据运维、组织架构的需求。

要厘清企业对下属单位的管控模式，例如企业管什么,如何管等，如图30-2-2所示。



**企业如何管理数据?**

◆首先要明确企业各层级对数据资产管控的范围，从业务功能、

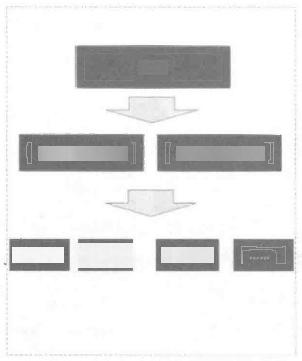
资产类别方面进行区分

企业

◆需要充分考虑企业产业多元化和业务单元分散管理的实际 现状问题



**企业管多深?分几层管理?**

职能管控单元 系 统 管 理 平 台

◆在管理功能方向进行清晰划分，需要明确管理的深度，对职

责到岗人员进行细分，可以根据下属单位的产业特性和

管理成熟度，将权限分层级进行下放，设立二级、三级的数据 资产管控模式，并将管理效果纳入绩效考核机制

通过什么管?管理手段如何配台?

下属单位

下属单位

下属单位

◆ 数据资产管理平台

◆ 人事管理

◆财务管理

◆信息标准委员会

◆ 管控机制和管控体系

**图30-2-2** **企业对下属单位的管控模式**

在理解企业数据战略管理目标和要求的基础上，通过调研与需求分析，可以全面了解企业 的管理现状和业务现状，准确把握现阶段面临的问题与挑战，明确未来企业数据战略的需求和 方向。

第30章 数据治理顶层架构规划与设计 **311**

下面以能源化工业企业的生产类数据管理为例进行具体介绍。

首先，要对能源化工企业在生产过程中所需要的物料、介质、耗材，包括大量的实时数据、 采集数据、各生产数据采集系统与总部集控系统汇总数据的现状进行分析，发现存在的关键业 务问题。

然后，针对存在的问题提出改善建议，提炼数据资产管理需求，对企业的数据治理现状进 行评估，总结数据治理的关键改进方向。

接下来，从IT 管控、IT 基础设施、应用系统、数据管理等方面对企业的IT 现状展开评估， 具体包括以下几个方面。

(1)企业现有信息建设和在 IT方面的历史投入，以及获得的成果；

(2)下属单位现有的信息技术应用及建设历史；

(3)企业数据资产分布、建设情况和投入历史；

(4)数据应用系统的主要执行标准；

(5)数据标准化、安全性和共享能力；

(6)IT 基础设施情况和对数据的开发及利用情况；

(7)IT 管控的组织结构、管理流程。

为确保数据治理规划的先进性，要对先进企业的数据治理模式、信息化发展趋势和领先的 实践进行研究，分析企业信息化建设成功的案例。

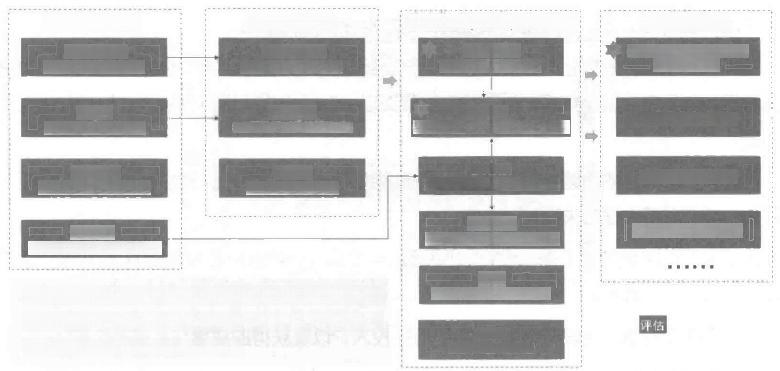
最后，要总结企业的数据战略、业务和管理现状，以及现有信息化技术，制定企业数据治 理架构和技术要求，以及实现数据战略落地实施的指导原则和总体的发展战略。

**30.2.2** **数据治理顶层设计要点**

在进行数据治理顶层设计时，需要研究国内主流的数据治理框架，对企业数据管理现状进 行充分调研，并形成翔实、可靠的现状分析报告；根据行业特点与企业业务实际，聚焦数据治 理，对企业数据治理开展顶层设计与体系规划，明确阶段目标与任务，有序推进，最终形成企 业数据治理的实施方案。

国内企业的数据治理理论发展过程是从通用体系/模型、单一领域体系、数据管理(治理) 完整理论到行业实践框架，具体内容如图30-2-3所示。

**312** 数据治理- 工业企业数字化转型之道



|  |  |
| --- | --- |
| 通用体系/模型  1509000  (玻量管理标准体系)  TQM  (全面质量管理体系) | 单一领域体系  15O 8000 (数据质螺标准体系)  TDQM  (全面数观质量管理体系) |

喻

MDA

(元模型蒙动架料)

CMMI

(软件能力减熟良集成模型)

**企业选择DMBOK、DCMM 以及《数据资产管理实践白皮**

书4.0版》中的实践框架作为研究对象，理论与实践结合，

**建立数据治理框架构建的理论基础**

**数据管理(治理)完整理论**

DMBOK

(数据管理知职体系)

DCMM

(效照能力成蒸后评估管理板型)

**DMM**

(数据管理能力减蒸度校型)

DSMM

(数联安全能力成系应模型)

**DGI**

数据治理研究所治理理论

**IBM**

数据治理要素极型

**行业实践框架**

**数据资产首理实践框架**

**(白皮书4.0)**

**工业企业实践框架**

**金融行业实践框架**



**互联网行业实践框架**

PSP/IQ

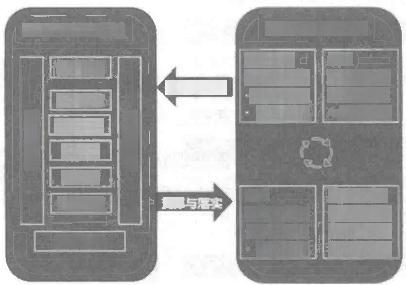
(产品和暖务续效核望)

图30-2-3 数据治理理论发展过程

**1.** **数据治理核心领域设计**

数据治理核心领域可分为两个方面： 一是数据质量核心领域；二是数据质量保障机制，如

图30-2-4所示。涵盖数据服务、主数据、元数据、数据质量、数据标准、数据安全等内容。



**数螺质量核心领域**

数据安全

保牌与规范

数据质置

数据指措标

主数据

元教螺

时床教据

数据模理

|  |  |
| --- | --- |
| **流程**  · **流程目标** · **流程任务**  **流程分级** | **π技术应用** · **技术规范** · **实施办法** · **支排平台** |

|  |  |
| --- | --- |
| **制度章程**  、规章制度 · **控办法** 考核机制 | 组织  ·组织架构 ·组织层次 · 组织责 |

**数据质量保障机制**

數播服努

散据标准

支挥

图30-2-4 数据治理核心领域

这些内容既有机结合，又相互支撑。下面具体介绍其中几种。

( 1 ) 数 据 模 型 。

数据模型是数据架构中重要的一部分，包括概念数据模型和逻辑数据模型。数据模型管理 是数据治理的关键。优秀的数据模型应该具有非冗余、稳定、 一致、易用等特征。逻辑数据模

型能涵盖整个企业的业务范围，能清晰地记录、跟踪企业的重要数据元素及其变动，并利用它

第30章 数据治理顶层架构规划与设计 **313**

们之间各种可能的限制条件和关系来表达重要的业务规则。

(2)数据标准。

数据治理对数据标准的需求可以划分为两类，即基础性数据标准和应用性数据标准。前者 主要用于在不同的系统间，形成对数据的一致理解和统一的坐标参照系统，是数据汇集、交换 及应用的基础，包括数据分类与编码标准、数据字典、数据资产地图；后者是为发挥平台功能 所提供的标准和规范，以保证数据的高效汇集和交换，包括元数据标准、数据交换技术标准、 数据传输协议标准、数据质量标准等。

(3)元数据。

元数据在数据资源目录、数据资产地图中的价值更大。 工业企业中的元数据种类繁多、形 式各异，对其进行集中管理后，需提供便捷的使用方式，才能发挥价值。通过用户数据视图、 元数据查询等功能提供的元数据细节信息，更适合于具体的人员、部门基于元数据开展相关工 作。对于企业中的管理人员等，由于这些用户的关注范围广泛，大量的细节信息不能有效满足 该类用户的需要。而数据资产地图一般用于在宏观层面组织数据，以全局视角对数据进行归并、 整理，展现数据变化情况、数据存储情况、整体数据质量情况等信息，适合为数据管理部门和 决策者提供参考。

(4)主数据。

主数据是企业核心的、最需要共享的数据。企业需要对主数据进行集中管理，并以服务的 方式把统一、完整、准确、具有权威性的主数据传送给企业内需要使用这些数据的操作型应用 系统和分析型应用系统。因此，对于主数据的管理，可以考虑运用主数据管理系统实现。在主 数据管理系统建设中，要从建设初期就考虑整体的平台框架和技术实现。

(5)数据质量。

数据质量将影响数据系统的应用程度。数据质量较差往往会造成开发出来的系统与用户的 预期大相径庭。数据质量也关系着分析型信息系统建设的成败。同时数据资源是企业的战略资 源，合理、有效地使用正确的数据能指导企业做出正确的决策，提高企业的综合竞争力。

数据质量管理包含对数据的绝对质量管理、过程质量管理。绝对质量即数据的真实性、完 备性、自治性，是数据本身应具有的属性。过程质量即数据的使用质量、存储质量和传输质量。 数据的使用质量是指数据被正确使用的情况。正确的数据如果被错误地使用，那么也不可能得 出正确的结论。数据的存储质量是指数据是否被安全地存储在适当的介质上，当需要数据时是

**314** 数据治理 工 业 企 业 数字化转型之道

否能及时、方便地取出。数据的传输质量是指数据在传输过程中的效率和正确性。

(6)数据服务。

数据服务是指针对企业内部积累多年的数据，研究如何能够充分利用这些数据，分析、优 化行业业务流程。数据使用的方式通常包括对数据的深度加工和分析(包括通过各种报表、工 具来分析运营层面的问题),以及通过数据挖掘等工具对数据进行深度加工，从而更好地管理服 务。通过建立统一的数据服务平台可以满足针对跨部门、跨系统的数据应用。通过统一的数据 服务平台可以统一数据源，加快数据流转速度，提升数据服务的效率。

(7)数据安全。

由于企业中重要且敏感的数据大部分集中在应用系统中，因此数据安全至关重要。如何避 免数据泄露和被非法访问是非常关键的问题。数据安全管理主要解决的就是数据在保存、使用 和交换过程中的安全问题，主要包括数据使用的安全性、数据隐私问题、访问权限统一管理、 审计和责任追究、制度及流程建立、应用系统权限的访问控制。

**2.** **重视数据标准管理**

数据标准管理组织是企业建立的以推动企业数据标准化工作为目标，负责并落实开展数据 标准管理工作全过程的组织体系。数据标准管理组织的设置遵循数据资产管理组织体系的相关 规定，按照在数据标准管理中承担的工作职责，可分为数据标准决策层、数据标准管理部门、 数据标准工作组等。

数据标准决策层：是企业数据标准的最高决策组织，主要职责是组织制定和批准数据标准 规划，审核和批准拟正式发布的数据标准，协调业务和 IT 资源，解决在数据标准编制、落地中 的问题，推进企业整体开展数据标准化工作。

数据标准管理部门：是企业数据标准管理的日常工作组织，主要职责是组织业务和 IT部门 编制数据标准并上报决策层审批；根据业务需求开展数据标准维护工作；根据数据标准化规划 组织业务和IT 部门开展数据标准落地工作。

数据标准工作组：由业务专家和IT 专家组成，负责在编制和推进数据标准落地工作中解 决具体的业务和技术问题。数据标准工作组在数据标准管理部门的领导下开展工作。

在绘制企业数据标准框架体系时，要重点结合企业业务的实际情况，从数据标准化的角度 制定数据资产管理标准、数据指标标准等，要涵盖所有相关制度和流程的编制及发布。具体包 括如下几项。

第30章 数据治理顶层架构规划与设计 **315**

(1)数据资产管理办法。

此办法用于研究并构建数据治理理论框架，明确数据治理管理的组织体系和职责，统一和 规范数据治理各过程域的管理内容，助力企业提升数据活力，高效运营，挖掘数据资产的价值。

(2)数据标准管理办法。

此办法用于加强企业对数据标准的管理和执行，规范数据标准管理流程，提高企业各业务 部门/中心、各分/子公司及其所属单位和技术部门对数据定义和使用的一致性和准确性，夯实数 据应用基础，促进信息资源共享。

(3)主数据管理办法。

此办法用于提升主数据管理能力，规范主数据管理工作，明确各责任主体及其职责权利， 构建全面的主数据标准化管理、质量管理和安全管理工作体系，打破信息系统之间的交互壁垒， 实现重要基础数据的充分共享、高度复用，有效地为业务提供真实、完整、 一致、规范的基础 数据，为数据管理和应用奠定基础。

(4)数据指标管理办法。

此办法用于确保数据指标治理工作有效开展，明确数据指标管理的组织体系及职责，规范 数据指标管理机制和流程，持续推进“体系化、规范化、指标化”的信息化建设，实现信息化 发展的量化管理和有效监控，促进信息化水平的提升。

(5)数据运维管理办法。

此办法用于建立数据运维标准体系、明确数据运维管理组织体系及其职责，规范数据采集、 数据处理、数据存储等过程的日常运行及维护，保证数据平台及数据服务的正常运行。数据运 维覆盖数据资产管理运维的全生命周期，确保数据资产管理运维工作人员在实际工作中“有法 可依、有制可循”,保障数据运维各项工作的有序开展。

**3.** **强化数据质量管理**

企业可以通过研究、参考 DAMA 数据管理框架、“数据质量管理十步法”等数据质量管理 理论，借鉴行业实践经验，同时通过现状调研和资料解读等方式，立足企业数据管理现状，设 计切合企业实际情况的数据质量管理框架。

结合企业数据质量管理提升方向，可以从管理组织、管理制度、规则框架、工作考核4个 方面，规划数据质量管理总体框架，推动数据质量的检测和提升。

**316** 数据治理——工业企业数字化转型之道

(1)管理组织。

数据质量管理组织架构应与数据治理体系建设同步。建议数据质量管理组织架构遵从数据 治理体系的组织架构设计，由决策管理层、组织协调层、执行层组成。决策管理层是数据治理 委员会，由网络与信息安全领导小组代行其职责；组织协调层是数据治理办公室，由网络与信 息安全领导小组办公室代行其职责；执行层包括业务部门(数据主题管理部门、使用部门和录 入部门)、下属单位和信息部门。

(2)管理制度。

通过数据管理制度流程的建设，可以将数据管理的理念与工作内容融入企业现有的制度管 理体系中，以保障数据管理工作的长效运转。为指导和规范数据质量管理工作的开展，需要从 办法、细则、流程和模板3个方面对企业内部质量管理过程、行为进行规范，制定数据质量管 理制度框架，从而提供数据质量保障。同时，数据质量管理制度框架作为数据管理的一部分， 需要从整体上与企业现有的数据管理相关的政策、制度规范、管理细则保持一致，遵循企业已 发布的数据管理制度、标准和细则，并根据数据质量管理的特性进行有针对性的细化和完善。

①办法。

办法主要包括章程、管控办法和考核机制。章程类似于企业的管理条例。该章程阐明数据 治理的主要目标、相关工作人员及职责、决策权力和度量标准。管控办法是基于规章制度与工 具而制定的可落地的操作办法。考核是保障制度落实的根本，要建立明确的考核制度和相应的 针对数据治理方面的考核办法，并与个人绩效相关联。

例如，数据质量管理办法主要明确数据质量管理组织架构，包括参与部门及各部门在数据 质量管理工作中承担的角色与职责、数据质量管理内容和步骤，并附有相应的流程和模板。

②细则。

细则是指对管理办法的细化。

例如数据指标标准定义细则是对数据指标标准的定义进行指导和约束，包括数据指标标准 定义及框架、数据指标标准分类和属性说明。

③流程和模板。

流程包括流程目标、流程任务、流程分级。可以根据数据治理的内容建立相应的流程，且 遵循本企业数据治理的规则制度。在实际操作中可结合所使用的数据治理工具，建立符合企业 的流程和模板。

第30章 数据治理顶层架构规划与设计 **317**

例如，数据质量管理流程和模板定义了数据质量管理各个方面的工作方法和步骤。建立数 据质量管理各项工作的流程和相关模板，可以规范各项工作的开展，明确各个相关部门在数据 质量管理中的分工和协作关系。配套的流程和模板主要包括数据质量规则管理流程和模板、质 量问题处理流程和报告模板、数据质量问题分析模板、数据质量监控流程和模板，以及数据质 量考核流程和模板。

(3)规则框架。

数据质量规则是保障数据质量的基本标准，是进行数据质量校验、度量、考核等工作的前 提条件。表30-2-1 是一个数据质量规则框架示例。

**表30-2-1** **数据质量规则框架**

|  |  |  |
| --- | --- | --- |
| **序** **号** | **属** **性** | **属性说明** |
| 1 | 数据质量标准编号 | 数据质量标准的唯一编号 |
| 2 | 数据标准标号 | 数据质量标准对应的数据标准项的标准编号，应当存在于数据标准中 |
| 3 | 数据标准名称 | 数据质量标准对应的数据标准项的标准中文名称，应当存在于数据标准中 |
| 4 | 数据质量标准描述 | 从数据质量标准评估维度的分类角度，根据评估指标，描述数据质量标准的内容 |
| 5 | 归口业务部门 | 数据质量标准的责任部门，负责指定、维护与解释数据质量标准。归口业务部门 与人力资源部门保持一致。注：这里所说的归口业务部门，以数据质量标准对应的 数据标准信息项的归口业务部门为准 |
| 6 | 业务分类 | 将数据质量标准按业务属性与特征进行分类。参照同行业的实践经验，数据质量 标准中的业务分类共包含6个，分别是人力资源域、财务域、物资域、基建域、生 产域、营销域 |
| 7 | 归属数据主题 | 数据质量标准衡量的数据对象对应数据域下划分的数据主题，与数据标准保持一致 |

(4)工作考核。

数据质量工作考核是数据质量管理制度和数据质量标准能够实施落地的重要保障，因而需 要定期评估各单位、各部门的数据质量管理水平，提升相关人员的数据质量管理意识，以及各 单位、各部门的数据质量水平。

应从数据质量管理水平和数据质量健康水平两个维度进行考核，促进数据质量管理制度、 标准执行落地。其中，数据质量管理水平是从组织与推进、制度建设、工作流程3个方面进行 评价，侧重于各单位、各部门的日常工作和业务过程考核，属于定性评价指标，适用于数据治 理办公室对数据治理工作进行总体评价；数据质量健康水平侧重于各单位、各部门的结果考核， 属于定量考核指标，适用于各业务部门考核本部门数据质量的健康水平。

**318** 数据治理——工业企业数字化转型之道

**4.** **编制企业数据资产目录**

通过数据资产管理可以将数据规范管理和数据处理有机地融合，实现对具体资源数据的元 数据描述，支持利用标准化的数据接口及形式丰富的图表展示，工具快速定制各类数据资产应 用，配合数据资产的全面评估，实现数据资产的“三全”管理：全生命周期管理、全流程管理、 全景式管理。

通过编制数据资产目录，可查询、追溯、共享数据。

(1)数据资产目录编制方法。

工业企业的数据资产目录编制方法主要有两种。

①方法一：系统视角

该方法以企业目前的核心系统为主，将系统功能模块分类，按系统数据主题、实体定义信

息、实体分类信息、数据相关方信息、技术信息构建数据资产目录，如图30-2-5所示。



数据主票

一级主题域

二级主题域

三级主塑域(选填)

**数据相关方信息**

实体分类信息

数据来源系统

数据来源部门

数据消费系统

数蛋消费路门

实体定义信息

实体名称 实体描述

数据共享程度

数据块型

技术信息

数据源表表名

系统

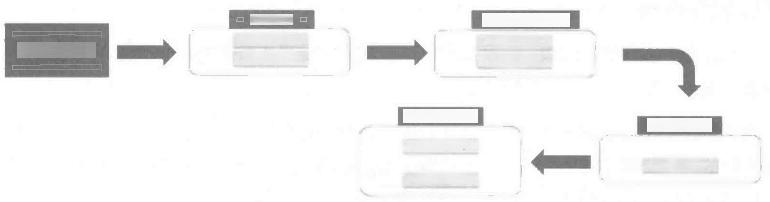
图30-2-5 数据资产目录编制方法——系统视角

②方法二：主题域视角

此方法首先构建企业统一的数据域主题(通常可按照战略发展、业务运营、管理支持抽象 一级主题域主题),将企业现有各系统按数据驱动方式，划分为各数据主题域。然后，抽象实体 定义信息、实体分类信息、数据相关方信息、技术信息构建数据资产目录，如图30-2-6所示。

哪

第30章 数据治理顶层架构规划与设计 **319**



数据实体定义信息

一级主题城

实体名称

二级主题城

卖体情述

数围相关方硫息

实体分关信息

对应的系统

数据类型

归属的部门

数据主题域

数据主题

**图30-2-6** **数据资产目录编制方法** **—** **—** **主题域视角**

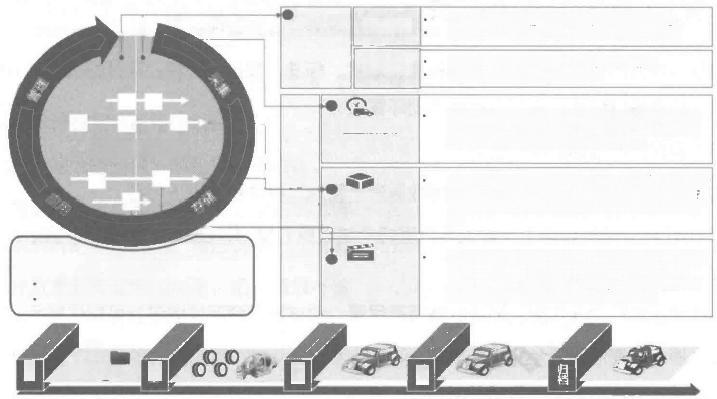
(2)数据资产目录可视化呈现。

从应用场景出发，利用数据治理工具，可以构建数据资产目录视图，使数据关系脉络化、

数据资产目录可视化。

根据数据目录的实体分布，可以构建数据资产地图，实现数据资产全生命周期、全流程、

全景式的可视化管理，如图30-2-7所示。



基础资产；未经过加工处理数据源、数据指标等 衍生资产：经过开发形成的新数据和数据指标等

业务领域：经营管理、生产营运、喜户服务 业务主题：采购、生产、销售等

基于元教据的数据资产全过程管理，采集、存储、应 用及管理过程的全记录与监控。权衡效率和需求之间 的关系，合理分级存储和保留、销毁数据

数 据

基于数据指标的数据资产溯源管理(数据血缘与影响分

析),数据来源，存储位置、处理方式、流转过程、安

全稽查规则，能追本潮源地发现所有资产的“前世今 生”

基于数据指标的资产全场景视图，从应用场景的堆度， 既有全局规划的管理者，也有关注细节定义的使用

者，还有加工、运维的开发者，提供多层次的图形化 展示，满足应用场景的图形查询和辅助分析

县

隆 7  型 蔡

元 数 据 管 理

数 据 血 缘 与 影 响 分 析

描述 、 管理 、 安全 、 稽查

资产 三全

数据资产管理

全生命周期

(时间)

数据资产分

类

全 景 式 (场景)

全流程 (空间)

产生方式

业务主题

管 理

**图30-2-7** **数据资产目录展示**

**5.** **规划数据治理平台**

企业通过规划数据治理平台，可以实现以下目标：

逐步建设数据资产库，实现数据资产整合与共享，推进基础信息共建与共享，加强安全、

生产、经营、管理等相关业务信息系统资源的横向整合和管理，建设企业统一的数据资产仓库

**320** 数据治理——工业企业数字化转型之道

及主数据管理平台。

打破原有“竖井式”的数据资产应用管理模式，实现异构数据的标准改造、实时交换、深

度共享、集中可用。

构建以企业核心数据中心为主，各核心板块数据中心为辅的两级数据中心，建立企业总部 和各板块的数据交换和共享机制，增加数据采集和存储的灵活性，提高数据灾备能力，保障企 业数据资产的安全。

数据治理平台要从以下5个方面来规划。

(1)构建统一的数据治理技术平台。

构建统一的数据治理技术平台包括以下内容。

实现基于统一云平台和大数据平台基础之上的数据治理平台，对企业战略发展、管理支持、 生产运营中的各类数据资源进行统一管理，逐步实现对大数据平台、数据仓库的统一管理。

开展数据资产全生命周期管理，提升企业的数据管控能力，营造良好的数据环境，支撑企 业的生产操作、经营管理、战略决策，推动企业数据价值的增加。

通过与数据治理的组织和职责、框架和模式、标准、规范、流程、审计配套，共同构建企 业统一的数据管理体系，实现对数据的闭环管理。

(2)构建全域数据汇聚平台。

构建全域数据汇聚平台，可以支持数据资产管理，具体包括以下内容。

对BW(商业信息仓库)、大数据平台进行全域数据汇聚，并在此基础上构建数据治理平台， 对企业数据进行系统性的数据治理。

通过数据资产管理功能，形成数据资产目录，通过数据资产地图进行可视化展示。

(3)规划数据治理平台的核心功能。

规划数据治理平台的核心功能包括以下内容。

建立符合企业现状的数据治理平台，确保实现对各类数据的集中管理。同时满足各类数据 管理的功能性和非功能性需求。

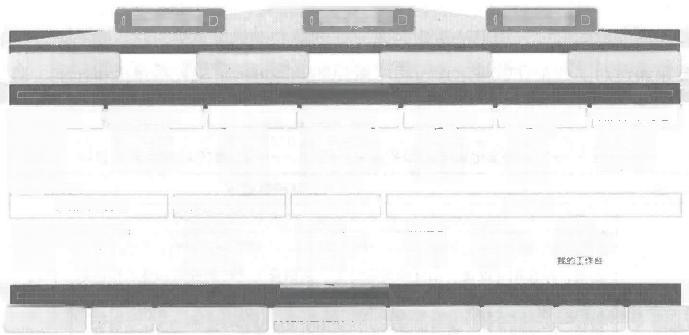
数据治理平台应包括数据标准管理、数据模型管理、元数据管理、数据资产管理、主数据

管理、数据指标管理、数据质量管理、数据安全管理、数据交换和服务管理、知识库管理等。

通过数据治理平台，提高企业的数据质量(准确性和完整性),保证数据的安全性(保密性、

完整性及可用性),实现数据管理的自动化、流程化、体系化，确保数据安全、准确，实现基础 数据的权威性、唯一性、准确性。

为实现数据治理总体目标，支撑数据管控工作，数据治理平台应具备如图30-2-8所示的功 能。



**数据治理总体目标**

救据质量提升

组织与职责 制度与流程 标准与规范 **常态与专项工作**

数据治理能力框架

**数据资产管理**

数据资产注册管理

数据被源目录管理

数据资产视压管理

数据资产施计分析

数期成本管理

数据价镇管理

**知识库管理**

知织库分型

如识库内者难护

内容审核发布

知识库检索

**数据治理技术平合**

**数据质量稽核库**

**数据指标管理**

指标数据库

指标体系

瘤标评价

海桐应用端标地图

**数据标准管理**

数据标准查询

数据标准准护

数据标准执行

数据标准分析

数据标准文精

数据标准导入导出

**数据安全词库** **元教城库** **主数据库** **知识库**

数据获取安全数据安全词库 般府数据识别数据院敏

统一认证 租户精离 角色认证 日志审计

**数据模型管理**

数据模理设计 模型差异棉核 数据模型变更

数据模型可视化

**数据质量管理**

质量需求管理

质量规则管理

质量任务管理

质量篮控分析

**主数据管理**

主数据整合

主数据偏码

主数据雄护

主数病分析

分发与共享

**元数据管理**

元模型地护

元数据采集

元数据雄护

元数搬分析

数据资产地图

数据采焦服务

数据交提服务

数据加工处理服势 数据共掌服务

数据变押服势

**数据安全管理** **数据交换和服务管理**

用户管理

功能菜单管理 系绕参数配置 公告管理

**数据标准库** **指标库**

角色管理

快捷蒸单管理 数据事典维护

权限管理

磨 内

日志管理 用户分析

**数据资产目录库**

数据资产可用

救据安全可

**系统管理**

**图** **3** **0** **-** **2** **-** **8** **数** **据** **治** **理** **平** **台** **功** **能**

(4)设计数据治理管理机制。

设计数据治理管理机制包括以下内容。

开展数据认责体系建设，成立数据认责管理组织，明确企业数据治理对应的责任主体及各 层级组织机构的职责分工、角色权限、人员配置与技能要求。

制定数据治理的管理制度、相关绩效考核制度、相关流程，明确数据认责定义，梳理企业 数据治理对应的责任主体。通过项目的实施，对公司管理人员、业务人员及IT 人员进行相关的 培训，提高全员对数据治理重要性的认识，明确数据资产是企业核心资产的意识。

(5)构建数据治理相关模型。

构建数据治理相关模型包括以下内容。

结合目前较为成熟、先进、稳定的模型算法，分别构建各专业及功能模块的数据分析模型 (如安全生产数据模型),推进大数据应用展现模式的创新。建立各专业业务视图，把 ERP及其 他专业系统内的关键数据通过数据集成平台进行获取，并在转换后存储在大数据仓库中，然后 进行多维度、即席式分析查询，以图形、图表、表格等多种展示形式进行综合展现。

**322** 数据治理——工业企业数字化转型之道

通过以上5个方面的规划，可以推进数据资产持续增值，驱动企业数字化转型，逐步改变 传统的以业务驱动数据建设的方式；通过以数据为中心的运营思维和管理思维的创新应用，逐 步实现以数据驱动企业的业务创新和效益提升。

**30.3** **成熟度评估**

数据治理顶层架构规划与设计在业内还没有完善的成熟度评估方法。本节所涉及的评估方 法，主要是通过对工业企业数据治理过程中常见的问题进行汇总和提炼而得出的，具体内容如 表30-3-1至表30-3-5所示。

**表30-3-1数据治理顶层架构规划与设计——数据管控域评估情况说明**

|  |  |
| --- | --- |
| **核** **心** **域** | **评估问题描述** |
| 数据管控 | 数据管理工作相关的岗位、职责不清晰，数据管理组织结构有待完善 |
| 数据管理工作缺乏统一授权，各部门间的责任和权力不清晰，跨部门间的数据管理工作协调 困难，存在相互推诿、相互冲突等情况，数据管理工作的成效和业务价值难以体现 |
| 数据治理工作缺乏企业层面的统筹管理，各部门分头开展数据管理工作，过于分散，无法开 展统一的协同工作，难以形成合力；存在不少的重复工作，造成资源浪费 |
| 企业各部门都发布了与数据管理工作相关的制度，但尚未形成企业级数据管理制度体系；数 据管理制度发布后，宣传和贯彻力度不强，且缺乏让制度得以落实的措施和手段 |
| 部门间数据管理工作的沟通渠道不通畅，工作流程不清晰，效率低下，数据管理工作难以得 到支持、认可和落实 |

**表30-3-2数据治理顶层架构规划与设计——主数据域评估情况说明**

|  |  |
| --- | --- |
| **核** **心** **域** | **评估问题描述** |
| 主数据 | 缺乏企业级的通用数据标准。跨专业数据标准不统一，同一业务对象在不同专业、不同系统 的数据不衔接、不匹配 |
| 实物编码标准不统一，实物编码在全生命周期管理的各阶段形成项目编码、物料编码、设备 编码、资产编码等信息，由于编码标准不一，导致资产信息难以关联匹配、跨专业数据共享、 应用困难等问题 |
| 主数据覆盖不足，应用不到位；已确定的主数据未在各专业进行有效应用，不同系统分散存 储，甚至重复存储同一主数据的情况仍然存在，部分主数据不一致、数据贯通困难 |
| 没有清晰定义核心主数据的权威来源，也未圈定核心主数据的范围，“一数多源”的现象较 为普遍，无法保证“一数一源”,存在数据不一致的现象 |
| 部分主数据没有明确的归口管理部门，未形成企业级的数据标准管理制度，无法指导数据标 准工作的开展，以及推进各项数据标准的构建与管理 |

**第30章** **数据治理顶层架构规划与设计** **323**

**表30-3-3** **数据治理顶层架构规划与设计——数据质量域评估情况说明**

|  |  |
| --- | --- |
| **核** **心** **域** | **评估问题描述** |
| 数据质量 | 未形成有效的数据质量管控体系，无法指导各部门开展数据质量管理工作 |
| 数据中心仅在数据接入及时性方面开展了数据质量的监控工作，对业务间数据的一致性、完 整性还未开展相应的质量监测工作，无法识别业务间数据的一致性问题，为跨业务域的数据分 析埋下数据质量隐患 |
| 缺少统一的数据质量工具支撑，各业务部门在数据质量管理工作中没有统一的数据质量工具 可使用，数据质量管理分散，不利于形成统一的数据质量管理习惯与数据质量支持库的沉淀 |
| 数据质量问题偏重于对数据生成后的清洗与校验，缺乏数据生成过程中的业务规则验证，导  致数据在源头就产生了质量问题 |

**表30-3-4** **数据治理顶层架构规划与设计——数据安全域评估情况说明**

|  |  |
| --- | --- |
| **核** **心** **域** | **评估问题描述** |
| 数据安全 | 数据安全相关的制度、条款以点状方式分布在信息安全管理办法中，未形成统一的、企业级 的数据安全管理办法，无法统一指导开展数据安全相关工作 |
| 没有明确的数据安全等级划分，业务部门无法确定对外共享的数据范围，数据需求分析方无 法获取分析所需数据，影响跨业务数据分析工作的开展 |
| 未识别敏感数据的范围及其在企业内的分布情况，对敏感数据的变更及流转缺少监控能力， 无法及时发现敏感数据变更与共享给企业带来的潜在危险 |
| 对外报送的数据缺少明确的企业级数据安全管理要求，可能导致潜在的数据安全风险 |

**表30-3-5** **数据治理顶层架构规划与设计——数据架构域评估情况说明**

|  |  |
| --- | --- |
| **核** **心** **域** | **评估问题描述** |
| 数据架构 | 数据模型应用的力度有待提升。人才储备比较单薄，难以支撑数据模型有效落地实施 |
| 数据架构 | 数据中心仍然以原业务系统为单位进行数据存储，并未在企业级数据模型的指导下对接入的 各业务系统数据按照主题域进行整合，不利于数据资产的开发和利用 |
| 尚没有企业级的、统一的数据资产目录，各部门无法准确掌握当前数据资产的现状 |
| 各业务部门之间以点对点的方式进行数据共享，存在“多头取数”的问题，没有充分利用当 前数据中心的数据资源和应用成果 |
| 数据与流程、岗位、系统之间的分布关系有待建立，需要进一步完善数据流转、交互与集成 的关系，并加强对数据分布的集中维护、定期更新、策略备份等管理工作 |
| 企业对数据集成与共享缺乏统一的管理，无法获得数据集成关系，造成数据重复、资源浪费 |
| 元数据管理分散在各个业务系统的数据模型中，缺少企业级统筹管理，需要针对元数据创建、 存储、整合、控制提供完整的支撑和管理架构 |

**324** 数据治理 字化转型之道

**本章精要**

本章从企业数据需求分析、顶层架构总体设计、应用场景验证3个阶段介绍了数据治理顶 层架构规划与设计的方法论。并从数据治理的核心域设计入手，系统地阐述了做好数据治理顶 层架构规划与设计需要关注的设计要点，为企业编制数据治理顶层架构规划与设计提供方法论 指导。

第31章

数据资产运营实施

**31.1** **实施内容**

数据资产运营是指对数据资产的确权、价值评估、共享流通、交易变现等活动进行管理的 过程(包括数据资产确权、资产购置、营销、服务、结算等),以及对数据资产成本管理(质量 评估、价值评估与定价等一系列评估)的分析、统计、监督等活动。目的是盘活企业数据资产， 通过数据开放、数据交易、数据合作等方式促进数据资产的流通和价值变现。

**1.** **数据资产成本管理**

数据资产成本管理是将数据资产成本计算方法落实到系统中进行管理，从而实现数据资产 成本自动计算，提高数据资产成本计算的效率。通过对数据资产进行成本分析，可以发现哪些 环节存在数据资产成本浪费的现象，并采取一定的措施，不断降低数据资产成本。

**2.** **数据资产质量评估**

数据质量的好坏直接影响数据价值的高低，所以，数据价值评估须结合数据质量评估结果。 利用演绎推算、内部验证、与原始资料对比、独立抽样检查、多边形叠加检查、有效值检查等 方法构建的数据质量评估模型，可以对数据的完整性、 一致性、重复性、及时性等进行检查， 并给出总的评估，最后形成数据质量评估报告。

**3.** **数据资产确权**

如果数据资产权属不明确，权益得不到保护，那么企业便不敢大规模地为客户提供数据服

**326** 数据治理——工业企业数字化转型之道

务。通过区块链技术的可追溯性，可以使数据从采集、交易、流通到计算分析的每一步记录都 留存在区块链上，无须第三方认证，即可确权、可溯源。区块链技术以点对点的方式实现安全 合规的数据交易，让每笔交易和数据都有确权证书。

**4.** **数据资产价值评估与定价**

数据资产价值评估是实现数据资产化的基础，也是数据资产定价和数据资产运营的基础。 要针对不同的评估目的和资产类别，设计适用的数据资产价值评估模型。评估结果分为货币化 和非货币化两种，从中既可以看到数据资产的价值，也能够了解构成数据资产价值的各部分， 从而可以分析影响数据资产价值的因素，以便于制订相应的数据资产增值计划。

企业要制定数据资产定价策略，并在系统中维护，还要设置相应的定价审批流程，严格控 制数据资产的价格，并将具体定价反馈至定价系统中，作为后续定价及数据资产价值评估的重 要输入。

**5.** **数据资产共享服务与计费**

数据资产共享是数据资产运营业务中最重要的内容，主要包括对内共享和对外开放两部分。

对内共享是指以最大化共享为原则，根据数据资产分类、分级及实际业务需要，在安全可 控的前提下，充分满足企业内部各业务部门的数据需求，范围涵盖明细数据、汇总数据及应用 数据。数据资产共享服务能方便用户申请、查询及使用所需数据。

对外开放是指对企业外部客户提供数据服务，包括数据封装、定价、上架、运行监控、计 费、下线管理等功能，支持数据租赁、合作等商务模式，并需要保证数据的隐私安全。

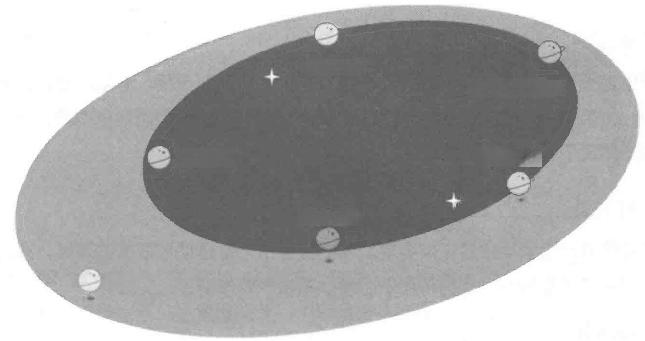
**6.** **数据资产运营分析**

通过采集数据资产运营过程中的数据，可以分析数据资产的处理效率，数据资产的热度、 广度及健康度，数据资产创造的价值及数据资产的投入产出比，以及数据资产共享服务情况、 数据资产经营效益等。

数据资产运营分析模块是数据资产管理价值的集中展示窗口。它既能让管理者看到数据资 产全景图，还能让管理者看到通过数据资产管理为企业节约了多少成本。数据资产运营分析模 块可以帮助业务人员快速定位问题、解决问题，也可以帮助技术研发人员提升研发效率、优化 任务、衡量自身价值。

**31.2** **步骤和方法**

数据资产运营项目的实施可分为5个阶段，如图31-2-1 所示。

**机会识别**

**现状评估与计划**

**评估复盘**

**项目执行**

**监测控制**

**常态运哲**

**图31-2-1** **数据资产运营项目实施**

**1.** **机** **会** **识** **别**

数据资产运营的目标是使企业能够满足其客户/用户的服务需求。企业应评估数据资产，并

根据客户/用户的需求和期望对其进行分析，从而挖掘数据资产管理的驱动因素和发展机遇。

数据资产运营应坚持以业务需求为驱动，以应用价值为导向。数据资产运营的内容不是越 全面越好，关键是挖掘并考虑如何满足业务需求、增加业务价值。该阶段的重点工作任务包括 解读业务战略、调研数据供需双方需求、制定数据战略并实施绘制路线图、明确利益相关方的 责任、制订相关资源计划、定义战略执行效果的评定标准等。

**2.** **现状评估与计划**

通过对数据资产运营能力成熟度评估，可以了解企业当前数据资产运营现状，识别数据资 产运营能力的不足之处，找准关键问题和目标差距，提出数据资产运营能力改进建议和方向， 规划未来数据资产运营管理路线图，制订数据资产运营管理行动计划，以便更好地利用数据资

产提高业务绩效。

**3.** **项** **目** **执** **行**

根据数据资产运营管理行动计划，企业可以建立相关数据资产运营专项管理项目，切实提

**328** 数据治理——工业企业数字化转型之道

升企业数据资产运营能力，其中包括但不限于企业数据资产全面盘点、数据资产目录构建、数 据资产成本核算、数据资产价值评估、数据资产开放与共享等能力，并以数据驱动业务为导向，

开发满足内外部客户需求的数据产品，提供相应的数据服务。

**4.** **监** **测** **控** **制**

企业要组建专门的检察和审计团队，制定一系列的策略、流程、制度和考核指标体系，来 监督、检查、协调多个职能部门的联合执法，从而优化、保护和利用数据资产，保障数据资产

在风险合规的情况下发挥最大的价值。

**5.** **评** **估** **复** **盘**

在完成数据资产运营项目后，需要事后评估数据资产管理项目取得的效果，分析数据资产

运营状况，总结经验和教训，从而有利于后续数据资产管理项目的实施。

**6.** **常** **态** **运** **营**

数据资产运营管理并不会伴随着项目的结束而结束。当项目完成后，需要将成果进行沉淀， 形成知识库，并完善相应的标准、制度和流程等。然后将数据资产运营管理转入常态运营工作， 变成日常例行的管理活动(包括数据资产目录日常维护、数据资产活性分析、数据资产安全管 控等)。企业只有在日常工作中维护有效的数据资产目录、保证数据的质量、规避数据安全风险 等，才能真正有效地挖掘企业数据资产价值，提升企业的竞争力。

**本** **章** **精** **要**

数据资产运营是数据资产实现价值的重要过程。本章主要从数据资产成本管理、数据资产 质量评估、数据资产确权、数据资产价值评估、数据资产服务与计费、数据资产运营分析等方 面阐述了数据资产运营的重要内容，同时给出了数据资产运营的实施方法，为企业实施数据资

产运营提供了一条可行的路径。

第 3 2 章

主数据管理实施

主数据管理的实施不是通过搭建一个主数据管理平台就能达到的，而是一项长期、复杂的 工程，涉及主数据管理体系、主数据标准、主数据管理平台、主数据质量和安全、相关系统的 升级及改造、数据清洗等多个方面。在项目实施中需依据快速见效、急用先建的思路，先整体 规划，以主数据模型和主数据标准为基础，以主数据管理平台为载体，来开展主数据管理专项 工作，确保主数据管理项目的成功。

**32.1** **实施内容**

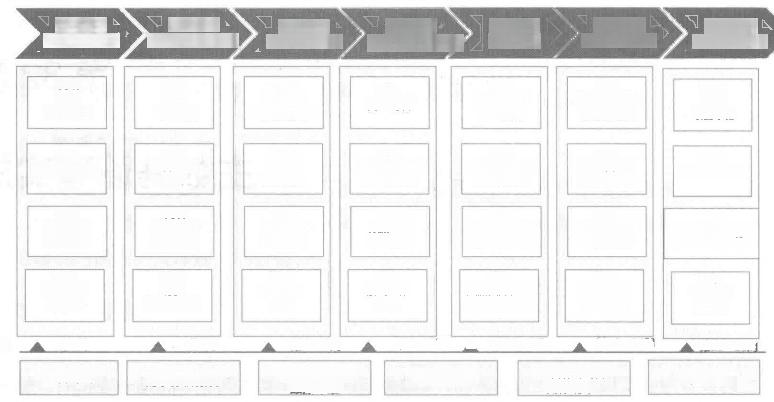
主数据管理实施的内容主要包括成立主数据管理项目建设期组织机构、调研主数据管理现 状、进行主数据识别、制定主数据标准、编制主数据管理办法、搭建主数据管理系统、集成主 数据管理系统和业务系统、建立运维期组织机构、规范主数据内容等。其中制定主数据标准是 基础，规范主数据内容是过程，搭建主数据管理系统是技术手段，建立组织机构和流程是前提 和保障。

**32.2** **步骤和方法**

**32.2.1** **实施步骤**

主数据管理实施一般包含项目准备、现状调研与分析、标准体系构建、主数据平台搭建、 数据清洗、数据服务集成、运营体系建立7个阶段，共28个实施步骤，如图32-2-1所示。

**330** 数据治理——工业企业数字化转型之道



第三阶段 标准体系构建

数

蓝图规划设计

主数据

标准制定

T11.

数据指标

标准制定

T12.

数据校型

设计

第7-14周

蓝图规划报告 数据标准

|  |  |  |
| --- | --- | --- |
| T14.  数据治理平台  客户化开发  数理  平台实施 | T18.  数振清洗规则  和标准建立  T19.  数据清洗、处 理和数据确认 | T22.  数据平台集 成接口开发  T23.  试点系统改造 和接口联调 |

T03.

标杆企业

对标分析

T04.

召开项目

启动会

时间 第2周

项目计划

|  |  |  |
| --- | --- | --- |
| T16.  系统整体  测试 | T20.  数据入库、理 立标准代码库 | T24.  系统上线及 用户培训 |

T06. 业务需求

调研

T07. 信息系统

需求调研

T08.

现状

和需求分析

T27.

制定数据切换路 径和推广应用策略

T28.

运维和项

目验收

第7~14周 A

需求规格书，

第6周

需求分析报告

第四阶段

主数据平台搭建

第二阶段

现状调研与分析

T21.

集成架构设计 及方案确认

第七阶段

运营体系建立

第一阶段

项目准备阶段

主数据代码库 图标清单

第大阶段 数据服务集成

T17.

数据清洗与

治理方案

T01.

制定

项目章程

T02.

组建项自团队

T26.

组建项目团队

第五阶段 数据清洗

T05.

制 定 调研方案

T25.

项制目定章程

T13.

需求确认

第15-24周

第15～24周

运营方案

第25-28周

T10.

图32-2-1 主数据管理实施步骤

**1.** **第一阶段：项目准备阶段**

项目准备阶段的主要工作是制定项目章程，即确定项目目标、实施范围、建设内容并制订 项目计划；确定项目的组织机构、人员配置、项目的组织资源管理及组织职责；选取组织架构、 业务范围类似的优秀企业作为标杆进行对比分析，取优补劣，为主数据项目建设提出方向；召 开项目启动会。主数据项目启动会非常重要，其既是动员会，又是分工会，同时也是培训会。 项目启动会是主数据项目良好的开端，项目启动会的顺利召开，可以起到事半功倍的效果，为 后续顺利展开工作奠定坚实的基础。

**2.** **第二阶段：现状调研与分析阶段**

现状调研与分析阶段的主要工作是制定落地、高效的调研方案，包含调研计划、访谈提纲 及项目调研问卷等；通过资料收集、业务需求调研、关键用户访谈等方式了解当前主数据管理现状 和主要需求，发现数据相关问题和关注点；同时对企业现有信息系统中的主数据标准、主数据质量

现状进行分析，找出需求点；结合企业数据管理应用现状、需求和存在的问题进行差异化分析。

**3.** **第三阶段：标准体系构建阶段**

标准体系构建阶段的主要工作是主数据治理蓝图规划设计和主数据标准制定。其中主要包 含4个架构体系：主数据标准架构、主数据管控架构、主数据质量体系及主数据安全架构。主 数据标准包含业务标准(编码规则、分类规则、描述规则等)和主数据模型标准。

第32章 主数据管理实施 **331**

**4.** **第四阶段：主数据平台搭建阶段**

企业要搭建主数据平台，通过数据标准文本发布、主数据模型建设、主数据管理流程建设， 实现对主数据的创建、审批、发布、修改、归档等全生命周期管理，以提高数据质量和改善信 息共享现状。

**5.** **第五阶段：数据清洗阶段**

数据清洗阶段的主要工作是根据项目范围和对象制订数据清洗与治理方案；建立数据清洗规 则和标准；根据主数据标准对历史主数据进行清洗、排重、合并、编码，保证主数据的完整性、 准确性和唯一性；最后形成一套规范、可信任的主数据代码库，建立整体的标准代码库。

**6.** **第六阶段：数据服务集成阶段**

将主数据管理平台与各个目标信息系统集成，可以实现主数据的采集、分发等交互操作， 从而最终实现将主数据服务于业务应用。根据系统集成的整体设计，企业要实现不同信息系统 与主数据系统的集成应用，其中涉及接口策略配置、属性映射配置、分发/订阅条件设置、日志 跟踪管理、数据同步管理、系统联调测试等。

**7.** **第七阶段：运营体系建立阶段**

企业要建设主数据管理运营组织，制定主数据管理办法、维护细则、应用考核规范等管理 规范，还要建立主数据运维体系。其中主数据运维体系由组织、制度、流程、知识库、平台组 成。企业通过组织各种方式的培训和交流，可以有效地传递知识。关键用户和内部 IT 人员要全 职参加项目建设，在实践中形成企业自己的实施和运维团队。

**32.2.2** **实施方法**

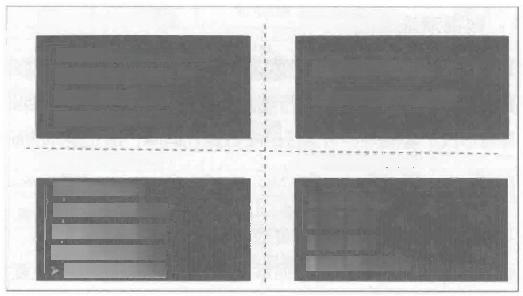
主数据实施应聚焦关键实施内容，选择合适的方法，按次序、分阶段合理有序地逐步推进。

**1.** **理需求**

通过现状分析及需求调研，企业可以对主数据管理的现状进行诊断分析。通过调查表格下发及 反馈、资料收集及分析、业务现状调研、关键用户访谈、信息系统数据应用现状摸查、对标标杆案 例研究和考察、公司知识库对比等过程，分析出用户对主数据建设的需求，以及与优秀企业的差距， 在数据标准、管理体系、数据质量、数据安全、数据全生命周期管理、数据平台应用等方面提出改 进建议。

**2.** **画蓝图**

在充分理解企业发展战略的基础上，企业可以根据调研分析及主数据管理能力评估结果， 按照系统的方法设计主数据蓝图。主数据蓝图主要包括4个架构体系：数据标准化架构体系、 数据管控架构体系、数据质量架构体系及数据安全架构体系(见图32-2-2)。



**数据标准化架构体系设计**

**数据标准化架构体系规划** **数据质量体系规划**

**平台建设路径规划**

→ **数据集成标准规范**

**数据管控架构体系设计**

> **数据管理模式**

> **数据管理制度制定**

**数据管理组织设计**

> **数据管理流程设计**

**数据管理工作考核**

**数据质量体系设计**

> **数据质量管理指标**

> **数振质量评价标准体系**

**数据安全架构体系设计**

**安全策端**

**安全组织**

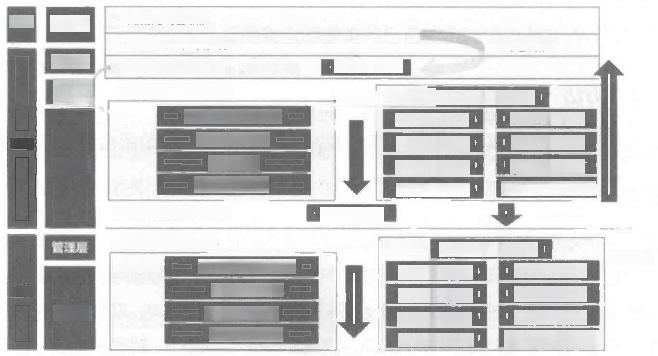
**安全技术**

**安全建设与运行**

图32-2-2 主数据蓝图设计中的4个架构体系

**3.** **定职责**

企业要建立主数据认责体系，还要确定主数据工作的相关各方的责任和关系，包括确定主 数据过程中的决策、管理、执行等活动的参与方和负责方，以及各方承担的角色和职责等(见 图32-2-3 )。



**集团信息化需理部**

**企业信息化管理要员会**

企业数据治理领导小组

企业数据治理额导工作组

业务部门| 数据治理业务主管

数据治理人员

业务专家

数据使用人员

子公司数据治理领导小组

子公司数据治理工作组

**数据治理技术主管**

业务条线

数振治理人员

数据录入人员

数据使用人员

|  |  |
| --- | --- |
| 数据标准人员  数据架构人员  生命周期人员  数振维护人员 | 数据质量人员  数概模型人员 数据安全人员  应用系练开发人民 |

散据标准人员 数据质量人员

政据架构人员 数据模型人员

生命周期人员 数据安全人员

政据维护人员 应用系统开发人员

组长及成员

日常治理 技术门政据治理技术主管

**组长及成员**

日常治理 技术条线

**指导晨**

**决策层**

**镇层**

数据治理业务主管

战略决策

组织协调

反馈沟通

**执行层**

**执行层**

领导监督

集团

**企业**

企业

图32-2-3 主数据认责体系

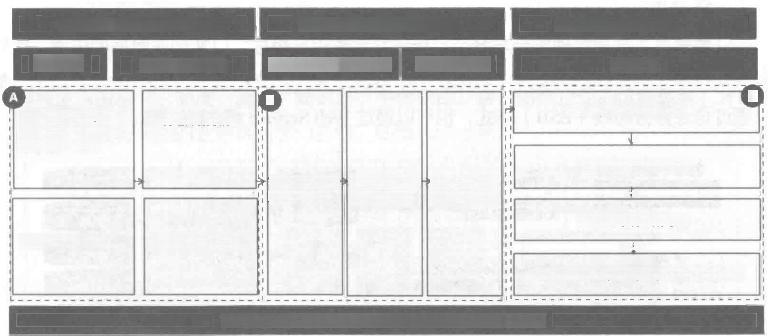
**4.** **定标准**

主数据标准是主数据管理工作的核心内容。通过主数据标准和数据指标标准，才能实现跨 组织、跨部门、跨流程、跨系统的数据集成和共享。主数据标准也是打通企业横向产业链和纵 向管控的数据基础。主数据标准一般包括主数据分类标准、描述标准、编码标准和管理标准。

**5.** **洗数据**

企业要按照主数据标准，梳理和清洗企业在经营活动中产生的各类数据，形成标准主数据 代码库。建立主数据代码库的过程是，按照一定的清洗规则对零散、重复、缺失、错误、废弃 的原始数据进行清洗，通过数据清洗保证主数据的唯一性、准确性、完整性、 一致性和有效性， 然后通过系统校验、查重及人工比对、筛查、核实等多种手段对主数据代码的质量进行检查， 以及通过数据清洗形成高质量的主数据代码库。

数据清洗工作分为3个阶段开展和推进，包括：①数据标准宣传培训阶段；②数据收集及 清洗阶段；③数据发布阶段，如图32-2-4所示。



阶段一：数据标准宣传培训阶段 阶段二：数据收集及清洗阶段 **阶段三：数据发布阶段**

工作启动 模板制定及培训 数据收集及规范化处理 专家评审阶段 数据发布

B

1.制定数据收集

模板

业

**3.生产环境数据导入**

业

**4.确认本次成果并制订下一步行动计划**

**组织与职责/数据工作程序/数据收集规范方法/数据标准规范**

**1.制作初始化导入模板**

工

**2.测试环境数据导入**

1.成立主数据管理组 织，明确职责，确认 专家评审人员。

.按照相同业 2.对数据进行 务合并数据 规范化处理

1.标准化专 家审核数据

2.组织宣传培训

2.制定清洗计划

C

图32- 2-4 数 据 清 洗 工 作 流 程

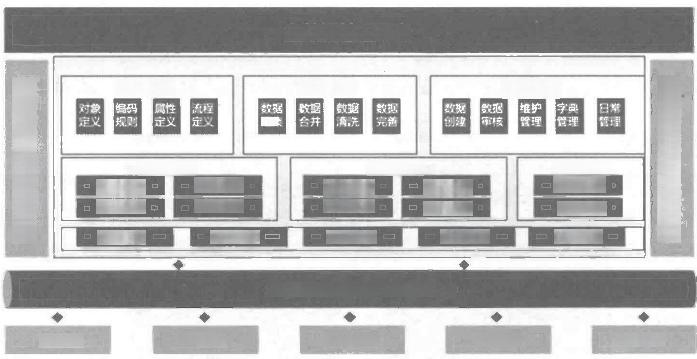
**6.** **搭平台**

企业要搭建主数据管理平台来发布主数据标准文本，实现主数据全生命周期管理、主数据 质量管理和主数据安全管控等。

主数据管理平台是企业数据规划、数据标准落地的载体，也是实现主数据统一标准、统一 规则的支撑，还是有效实施主数据全生命周期管理和数据服务的平台，以及实现数据从产生到

**334** 数据治理 工业企业数字化转型之道

应用，分层协同、全面治理的核心，如图32-2-5所示。



**主数据管理平台**

主数据管理核心系统(MDM)

数据整合

汇集

|  |  |
| --- | --- |
| 数据服务 | 基础功能 |
| *政据分发* 订阅服务 | 用户管理 日志管理 |
| 查询服务 异常处理 | 权限管理 系统管理 |
| 排重引摩 流程引肇 | 检索引擎 分发引擎 |

**企业服务总线(或共事服务接口)**

OA 系统 财务系统 人力系既 **ERP系统** **其他系统**

标准管理

祝准维护 标准查询

高速缓存

主数据标准体系

主数据管理体系

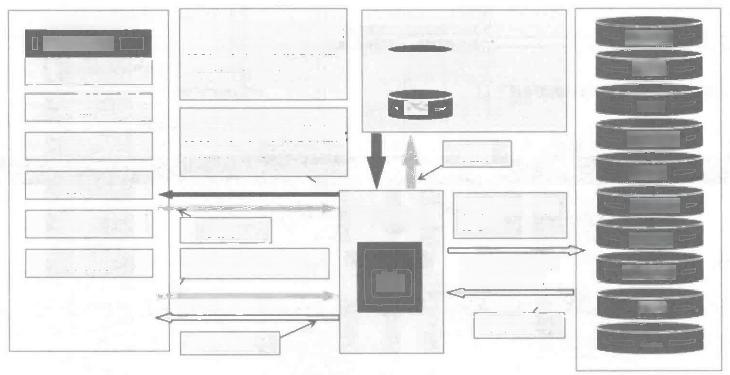
**数据建模**

**数据管理**

**图32-2-5** **主数据管理平台的功能架构**

**7.** **接** **服** **务**

企业需要将主数据管理平台与各个目标信息系统进行集成，以实现主数据的申请、审核、 分发等交互操作，从而最终实现主数据在多个系统之间的共享和统一，如图32-2-6所示。集成 可以通过企业服务总线(ESB) 方式，也可以通过 Web Service 和 XML 方式。



|  |  |
| --- | --- |
| **通过Web** **Service** **将各业** **务管理系统基本信息传输** **大数据服务总线**  **通过Web Service 将各业** **务管理系统基本信息传输大** **数据服务总线** | **业务管理系统**  积 **SRM**  人员 **HR系统**  **回执** |

**客户**

**回执**

**定时提取主数据信息**

**ESB**

信息转换成XML 格式

返回日志

保存日志

ERP系统

WMs

MES

SRM 系统

I 系统

移动APP OA 系统

EAM系统

LIMS

**主数据管理平台**

**主数据管理**

**物料**

**产品**

**人员**

传输主数据到目 标系统

信息转换成目标 系统接收格式

**供应商**

**组织机构**

**数据交换总线**



图32-2-6 主数据管理平台集成架构

**8.** **建体系**

主数据管理平台上线运行后，企业需要成立数据标准化运营组织，明确各岗位的职责，结 合 企 业 的 实 际 情 况 制 定 主 数 据 管 理 制 度 、 管 理 流 程 及 维 护 细 则 ， 以 及 建 立 主 数 据 运 维 体 系 ， 为

主数据的长效、规范运行奠定坚实的基础，如图32-2-7所示。



总体介绍该管理办法的内容、指导原则、 适用范围等

介绍该管理办法执行的组织和管理职责

介绍该项管理涉及的所有流程、制度；

对各个流程进行详细的介绍，并通过附件 的方式对涉及的流程图及岗位职责进行明 确，以便该项管理办法能够落地实施

对该项管理办法涉及的系统进行介绍

附则部分

主数据管理总则

主数据标准管理组织与职责

主数据标准的制定

主数据标准的评审

主数据标准的发布

主数据标准的执行

主数据标准的变更

主数据标准的复审

试行主数据标准的管理

主数据标准管理平台和版本管理

附则

指导原则

组织与职责

**管理流程**

及

**工作模板**

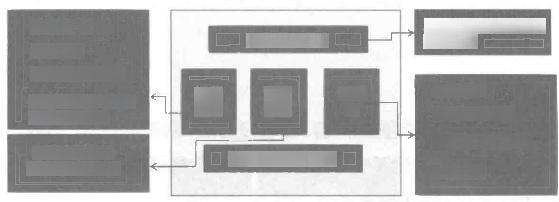
技术规范

附则及附录

**图32-2-** **7** **主** **数** **据** **管** **理** **制** **度** **规** **范**

运维体系由组织、制度、流程、知识库、平台组成。建立适合企业的运维体系，才能确保

主 要 的 数 据 管 理 有 对 应 的 业 务 牵 头 部 门 负 责 ， 如 图 3 2 - 2 - 8 所 示 。



总部-分/子公司二级 组织保障

维护细则

**提报审核指南**

**常见问题100例**

**规范用语**

**数据管理办法**

维护细则

应用和考核规范

应急预案运营管理 关键指标考核

各类主数据和 数据质保护流程

数据运维组织

运维

流程

**数据治理管理平台**

知识 库

运营 制度

图32-2-8 主 数 据 运 维 体 系

**9.** **促应用**

主 数 据 应 用 管 理 是 保 障 主 数 据 落 地 和 主 数 据 质 量 非 常 重 要 的 一 环 。 主 数 据 应 用 管 理 主 要 包 含 明

确 管 理 要 求 、 实 施 有 效 的 管 理 、 强 化 保 障 服 务 ， 以 及 转 化 和 切 换 存 量 系 统 主 数 据 代 码 等 内 容 。

**336** 数据治理——工业企业数字化转型之道

(1)明确管理要求。

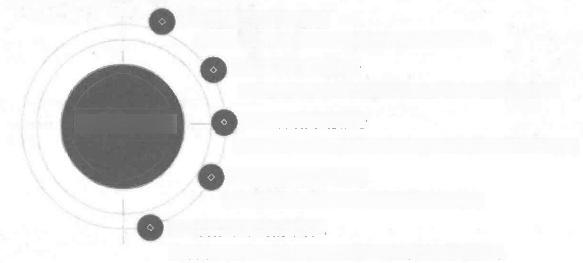
企业要制定主数据应用管理制度规范，对主数据的应用范围、应用规则、管理要求和考核 标准做出明确规定，并以此为依据，对主数据应用进行有效管理。

(2)制定主数据切换方案和推广应用策略计划

企业要按照已建系统、在建系统、待建系统3种情况制定系统切换方案，使项目成果在未 来2~3年能在企业的各层级信息系统中得到全面的应用和推广。

(3)主数据的5种主要应用方式

主数据的应用方式主要有5种，如图32-2-9所示。



**1.** **数据订阅/分发服务**

通过集成平台完成主数据与所有用友系统的数据交互

**2.** **主数据查询服务**

主数据将所有数据的查询页面封装成服务，供应系统调用

**系统集成方式** )**3.** **主数据申请服务**

主数据将所有数据的查询申请功能封装成服务，供应系统调用

**4.** **数据调用API服务**

指定调用类型，通过关键字查询数据详细信息

**5.** **数据公共数据资源池**

开放数据治理平台基础库只读权限，各应用系统可直接调用

图32-2-9 主数据的5种应用方式

**本章精要**

本章详细介绍了主数据管理实施的内容、步骤及方法，系统地讲述了现状调研、标准制定、 数据清洗、平台建设的过程，体现了主数据管理标准制定是基础，规范数据是过程，建设平台 是手段，建立组织和流程是前提和保障的核心思想。

规划 ： 设计

管 理

系统运维

第33章

元数据管理实施

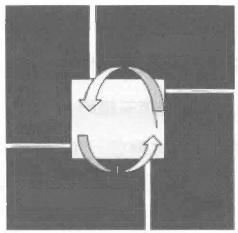
元数据管理的实施对企业具有非常重要的意义，可以协助企业管理数据资产，形成统一数 据地图，有助于解决企业的数据孤岛问题。元数据管理的实施也为各部门提供业务数据(如指 标、报表定义和统计口径),帮助业务人员理解和使用数据。因此，元数据管理的实施非常重要。

**33.1** **实施内容**

元数据实施的内容主要包括元数据的需求调研、定义元数据框架、建立元数据标准、搭建 元数据管理平台、创建和维护元数据、分发和交付元数据等。

**33.2 步骤和方法**

元数据管理的实施步骤主要分为以下4个阶段，如图33-2-1所示。

**业务分析**

**MetaData**

**系统设计、实现**

**图33-2-1** **元数据管理的实施步骤**

**338** 数据治理 教字化转型之道

**1.** **第** **一** **阶段：规划、设计阶段**

在规划、设计阶段主要是制定元数据管理战略。元数据管理战略是关于企业的元数据管理 目标的说明，同时也作为开发团队的参照框架。元数据管理战略的主要关注点是理解企业的关 键业务驱动力、存在的问题和信息需求，并达成共识。制定元数据管理战略的目标在于理解当 前环境是否能满足企业现在与未来的需求。元数据管理战略的目标定义了企业未来的元数据架 构，同时其中会规划分阶段的实施步骤，以实现企业的元数据建设愿景。

**2.** **第二阶段：业务分析阶段**

(1)需求调研和理解。

元数据管理战略必须反映对企业的元数据需求的理解。收集企业的元数据需求的目的包括： 确认企业需要元数据管理环境、设定元数据管理范围和优先级、指导元数据建模、指导建立元 数据内部标准、指导提供基于元数据的服务，以及预估和验证人员需求等。元数据需求是通过 与企业中的业务人员和技术人员进行沟通而获得的，并且会影响对特定人员的岗位角色、职责、

挑战等进行的分析。

(2)建立元数据标准。

企业要根据需求调研，并参照国际标准、国家标准、行业标准、共识标准，形成自己的元

数据标准。在此过程中，理解行业标准或共识标准对企业进行元数据管理实施是非常重要的。

(3)制定元模型。

然后企业根据确定的元数据标准对元数据进行分类，并确定各类标准的元模型。元模型的

定义应遵循标准化、国际化的 CWM (公共仓库元模型)。

元模型基本包括以下几类：

①管理类元模型：包括系统资源、人员管理、任务管理等元模型；

②技术类元模型：如关系型数据库、OLAP、 接口、ETL、 ERWin 等元模型；

③业务类元模型：如指标、KPI、 报表等元模型；

④编码规则：如信息分类及编码模型规则。

**3.** **第三阶段：系统设计、实现阶段**

(1)搭建元数据管理系统。

第33章 元数据管理实施 **339**

搭建元数据管理系统，可以实现以功能：

对元数据的采集、创建、存储、整合与控制等流程进行集合；

实现元数据版本控制追溯、元数据血缘分析、影响分析等；

为企业绘制数据地图、统一数据口径、标明数据方位、分析数据关系、管理模型变更。

(2)建设及管理元数据存储库。

元数据存储库形成的过程是在企业内部或外部采集并存储元数据的过程，即把从元数据来 源库中抽取到的元数据，与相关的业务元数据和技术元数据进行整合，最终存储到元数据存储 库中。元数据的抽取有多种方式：可以使用适配程序、扫描程序、桥接程序或者直接访问数据 存储库中的元数据。

(3)元数据的交付和分发。

元数据的交付和分发就是将存储库中的元数据分发到最终用户和其他需要使用元数据的应 用或工具中。

常用的元数据交付工具和方式包括：

①元数据内网：提供浏览、查询、搜索、报告和分析功能。

② 报告、术语表、其他文档及网站。

③数据仓库、数据集市和商务智能工具。

④建模和软件开发工具。

⑤ 消息传输交换。

⑥应用程序。

⑦外部组织接口方案。

**4.** **第四阶段：系统运维管理阶段**

对元数据的日常运维管理是保证元数据持续优化的基础，需要将元数据管理战略集成到元数据 管理的生命周期中，确保能及时收集变更过的元数据。企业要制定符合企业管理模式的制度和流程， 对元数据的增加、删除、修改等操作进行管理，还要针对业务元数据、管理元数据、库表结构 元数据等制定不同的运维策略。同时企业要开展元数据治理工作，建立正式的角色和职责并分 配专用资源(特别是在大型或关键业务领域)。元数据治理过程本身依赖于可靠的元数据，因此

**340** 数据治理——工业企业数字化转型之道

负责元数据管理的团队可以在创建和使用元数据的过程中对元数据管理原则进行验证和测试。

**本章精要**

对企业的技术人员而言，元数据管理的实施是通过将分散、存储结构差异大的信息进行描 述、定位、检索、评估、分析，实现了信息的结构化，为实现自动化处理创造了可能，从而可 以大大降低数据治理人工成本；对企业的业务人员而言，元数据管理通过对业务指标、业务术 语、业务规则、业务含义等业务信息进行管控，可以协助业务人员了解业务含义、行业术语和 规则、业务指标和影响范围等。正因如此，元数据管理已经成为很多大型企业数据治理项目的 重要组成部分。

第34章

数据指标管理实施

本章将介绍如何通过构建数据指标管理工具， 推进企业数据指标标准化工作落地

通过构建数据指标管理工具可以实现企业经营数据精细化，使数据指标“数出一家，政归 一门”,使企业能够“准确”“智能”“敏捷”地消费数据，让数据高效赋能业务，提升业务协同 和创新能力，充分挖掘数据资产的价值潜力。通过强化数据指标管理工具，可以在企业生产经 营等全领域、全过程中推行数据指标规范化、数据标准化和精细化管理，提高数据质量。

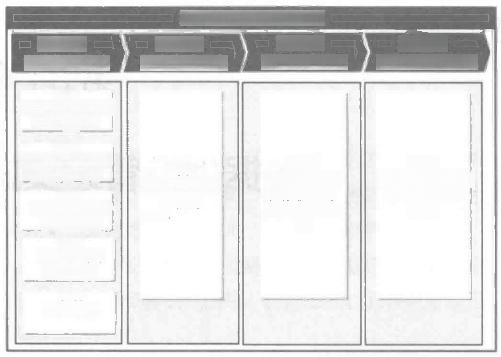
**34.1** **实施内容**

数据指标实施是一个系统工程，通常分成三期实现(见图34-1-1):一期为梳理实际指标； 二期为完善实际指标、抽象基本类指标；三期为配合平台落地。

其中的难点是收集和整理分散在企业各处的指标。企业要通过导出系统指标、搜集手工报 表指标、问卷调查和现场访谈等形式，尽可能全面地搜集数据指标，并按统一模板进行梳理、 归纳和标准化，形成完整的指标库。

抽象基本类指标

**342** 数据治理 — —工业企业数字化转型之道



第一期：梳理实际指标

第三步：

制订分类体系

T01.

系统指标梳理-分

析导出指标

T02.

收集、继承已编

制指标成果

T03.

问卷反馈填报

现场访谈

T04.

收集已编制

指标成果

T05.

指标成果

T08.

梳理形成实际指

标库清单属性 描述

T07.

在原先指标标 准规范之上， 形成指标库分

类

T06.

借鉴央企指标

库实践

√

第四步：

编制指标库清单

第二步：

领先实践研究

第一步：

各类指标收集

**图34-** **1-** **1** **数** **据** **指** **标** **实** **施**

第二期：

第三期：

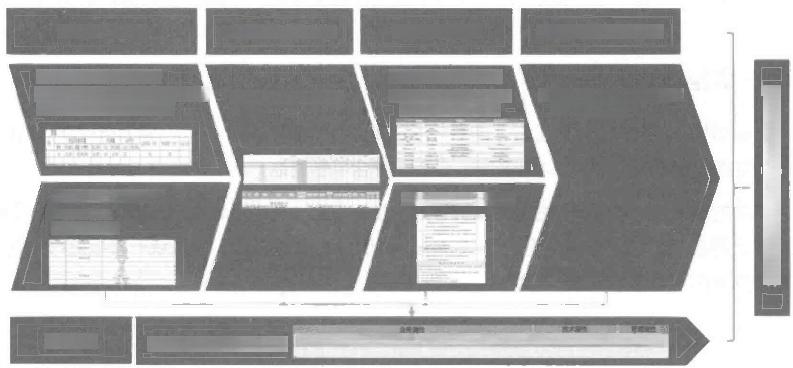
配合平台落地

完善实际指标

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | |  | |
| T11.  配合数据指标 平台实施落地 |  |
|  | T09.  完善一期梳理  的实际指标  T10.  基于实际指标， 抽象基本类指  标 |  |

**34.2** **步骤和方法**

各类数据指标收集的工作共包括5步，如图34-2-1所示。



**1** **梳理系统指标** **2梳理手工报表指标** **3收集访谈指标** **4收集已编制指标成果**

3-1调查问卷收集指标：

√调直问卷，资料收集等

形式收集指标及维值

**1-2梳理系统指标堆度：**

√通用维度；

√非通用维度

**模板梳理** **√统一按照模板进行梳理** 一生二章工名业期度文处P 度 + 位 名 主 了

**1-1梳理系统及报表相标；**

√导出集团统建系统中各类统计表。 √收集各类报表涉及的指标。

√借感，继承目前已编制的数 据标准指标成果

√收慧。梳理手工报表

指标

5 汇 总 指 标 ， 形 成 完 整 指 标 库

**3-2现场访读收集指标：**

**图34-2-** **1** **各** **类** **指** **标** **收** **集** **的** **工** **作** **步** **骤**

**1.** **第一步，梳理系统指标**

(1)梳理系统及报表指标。

首先导出企业各个系统中的各类报表指标，其中主要涉及两类指标。

第34章 数据指标管理实施 **343**

①系统产生的指标：集团总部统建的 ERP、CRM、SRM

②报表产生的指标：各类报表涉及的指标。

等系统中的指标；

(2)梳理系统指标维度。

对各系统中的指标维度进行收集、整理(例如图34-2-2 所示的例子),一般包括以下两种 维度。

①通用维度：包括6个基础维度(时间、空间、组织、业务板块、业务阶段、统计口径), 可凭经验及代码表获取；

②非通用维度：通过多种途径获取，包括但不限于访谈、分析数据等，并按重要性进行

排列。

|  |  |  |
| --- | --- | --- |
| 来源 | 维度名称 | 维度明细 |
| 物流管理部 | 客户类型 | 新客户 |
|  |  | 老客户 |
|  | 业务板块 | 合同物流 |
|  |  | 工程物流 |
|  |  | 化工物流 |
|  |  | 冷链物流 |
|  | 业务区域 | 东欧及中亚 |
|  |  | 东南亚 |
|  |  | 东亚 |
|  |  | 非洲 |
|  |  | 南亚 |
|  |  | 欧美澳 |
|  |  | 中东 |
|  | 车辆种类 | 轴线车 |
|  |  | 平板车+普通车 |
|  |  | 叉车+吊车 |
|  |  | 驳船+拖船 |

**图34-2-2** **梳理系统维度示例**

**2.** **第二步，梳理手工报表指标**

由于部分指标存在于手工报表中，所以对手工报表的收集、整理、分析也是一项必要的工作。

**3.** **第三步，收集访谈指标**

(1)通过调查问卷收集指标。

可以通过下发调查问卷、资料收集等形式，收集企业各部门、各板块、各服务支持中心要 求二级子分公司、三级单位填报的用来满足业务分析和管理需要的指标。

(2)通过现场访谈收集指标。

**344** 数据治理——工业企业数字化转型之道

可以通过现场访谈的方式，并根据访谈提纲引导各部门核心骨干回答相关问题来收集相关 指标，例如表34-2-1所示的例子。

**表34-2-1** **访谈示例**

|  |  |
| --- | --- |
| **序** **号** | **访** **谈** **示** **例** |
| 1 | 部门如何对数据指标进行管理和使用?(例如是否有专职/兼职人员管理指标?如何对数据指标内容 进行校验、审核? |
| 2 | 是否对指标进行了统一定义?如何发布、共享数据指标给相关部门 |
| 3 | 通常使用哪些信息系统?分别关注这些系统的哪些指标 |
| 4 | 经常编制哪些统计报表(含电子、纸质)和分析报告?分别关注哪些指标 |
| 5 | 您认为以上指标是否能涵盖所有业务范围，并达到业务监管的要求?如不能，请列举不在系统中但 实际需要的指标 |

**4.** **第四步，收集已编制指标成果**

在各类指标收集的过程中，要借鉴和收集已编、在编项目标准中的基本类指标，并收集项 目过程中的报表，以形成实际指标。

**5.** **第五步，汇总指标，形成完整指标库**

在将上述指标按统一模板进行梳理、汇总后，形成完整的指标库。

**34.3** **模板**

企业可根据自身业务的管理需要，按照指标体系框架进行三级指标的细化、完善和扩展。

指标体系扩展的基本步骤如下。

(1)各职能部门根据自身的管理需求，提炼数据指标，明确数据指标定义。

(2)依据指标体系分类框架，确定三级数据指标归属。

(3)详细定义数据指标，如数据指标项的来源、数据格式、计算公式、取数口径等。

(4)数据指标经审批后进行发布。

**34.3.1** **数据指标项定义**

数据指标包含各类统计数据的指标分类、指标代码、指标名称、指标定义、计算逻辑、指

标单位、分析维度、取数频率等。

●指标分类：针对企业业务将数据指标分为不同的类，如在一级指标下分为二级指标，在 二级指标下分为三级指标。

●指标代码：数据指标对应的代码。

●指标名称：数据指标的正式名称。

●指标定义：对指标进行的解释说明。

●计算逻辑：若为计算指标，则需列出计算公式。

●指标单位：指标所对应的计量单位。

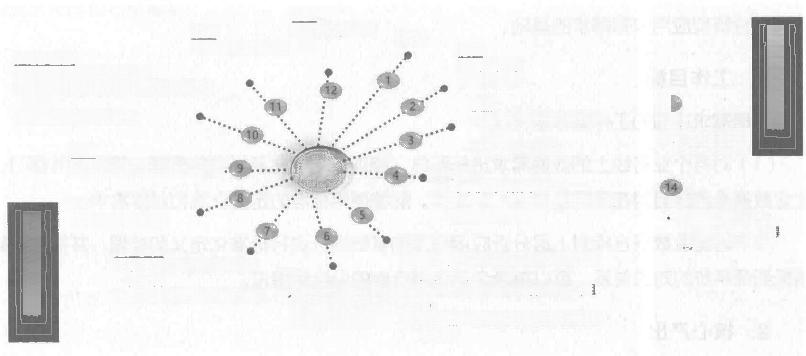
●分析维度：该指标主要的分析维度，通常包含物料维度、空间维度、时间维度、数据层 次等多个维度。

●取数频率：该指标在管理上需要提取数据的频率，单位可以为班次、日、旬、月、季度、 年等。

**34.3.2** **形成指标卡片及指标模板**

对数据指标进行梳理后，可形成指标卡片及指标模板。

指标卡片如图34-3-1所示。



**用途**

**对象** ·損标的管理用

·指标的使用对象 途”

·财务指标

·运营指标

**类别**

**·不含**维度与口径的指标 名称

·如：销售收入

特度定视角的指标颗粒

**·** **如**：月、季、年

**维度层次**

**·某类别指**标按不同维度的组合关系 与先后关系

·如：某地区某产品的销售收入(组 合]

·如：先按地区分析，再按产品分析 (先后)

**展示**

·图形图表等特定展现方或。

**频度**

**口径组合**

·对比分析、趋势分析等口 径组合

·如：预算数与实际数比较

·如：差旅费具体包括哪些种类的支

·旅费计算公式

国

来 源

· 指标数据来源 ( 哪 个 系 统 )

如：产业、客户、地区、品牌、 组织、项目……

口径

。预算数、预测数 。实际数，历史 数、对标数……

**定义与公式**

·指标的内涵与计算公式

摸标是丢工填报， 语是系统集成联得

信是的累级规

仕知的业务规

获得的方式

指标

属性

**名称**

**图34-** **3-** **1** **指** **标** **梳** **理** **—** **—** **指** **标** **卡** **片** **(** **示** **例** **)**

指标模板中包括指标代码、指标名称、业务定义、主题、数据需求提出者、数据使用者、 计量单位、上报频度、上报时间、计算公式等内容，如图34-3-2所示。

**346** 数据治理——工业企业数字化转型之道

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 据标  **代码** | 指标 名称 | 业务定义 |  |  | 蠢 名 器 生 级 | 国 查 | ：疆 | 發 舞 的 人 | 计解单位 | 数据 类型 | 单位  粒度 | 辅度 长度 | 上报 预度 | 报间  母上 | 时间 粒度 | 数据 展次 | 映射类型 | **计算**  公式 | **语**  **数挑**  **表**  **名称** |
| 1011  0019 | 。 | 报告期内 重整装置 计划完成 处理量 | 生产 | 覆道指群悲 ： | 要者置直 |  | 置能 |  | 吨 | 数字 | 集E团  公 司 / 级 司  疆 | 2 | 日/月 /年 | 月报表 每月3日 前/调度 日报统 计前一 天数据 | 月计划/ 年计划 | 业报  量  企上 | 接射  直映 |  | 炼化公司生产 指标完成月报 表l/调度日报 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**图34-3-2** **指标模板**

**34.3.3** **数据需求规划**

数据需求规划是对业务数据需求进行汇总、整理、指标化、定义标准化及指标分解，并将 其记录到企业数据仓库的基础指标映射关系中的过程。要在总部层面建立有效的数据共享和抽 取机制，从而为企业提供一套完整、统一、没有歧义的总部业务分析指标。

数据需求规划是企业数据仓库和分析型应用建设的基础工作，而企业数据仓库的建设，又 是满足分析型应用数据需求的基础。

**1.** **工作目标**

数据需求规划的工作目标是：

(1)对各个业务线上的数据需求进行汇总、整理、指标化及标准化定义，使各业务部门、 企业数据仓库项目组在建设相关应用系统时，能够使用标准的定义交流和对接需求。

(2)为企业数据仓库对上层分析应用提供的基础指标进行标准化定义和梳理，并指定基础 指标到源系统的对应关系，可以指导企业数据仓库的设计和建设。

**2.** **核心产出**

数据需求规划的核心产出包括以下内容：

(1)数据分析框架：用于指导企业分析型应用的分步建设。

(2)数据需求规划清单：用于提供经过汇总、标准化定义、无歧义的业务数据需求指标项。

第34章 数据指标管理实施 **347**

(3)数据标准清单：用于指导企业数据仓库基础指标的设计。

**3.** **主要收益分析**

数据需求规划可以为企业带来以下收益：

(1)提供部门及跨部门的业务数据需求，避免在建设分析型应用时产生不必要的重复建设 和资源浪费。

(2)提供业务数据需求的标准化定义，避免在跨部门对接业务需求时产生歧义和重复劳动。

(3)支持不同业务线、不同项目组在基于企业数据仓库建设各自的分析型应用时，对企业 数据仓库所能够提供的指导原则。

(4)明晰企业数据仓库基础指标到数据源的对应关系。

**本章精要**

本章详细介绍了各类指标收集及梳理的工作步骤，以及形成指标模板、通过数据指标平台 落地的完整过程。通过数据指标项目的实施可以对企业的各个业务线指标需求进行汇总、整理、 指标化、标准化定义，使企业各部门、项目组在建设相关应用系统时，能够使用标准的数据指 标定义交流和对接需求。

**第35章**

**数据质量管理实施**

科学合理地开展数据质量管理是发挥数据价值的基础，是实现数据业务应用的前提和保障。 但数据质量管理不是一蹴而就的，而是需要企业持之以恒地进行。这一工作将数据、数据拥有 者、数据管理机构、数据使用者、数据管理平台有机地整合在一起。依据数据在其生命周期的 各个阶段的特性，建立数据质量控制机制，可以及时发现数据质量问题，不断改善数据质量， 从而提升数据的可用性，实现数据更大的商业价值。

数据质量管理实施策略主要包括以下内容。

(1)严控数据的采集、录入、审核、传输、存储、清理、发布等方面的操作。

(2)统一数据口径，制定数据标准规范。

(3)确保数据在各个环节的一致性，避免数据指标体系相互冲突。

(4)做到一次采集，多处使用。

本章将以数据质量管理功能实施为核心，介绍其中的实施内容、步骤和方法。

**35.1** **实施内容**

数据质量管理实施内容包括如下4个方面。

(1)了解数据：通过业务部门和IT 部门之间的努力协作，评估数据内容、结构、字段、记 录，以及与系统之间的关系。

(2)设计数据标准：了解企业存在的数据及服务的需求，编制长期按计划施行的跨系统标 准定义和规则。

数据报告及记分卡

第35章 数据质量管理实施 **349**

(3)自动化数据质量：将标准纳入数据集成、迁移和实时流程；补充、扩充数据；使用第

三方提供的用户统计信息、企业和行业标准信息。

(4)监控数据质量：要评估结果并不断对规则进行系统性调整，以满足新的业务需求。

**35.2** **步骤和方法**

本节介绍数据质量管理实施的步骤和方法，其中详细介绍了数据剖析、数据质量诊断、数 据处理规则、数据质量优化、数据质量监管等，图35-2-1 为数据质量管理实施应用架构图。

数据管理员

IT专员



数据分析员

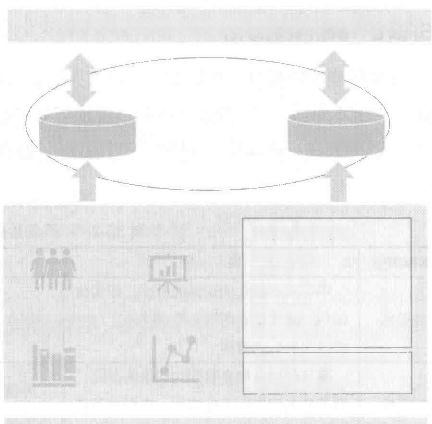
业务分析员

管理

部署

开发

发现



ETL,MDM,CRM,ERP,DW,BI,C++,Java,Apls

实时处理

数据质量处理程序

●数据验证

●数据转化

●数据清洗

●数据优化

●数据监控

规则，程序，实用工具，国 家，处理，内容

财务数据 地址/地理数据

群体数据 产品数据

批处理

数据剖析及数据发现

**图35-2-1** **数据质量管理实施应用架构图**

**35.2.1** **数据剖析**

数据剖析是通过剖析企业特定的数据集，挖掘元数据以获取统计信息，从而帮助企业全面 了解数据，并确定这些数据的可用性的过程。 一般数据剖析主要分析的内容有最大值、最小值 及其他基本统计数据；字段、数据类型和模式/格式的频率数；符合预期的值。其他高级数据剖 析还对字段之间的关系进行分析，如单个数据集的字段之间和单独数据集的多个字段之间的依 赖关系。数据剖析主要包括内容。

**350** 数据治理——工业企业数字化转型之道

(1)数据内容及背景分析。

(2)数据结构及路径分析。

(3)数据成分及业务规则合规分析。

(4)数据间关系及相关资源匹配。

(5)识别数据转化机制。

(6)建立数据有效性及准确性规则。

(7)校验数据间的依赖性。

**35.2.2** **数据质量诊断**

数据质量诊断是基于数据剖析的结果来审核数据的质量，用来发现数据可能存在的异常和 问题，为根本原因分析、数据纠错和错误预防提供基础。数据质量诊断主要从准确性、 一致性、 完整性、相关性、有效性、即时性等方面评估数据和质量，发现潜在的数据问题，如表35-2-1 所示。

**表35-2-1** **数据质量诊断维度**

|  |  |  |  |
| --- | --- | --- | --- |
| **指标类型** | **说** **明** | **衡量标准** | **备** **注** |
| 完整性 | 不存在或缺失的记录的占比。即实体  的每个属性都有明确的值，不存在“空” 或“未知”的属性 | 样本记录的空值率 | =空值记录总数/样本总记录数 ×100% |
| 相关性 | 满足外键参照完整性的记录的占比， 对于数据库中的某些实体，它们的存在 可能要依赖于其他的实体 | 外键无对应主键的样 本记录比率 | =外键无对应主键的样本记录总 数/样本总记录数×100% |
| 唯一性 | 满足主键唯一性约束的记录的占比， 即一个表中的一组属性的值是唯一的 | 1-主键的重复率 | =1-主键重复样本记录总数/样 本总记录数×100% |
| 有效性 | 满足在值域和数据有效范围内定义的 记录的占比，即实体属性的值要在用户 定义的有效范围之内 | 1-样本记录异常值比 率 | =1-超出值域的异常值样本记录 总数/样本总记录数×100% |
| 及时性 | 满足业务应用的时间要求的记录的 占比 | 满足时间要求的样本 记录的比率 | =满足时间要求的样本记录总数/ 样本总记录数×100% |
| 真实性 | 真实记录的占比，即数据库中的实 体必须与对应的现实世界中的对象是 一致的 | 样本记录中真实记录 的比率 | =1-样本记录中失真记录总数/ 总样本记录数×100% |

**第35章** **数据质量管理实施351**

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| **指标类型** | **说** **明** | **衡量标准** | **备** **注** |
| 精确性 | 满足业务精度需求的记录占比 | 样本记录满足业务对 精度需求的比率 | =样本记录中满足业务精度需求 的记录总数/样本总记录数×100% |
| 一致性 | 与其他系统(或者系统内部)一致的 记录的占比 | 样本记录与其他系统 的匹配率 | =1-样本记录中不同意义的记录 总数/样本总记录数×100% |
| 可理解性 | 含义明确和易于理解的记录的占比。 即表明记录本身的含义是否简单、明确 | 样本记录易于理解的 比率 | =1-样本记录中费解的记录总数 /样本总记录数×100% |
| 可用性 | 可获得、可满足业务使用的记录的占 比 | 样本记录可获得、可 满足业务使用的比率 | =样本数据可获得的记录总数/样 本记录总数×100% |

**35.2.3** **数据处理规则**

数据处理规则是基于企业本身的特殊需要而创建、维护的一套企业专属的业务规则。其适 用于企业长期的数据质量管理，并能随着企业的发展而添加、变更或删减某些业务规则，确保 企业数据符合决策需要。可以基于以下3个方面快速创建或提供开箱即用的数据处理规则，全 面规范数据呈现状态，促进企业进行有效的数据管理：

(1)数据属性业务规则(如字段);

(2)数据实体业务规则(如库、表);

(3)企业特定业务规则(如行业规范)。

**35.2.4** **数据质量优化**

数据质量优化是帮助企业将原本杂乱无章的数据转化为有价值的信息的过程。针对各种复 杂的企业环境，通过数据质量优化可以实时批量改善数据的可靠性、可用性及业务适用性，而 且这个过程对于企业中的任何数据系统或应用程序都适用。可以从以下5个方面优化数据质量， 输出有价值的资产。

**1.** **数据清洗与标准化**

数据清洗与标准化可以帮助企业统一和规范数据的各个方面，制定跨行业通用及行业定制 的数据标准，规范数据的采集、录入、传输、处理等过程，进一步更正、修复企业系统中的错 误数据，并对数据进行归并整理，从而使企业应用数据更方便，企业内的信息交流更高效，加 快数据变现的速度。

**352** 数据治理——工业企业数字化转型之道

**2.** **数据匹配**

通过使用在线数据匹配工具，可以轻松地把企业中不同来源的数据匹配到统一的编码下。

数据匹配可以帮助企业解决以下问题：

(1)在系统集成时进行数据合并与编码统一，保证系统的一致性；

(2)防止企业重复采购外界数据，减少采购成本；

(3)提取合适的数据用于企业营销项目，最大化营销效果；

(4)通过将新导入的数据与现存的数据进行匹配，减少重复数据的产生；

(5)帮助客户发现数据关联组织，在最大程度上控制风险。

**3.** **数据校验和补充**

基于企业对数据完整性的需求，可引用具时效性的第三方权威数据资源，如人口统计信息、 地理编码、邮政编码等。通过比对第三方数据与企业的原始数据，可以确保数据的完整性与有 效性。

**4.** **数据的查重/删重**

通过清理企业信息系统中近似重复的数据(如客户数据、供应商数据等),可以保证数据的 一致性和正确性，以使企业信息系统进行正确的决策。为了提高数据质量，必须要查找重复数 据和删除重复数据，确保企业可以高效地使用数据。

**5.** **关联与统一服务**

基于企业数据库的通用内容，企业可以对各个数据源进行数据关联，识别记录之间的相互 关系，进而确定关联规则以符合企业的特定需求，并通过数据关联获得最佳数据记录，形成统 一视图，让企业对数据对象有更深的认识。

**35.2.5** **数据质量监管**

数据质量监管主要应用于数据质量优化流程中。企业在完成数据质量优化完成后，还需进 行持续的数据质量监管(往往通过数据报告与记分卡来监管),以便及时管控数据质量。数据质 量监管主要用于以下情况：

(1)监控一段时间内数据质量的合规情况；

第35章 数据质量管理实施 **353**

(2)展示数据质量优化成果；

(3)及时检测数据质量问题；

(4)为数据质量投入成效提供依据。

**35.2.6** **实施数据质量管理时需注意的问题**

企业在实施数据质量管理时需注意以下问题。

(1)从各个业务管理主题对数据质量的需求出发，推动数据质量管理工作。

(2)把分工协作的原则作为设计数据质量管理组织架构、工作流程和沟通机制的基础。这 种分工协作的形式具体体现在如下两个方面。

①业务部门负责评估本业务管理主题范围内的数据集合在数据质量上的可用性，并提出数 据质量的要求；

②由数据质量管理团队牵头协调，协同技术部门和其他相关人员执行针对数据质量问题 的优先级划分/评估、解决方案的制定与选择、解决方案的实施、问题解决状态的审核等多项实 际工作。

(3)数据质量问题的解决情况必须根据业务的实际需要来衡量，根据统一的优先级评估机 制，分阶段解决已发现的数据质量管理问题。

(4)数据质量管理不能代替系统开发的测试工作。

(5)数据质量管理工作流程的起点是数据分析人员定期提交数据质量问题报告和系统管理 人员定期提交数据质量报告。

(6)随着质量管理工作的具体开展，可能会逐步增加新的工作内容，现有的工作内容可能 会有调整，并落实到流程的各环节中，例如：

①根据实际状况对流程的各环节的调整；

②更新后的数据质量标准、指标和参数的使用；

③数据质量问题解决方案的制定、选择、确认与实施；

④对各工作环节需要的数据质量问题报告的适当修订；

⑤数据质量改善状况评估等内容的逐步引入。

**354** 数据治理——工业企业数字化转型之道

**本章精要**

数据的质量问题从某个角度反映出企业存在的一些问题，而产生问题的原因可能是数据流 动，可能是业务流程，也可能是管理问题等。对数据质量问题的分析可以帮助企业找到问题的 源头。而高质量的数据对企业的管理决策、业务支撑都有极其重要的作用。

数据质量是数据治理中非常重要的一把标尺，而数据治理又是当今企业的首要战略之一， 只有持续地进行数据质量实施、改进，才能推动数据治理体系的完善，为企业数据战略提供坚 实的保障。

第36章

数据安全管理实施

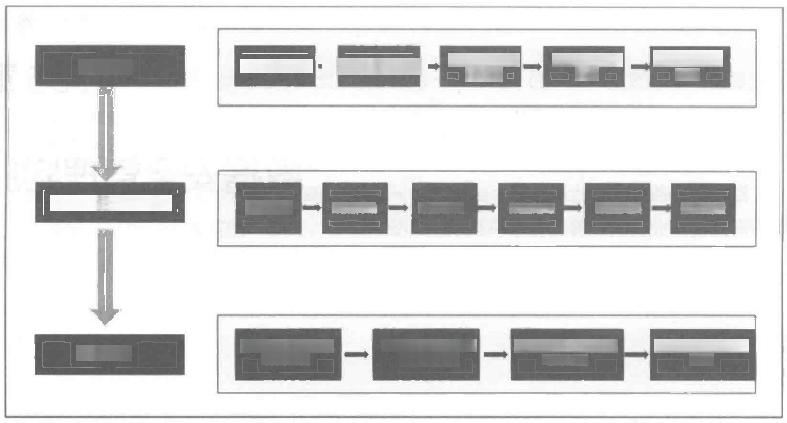
**36.1** **实施内容**

数据安全能力的构建步骤一般是“建立组织架构→梳理应用需求→梳理数据资产→ 引进数 据安全平台技术→建设数据全生命周期管控→支持数据应用→维护数据运营”等。数据安全管 理以数据本身为导向，分布在数据安全能力构建的多个环节中。本章主要围绕数据安全管理， 具体阐述其实施步骤、实践模式，以及数据安全管理成功的要素。数据管理能力成熟度不同的 企业开展数据安全平台建设的具体步骤和实施内容要根据自身情况而定。

**36.2** **实施步骤**

数据安全管理可按照统筹规划、数据全生命周期监管、稽核检查三个阶段来实施，每个阶 段对应的管理职能如图36-2-1所示。

**356** 数据治理——工业企业数字化转型之道



盘点数据流转

全过程

**数据全生命周期监管** 数据采集 数据传输 数据存砷 数据处理 数据交换 数据销吸

数据安全管理流程

检查和改善

灵活配置数据存储集 略的检查

数据安全管理定期 检查

标准制定，标准执 行的检言

评估数据管理 水平

制定数据标准 规范

制定管理制度

建立组织体系

**稽核检查**

**统筹规划**

**图36-2-1** **数据安全管理各阶段的管理职能**

**1.** **第一阶段：统筹规划**

数据安全管理实施的第一阶段是统筹规划的过程，即制定数据安全管理战略规划，明确数 据安全管理目标。该阶段的成果是后续工作的基础。

此阶段的工作一般分为以下三步。

第一步，建立组织体系，并根据企业自身情况制定数据安全管理制度。即建立一套独立、 完整的关于数据安全管理的组织机构，明确各级角色和职责，确定兼职/专职人员，保障数据安 全管理的各项管理办法、工作流程的实施，从而推进工作的有序开展，并逐步打造管理及技术 方面的专业人才团队。

第一步的主要交付物包括《数据安全管理规划》《数据安全管理认责机制》《数据安全管理 工作指引》《数据安全管理考核评价办法》。

第二步，结合业务盘点数据流转全过程，评估企业当前的数据管理水平。盘点数据流转全 过程是开展数据安全管理工作的前提之一，企业需要分析企业战略及业务现状，盘点企业内、 外部数据现状，确立数据安全管理目标，并逐渐实施需求调研、数据流程盘点、数据采集及汇 聚等专题任务。与此同时，企业还要了解企业数据来源、数据采集手段和硬件设备情况，以评 估自身数据安全管理能力，规划未来数据安全管理能力成熟度的提升方案。

第二步的主要交付物包括《数据流转全过程盘点清单》《数据安全管理现状评估》等。

第36章 数据安全管理实施 **357**

第三步，制定数据标准规范。企业在建立组织架构、制度体系和盘点数据流程的基础上， 结合国际标准和行业标准，围绕数据资产全生命周期管理，制定相关的数据规范体系(包括数 据采集认证和风险评估标准、数据传输加密标准、数据存储标准、数据授权与使用标准、数据 安全共享交换标准、数据销毁标准等),可以使得数据管理人员在工作中有明确的规则可依。企 业应逐步推动相关数据规范和标准的工作建设，使数据可以有效地汇聚和应用，切实保障数据 安全管理的顺利实施。

第三步的主要交付物包括《数据全生命周期管理标准办法》等。

**2.** **第二阶段：数据全生命周期监管**

如果说第一阶段的重点在于对数据流转全过程的定义、规划、梳理，那么第二阶段的重点 就是对第一阶段得到的成果的落地实施。

首先，在搭建数据安全管理平台、完成数据汇聚工作的基础上，根据企业自身存量数据基 础和增量数据预估，建设或采购必要的数据安全工具，或引入第三方工具以支撑数据管理工作， 切实建立起企业数据安全管理能力。

其次，建立数据安全管理体系，防范数据安全隐患，执行数据安全管理职能。在第二阶段， 需要从数据安全管理的相关业务、技术部门的日常工作流程入手，切实建立起企业数据全生命 周期管控能力，包括从业务角度梳理企业数据，检测数据标准实施情况，保证数据标准规范在 企业信息系统生产环境中真正得到执行。针对关键性数据安全管理工作，可以借助管理工具， 建立数据全生命周期的管理流程，保证相关工作都有专人负责。

第二阶段的主要工作目标是为企业打造核心的数据安全管理能力，同时为企业中的数据安 全管理部门形成数据管理的工作环境，实现企业数据安全的可管理、可落地。

本阶段的主要交付物包括《数据安全管理办法》《数据安全管理实施细则》等(包括数 据采集管理、数据传输管理、数据存储管理、数据处理管理、数据交换和共享管理、数据销毁 管理等)。

**3.** **第三阶段：稽核检查**

稽核检查阶段是保障数据安全管理实施阶段涉及各管理职能有效落地执行的重要一环。在 这个阶段主要检查数据标准执行情况、监管数据生命周期等。

另外，在此阶段需要抓好四个“常态化”。

**358** 数据治理——工业企业数字化转型之道

一是对数据安全标准管理检查的常态化。数据安全标准管理是企业数据安全管理的基础性 工作。通过数据安全标准管理的实施，企业可实现对数据的统一运营管理。对数据安全标准管 理的检查主要从标准制定和标准执行两个方面进行。标准制定的检查主要围绕与国家标准、行 业标准的一致性，同时参考与本地标准的结合性。标准执行的检查主要围绕标准的落地情况， 包括数据安全标准的创建和更改流程的便捷性、数据安全标准使用的广泛性。

二是对数据安全稽核的常态化。企业要建立一套良性循环、动态更新的数据安全管理流程，制定符 合业务目标的稽核规则，明确在数据全生命周期管理各个环节中提升数据安全的关键点，持续评估和监 督数据安全与数据安全服务水平，不断调整、更新数据安全管理程序，以保障企业的数据安全。

三是灵活配置数据存储策略的常态化。数据生命周期管理的目标是完全支持企业业务和服务的需求， 根据数据对企业的价值对数据进行分类分级，形成数据资产目录，然后制定相应的策略，以最优的服务 水平和最低的成本，将数据转移到相应的存储介质上，争取以最低的成本提供适当的数据保护、复制和 恢复措施。企业借助数据生命周期管理，不但能够在整个数据生命周期内充分发挥数据的潜力，还可以 按照业务要求快速对突发事件做出反应。

四是数据安全检查的常态化。在大数据时代，数据资产更容易面临被泄露、篡改、窃取、 毁损、非法使用等风险。企业应建立数据资产及相关信息系统的保护体系，合规采集数据、应 用数据，依法保护客户隐私，提高员工的数据安全意识，以及定期进行数据资产安全检查，以 保证数据的完整性、保密性、可用性。

本阶段的主要交付物包括《数据安全管理稽核办法》《数据安全管理问题管理办法》等。

**36.3** **实践模式**

工业企业的数据安全管理一般都参照“统筹规划→数据全生命周期监管→稽核检查”三个 阶段来实施。另外，工业企业还可以选择两个维度进行实践：其一是组织模式，包括自上而下 的顶层设计模式和自下而上的各个击破模式；其二是切入模式，包括生产系统优先模式和数据 系统优先模式。

**36.3.1** **数据安全管理的建设策略**

工业企业数据安全管理的实施策略主要包括自上而下和自下而上两种模式。如果企业将数 据安全管理纳入自己的战略规划中，且企业高层领导拥有较大的决策权，那么可以采用自上而 下的实施策略，并结合企业业务的发展目标制定长远的数据安全管理规划。如果企业的数据管

第36章 数据安全管理实施 **359**

理部门具有一定的独立性，并且具备专业技能和相关经验，那么可以采用自下而上的实施策略， 以探索数据安全管理需求为驱动力，推动企业数据安全管理的逐步完善。

自上而下的实施策略的显著特点是“规范、标准先行”,项目通常由数据安全管理咨询项目 开头，这种模式通常有以下两项重要活动。

●评估数据安全管理水平：通过自评估或者专业机构进行数据安全管理评估，将帮助工业 企业在自上而下实施之前了解当前自身数据安全管理的现状，明确存在的问题和潜在的 挑战，规划适当的数据安全管理蓝图。

●建设数据安全管理体系：数据安全管理工作是一项跨业务、跨部门的系统工程，数据 安全管理从上而下实施需要高度依赖高层管理人员的支持和职能集中的数据安全管理 组织。企业通过建设数据安全管理体系可以明确管理战略、制定管理制度、搭建组织架 构等，以工业企业的全局视角推进数据安全管理的实施。

自下而上的实施策略的显著特点是“问题导向、系统建设先行、快速见效”。此实施策略以 解决各生产部门和业务系统数据管理中的问题为出发点，通过使用成熟的数据安全管理工具， 快速搭建数据安全管理平台，实现对问题的逐个击破，并逐渐探索出全面的解决方案。

**36.3.2** **数据安全管理的切入方式**

数据安全管理的实施， 一般从生产系统或数据系统入手。从生产系统入手的常用实施方式 包括企业数据模型建设方式和主数据建设方式。从数据系统入手的常用实施方式包括统一数据 平台方式和数据集市方式。

工业企业在选择不同的实施方式时，应考虑数据对信息系统的重要性及企业目前的数据管 理水平。如果数据是企业重要的业务资源，同时企业已经具备了一定的数据管理专业水平和经 验，则可以采用从数据系统入手的实施方式，通过修复数据管理漏洞、提升数据服务应用水平， 推进数据管理能力建设。如果企业的数据管理水平并不成熟，那么选择从数据管理系统入手的 实施方式有些冒险，而从生产系统入手的实施方式则较为稳妥，也易见成效。

**36.3.3** **工业互联网云平台的数据安全**

当工业互联网平台采用私有云模式时，安全可控性成为企业重要的考量因素之一，企业 需要将云安全保护措施纳入数据安全管理范围内。为确保上云系统中的数据的保密性和完整 性，建议采用数据摘要机制来校验数据的完整性。具体方法是数据发送方使用散列函数(如

**360** 数据治理——工业企业数字化转型之道

SHA、MD5 等)对要发送的数据进行摘要计算，得到数据的鉴别码并发送给接收方，然后再将 数据与数据摘要进行打包后插入身份鉴别标示，发送给数据接收方。数据接收方接收到数据后， 首先确认发送方的身份数据，然后将数据解包后重新计算，将得到的鉴别码与收到的鉴别码进 行比较，若二者相同，则可以判定数据未被篡改，数据的完整性没有受到破坏。通过上述方法， 可以满足应用系统对数据完整性校验的需求。而对于用户数据，特别是身份鉴别信息的数据保 密，应用系统通常采用密码技术进行数据加密，以实现鉴别数据的存储保密性。

在传输过程中，主要依靠 VPN 系统来保障数据包的数据完整性、保密性、可用性。VPN 系统的组建主要采用两种方式：基于IPSEC 协议的 VPN 和基于SSL 协议的VPN。

为保证工业互联网私有云平台中的数据的抗抵赖性，最有效的方法就是采用数字签名技术。 通过数字签名及签名验证技术，可以判断数据的发送方是否是真实存在的用户。数字签名是不 对称加密算法的典型应用。数字签名的应用过程是数据发送方使用自己的私钥对数据校验或对 其他与数据内容有关的变量进行加密处理，完成对数据的合法“签名”;数据接收方则利用对方 的公钥来解读收到的“签名”,并将解读结果用于对数据完整性的检验，以确认签名的合法性。 同时，通过对“签名”的验证，可以判断数据在传输过程中是否被更改，从而使得数据发送方 不能对发送的数据进行抵赖，证明发送的数据是完整的，实现数据的抗抵赖性和完整性需求。

当工业互联网平台采用公有云模式时，为了保证数据安全需要采用服务水平协议(SLA), 其中应包括对要提供的服务及其预期服务和可靠性水平的描述、衡量服务指标的定义、各方的 义务和责任、补救措施或未能满足的处罚等，以及如何添加或删除指标的规则，还需要明确提 供的服务细节，例如正常运行时间和故障响应、测量标准和方法、报告流程和解决流程的定义， 以及保护客户尽可能免受因服务级别违规而导致的第三方诉讼的赔偿条款等内容。

**本章精要**

数据安全是数据治理成功的重要因素之一，本章主要内容包括建立数据安全组织架构、数 据安全应用需求梳理、引进数据安全平台技术、建设数据全生命周期管控等，并基于实践经验， 从数据安全实施步骤、数据安全实施模式进行了阐述，最后提出了要根据数据管理能力成熟度 的不同而开展差异化的数据安全管理工作。

第37章

数据治理常见误区

众所周知，数据治理是一项长期、繁杂且持续的工作，可以说是数据管理领域中的脏活、 累活，很多时候数据治理实施商做了很多工作，客户却认为没有看到什么成果。大部分数据治 理咨询项目都能交上一份让客户足够满意的答卷，但是当把咨询成果落地时，却因为种种原因， 很可能得到意想不到的结果。如何避免这种情况发生，是每一个进行数据治理的企业值得深思 的问题。

在数据治理过程中，数据治理实施商会提出数据治理的各种建议：有的提出覆盖数据全生 命周期的数据治理；有的提出以用户为中心的自服务化数据治理；有的提出减少人工干预、节 省成本的基于人工智能的自动化数据治理。在面对这些建议时，企业一方面要对自身的数据治 理现状有清晰的认识，对数据治理的目标有明确的诉求；另一方面还要知道数据治理中各种常 见的误区，只有避免进入这些误区，才能把数据治理工作真正落到实处，做到数据更准确，数 据更好取，数据更好用，真正地用数据提升管理和决策水平。

**1.** **误区一：企业需求不明确**

企业在进行数据治理时，必定是看到了企业数据存在种种问题。但是做什么?怎么做?范 围多大?先做什么后做什么?达到什么样的目标?业务部门、技术部门、厂商之间如何配合…… 很多企业其实并没有想清楚。所以，企业的数据治理难在找到一个切入点。

**2.** **误区二：数据治理是技术部门的事**

产生数据问题的原因往往来自业务，例如，数据来源渠道多、责任不明确，导致同一份数 据在不同的信息系统中有不同的表述；业务需求不清晰、数据填报不规范或缺失，等等。很多

**362** 数据治理——工业企业数字化转型之道

表面上看像技术造成的问题，如在ETL 过程中由于某编码变更导致数据加工出错，影响报表中 的数据正确性等，其本质还是由业务管理不规范造成的。但是，大部分企业认识不到造成数据 质量问题的根本原因，只想从技术维度单方面来解决问题，这样的思维方式导致企业在规划数 据治理时，根本没有考虑到建立一个涵盖技术部门、业务部门、管理部门的强有力的组织架构 和能有效执行的制度流程，导致效果大打折扣。

**3.** **误区三：大而全的数据治理**

出于对投资回报的考虑，企业往往想做全面覆盖企业和各板块的大而全的数据治理项目： 从数据的产生到加工、应用、销毁，实现数据的整个生命周期管理；从业务系统，到数据中心、 数据应用，里面的每个数据都能被纳入数据治理的范围。但是，广义上的数据治理范围非常大， 包括很多内容，想在一个项目里就全部实施通常是不可能的，而是需要分期、分批地实施。因 此，数据治理要从核心的系统、最重要的数据、最容易产生问题的地方开始着手做，才能见到 成效。

**4.** **误区四：数据治理工具是万能的**

很多企业都认为，数据治理就是花一些钱，买一些工具，并且认为工具就是过滤器，数据 经过过滤器，就没问题了。而结果是一方面工具越来越多；另一方面工具在实际上线后，功能 复杂，用户不愿意用。其实这种想法是一种简单化的思维。数据治理本身包含很多内容，如组 织架构、制度流程、成熟工具、现场实施和运维等，工具只是其中的一部分。企业在进行数据 治理时最容易忽视的就是组织架构和人员配置，但实际上所有的活动流程、制度规范都需要人 来执行、落实和推动，没有对人员的安排，后续工作很难得到保障。因此，建议企业在进行数 据治理时将组织架构放在首位。有组织的存在，就会有人去负责任，以人为中心的数据治理工 作，才更容易推动、持续做好及推广落地。

**5.** **误区五：数据标准难落地**

很多企业一谈到数据治理，马上就说自己已经有了很多的数据标准，但数据治理为什么还是 没有显著成效?数据质量为什么还是很差?这其实混淆了数据标准和数据标准化，数据标准化不 仅需要制定各项数据标准，而且需要做好各项标准的落地，将标准贯彻到数据治理的全过程。

**6.** **误区六：数据质量不能闭环管理**

平台建立起来了，数据质量的检核规则也配置好了，数据质量问题也找出很多，但是半年

第37章 数据治理常见误区 **3 63**

之后、 一年之后……同样的数据质量问题依旧存在。造成这种问题的根源在于没有形成数据质 量问责的闭环。要做到对数据质量问题的问责，首先需要做到对数据质量问题的定责。定责的 基本原则是“谁生产谁负责”:数据是从谁那里出来的，谁就负责处理数据质量问题。定责之后 是问责，问责之后是整改和反馈，然后是对数据质量问题的新一轮评估，直至形成绩效考核和 排名。只有形成这种工作闭环，才能真正提升数据质量。

**本章精要**

本章主要介绍了数据治理中各种常见的误区，只有避免进入这些误区，才能把数据治理工 作真正落到实处，做到数据更准确，数据更好取，数据更好用，真正地用数据提升管理和决策 水平。

**参考资料**

[1]DAMA 中国分会翻译组.DAMA数据管理知识体系指南(原书第2版).北京：机械工业

出版社，2020

[2] 《数据资产管理实践白皮书(4.0版)》,中国信息通信研究院



**第** **5** **篇** **案例篇**

**第38章**

电力行业：夯实数字化转型基础——

南方电网数据资产管理行动实践

**38.1** **背景介绍**

**1.** **公司介绍**

中国南方电网有限责任公司(以下简称“南方电网”)是大型国有供电企业，负责南方五省 (自治区)(广东、广西、云南、贵州、海南)的电网投资建设、运营管理及相关的输/配电业务。 在日常的生产经营管理过程中，南方电网产生了大量数据(包括一体化管理系统数据、电网生 产运行数据，以及引用的部分外部数据)。这里面既有周期性的统计数据，也有大规模的实时数 据。据测算，整个南方电网拥有的数据量已经超过了5PB。 这些海量的用电数据、跨行业数据 和客户数据都蕴藏着巨大的商业价值。南方电网以创新驱动为引领，通过“全要素、全业务、 全流程”的数字化转型战略要求，在推动全社会能源资源优化配置中发挥着积极的作用。

**2.** **信息化现状**

经过多年的建设和实践，南方电网开拓出了一条“信息与技术深度融合、自主可控、可持 续发展”的信息化创新之路。南方电网的信息系统已基本覆盖生产、运营、管理、服务等领域 的核心环节，基本建成了一体化、现代化、智能化的企业级信息平台，实现了数据全覆盖、全 网异地数据灾备，以及以主数据为核心的数据资源管理、以人为本的多渠道用户交互等，进一 步夯实了信息化基础，加强了应用系统之间的信息共享与集成。

在数据标准化及治理方面，南方电网主要取得了以下成果。

**366** 数据治理——工业企业数字化转型之道

(1)构建和完善数据管理制度体系。

南方电网编制了《中国南方电网有限责任公司数据资源管理办法》《中国南方电网有限责任 公司数据质量管理办法》并正式下发执行，初步构建了“五横三纵”的管理体系，从主数据管 理、数据质量管理、数据编码管理等方面明确了数据资源管控的要求。

(2)数据治理管控能力实现初步落地。

在数据治理方面，南方电网依据《中国南方电网有限责任公司数据质量管理办法》中的要 求，以质量提升指标为具体导向，将数据质量作为一项常态化的工作开展，这对提升南方电网 整体的数据质量起到了重要的推动作用。

(3)数据管理技术支撑能力初步建立。

南方电网建设了数据资源管理平台，其中具备了数据标准(编码)管理、数据质量管理和 主数据管理等主要功能，并在数据管理的几个重要领域中初步建立了技术支撑能力，为南方电 网开展数据治理工作和提升数据质量提供了重要保障，也是后续向数据资产管理平台演进的基 础。

(4)数据协同和数据应用得到提升。

南方电网的数据治理初见成效，数据质量有所提升，从而为上层的业务应用提供了更优质 的数据基础，为客户提供了更好的服务。

在取得成绩的同时，南方电网的数据资产管理工作仍然需要持续改进和提升，主要体现在

以下6个方面。

(1)整体工作长远规划需要改进。

南方电网的数据资产管理战略规划工作的统一性和长效性还需要提升，虽然在其信息化整 体规划中有数据资产管理相关的内容，但篇幅较少，数据资产管理战略规划内容尚待逐层细化， 连续性和一贯性还需要提升，战略目标还需要进一步明确和清晰。

(2)组织人员专业化程度有待提升。

尽管部分子公司设立了独立的数据资产管理部门，但南方电网全集团专业化组织体系尚待 完善，由上至下的领导力和执行力需要强化，数据资产管理工作开展的效率需要进一步提升。 同时，数据资产管理人员的专业能力有待增强。

(3)制度规范覆盖领域有待完善。

第38章 电力行业：夯实数字化转型基础——南方电网数据资产管理行动实践 **367**

南方电网的数据资产管理制度规范体系还在逐步构建中，尚有多个领域的管理办法和相关 细则亟待补充，而且在省公司、地市公司层面的指导规范需要进一步补充和完善，相应体系需 要进一步健全。

(4)数据质量需要进一步提升。

南方电网的数据资产管控机制的长效性和闭环管理仍需完善，数据质量发现及评估的时效 性仍有待提高，数据管控工作仍然存在“边污染，边治理”的问题，数据的供给能力亟待提升， 以便满足数据业务的发展需求。

(5)数据协同能力还需增强，数据价值还需充分释放。

数据资产需要在充分流通、融合之后才能产生更大的价值，“闭门造车”必然会造成机制的 僵化和价值的流失。南方电网当前数据资产价值的实现方式还需要多元化，信息和业务部门的 协同机制还需要健全，体系化的数据运营手段还需要进一步增强。

(6)数据安全保障能力有待完善。

数据安全是信息安全的重要组成部分，且其重要性日益凸显。但目前南方电网的数据安全

防护体系尚处在初级阶段，数据安全保障能力尚有较大的提升空间。

综上所述，虽然南方电网在数据资产管理方面取得了一定的成果，但是还存在需要进一步 改进的地方，尤其是基础数据治理，需要完善长期性、连续性的工作规划和科学合理的工作流 程。这些问题制约和影响了南方电网的数据资产管理工作健康、有序地发展。因此，为了有效 维护和整合数据管理已有成果，明确南方电网的数据资产管理的整体战略规划，计划和组织数 据资产管理的各项工作，南方电网数字化部制订了历时两年的南方电网数据资产管理行动计划， 全面部署数据资产管理工作，全方位地保障相关工作有序推进。

**38.2** **项目实施**

**1.** **工作目标**

南方电网的数据资产管理行动计划依据南方电网的“十三五”规划并结合公司数字化转型

的战略要求，建立了以下工作目标。

(1)构建覆盖全面、职能完备的数据资产管理体系框架。

(2)全面提升数据质量，优化数据服务，促进数据的共享和流通，加速跨专业的数据融合，

**368** 数据治理——工业企业数字化转型之道

强化数据的安全保障，推动数据的对外开放，探索数据的合作运营。

(3)实现全公司、全业务、全领域数据资产的可见、可用、可管，对内支撑公司业务的协 同和高质量发展，对外培育电力数据生态环境，促进数据价值的全面释放， 推动“数字南网”

建设，支撑公司向智能电网运营商、能源产业价值链整合商和能源生态系统服务商转型。

**2.** **工作框架**

南方电网数据资产管理行动计划的总体思路是：以价值为导向，做好顶层设计；从易到难， 夯实高质量数据基础；勇于探索数据价值变现模式，确保公司数据资产“看得见，管得住，用

得着”。

南方电网在充分借鉴国内外数据资产管理先进理论的基础上，结合公司发展现状及面临的 实际问题，提出“南方电网数据资产管理总体框架”(以下简称“数据资产管理总体框架”)。此 数据资产管理总体框架(见图38-2-1)明确了数据资产管理工作在公司数字化转型进程中的价 值定位和支撑作用，并针对公司数据资产管理体制的构建，提出了包括职能活动和保障手段在

内的一整套运作体系。



**图38-2-1** **数据资产管理总体框架**

数据资产管理总体框架以数据资产化管理的理念为核心，以保障数据安全为前提，紧紧围 绕数据价值最大化的业务目标，着眼于多元数据融合，重点建设“运营保障、数据供给、价值 创造、治理管控和技术支撑”五大能力，打造先进的数据资产管理体系，最终通过数据和服务 的良性循环，实现数据资产的“内增效”和“外增值”。

**3.** **工作内容**

南 方 电 网 的 数 据 资 产 管 理 行 动 计 划 主 体 内 容 包 括 “ 搭 框 架 、 重 治 理 、 夯 基 础 ” 三 大 类 工 作 。

(1)搭框架：构建完善的数据资产管理体系。

①开展数据资产管理专项规划，指导整体工作的有序开展。

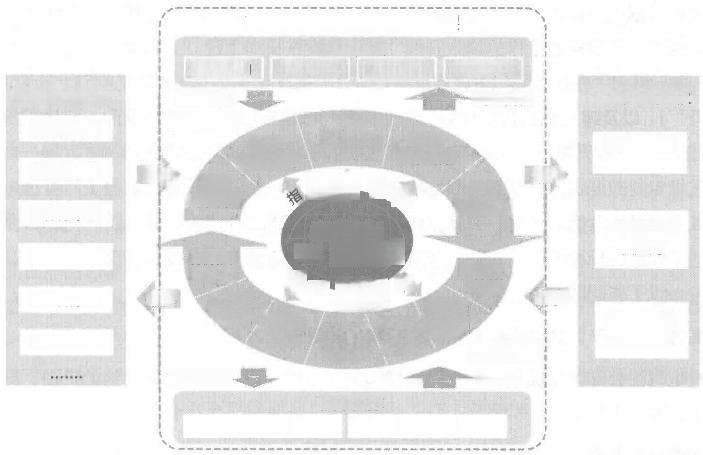
南方电网通过开展数据资产管理专项规划，紧扣公司数字化转型的战略要求，并充分考虑 公司现状，又提出了具有南方电网特色的数据资产管理框架，如图38-2-2所示。

南方电网的数据资产管理框架定义了一套环环相扣、务实可行的管理体系，其主要由职能 活动和保障手段两部分构成。

●职能活动描述了数据资产管理的具体工作，包括数据战略、数据治理和数据运营3个领 域，通过界定各项活动的职能和内在联系，相对完整地覆盖了公司在规划期内要实现的 数据资产管理工作方向。

●保障手段则定义了确保职能活动有效开展所应具备的前提条件和支撑能力，包括组织管 理和技术支撑两个领域，通过与职能活动相结合，能够有针对性地提出各种细化管理要

求 ， 确 保 执 行 过 程 准 确 到 位 。

业务数字化

营销数字化

物资数字化

财务数字化

综合数字化

办公数字化

生产数字化

**南方电网数据资产管理框架**

组织管理

制度规范 管理机制

人才培养

组织架构 规范保障

调整发展

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 数  产生 | 据运营  数据 服务管理  数据  需求管理 | 数据  数据 共享开放  运维管理 | | | 服务 |
| 易与货  数 据 战 略 |  | 数据  效益评估 |

规划 实 油 评 估

数据安 全管理

省 导 *与监* 楼 需求

促进

数据质 量管理

数据目

主数据 数据架 数据标录管理

管理 构管理 准管理 数 据 治 理 IT能力

支撑需求

技术支撑

管理机制

制度规范

公司转型升级

智能电网运

营商

能源价值链

供应商

能源生态系

统供应商

图38-2-2 南方电网的数据资产管理框架

**370** 数据治理 工 业 企 业 数字化转型之道

②优化数据管理组织，建立协同联动机制。

为加强对公司数据资产管理的集中领导和统筹组织，进一步优化数据资产管理组织体系， 南方电网成立了由首席信息官“挂帅”的数据治理和跨业务协同专项工作组，各业务部门、分/ 子公司的主要负责人为小组成员，负责推动公司数据资产管理的重大事项，协调部门之间、分/ 子公司之间的数据资产管理工作的有序推进，具体包括以下工作。

●构建专业的数据资产管理组织体系。在现有的基础上，完善和优化数据资产管理组织体 系，有序推进“网一省一地”三级专业数据资产管理组织的改革，统筹管理、执行、落 实全网数据资产的规划、治理和运营工作，初步形成组织紧密、运行高效、技术过硬的 专业团队。

●加强数据资产管理落地工作的网络建设。南方电网数字化部加强对总部数据资产管理执 行工作的监督和管理力度；分/子公司信息中心负责全面落实本单位数据资产管理的具体 执行工作。同时，南方电网积极探索构建数据资产管理虚拟化团队，盘活现有信息化团 队力量， 一方面培养一批既懂业务又懂技术的综合数据人才，操作和管理一线数据，对 接和落实省级公司的工作要求；另一方面强化与业务部门的沟通和联系，提升工作执行 效率。

●深化数据资产管理工作的组织协调机制，形成企业级数据资产管理协同工作机制，推动 数据资产管理各项工作的高效开展。南方电网在横向上建立业务部门与数字化部门的协 同联动机制，在业务部门设置专业接口人员(充当业务部门与数字化部门之间沟通的“桥 梁”),以及建立包含数字化部门和业务部门数据接口的虚拟工作团队，明确职责分工， 细化目标，提升数字化部门与业务部门的协同效率：在纵向上建立专业的“网—省一地” 三级专业数据资产管理组织联动机制。“网—省一地”三级专业数据资产管理组织明确 数据资产管理组织机制，并建立内部数据资产管理工作虚拟团队，统一内部数据资产管 理工作的目标和要求，协调资源配置，及时处理专业问题，确保内部高效协作并及时解 决一线工作问题。

③ 加强数据资产管理的宣传及培训，培育数据管理人才。

南方电网在公司内部加强数据资产管理知识体系的建设，建立了多维培训机制，具体包括 以下3个方面。

●邀请行业专家讲解大数据战略和前沿技术，提升各级员工的技术技能，以及员工对数据 资产管理的整体认识。

第38章 电力行业：夯实数字化转型基础——南方电网数据资产管理行动实践 **371**

●建立和完善数据资产管理专业人员的提升和选拔机制，加快公司级数据资产管理人才的 培养；通过考试、认证等方式选拔人才，为数据资产管理的深化推进夯实基础。

●加强和持续推进公司的大数据文化建设，培养员工的数据资产管理思维，为数据资产管 理工作的开展贮备知识、技术和人力资源。

④完善数据资产管理各项制度，扎紧制度规范“篱笆”。

南方电网根据大数据技术的发展形势及公司业务的实际情况，在现有制度的基础上持续完 善数据资产管理的各项制度，建立“1+N+n” 的数据资产管理基本制度体系，扎紧制度规范“篱 笆”,为全面实现“依法治数，依法管数”打下坚实基础，具体包括以下内容。

首先，制定统一的数据资产管理办法。结合多方意见，南方电网编制了《中国南方电网有 限责任公司数据资产管理办法》,并作为公司正式制度颁布并执行。

其次，编制数据资产管理各领域的管理办法。南方电网编制了数据资产管理各领域的管理 办法(主要包括《元数据管理指导意见《主数据管理指导意见《数据共享开放指导意见》等), 明确具体的管理流程和详细内容，确保数据资产管理相关人员在实际工作中“有法可依”,保障 数据资产管理中的各项工作在总体框架下有序开展。

第三，优化并完善若干配套管理制度和细则。南方电网在 “1+N” 的数据资产管理主体制 度下，优化并补充了n 个相关配套的专项管理实施细则，规范在数据认责、项目保障、应用管 理等方面的具体要求，提升数据资产管理制度体系的“健壮性”,推动实现数据资产管理“全面 管理、细化管理、专业管理”的阶段性目标。

⑤ 引入数据管理成熟度国际标准，开展数据管理水平评估。

南方电网引入数据管理能力成熟度评估模型，并由第三方专业评估单位对公司(包括总部) 的10个一级单位开展数据管理能力水平全面评估并准确定位，具体包括以下工作。

●全面厘清各单位数据资产管理能力现状及所处的发展阶段，同时找出各单位在数据资产 管理能力方面的短板，以及存在的突出问题，然后形成数据资产管理能力成熟度评估报 告并提出改进建议，为各单位后续的数据资产管理工作指明方向。

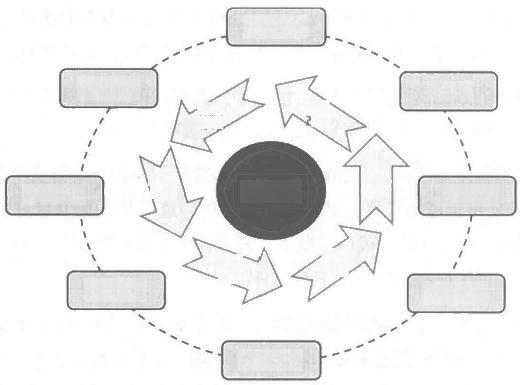
●各单位基于评估报告，根据本单位在数据资产管理能力方面存在的短板或突出问题，有 针对性地自主开展数据资产管理能力专项提升工作，并制定专项提升方案(既承接公司 数据资产管理行动计划，又兼顾本单位的实际要求),以指导本单位数据资产管理工作。

(2)重治理：开展全面的数据资产梳理和标准化。

**372** 数据治理——工业企业数字化转型之道

①开展全业务元数据梳理，实现元数据标准化。

图38-2-3所示为南方电网以元数据管理为核心的数据治理模式。南方电网组织了各业务部 门、分/子公司进行科学分工，采用“专业责任制”,对全领域、全业务(管理信息化+计量自动 化+调度自动化)的元数据进行全面梳理，提炼关键信息和关联关系，形成全公司完整的元数据 清单；在此基础上，根据《元数据标准规范》对元数据进行标准化，形成静态的标准化元数据 清单，以供各系统参照并进行实际调整。



数据质量

数据运维

数据

存储

元数据

数据

归档

数据开放

数据应用

]数据[ 规划

数据模型

数据 产生

数据安全

数搪 使用

数据共享

\数据 迁移

主数据

**图38-2-3** **以元数据管理为核心的数据治理模式**

目前，南方电网各单位都已完成资产、生产、营销、财务、人力资源、协同办公、审计、 等 1 0 个一体化企业信息管理系统的元数据梳理，并提交元数据清单，总体覆盖度和规范度达 到100%。部分单位在计划外还开展了IT 域、OMS 等系统元数据的梳理工作。

②制定数据资产管理标准规范，构建数据标准体系。

南方电网以6大业务域为基本范围，明确企业级数据标准的分类规则，编制元数据、主数 据、数据接口、数据模型、数据安全等多个领域的标准规范(包括《元数据标准规范》《主数据 标准规范》《数据质量标准规范》等),已初步建立统一的、覆盖数据资产管理全生命周期的数 据资产管理标准体系，以指导、规范公司元数据、主数据等多个领域工作在局部、分步地落地 实施，为实现数据资产管理标准化和后续工作的有序开展建立了坚实的基础。

③统一主数据全网定义，明确唯一来源。

南方电网全面梳理了现有主数据，充分结合相关人员的实际需求和建议，形成公司主数据

第38章 电力行业：夯实数字化转型基础——南方电网数据资产管理行动实践 **373**

清单。

为加强主数据的一致性，南方电网在集团公司范围内开展主数据一致性治理。其遵循“清 存量、控增量”的思路，在确定各类主数据唯一来源的基础上，对来源系统、主数据管理平台 和消费系统三方的一致性进行检查，并对不一致的主数据问题逐个解决。南方电网共完成19 类核心主数据的存量问题整改，有效地提升了主数据的一致性(平均一致性达到了98.48%)。

南方电网扩展了主数据统一应用的范围。其中包括进一步拓展主数据在全公司的应用范围， 在现有6大业务系统的基础上扩大覆盖公司的数据中心及各分/子公司的个性化系统，确保在全 网范围内实现主数据的统一应用。

南方电网优化了主数据应用生效机制。其中包括优化主数据生成、更新和归档机制，实现 对主数据对象、属性、责任部门等信息的增加、删除、变更管理，加强对主数据的读/写和修改 权限管理，保证主数据修改的一致性和稳定性，并在实际工作中“以用促改”,不断优化主数据 服务内容和使用机制，提升主数据的可靠性和即时性，有效支撑了各级单位的日常工作。

④编制企业级数据资产目录，构建企业数据资产视图。

南方电网结合对各系统的元数据的全面梳理，提炼关键信息和关联关系，并根据数据资产 目录的构建要求，提供相关业务信息、管理信息等内容。各分/子公司开展本单位的数据资产目 录编制工作，经过汇总及整理形成全公司统一的企业级数据资产目录，作为公司数据资产应用、 维护、运营的基础。在企业级数据资产目录的基础上，公司针对内共享数据和对外开放数据的 实际需要，增加共享类型、共享方式、共享权限等具体管理信息，明确可开放数据的敏感程度 及相应的安全保障措施，形成公司数据共享目录和数据开放目录。

在企业级数据资产目录基础之上，南方电网初步构建了较为完备的企业级数据资产视图， 实现数据从源头到应用层面链路的全景可视；搭建业务部门与技术部门之间的桥梁，规范了业 务部门的数据需求；同时为技术部门提供数据使用手册和指引，便于技术人员开发数据需求； 还为公司决策层提供了全网数据资产总览，方便决策层全面掌握公司现有数据资产的分布状况， 为公司战略和业务转型提供必要的支撑。

⑤开展数据质量管理提升工作，实现“正本清源”。

南方电网全面梳理以业务领域为维度的数据质量，构建了统一、科学、合理的数据质量体 系，明确了数据质量管理的战略目标与实施思路，具体包括以下内容。

一是有步骤地清理存量问题数据，加强在源头管控数据质量，推动数据质量管理从事后检 查逐步向事前、事中管控转变，初步形成从源头管控数据质量的全过程闭环机制。

**374** 数据治理 字化转型之道

二是基于大数据平台，提升数据质量分析及精细化管理能力，确保入库数据符合质量要求。

三是本着“业务导向”的基本原则，紧密结合业务部门关注的高价值数据，提升核心数据

的质量，确保重点数据的准确、可用。

四是建设数据质量分析及精细化管理看板，为数据质量问题统计、查询提供技术支撑；持 续完善数据质量分析方法，优化数据质量统计、查询功能，全面提升数据质量。数据质量管理 提升工作具体包括如下内容。

在业务协同场景治理方面，南方电网遵循“一场景， 一方案”的思路，全面开展213个 协同场景治理，大大提升了业务协同质量和效率，有力支撑了公司业务的“横向到边”及“纵 向到底”。

(3)夯基础：全方位提升数据资产管理的技术支撑能力。

①启动数据资产管理平台建设，全面提升数据资产管理的技术支撑能力。

由于南方电网在数据资产管理方面的技术支撑能力较弱，仅实现了针对数据编码、主数据、 数据质量的基本管理能力。因而，南方电网在整合现有数据资产管理平台中实用性较好的功能 基础之上，升级建设数据资产管理平台，新建或加强元数据、数据质量、数据开发、数据运维、 数据共享、数据开放等各方面的管理支撑能力，形成科学化、体系化、实用化的数据资产管理 技术支撑，覆盖数据资产全生命周期，为公司数据资产管理提供了坚实的技术支撑能力，具体 包括以下内容。

●元数据管理和应用能力。建立元数据采集、存储、维护、查询等基础能力，可以实现跨 平台的元数据采集和管理；实现元数据的血缘分析、影响分析等应用功能，以及依据元 数据标准的检查和稽核功能；完成面向多层级、多角色的元数据管理相关功能。

●电子化数据资产目录。将系统数据及相关信息进行必要的处理和提炼，以目录的形式展 现，并提供便捷的数据注册、发布、管理、获取等技术途径，让数据需求者自主式、自 助式查询、获取和使用数据；与数据资产管理系统已有功能进行整合，在数据资产管理 平台形成统一界面或统一入口。

●数据状态评估能力。实现对一定范围内数据状态的评估能力，构建包含数据标准化程度、

存储规范化程度、数据冗余度、数据使用情况、数据处理效能、空间使用情况、系统利 用率等指标在内的自评估指标体系，并基于此形成对数据运行现状潜在风险的主动分析 及发现能力。

●数据资产管理业务流程线上流转能力。让相关数据资产管理业务从线下升级为线上管理，

第38章 电力行业：夯实数字化转型基础——南方电网数据资产管理行动实践 **375**

让数据资产管理业务流程实现全流程可监控、可追溯，大大提升数据资产管理的效率。

② 加强数据安全管控能力建设及数据安全保障。

南方电网在其“十三五”信息化规划及数据资产管理专项行动计划中，对于信息安全体系 建设提出了明确要求：构建关键领域和敏感信息“进不来、拿不走、打不开、赖不掉”的四道 信息安全防线，实现全方位信息安全防御。当前，南方电网的各单位均已逐步完成数据加密系 统、数据脱敏系统、数据审计系统的安装部署及验收，实现公司核心数据入口加密、出口脱敏、 全程审计等数据安全防护措施，在不影响业务、不降低性能的前提下，为企业数据的安全防护 构建了初步的综合解决方案。

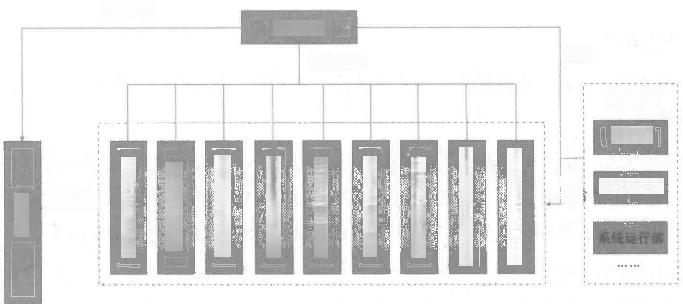
●数据加密系统。南方电网对生产区及测试区存储的重要数据进行加密，确保数据以密文 的方式存放于数据库中，避免和消除了数据库中由明文存储引起的数据泄露隐患。

●数据脱敏系统。南方电网针对不同敏感程度的数据制定了不同的数据脱敏策略，明确具 体的加密方式、稽核环节等；初步建立了“网一省”两级数据脱敏系统，具备数据脱敏 服务能力，防止隐私数据泄露。

●数据审计系统。南方电网已完成将重要信息系统接入数据审计系统，精确、全面监测数 据库超级账户、临时账户等的数据库操作，可以及时发现违规操作并实时报警，基本达到 全过程审计和安全问题可追溯的要求，满足相关审计方面的要求。

**4.** **组织保障**

南方电网数字化部通过组织各相关单位，于2018年年初建立了数据资产管理专项行动计划 工作团队，负责各项工作任务的具体实施，其工作机制如图38-2-4所示。



门 部

部 部 源

务 场 资

业 市 力

人

合

配

深细供电员工作组

广州供电局工作组

海南电网工作组

费州电网工作细

云南电阙工作组

广西电网工作组

广东电隔工作组

指高压工作组

调峰调与工作组

行

执

合

配

勤册院

统筹指导

数字化部

任务下达

配合

**图38-2-4** **数据资产管理专项行动计划工作机制**

**376** 数据治理 教字化转型之道

其中，核心工作团队主要由南方电网数字化部、分/子公司工作组、业务部门工作组和南方

电网数字电网研究院有限公司(以下简称“数研院”)组成。

(1)南方电网数字化部。

数字化部是本次项目的牵头负责单位，负责统筹整体工作，部署具体工作，并对各分/子公 司工作组的实施工作提供专业指导；同时其作为总体协调单位，负责协调业务部门与信息部门、 各分/子公司之间相关工作配合事宜。南方电网数字化部还负责公司相关各项工作的推进，并将 相关工作任务明确下达给相关的单位具体落实和执行。

(2)分/子公司工作组。

各分/子公司由本单位的数字化部门牵头，协同业务部门成立联合工作组，负责本单位所负 责或配合开展的工作任务的实施。分/子公司工作组接受南方电网数字化部的统一工作部署和指 导，并及时将在实施过程中发现的问题反馈给南方电网数字化部。

(3)业务部门工作组。

各级业务部门工作组负责配合在本部门范围内的各项工作任务。

(4)数研院。

数研院作为具体执行单位， 一方面执行公司层面的具体任务，另一方面在各分/子公司各项 任务的实施过程中，协助各分/子公司开展各项任务。

公司各部门、分/子公司协同推进各项工作，尤其是分/子公司的数字化部门和业务部门积极 开展协同合作，建立了职责清晰、沟通顺畅、办事高效的协同工作机制，确保了各项行动计划 的落地，共同推进南方电网数据资产管理工作的全面落实。

**5.** **过程管控**

在项目启动前，南方电网数字化部充分调动相关方的工作积极性，为各项工作的有序开展 建立思想基础。

在项目启动时，南方电网数字化部组织各业务部门及分/子公司工作组召开数据资产管理专 项行动计划贯彻及宣传会议，向员工传达公司领导对本次项目的工作要求，明确各项工作的职 责分工和目标要求，确保各业务部门和分/子公司切实配合相关工作的实施和开展。

在项目实施中，南方电网数字化部建立了周报和月度例会制度。各分/子公司工作组作为各 项工作的具体负责人和执行人，每周需要及时将本周工作进展提交给南方电网数字化部，南方

第38章 电力行业：夯实数字化转型基础——南方电网数据资产管理行动实践 **377**

电网数字化部由专人审核并汇总全平台当周的工作情况。南方电网数字化部每月组织一次例会， 各分/子公司工作组负责人介绍本单位当月的工作进展，讨论存在的问题和困难，协调各分/子公 司之间需要协同的事项等。

在项目实施后，南方电网数字化部组织各分/子公司工作组开展总结工作，对取得的经验和 待改进的地方进行总结，为后续不断提升数据资产管理能力做好闭环管理工作。

**38.3** **项目成果**

此次南方电网在全公司范围内推动数据资产管理实践工作，在框架体系、专业能力、支撑 平台三个方面都获得了一系列成果，具体介绍如下。

(1)普及数据资产管理文化，提升员工的数据资产管理基本认知。

数据资产管理不仅需要依靠数字化部门，还需要全公司各个部门的通力合作和参与，因此， 培养公司的数据文化，提升领导和业务部门对数据管理工作的认知极为重要。通过本次项目， 南方电网的各业务部门和分/子公司均参与其中，有的还负责具体事项，有效地提升了业务部门 对数据管理的认知，培养了数据管理工作氛围，也为各项工作持续开展建立了思想和认知基础。

(2)完成全平台数据盘点，实现公司数据“颗粒入仓”。

本次项目最重要的工作之一就是对全平台系统进行元数据梳理。在相对有限的时间内，通 过充分发动各方面的技术力量，南方电网数字化部高效统筹，基本按时完成了元数据梳理工作， 第一次直观、详细地呈现公司的数据“家底”,并在此基础上构建企业级数据资产目录，完成了 数据从“看得到”到“看得懂”的跨越，实现公司数据“颗粒入仓”。

(3)建立管理制度和标准规范，明确管理基线。

开展数据资产管理的基础是建立数据资产管理制度和标准。通过本次项目，初步形成了 “1+N+n” 的数据资产管理制度体系和各领域的标准规范，基本覆盖了公司所有正在开展的数据 资产管理活动，为全平台统一管理、统一标准、高效协同建立了良好的基础，同时也为各分/ 子公司开展本单位的具体工作要求的细化和落地执行明确了基本原则和管理基线。

(4)围绕主数据，初步解决核心数据共享及协同问题。

主数据是公司的“黄金数据”,也是各部门之间普遍共享的数据。本次项目重点关注主数据 管理，在原有主数据的基础上开展了主数据唯一来源定义，解决了数据层面和系统层面的多个 问题，极大地提升了主数据的一致性和应用效果，初步解决了核心数据的共享和协同问题。

**378** 数据治理 字化转型之道

(5)以元数据管理为核心的数据资产管理技术支撑能力体系基本形成。

通过本次项目，南方电网升级建设数据资产管理平台，夯实元数据管理功能，构建了自动 化、可视化的数据资产目录，并且优化和固化数据申请、获取流程，基本搭建起包含元数据管 理、数据标准管理和数据质量管理“三驾马车”的数据资产管理技术支撑体系，从客观上具备 了以元数据管理为核心的数据资产管理技术支撑能力。

(6)初步建立数据安全技术能力。

数据安全已经成为当前最受关注的领域之一。本次项目建立了数据加密系统、数据脱敏系 统和数据审计系统，使公司初步具备了针对数据本身的安全防护能力，也为公司开展数据共享 和对外开放数据建立了基础的“安全防线”。

(7)夯实数字化转型基础。

南方电网把数字化转型作为实现企业总体战略目标的重要路径之一，这给数据资产管理提 出了更高的要求。通过本次项目，南方电网落实各项数据资产管理举措，充实数据资产管理框 架，使之成为保证公司数据资产安全、高质量、合理配置的工作体系，有效支撑南方电网达成 数字化转型的战略目标，并确保执行过程的准确到位和协同一致，夯实了数字化转型基础。

(8)参与国家与行业的专业交流，获得广泛认可。

南方电网通过在全公司推动数据资产管理实践工作，使得公司的数据资产管理水平和能力 获得了全面的提升，取得的成果也受到了国内外专业组织机构的高度评价和表彰。

**38.4** **项目亮点和洞察**

本次数据资产管理项目的亮点主要有以下7个方面。

(1)顶层设计驱动任务落地。南方电网以公司数据资产管理体系框架和发展蓝图为依据， 结合当前业务部门的需求及迫切需要解决的数据问题，拆分并制定切实可行的工作任务，解决 实际问题，体现了数据资产管理工作的价值。

(2)数据管理制度体系覆盖全流程。本次项目从公司层面制定并下发了一系列数据资产管 理制度，初步形成了“1+N+n” 的数据资产管理制度体系，内容覆盖数据资产管理的各个领域 和数据生产全过程，为南方电网的数据生产活动建立了明确的基线，为进一步加强实时数据管 理建立了基本原则和依据。

第38章 电力行业：夯实数字化转型基础——南方电网数据资产管理行动实践 **379**

(3)点面结合确保高效推进。基于“总部抓总，分/子公司做实”的基本原则，南方电网数 字化部统筹各项工作，并直接负责核心、重要的工作，各分/子公司结合自身的实践经验和优势， 分别牵头负责相应的工作，确保各项工作的有序开展，达成预期目标。

(4)管理与技术“两手抓”。数据资产管理不仅仅是技术问题，更是管理问题。本次项目坚 持“管理与技术并重”的基本方向，着重解决了一批目前公司在数据资产管理方面尚未理顺的 问题，再通过技术能力的升级进一步提升管理效率，形成技术、管理互相促进的良性循环。

(5)打破数据“黑匣子”。针对“数据看得见摸不着，数据问题难以定位”等问题，南方电 网数据资产管理平台首次将公司数据资产从系统后台的“黑匣子”转变为前端可查可看的数据 字典，实现数据的可看、可查、可管、可用，为公司数据资产共享开放打下坚实的基础。

(6)构筑以元数据为核心的数据资产全生命周期管理模式。针对“数据资产管理如何实现 端到端的有效管控”等难点，南方电网通过深入开展元数据管理，以元数据为核心，通过覆盖 事前、事中、事后的元数据标准和质量管控，实现数据资产的全生命周期管理，同时与业务系 统衔接，形成良性管理闭环。

(7)实现一站式数据服务。针对“数据资产管理容易出现‘两张皮’的现象”,南方电网通 过全面打造数据资产管理业务流程线上流转能力，打通了技术与业务的关联关系，实现了基于 数据资产目录的管理工作流程，使得数据资产管理全程可查、可控、可追溯，大大提高了管理 效率。

**38.5 数据治理愿景**

南方电网将在现有成果的基础上，以解决业务问题为目标，围绕“筑基、创新和生态”主 基调，持续深化改进和开拓，其未来的数据治理愿景具体包括以下内容。

(1)筑建牢固的基础，通过数据资产管理平台，进一步夯实数据治理基础，助力公司高质 量发展。

(2)做好创新，通过与先进的大数据、人工智能等技术融合，进行数据资产管理能力创新， 助力公司业务创新。

(3)做优产业生态，用平台思维，通过数据资产管理平台，输出南方电网数据资产管理能 力，有效整合上下游企业数据，实现产业链上业务数据的融合，以及创新业务合作模式及数据 应用和服务，提升产业链整体数据价值。

**第39章**

电力行业：支撑集团产业数字化转型——

国家电投集团数据治理实践

**39.1** **背景介绍**

**1.** **公司介绍**

国家电力投资集团有限公司(以下简称“国家电投集团”)是一家以电为核心、 一体化发展 的综合性能源集团公司。

国家电投集团于2017年12月成立了集团大数据中心，由集团科技与创新部领导，是集团 大数据工作的管理支撑机构与业务运营平台。大数据中心旨在推动新兴信息技术与传统能源电 力行业的深度融合，促进技术与商业模式的变革。其当前四条工作主线是“搭建平台、汇聚数 据、开发应用、运营服务”。

**2.** **信息化现状**

国家电投集团多年来比较重视信息化工作，为促进企业发展和管理提升，其大力推进信息

化建设并取得了明显成效。

根据国家电投集团“十三五”信息化规划，由集团信息公司承担全集团统建信息系统(集 团化系统)的建设与运行维护工作(包括应用平台类、数据平台类、云平台类、信息安全类和 其他类5个类别共50多个信息化系统，覆盖全集团的各层级单位)。

各二级单位在使用集团统建信息系统的同时，还根据各自的需求，逐步建立了各自的业务

第39章 电力行业：支撑集团产业数字化转型——国家电投集团数据治理实践 **381**

信息系统(自主化系统)。各二级单位自建的各类业务信息系统约230个，分别应用在办公文档 管理、财务管理、基础应用等方面。集团化系统和各单位自主化系统共同为二级单位的信息化 提供了良好的支撑。

**3.** **数据资产管理现状**

国家电投集团的数据资源主要包括以下三大部分：

(1)经营管理类数据，基本为结构化数据，存在于集团统建信息系统及各二三级单位自建 系统里面。

(2)生产类数据，以时序数据为主，覆盖集团火电、水电、风电、铝业、煤炭、金融等产 业板块，其中新能源方面的数据比较多，接入集控中心的数据点约是传统能源的两个量级以上。

(3)非结构化数据，包括在设计、科研、实验等过程中形成的文本、图纸、图片、视频等 数据，这对研究院、设计院与工程公司来说是核心的资产。目前在运、在建的54个统建信息系 统中的数据总量接近100TB, 加上二级单位约230个自建信息系统中的数据及各类生产数据， 现存数据总数量超过了1PB。

目前国家电投集团在数据资产管理方面主要存在以下不足。

(1)国家电投集团各层级单位对数据资产管理不够重视，目前普遍没有把数据当作资产来 管理。尽管大家已经意识到了数据的重要性，但是各单位相应的数据资产管理职责尚未完全落 实到位，基本都未建立权责清晰的管理组织与制度流程，专职、兼职人员配备严重不足。

(2)数据的管理、技术标准体系不完善，长期存在短板，相关的制度流程还很不完善。国 家电投集团建设了数百个信息化系统，但是从集团总部到二级单位层面，普遍没有发布数据标 准，大多是直接采用相关的信息化国际标准与行业标准。因为没有完善的数据标准，导致数据 质量不高，不能很好地支撑数据分析应用。

(3)数据质量管理水平有待进一步提升。目前国家电投集团的标准体系不健全，数据质量 管理缺乏标准依据，有的系统中的数据完全依赖人工填报，缺乏有效的稽核机制；管理制度和 流程不完善，数据质量管理工作的开展缺乏常态化机制，跨部门的数据沟通困难；数据质量管 理人员配备不足、知识与经验不够、监管方式不全面等；缺乏数据质量管理工具支撑，目前多 依靠人为干预；数据质量评价规则库不健全，数据质量的评估过多依靠专家经验。

(4)跨单位之间的数据资源共享困难。目前国家电投集团的大部分业务系统按业务线垂直 建设，“烟囱式”的业务系统导致跨单位之间的数据资源共享困难；集团层面尚未建立业务数据

**382** 数据治理——工业企业数字化转型之道

开放、共享的机制，这在一定程度上限制了数据价值的挖掘，特别是设备故障样本数据。

(5)数据价值评估体系未建立。鉴于实际困难，国家电投集团拟建立有偿的机制，推进集 团内部数据的共享与交易，但因数据价值评估体系尚未建立，难以实施和推动。

(6)数据资产管理专业人才储备不足。国家电投集团虽已明确由大数据中心承担总部层面 的数据资产管理职责，由各产业创新中心承担产业内的数据资产管理职责，数据资产管理的职 责已落实到部门，但部门内的岗位设置和人员配备尚未完全到位，导致集团的数据资产管理专 业人才储备不足，部分二级单位缺乏全职的数据资产管理人员，以致数据资产管理工作对外部 运维单位的依赖过大。

**4.** **数据治理工作背景**

国家电投集团于2019年从“管资产、管运营”转变为“管资本、管战略”,全集团实施“区 域化、专业化”管理，相继成立了光伏、风电、水电、火电、综合智慧能源、环保等产业创新 中心，并明确各产业创新中心是集团在本产业的大数据分中心。随着集团管理变革的推进，集 团总部将权力进一步下放，对高质量数据的需求进一步增强，国家电投集团迫切需要建立集团 数据治理体系，不断提高数据质量，以满足集团经营管理与安全生产的需要，具体包括以下几 个方面。

(1)集团数据资产管理的需要。自从进入数字经济时代，数据已经逐渐被当作企业的重要 资产来管理。国家电投集团明确要求加强对全集团数据资产的管理，参照国际先进的数据治理 理论，建立集团数据资产管理体系，推进集团数据的共享与运营，发挥数据价值。

(2)集团数字化转型发展的需要。近些年，能源、电力企业的数字化转型明显加速，国家 电投集团也正在大力推进数字化转型。2020年，国家电投集团编制了集团数字化发展战略规划， 作为对数字化转型发展的支撑。作为“数据一信息一知识一智慧”价值链的基础，高质量的数 据是集团数字化转型的根本保障，因此集团需要建设并不断完善数据治理体系与培养相关人才。

(3)提高电厂核心竞争力的需要。随着新兴信息技术与传统能源技术的深入融合，国家电 投集团各二级单位都在探索及实践大数据与人工智能技术在电厂安全生产、经营管理中的应用， 其实施了不少大数据项目，并在项目中开展了数据治理工作，确保大数据应用开发、测试所需 要的高质量数据的需求。

第39章 电力行业：支撑集团产业数字化转型——国家电投集团数据治理实践 **383**

**39.2** **数据治理工作实践**

针对集团数据资产管理存在的主要不足，综合参考国际数据管理协会的《DAMA 数据管理 知识体系指南(原书第2版)》、国际标准《数据管理能力成熟度评估模型》《数据资产管理白皮 书(4.0版)》,以及国内能源、通信、IT 企业的数据治理实践，国家电投集团结合《国家电投 集团大数据建设总体方案》,采用自上而下+自下而上的方法，开展集团大数据治理体系的建设。

自上而下：自上而下是指从集团层面整体设计数据治理的组织机构、管理制度与流程，构 建集团数据标准，确保集团数据治理工作的规范性，逐步完善集团数据资产管理。国家电投集 团于2019年实施了“国家电投集团大数据治理体系规划咨询项目”,从宏观上设计集团的数据 治理组织机构、制度与流程，覆盖了元数据、主数据、数据标准、数据质量、数据安全、数据 模型、数据共享等主题域，制定并发布了数据管理办法，形成元数据、主数据、数据标准、数 据质量、数据模型管理与技术标准(待发布),以推进集团数据治理工作，为二级单位实施提供 参考。

自下而上：自下而上是指各二级单位根据工作需要，以应用为驱动，开展数据治理相关工 作，并以点带面，逐渐建立本单位的数据治理相关制度流程。自2017年起，集团各单位根据自 身的业务需求，结合数据应用分析，开展了数据治理工作。集团所属公司五凌电力、广西公司、 江西公司、云南国际、黄河公司、上海电力、中电国际、远达环保、信息公司、AE 公司、上海 成套院等，在水电、风电、光伏、火电、环保、经营管理方面开展了很有意义的数据治理实践 工作，并取得了很多经验与教训，特别是在光伏、风电与水电领域，在后面将重点介绍集团内 相关案例。

**1.** **自上而下——集团数据资产管理体系规划**

(1)主要内容。

2019年8月，集团大数据中心正式实施“国家电投集团大数据治理体系规划咨询项目”, 开展集团大数据治理体系顶层设计。其中主要包含构建集团数据治理组织体系、制定管理制度 与流程、编制数据资产管理总体规划及实施路线图(覆盖元数据、主数据、数据标准、数据质 量、数据安全与数据模型6个能力域),并在大数据中心、五凌电力中实施试点验证，根据试点 验证结果完善规划设计，并给出集团数据治理工作需求分析。

(2)建设目标。

集团数据资产管理的总体目标为建成相对完备的数据资产管理体系，建成集团数据资产管

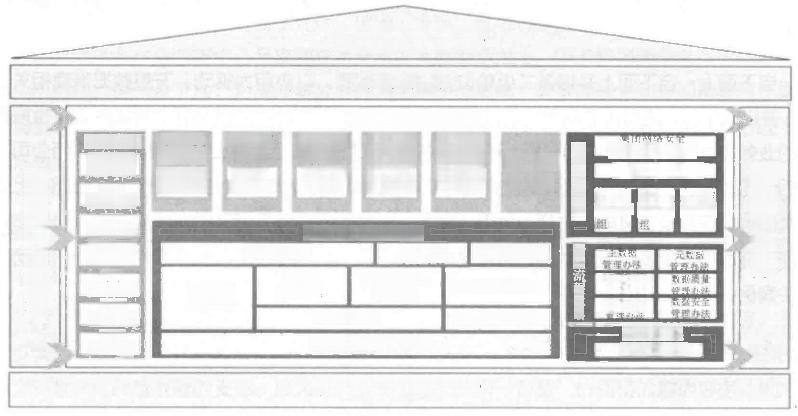
**384** 数据治理——工业企业数字化转型之道

理平台，建立统一的数据标准，完善主数据管理及元数据管理，深化数据模型设计与应用，持 续提升数据质量，不断强化数据安全，以数据资产目录建设与管理为抓手，研究并建立数据估 值模型，探索集团内部共享数据和对外开放数据的数据资产运营模式，为集团大数据分析应用

奠定坚实的基础，实现集团数据资产的增值、保值。

(3)体系架构

如图39-2-1 所示，国家电投集团数据资产管理体系架构主要包含1套标准体系、1个管理 平台、3套保障体系、6个管理域，以确保完成集团数据资产管理的总体工作目标。其中，管理 平台是为了支撑数据资源管理工作的顺利开展和落地实施，标准体系是为数据管理体系的建设 和落地实施提供规范的参考依据，保障体系是为了支持管理域而实现的一些辅助的组织架构和 制度体系，管理域是指落实数据资源管理的一系列具体行为。



**集团大数据战略**

大数据应用

大数据治理体系

数据质 量管理 数据质量 专题规划

标准体系

数据资产管理平台

平台工其

标准体系

管理

标准体系

安 全

标准体系

应 用

标准体系

大数据基础设施平台

与信息化领导木组

数据治理办公室

管理数 产业数 数据治 据治理 据治理 理专家

组

标准体系

基础数据

标准体系

数据

技术

标准体系

数据集成工具

数据全生命周期管理工具

数据模 型管理

数据模型

专题规划

主数据

管理

主数据

专题规划

元数据 管理

元数据专 题规划

数据安 全管理

数据安全

专题规划

数据信息资源池

数据共享 与服务

组 织 保 隆 作 素 管 理 程体载

专家知识库

数据标准 管理办法

数据模型

评 价 体 系 评价指标体系

数据交换

共享平等

数据资源目录

管理工具

元数据

管理工具

数据标准 管理工具

数据资质量管理工具

数据安全管理工具

数据服务平台

数据分析工具

图39-2-1 国家电投集团数据资产管理体系架构

需要说明的是，鉴于数据标准的重要性，以及国家电投集团的数据标准是其长期弱项，国 家电投集团在数据资产管理体系架构中单独将数据标准分列出来，作为近两年的重点工作进行 推进。根据国家电投集团现有的数据管理情况、在已有系统中推动数据管理和应用的初步实践、 未来集团信息化发展的趋势，国家电投集团规划并设计了国家电投集团大数据标准体系框架，

如图39-2-2所示。

其中，基础的、通用的数据标准以直接采用国际标准、行业标准为主；与企业产业特点密



"

重重

:

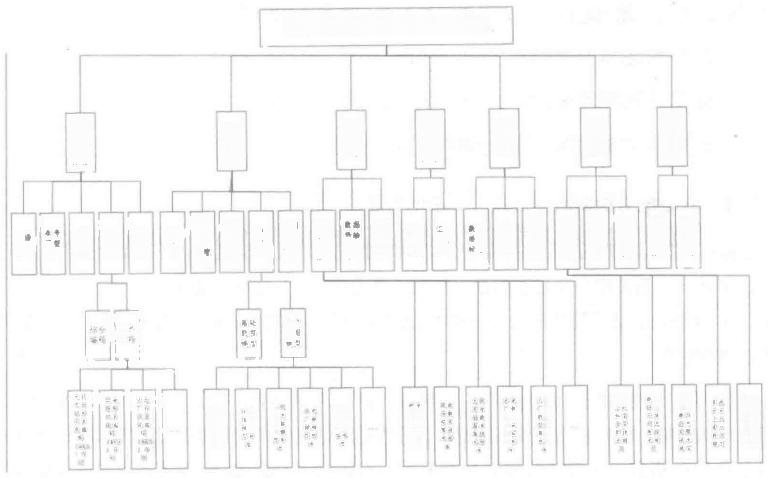
础 据 准 基 默 标 件

白 具 市 事 平 工 粗 悟

重通或重黑

目

第39章 电力行业：支撑集团产业数字化转型——国家电投集团数据治理实践 **385**

切相关的数据模型及标准，在借鉴国际标准、行业标准的基础上，编制落地的集团标准。

国家电投集团大数据标准体系

一级

架构

出

髻

体系



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 二级  架构 | 米  解 |  | 准 | 器  标 难 | 色言喜 数\* 常 | 主数    难 | 数 生\* 周角 管  准 | 出 | 数想解 | 数  数 使    标准 | 壹 唐    准 , | 述 | 被才术  准 | |  |  | | --- | --- | | 数据 运结  才  重 量  平 找  具 未  标准  标难 | 出 |     事 重 指  放思中重 开 壹 重 毒  重数争样  藏 通 色 重 |

三级

架构

通



四级

:

架构



\*





?

**图39-2-2** **国家电投集团大数据标准体系架构图**

(4)经验教训。

从集团层面自上而下地整体规划和设计数据资产管理体系的优点是：可以从整体视角规划 集团的数据治理工作，相对比较成体系，架构比较完整，有利于后续的分步实施；缺点是跨部 门、跨单位的组织协调工作量大，难度高，若做得不好，则很容易成为一纸空文。因此，国家 电投集团一是在规划过程中深入了解总部、各二级单位在经营、生产中的数据痛点问题，在各 个阶段广泛征求各层面的意见，求同存异，形成共识，并持续改进和优化；二是在落地实施过 程中，结合集团中不同单位、不同业务的特点，聚焦创造价值，灵活推进，优先在具备条件的 业务板块与二级单位中落地实施，绝不搞“一刀切”及强制执行。

**2.** **自下而上——集团生产时序数据治理实践**

电力行业是重资产行业(主要以设备形式体现),因此，电力企业在生产经营过程中对设备 全生命周期管理非常重视。国家电投集团多年以来积累了大量的生产数据(主要为时序数据)。 开展时序数据的治理与分析工作，对国家电投集团的电力资产经营管理、安全生产与运营优化 有着重大的价值。电厂生产数据来源整体可以被归结为三大类：设备及其在生产、监测、维护 中产生的数据；生产业务及管理数据；外部接入与生产环境数据。

**386** 数据治理 工 业 企 业数字化转型之道

下面重点介绍国家电投集团的3个二级单位在生产时序数据治理方面的典型案例，其中成 功的经验与失败的教训可供相关人员借鉴与参考。主要案例如下：

(1)五凌电力数据治理实践——水电领域。

(2)黄河公司数据治理实践——光、风、水领域。

(3)云南国际数据治理实践——风电领域。

**39.2.1** **五凌电力数据治理实践——水电领域**

五凌电力是国家电投集团的水电业务板块中的主要二级单位之一，下辖数十个部门及三级 单位。2015年，五凌电力规划了数据平台建设咨询项目，并编制完成《五凌电力企业公共信息 模型(WL-CIM)》 及《五凌电力企业基本通信结构规约(WL-BCS)》, 为五凌电力开展数据治 理提供数据标准。

2018年，五凌电力开始建设集团水电产业创新中心大数据平台，将 WL-CIM 和 WL-BCS 运用在数据采集、数据编码、数据模型构建等数据治理工作中，并根据应用需求，构建以设备 为核心的CIM 标准，将CIM 标准与BCS 标准、KKS 标准充分融合，基本实现时序数据与关系 型数据在大数据平台上的整合。

截至2019年，五凌电力已完成五凌电力生产业务关系型数据、12个常规水电厂集控监控 数据、2个试点电厂场站侧时序数据、水调系统数据的采集、编码、数据模型的构建；建成了 公司水电站智能决策应用系统，建立了设备故障诊断与健康评估等模型，实现了部分设备的状 态精准评估、设备故障智能预警等功能。

**1.** **工作背景**

五凌电力基本建成并打通了公司的信息化系统，覆盖了公司的核心业务，有效支撑了公司 的生产经营活动，将“大营销、大生产、大数据”作为其“十三五”信息化战略重点。通过 ERP 等系统的建设，五凌电力已经初步具备数据分析的功能。但由于各信息系统的核心需求、业务 模式、方案架构和技术路线都不相同，还存在以下问题：

(1)公司的ERP等系统是由不同开发商在不同时间建设的，数据彼此“异构”。

(2)公司通过各种信息系统沉淀了大量的数据，但数据应用仅停留在统计分析阶段，未能 对其进行数据挖掘。

(3)业务数据散落于各系统中，存在数据孤岛，跨系统分析难度极大。

第39章 电力行业：支撑集团产业数字化转型——国家电投集团数据治理实践 **387**

(4)各业务系统的数据格式众多，数据统计口径不一，难以进行数据整合。

(5)数据定义、标准不统一，企业内部的各部门之间难以打通和共享。

(6)缺乏对外部数据的利用。

**2.** **工作思路**

五凌电力的数据治理工作思路包括以下内容：采用元数据模型对企业管理数据进行规范和 描述；参考 CIM 标准，针对企业自身需求进行裁剪和扩充；编制 WL-CIM 标准，描述企业数 据的结构及数据之间的逻辑关系，用于对电力企业数据资产的管理及企业系统之间的数据交换 与系统集成。

采用基本通信结构对企业生产量测数据进行统一建模；参考 IEC-BCS 标准，针对企业自身 需求进行裁剪和扩充；编制 WL-BCS 标准，描述企业设备的量测数据类型、名称定义及相关参 数，使相关系统对数据的定义有着公共、统一的认识。

**3.** **工作方案**

五凌电力的 WL-CIM 标准以五凌电力各业务域的数据资产为输入，遵循数据建模原则，对 五凌电力的信息模型进行设计，扩展了数据类近1000个、数据属性5000多个，实现了各核心 业务域的数据标准全覆盖。具体编制步骤如下。

(1)以数据资产卡片为输入，导入CIM 标准(IEC61970 能量管理系统CIM、IEC61968 配 电网 CIM 标准)。

(2)根据现有业务对CIM 标准进行合理修改，复用已经存在的模型。

(3)依照IEC 建模规范和基本词汇表，对未定义的数据资产进行命名及属性定义，通过辅 助工具进行规范性检查。

(4)WL-CIM 由完整的一组包组成，沿用IEC-CIM包结构，对业务扩展部分按照业务域一 子业务域划分包，例如：计划经营一BusinessPlan, 项目建设一Project 等。

(5)采用 CIM 标准设计工具对 CIM 进行抽象、关联定义，以此构建数据的关联关系和血 缘关系。

(6)参考电厂标示系统(KKS) 编码对关键系统、设备、部件进行精细化建模。

(7)根据 CIM 标准对非结构化数据定义非结构的 CIM, 例如：工作报告模型、指导性文件 模型。

**388** 数据治理——工业企业数字化转型之道

五凌电力的 WL-BCS 标准以五凌电力各类型发电场站监测数据为输入，遵循企业 KKS 进 行研制，扩展了逻辑数据节点，实现了对各量测数据的统一、规范的编码，具体编制步骤如下。

(1)研究 IEC-BCS 标准，包括 IEC 61850-7-4 变电站、IEC 61850-7-410 水电厂、IEC 61850-7-420分布式能源、IEC 61400-25风电场等部分。

(2)定义若干个基本物理量(如振动、温度、流量等),对各系统量测数据进行聚类分析， 重点识别“同义不同名”的数据。

(3)依据 KKS 标准定义设备逻辑结构及 KKS 数据对象，以及进行数据分类(例如分为公 共信息、统计信息、状态信息、模拟信息、控制信息、设定值信息)。

(4)与IEC-BCS 标准进行查找对照，找到已定义标准的数据对象并引用。

(5)对于IEC-BCS 标准未定义的数据对象，按照标准约定的习惯与方法进行扩展。

(6)在编制 WL-CIM 和 WL-BCS 标准的过程中，通过设计者自查、项目组审查、专家评审 等方式，保证 WL-CIM 和 WL-BCS 标准设计的规范性、合理性。

**4.** **工作成果**

(1)WL-B CS 标准适应性设计及落地。

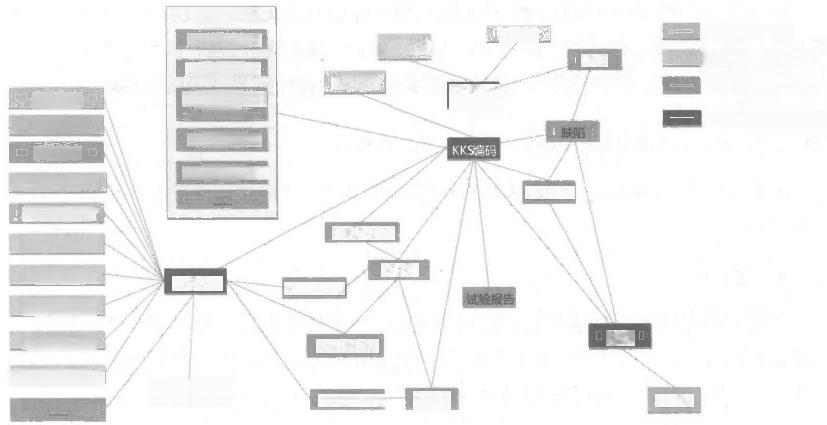
在实际编码中，五凌电力对 WL-BCS(水电)标准进行适应性落地设计，增加了扩展属性。 经过完善后的 BCS 标准采用四级结构形式： 一级为组织编码，引用五凌电力确定的水电厂组织 代码；二级为系统设备分级标示，引用五凌电力 KKS 编码；三级为标准量测名称，引用 BCS 标准；四级为扩展属性：引入数据源+部位+算法属性+单位+数据类型的形式。

五凌电力将WL-BCS 标准与《五凌电力有限公司水电厂标识系统(KKS) 编码导则》深度 融合，在大数据平台建立以厂站为基础，以KKS 为中心的 BCS 数据模型，提供用于数据共享、 数据展示、数据挖掘的数据对象。

(2)WL-CIM 标准应用落地。

五凌电力构建了以设备为核心的CIM 标准，将CIM 标准与BCS 标准、KKS 标准充分融合， 实现了时序数据与巡检、定期工作、缺陷、设备异动、工单、定值修改、试验报告等 ERP数据 关联集成，便于全面掌握设备信息。整体模型结构如图39-2-3所示。

第39章电力行业：支撑集团产业数字化转型——国家电投集团数据治理实践 **389**



**量测数据**

监控系统数据

代表原始数据不完整

水调系统数据

调速器曲磁系

统数据

运行日志

振摆系统数据

运行台账

油色谱系统数据

定期工作

调度指令记录

动火工作票

技术监督报表

应急预案演练

标准工作票

应急预案计划

技术监督计划

检修技术改造项且

标准操作票 操作题

代表原始数据完整且 已菜集

代表核心数据且已采集

代表数据未采集

技术方案(报告)

生产电理数据

介入工作票

定值修改

设备异动

水工日志

场站编码

设备综号

工作票

停复役

物料等

工单

巡检

**图39-2-3** **五凌电力整体模型结构图**

(3)数据采集及编码。

五凌电力已完成12个常规水电厂集控时序数据的采集及 BCS 编码对标，以及完成五凌电 力水调系统及 ERP系统的巡检、定期工作等70余张结构化数据表的采集及 CIM 标准化。

(4)数据清洗。

五凌电力根据数据类型及结构对采集到的各类数据进行清洗：对时序数据的缺失值采用插 补法进行补充；对异常值采用基于正态分布的离群点检测进行过滤；对关系型数据采用基于模 型检测的方式进行清洗，为企业应用提供有价值的数据对象。

(5)数据应用。

五凌电力以 BCS 及 CIM 标准化的数据为基础，在大数据平台上构建了多种应用，以服务 于安全生产管理。其中4层驾驶舱、设备全景信息展示、设备全景视图、三维视图、设备远程 状态监测等功能已经全部投入使用，后续会陆续开发设备故障诊断与状态检修智能决策等功能。

**5.** **工作价值**

WL-CIM和 WL-BCS标准的建立及标准化应用较好地解决了业务层面的数据资产向计算机 相关模型的转换，统一了各业务系统中关系型数据、量测数据，为数据采集、集成、编码、应 用及数据资产模型构建提供了坚实的基础。

**390** 数据治理——工业企业数字化转型之道

WL-CIM 和WL-BCS 标准的建立及实践为国家电投集团的大数据治理探索了方向，有利于 集团通过数据驱动激活存量资产潜在价值，实现增量资产高质量地发展，从而创造更多的价值 以提升企业竞争能力，对集团的数据资产管理体系构建及智慧化企业建设具有重要的意义。

**39.2.2** **黄河公司数据治理实践——光、风、水领域**

黄河公司是国家电投集团中最重要的清洁能源公司之一，其2019年的发电量突破600亿 千瓦时。

**1.** **工作背景**

为促进黄河公司旗下非电产业及相关业务的开展，提升各业务领域的生产效率，黄河公司 构建了西宁新能源生产运营中心智能运维平台和西安新能源生产运营中心智能运维平台(以下 简称“生产运营中心”),以提升电厂资产运营效率。

黄河公司通过结合此平台建设，实施数据治理工作。新能源企业因为生产设备多，其数据 测点远比传统的发电厂多得多，数据治理的工作量也大大增加。

**2.** **工作思路**

因为黄河公司的新能源场站比较多并且多建设于不同时期，不同场站的设备厂商、设计方、 施工方都不一样，各场站的量测点数据原始描述差异很大，以及在前期的建设生产中并没有考 虑后期的数字化智能运维，很多与数据相关的工作都需要重新开展，这些导致了开展数据治理 的工作量非常大，难度也很高。并且由于新能源数据量巨大，数据治理工作的组织难度也大， 还要优先保障安全生产，现场往往很难有足够的人力来配合数据治理工作的开展。对于场站的 KKS 编码，由于传统的人工核对数据点表的工作效率太低，需要定制开发数据编码工具，进行 自动编码与稽核。

黄河公司的数据标准化建设总体思路如图39-2-4所示。

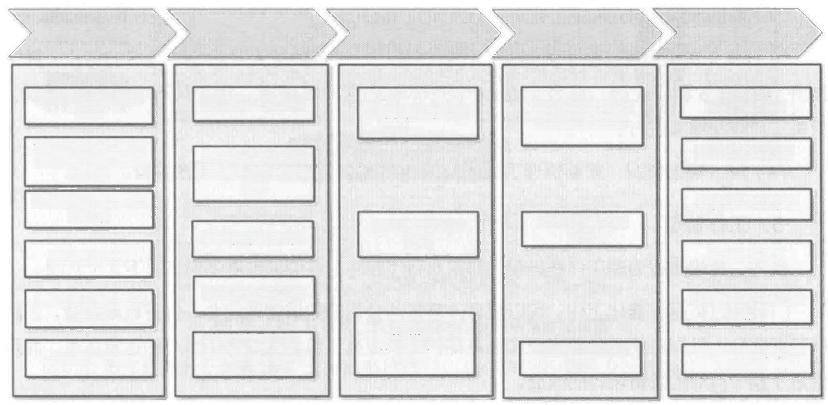
**3.** **工作方案**

黄河公司以生产运营中心为主导开展数据治理工作，组织第三方服务商东华软件和场站人 员共同实施数据标准化建设，并优先在1～2个场站中进行试点验证，然后逐步推广。其主要经 历了以下步骤。

(1)场站资料收集及梳理。根据项目的实际情况对数据资产进行调研及整理，包括场站基

第39章 电力行业：支撑集团产业数字化转型——国家电投集团数据治理实践 391

础信息、各个场站样板机基础信息、设备参数手册、场站接线图册等。



梳理设备树，进行

KKS编 码

HH-CIM

KKS 编码导则培训

(生产管理域)

HH-BCS ( 光 伏 )

码导则

光伏场站KKS编码

KKS 编码工具制作

KKS 编码工具

各场站时序数据标准

映射表

量测数据逐一编

以场站人员为主，

生产管理数据标准

映射表

BCS 标准数据编码设 计方案

BCS标准初稿

编 制CIM标准建模设 计方案

CIM 标准初稿

数据标准设计方案- 总体

黄河新能源发电场站

(KKS) 标识系统编

典型风电、光伏场

站设备信息梳理及 编码

全面梳理新能源场 站KKS编 码

管理数据标准化

映射

生产管理系统关系 数据表结构收集

码，形成标准化

映射

HH-BCS 适应性 改造

HH-CIM适应性 改造

BCS 标准细化设计

CIM 标准细化设计

数据标准方案设计

KKS 编码工具培训

时序数据点表收集

输出成果物

**图39-2-4** **黄河公司的数据标准化建设总体思路**

(2)按照“系统一设备一部件”划分，对场站进行建模，并根据已经建好的设备模型及场

站的实际情况对场站进行实例化配置，生成场站侧的 KKS编码。

(3)根据集团及其他二级单位的CIM 标准，并结合黄河公司实际情况进行定制化修改，形 成 HH-CIM 标准，在平台落地实施。

(4)接入ERP 等生产系统中的设备台账及KKS 编码等数据，形成黄河公司CIM 实例。

(5)依据黄河公司CIM 实例，以集团及其他二级单位实践的BCS 标准为基础，形成HH-BCS 标准，并在平台落地实施。

(6)依据 HH-BCS 标准，将数据接入平台。

(7)梳理现有 ERP 等生产、管理信息系统中的结构化数据和非结构化数据，制定转换策 略，按CIM 标准接入。

(8)对HH-CIM 、HH-BCS 标准及平台上的数据模型进行日常维护。

**4.** **工作成果**

尽管难度巨大，黄河公司的数据标准化工作依然取得了阶段性的成果。

(1)数据接入情况：青海区域共接入19个光伏电站、5个风电场、13个水电厂，共计

**392** 数据治理 工业企业数字化转型之道

约980万个测点。到2020年，预计数据测点可达4000万个。

(2)数据标准编制：形成了黄河公司的HH-CIM和 HH-BCS 标准，并且正在逐步落地实施。

(3)KKS 实施情况：由于黄河公司的新能源场站设备数量大、设备种类多的特点，其定 制开发了KKS 编码工具，已经完成23个光伏电站的KKS 编制，11个风电场的KKS 编制， 相关工作还在继续进行。

(4)BCS 实施情况：重新梳理了黄河公司BCS 基础词汇表及标准属性编码。

**5.** **工作价值**

黄河公司的数据治理工作在行业内非常具有代表性，其价值主要体现在以下3个方面。

(1)通过数据标准化工作，可以显著改善黄河公司的数据管理能力，提升数据质量，为后 续开发各种大数据应用工具奠定坚实的基础，不仅提高了黄河公司电力资产的运营效率，而且 促进了整个行业的设备设计的改进。

(2)为国家电投集团光伏产业的数据资产管理打下了坚实的基础，为集团其他二级单位光 伏场站的数据治理提供了参考价值。

(3)黄河公司接入的新能源场站数据测点多，面临的工作困难及工作量巨大，给数据治理 带来了很大的挑战，其中的经验与教训，为集团内部的公司、业内同行提供了很好的参考价值。

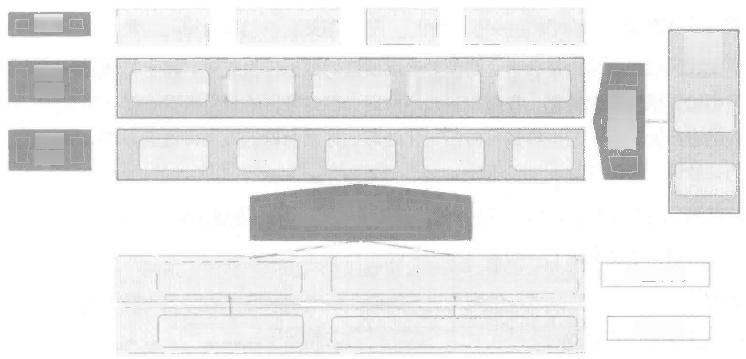
**39.2.3** **云南国际数据治理实践——风电领域**

云南国际以新能源装机业务为主，其地处高原、高寒、高海拔的高山地区，运行工况复杂， 日常运维成本较高。为降低成本，提高效率，云南国际建设了风电机组故障预警及智能分析系 统，并依托该系统建设，开展了数据治理相关工作。

**1.** **工作背景**

高原山地的风电场面临的问题是地势复杂、机组分散、运维成本高，运行数据价值无法充 分利用，风机功率曲线跌落原因也难以发现；大部件一旦损坏，更换周期长，无法实现故障预 警，尤其是齿轮箱、发电机等大部件发生故障会严重影响了风电场的安全运行与经济效益。云 南国际与远景集团合作，创新采用数据挖掘及机器学习等技术，开发了140多个风电领域算法 模型，对风机控制系统、大部件等开展监测及诊断，实现了风机异常预警。其系统架构如图39-2-5 所示。

第39章 电力行业：支撑集团产业数字化转型——国家电投集团数据治理实践 **393**



监视 预警 预测 其他

外部 数据

采样数据 统计数据 事件数据 数据标签 **数据质量**

**数据库**

订阅 抽取 转换 **加载** 合并

文件

规约画道

(带时标、有序、容错)

|  |  |
| --- | --- |
| 实时采集数据  实时采集数据 | 历史运行数据  (风机、测风塔、升压站、关口表………)  历史运行数据  (风机、测风塔、升压站、关口表…… |

应用

数据 服务

云端 处理

**前置采集**

**数据源**

**数据**

批绿 集成 通道

**图39-2-5** **风电机组故障预警及智能分析系统架构图**

风机故障预警及分析在实践中非常具有挑战性，具体体现在以下几个方面。

(1)数据获取困难。各风机厂家的后台监控系统、CMS 等辅助监控系统均为各厂家独立开 发，且在数据获取层面设有多重技术壁垒。需要逐一与风机设备厂家、辅助监控系统厂家开展 技术、商务沟通，打通数据获取通道，这样才能获取大数据分析所需要的实时数据。

(2)数据标准杂乱。获取到的数据存在数据标准不统一、采集点不统一、点表不统一、分 辨率不统一、采集频率不统一、时序不统一等问题，需要在现场核实并在数据中台中对数据进 行统一治理。

(3)数据质量差。因现场传感器、传输通道、原有系统存在缺陷等问题，导致获取后的数

据质量较差，在数据中台中必须完成高效且精准的数据清洗。

(4)数据值域差异化严重。不同风机厂家、不同机型、不同硬件配置下的同一个测点会因 零部件配置不同而造成值域不同，无法进行同一平台、同一模型的数据分析。例如，同一家发 电机厂家生产的同型号发电机，因轴承供货商不同，在同载荷、同转速的工况下，正常运行温 度区间、温度预警区间、运行温度上限各不相同，需要结合设备实际情况进行差异化的识别、 分析。

(5)对风机零部件级别的健康度预警在行业内没有成功案例，跨厂家、跨机型的大数据整 合及智能预警应用更是空白，没有可借鉴的经验。

**394** 数据治理 工 业 企 业数字化转型之道

**2.** **工作思路**

要实现风机故障预警及分析，必须要有高质量的数据作为支撑。鉴于不同的风机厂家，不 同系统中的数据存在差异化，在数据进入平台后，首先要做的就是数据治理：对缺失的数据或 者脏数据进行处理，同时也将数据做归一化处理。数据治理的效果将直接影响风机故障预警及 分析的有效性、准确性。

云南国际的整体数据治理工作围绕数据采集标准化、模型标准化和应用标准化3部分展开。

(1)数据采集标准化。

通信规约标准化：采用统一的规约进行通信。云南国际统一采用104规约和南瑞集控系统 交互，并通过设定固定的刷新频率发送遥测数据到大数据平台。

传输存储策略标准化：包含场站到数据平台的传输、数据转发、数据过隔离的传输和数据 存储等的标准化。

(2)模型标准化。

设备点表标准化：根据IEC61400-25 和 IEC 61970标准，建立风机、光伏、箱变、升压站 监控、风功率预测、测风塔、电量等数据点表标准，其中风机按照双馈和直驱建立分类点表， 保证风机发送的是标准点表数据，确保设备数据的完整性。

设备状态量标准化：建立统一的9种风机状态：正常发电、限功率、环境待命、服务状态、 用户停机、技术待命、故障停机、电网故障、通信中断，解决不同厂家风机状态不一致的问题。

设备模拟量标准化：通过建立标准的设备模型，映射不同风机厂家的数据，方便统一分析。

(3)应用标准化。

应用接口标准化：提供 APIM 管理系统，其对外提供标准的数据接口。

指标分析标准化：统一企业指标库、数据算法、应用场景及数据应用流程规范等。

**3.** **工作方案**

云南国际的数据治理工作方案主要包含以下内容。

(1)数据生命周期管理，针对不同类型的数据选择不同的存储策略，例如对于关键的风速、 功率等数据进行时序存储并定期清理，报表数据按照日、周、月、年的维度存储等。

(2)给关键数据编码，并作为元数据(固有属性，比如五街1号风机)进行存储，元数据 与相关设备绑定作为数据的属性标示(比如风速)。

第39章 电力行业：支撑集团产业数字化转型——国家电投集团数据治理实践 **395**

(3)根据硬件情况制定合理的存储策略，包括存储时间长度和存储周期间隔(根据硬件、 空间扩展，空间划分)。

(4)所有的数据增加质量标示，对于脏数据增加属性标签，以便上层应用在使用数据时能 够快速识别可用数据。

(5)加强数据安全管理，对数据的增、删、改、查操作都建立相应的权限管理。

(6)开展数据审计，对数据的操作留存记录，所有的操作都可查、可审计。

对于新能源电站风机、箱变系统等产生的数据，因设备庞杂，需要进行完整性和有效性评 估。具体采用的主要方法有以下3种。

(1)数据清洗及插补。针对核心测点数据(如风速、电量),提供完整的数据插补(后10 分钟的风速、电量数据),其余测点数据可以根据实际业务场景自行插补。

(2)风速插补将对后10分钟的数据进行数据清洗、维护、插补，提供故障数据、无效数据 等失效模式识别。

(3)对于无法断点续传的数据，基于累计发电量测点的数据接入后，系统将提供后10分钟 的电量数据插补服务。针对不同数据源的电量数据，将提供不同的具体数据服务。

**4.** **工作成果**

云南国际通过搭建的风电场数值模型，对孤立数据、错误数据、缺失数据进行了清洗；利 用风机各运行数据之间的关系、风机运行数据和风机状态之间的关系，对冗余数据进行了状态 估计，过滤了错误数据。

目前云南国际已经实现的风电系统中常见的数据清洗类型包括电量跳变、风机状态跳变、 风速与功率跳变、数据断点重传可能导致的重复记录识别等。

云南国际通过断点续传、边缘计算和数据清洗等技术，确保数据完整率遥测量达到100%, 覆盖率达到100%、数据有效率达到95%、数据一致率达到100%、数据及时率达到99%以上， 全面提升数据质量。

同时，对于数据是否在业务上满足要求、是否可用进行了标记，将不满足业务要求的脏数 据(主要包括无连接时间段内的数据、因冰冻等造成的不可用风速数据、不可用的限功率状态 数据等)标记出来，以满足后续数据分析应用的需要。

**396** 数据治理 工业企业数字化转型之道

**5.** **工作价值**

云南国际对集控平台采集的大量数据进行分析，结合数据挖掘及机器学习技术，采用了140 多个风电领域算法模型，能够识别并诊断风电机组变桨、偏航等控制系统亚健康问题，提高发 电性能，并能对风机主轴、齿轮箱、发电机等大部件进行早期预警服务，分析及发现风机主轴 承、齿轮箱等机械传动链存在的问题，评估部件健康度，及时提出维护与检修建议，从而有效 降低设备因停运造成的损失等。云南国际打通了将预警发现的各类故障信息以工单的方式推送 至现场，现场结合实际情况统筹人员、物料和备件的分配，实现生产管理的集约化和共享化。

云南国际通过借助海量现场端的生产数据，结合风光一体化大数据分析系统，对平台所接 入的新能源场站进行远程集中状态监测、故障诊断、发电数据分析预测，从而提高智能化运维 水平，减少由于设备故障、功率曲线跌落等原因造成的电量损失。

**39.3** **经验总结**

在各二级单位的数据治理项目的实践过程中，国家电投集团取得了以下成功的经验。

(1)生产部门与信息化部门紧密结合，甚至一体化办公。生产系统的数据治理需要信息化 部门与生产部门协作，例如黄河公司、云南国际的数据治理均是生产部门牵头的，而五凌电力 的数据治理则是信息化部门主导的。

(2)标准先行，以支持“云边协同”为目标，努力做到全集团的数据标准统一。

(3)以设备为中心，并基于中文语义开展数据治理，设备编码遵循 KKS编码规范，测点与 KKS 编码挂接，以设备一测点模型提供数据服务。

(4)大数据平台提供标准化的数据接口，源端测点编号的改变不应影响到大数据平台下游 应用系统的使用。

在以上二级单位开层数据治理项目的过程中，面临的主要难点如下。

(1)场站侧未建设 KKS, 且缺少信息化设备清单。

(2)BCS 字典因涉及语义翻译，维护较为困难，质量难以控制。

(3)各场站原始测点描述、设备描述差异较大，难以批量匹配，且场站侧运检设备更新后， 并不会及时通知信息化部门。

(4)人工操作中易存在属性匹配错误。

第39章 电力行业：支撑集团产业数字化转型——国家电投集团数据治理实践 **397**

(5)BCS 标准化工作效率提高困难。

(6)难以调集各场站人员集中办公，参与数据标准化工作。

在发电侧，时序数据治理的成功有以下几个关键点。

(1)集团统筹建设，确保规则、层级和字典统一。

(2)集团统筹立项，按照型号将设备一测点模型沉淀为数据资产。

(3)建设半/全自动化的编码匹配工具，提高数据编码效率。

(4)从源头把握数据质量，在场站基础建设、设备采购等阶段，完成 KKS 建设，在监控系

统建设阶段，完成 BCS 建设。

(5)成立专项数据治理小组，利用各场站人员熟悉现场情况的优势，保质保量完成数据编 码工作。

**39.4** **总结与展望**

国家电投集团将数据作为重要资产进行管理，大力推进数据治理工作。他们深刻认识到不 做好数据治理，数据质量难以支撑基于大数据分析的应用，不开展基于大数据分析的应用，数 据就是负资产。

国家电投集团采用“自上而下+自下而上”相结合的方法，进行数据治理工作，特别是在生 产侧时序数据治理方面，做了一些很有意义的尝试和实践。国家电投集团逐步在集团总部各部 门、各产业创新中心及具备条件的二级单位中，按照集团数据治理体系规划，进一步落实数据 治理组织，制定并发布集团数据质量、数据安全、元数据、主数据、数据模型、数据共享等管 理制度流程或标准规范，建立了集团数据资产目录，实施了一批数据治理项目，并推动建立集 团数据共享机制，优先推动生产侧时序数据的共享。

大力推进数据标准化工作，是做好数据质量管理的前提。国家电投集团努力构建以设备管 理为核心的数据质量管控体系，优先在风、光、水电领域，编制并发布集团基本通信结构、公 共信息模型、设备编码等数据标准与规范。国家电投集团在水、风、光领域的主要二级单位中 建立起数据质量管控体系，提升数据质量，支撑企业的日常安全生产与经营管理，以及基于大 数据分析的应用开发。

**第40章**

**能源化工行业：数据治理助百年油企数字化转型**

**40.1** **背景介绍**

**1.** **公司简介**

陕西延长石油(集团)有限责任公司(以下简称“延长石油”)是集石油、天然气、煤炭等 多种资源高效开发、综合利用、深度转化为一体的大型能源化工企业。其产业主要覆盖油气探 采、加工、储运、销售、石油炼制、煤炭与电力、工程设计与施工、新能源、装备制造、金融 服务等领域。

**2.** **信息化现状**

延长石油围绕自身的发展战略和主营业务开展信息化工作，坚持“五统两分”原则，积极 推进以生产、经营为主线的信息化建设与深化应用工作，其信息化现状如下。

(1)在“十二五”期间基本建成总部统一的生产经营平台。

(2)稳步推进总部和下属企业两个层面的信息系统建设，部分实现了总部层面信息系统的 集成和下属企业生产层面信息系统的集成。

(3)完善了计算机网络及信息安全系统，强化了信息标准化工作，初步建立了信息安全和 信息化标准两个体系。

(4)ERP、 生产营运指挥、油藏描述与模拟、加油IC 卡等系统的建设和应用达到行业领先

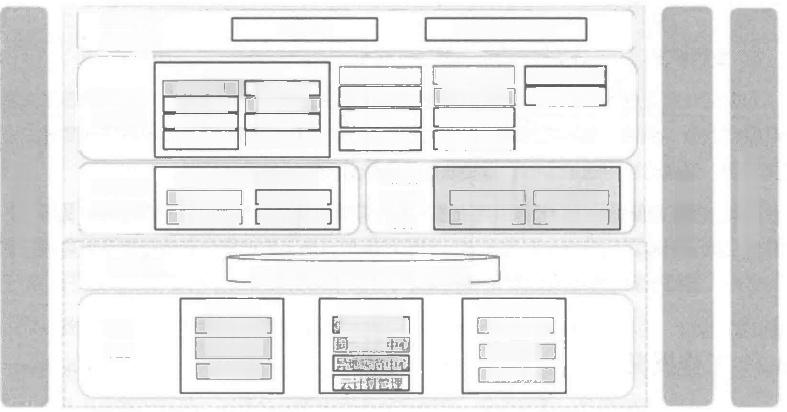
水平。

黑体海时化息信

(5)成品油物流配送、智能制造与集成应用等方面取得了突飞猛进的进展。

延长石油信息化总体架构包括四大应用平台、三大体系、两大支撑(见图40-1-1),具体介 绍如下。

(1)四大应用平台：包括决策支持应用平台、经营管理应用平台、生产运营应用平台、科 研协同应用平台，涵盖所有支持生产、经营和科研业务的应用系统具体介绍如下。



决策支持

应用平台

经营管理

应用平台

生产运营

应用平台

平台

**基础数据**

**基础**

**设施**

|  |  |  |
| --- | --- | --- |
| 网络系统  重团专网  总部内网  无理网 | 云数据中心  团数据中心  城灾备部 | 基础应用系统  现频会议  邮件系统  即时通信 |

|  |  |
| --- | --- |
| 法律事务管理 办公白动化 综合协同管  客户关系管 | 内控审计管现  干部管理 |

全面预算管理

财务共享服务 电子商务 物滤管理

科研协

同应用

科技创新管理

到 持 科技管理

知只管理 技术合作

|  |  |
| --- | --- |
| 目管计理  质量管理 | 修传管理 设备管理  人力资源管理 |

数据仓库 生产经营决策支持

ERP 相关应用

财务管理 物封管理

主数据(单位/物料/财务/人事/指标等)

应角构案 应含排程

生产调度管理与应急指挥

亚度管理

生产报表□

信 息 化 标 准 体 系

信息安全体系

**图40-1-1** **延长石油信息化总体架构图**

决策支持应用平台将数据仓库、知识管理、辅助决策等应用系统形成一个整体，为企业的 生产经营综合分析、预测预警、决策支持业务提供支持。

经营管理应用平台以ERP 系统为核心，汇集全面预算管理、法律事务管理等应用系统，全 面支持企业的财务、采购、物流、销售、投资、人力资源、行政办公等各项经营管理业务，推 进企业各项业务的协同发展。

生产运营应用平台面向延长石油的核心价值链，集合生产调度管理与应急指挥管理等应用 系统，支持油气勘探开发、油气储运、炼油化工、产品销售等领域的各项生产管理业务。

科研协同应用平台涵盖科技管理、知识管理等应用系统，支持企业的科技管理、知识管理、 情报管理、基础研究、专业研究等各类科研业务。

(2)三大体系：包括信息化管控体系、信息化标准体系、信息安全体系，从组织管理、标 准规范、信息安全3个方面为信息化建设提供整体保障。

**400** 数据治理——工业企业数字化转型之道

(3)两大支撑：包括基础数据支撑和基础设施支撑，从源头数据、基础建设等方面对延长 石油信息化建设提供稳固的支撑。

基础数据涵盖勘探开发源头数据、炼油化工生产实时数据、地理信息数据等，为多个应用 系统提供基础数据支撑。

基础设施是服务于信息系统运行和信息交互的公共物理设施，涵盖网络系统、云数据中心、 基础应用系统等内容。

**3.** **数据标准化及治理背景**

为促进业务发展，多年来，延长石油总部及各级企业建立了多个信息系统，覆盖了大多数 生产经营业务。但是，集团总部和各级企业的协同工作始终难以实现，如系统无法集成、业务 无法互通，造成这种问题的根本原因是缺乏各业务层面的信息数据标准。

因此，做好数据治理、统一管理数据源头成为实现各信息系统集成的基础。而实现跨业务、 跨二级企业、跨应用系统统一对数据进行组织和规划，提高数据集中存储和跨系统之间数据共 享的效率，是延长石油信息化工作的迫切需求。

**40.2** **工作概况**

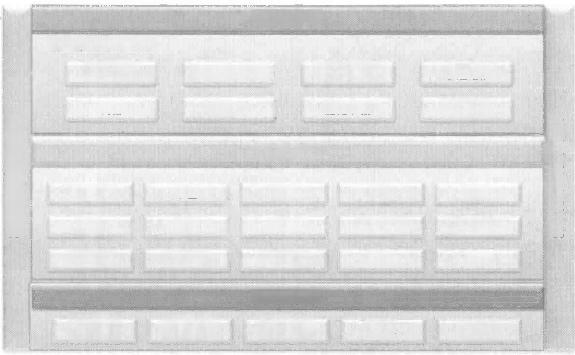
**1.** **建设历程**

延长石油特别重视信息标准化建设工作，并于2015年开始进行数据标准化建设。其数据标 准化建设工作按照“业务部门专业牵头，信息部门综合管理， IT 队伍技术支持”的模式，共建 立了10大类信息标准代码，广泛应用于 ERP 及相关信息系统中，为信息系统的建设、集成和 信息共享打下了坚实基础，也在油气勘探开发、工程建设、物资采购、炼化生产、销售等企业 管理、经营活动中发挥着越来越重要的作用。在2016年和2017年，延长石油分别在下属单位 (炼化公司和榆林能源化工有限公司)建设专业主数据标准，同期启动数据仓库的一期建设工作。 2019年，延长石油启动大数据服务平台项目建设。

**2.** **信息化标准体系框架**

延长石油的数据标准体系框架由技术标准、数据标准、应用标准、信息化标准管理规范和 信息化标准管理系统5个部分组成，如图40-2-1所示。

第40章 能源化工行业：数据治理助百年油企数字化转型 **401**



**应用标准**

业务流程模板和极型

管道运输

**矿业**

科研创新

**数据标准**

**项目**

**数据指标**

**工程**

技术标准

软件技术

入事

文档

管道运输

通用基础

物料

销售采购

**财努**

**安金环保**

**炼化**

熵油及化工

工程装备

**合同**

**勘探开发**

焦团通用

销售

信息化标准管理规范

信息化标准管理察统

**油气勘探开发**

基础设施

信隐安全

**大数据**

**单位**

SOA

**图40-2-** **1** **延** **长** **石** **油** **的** **数** **据** **标** **准** **体** **系** **架** **构** **图**

(1)技术标准由基础设施标准、信息安全标准、软件技术标准、SOA 标准、大数据标准共 5个部分组成，主要采用国家标准(包括网络、硬件、软件、系统集成、数据处理等一系列标 准),其属于规范指导集团信息化建设的基础性、全局性、总纲性的信息技术标准。

(2)数据标准由9类集团通用标准、5类板块专用标准和数据指标标准共15类标准组成。

●集团通用标准包含通用基础、单位、人事、财务、项目、物料、合同、文档、安全环保 共9大类数据标准，主要包含的内容有编码规则、分类规则、数据模型。

●板块专用标准包含销售采购、勘探开发、管道运输、炼油及化工(以下简称“炼化”)、 工程共5大类数据标准，主要包含的内容有专业数据的编码规则、分类规则、数据模型 (以下简称“数据结构表”,包含属性定义和描述模板)。

●数据指标标准主要包含各类统计数据的指标主题、指标名称、指标定义、指标单位、计 算方法、数据来源、分析维度、提报频度等。

(3)应用标准是信息系统中的业务流程模板和模型，涉及集团通用、油气勘探开发、管道 运输、炼油及化工、工程装备、矿业、销售、科研创新共8大类业务，主要包含数据及表单格 式、信息系统中业务流程和信息系统的部署及使用规范。

(4)信息化标准管理规范是信息化标准管理各项制度的集合，其中规定了信息化标准工作 的组织机制和管理制度、规范、办法和细则，用于对信息化标准的各项工作进行指导和规范。

(5)信息化标准管理系统是发布信息化相关标准文本、数据全生命周期管理的重要平台，

用于实现企业内部标准数据资源共享服务，为信息化标准工作提供重要支撑。

**402** 数据治理——工业企业数字化转型之道

**3.** **工作目标**

延长石油的工作目标是构建以云技术架构为支撑，以共享服务为建设方向的数据标准化管 理平台，为企业的信息系统建设和深入应用提供标准和规范保障，为集团各部门、各企业、各 系统提供高质量的标准数据，推动标杆企业信息系统的深度集成、数据共享和深化应用，具体 包括如下内容。

数据标准化共享服务为集团 ERP 系统、预算系统、人力资源系统等多个核心系统提供了基 础数据共享服务，为用户提供随时随地、唯一源头的数据资源共享服务，实现核心主数据的在 线管理维护。

延长石油通过完善供应链管理等核心业务模板，以及建立资产、设备、物料、内部单位、 外部单位等标准代码实现了以下目标：

建立炼化企业生产物流、能源和主要生产运行单元的数据模型；

参照国家信息安全标准和相关国际标准，完善信息系统安全等级保护、身份认证及IT 基础 设施相关标准；

建成和完善经营管理平台、生产营运平台主要信息系统的核心业务模板；

建成以信息标准代码、数据模型为代表的信息资源类标准；

建立信息标准管理与应用规范，以及统一的信息化标准管理平台，实现了信息标准的统一 管理，支撑各个层面的应用集成和信息共享。

延长石油采用先进的信息技术，实现企业的经营管理、生产执行、操作控制各层面的信息 系统集成，以及集团公司和下属各单位重要信息系统的全面集成，从而实现以数据源头一次采 集、统一处理、按需共享，全面支撑集团的业务协同和一体化发展战略。

**4.** **实施方法**

通过多年的项目实践，延长石油形成了一套全面、行之有效的企业数据治理体系框架，如 图40-2-2所示。延长石油的数据治理体系框架以数据治理组织架构为引领，以数据管控内容为 支柱，以数据管理办法、流程和系统为基础，立体化地指导企业开展数据治理工作。



劣躁



内 容

第40章 能源化工行业：数据治理助百年油企数字化转型 **403**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 总体规划 | 统筹规划标准定义工作  ●标准体系架构  ●标准定义规划  ●确定每年的标准定义目标及工作开展形式 | | | 统筹规划执行工作  ●制定标准实施原则和实施程序  ●整体统筹规划，确定改造方案、阶段执行重点  ●确定每年的标准执行项目 | |
| 标准编制和发布  √组织落实标准编制的  业务牵头部门  √组织专业团队与业务牵头  部门共同成立标准化项目组， 完成信息标准的定义工作  √组织相关企业和职能部门 的专家对形成的标准进行 审查，审查通过后按相关  流程进行发布 | | 标准宣贯和执行  √组织落实具体的标准执行 工作  √对新项目的标准执行过程进 行控制。如立项阶段标准使 用情况的检查与控制，验 收阶段标准使用程度的检 查与控制  √对已建项目标准的执行情况 进行日常监控检查  √定期培训与推广 | | 标准维护  √标准日常维护  √关注变更动因，记录变更需 求  √遵循信息标准管理办法，按修计 流程对标准进行修订和变更  √对已发布的信息标准进行定期 审查 | 标准监控  √标准日常监控  √标准实施的监督检查以及 标准体系的评价和改进  √监控标准需求，发现查异和 新需求，分析后记来到需求 管理库  √通过信息标准管理平台对数 据质量、申请、审核情况、 修订情况等进行监控 |
| 管理控制 | 组织结构  √确立标准管控组织结构 及其职能 | | 管理办法和技术规范  √信息标准制定、修订、审查、发  布的管理办法及流程  √各类信息标准的管理维护细则等 | | 标准考核体系  √考核办法  √考核内容 |

**图40-2-2** **延** **长** **石** **油** **数** **据** **治** **理** **体** **系** **框** **架**

同时，延长石油也形成了数据治理方法论，得到延长石油集团内部领导的一致认同，并取 得了不错效果，如图40-2-3所示。

战略、方法



1.主数据标准化规则



建制度

●建制度：建立《主数据标准化管理 办法》,让主数据管理工作有法

可依

定标准：

●定标准：制定《主数据标准化管 理标准规范》,分类标准、描述 模板、编码规则、数据模型等，

让主数据管理有章可循

设组织：

●设组织：设立主数据管理组识，

专职负责主数据管理及标准维护， 确定主数据拥有管理权力，让主

数据管理由组织监督执行

理流程：

●理流程：梳理数据维护及管理流程。 建立符合实际情况的管理流程，

保证主数量标准规范有效执行

|  |  |
| --- | --- |
| 2.数据清洗  数据清洗： | KPI  **工具**  3.标准贯彻  5.数据质量治理  4.平台搭建  建制度： |

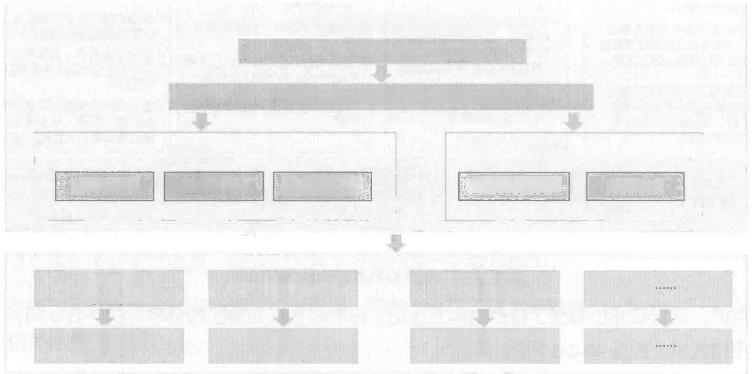
|  |  |
| --- | --- |
| ●数据清洗：根据标准规范对历史数据进行清洗， 排重、合并、编码、保证数据的完整性、准确 性和唯一性  标准贯彻：  ●标准贯标：主数据标准制定完成、数据清洗、 编码完成，需要对已上线、在建等业务系统. 根据系统所处阶段及重要性的不同，采用  “完全、映射，择机”不同的策略进行标准 数据的导入。  搭平台：  ●平台搭建：通过主数据标准规划产出的标准规范、 管控组织与管控流程，通过平台进行落地，并对主 数据的采集，存储管理、共享提供统一的管理 工具 | 数据治理：根据《主数据标准化管 理办法》进行数据标准的管理和 维护。保证标准规范的适成性和 健全性；根据标准规范制定数据质 量的考核机制，保证数据质量持 续改进  ●数据冗余：未被修改、未被使用 数据的持续时间  ●数据质量评估：关键信息完整率、 历史修正率  ●数据使用维护；申请者、使用者、 申请天数、修改次数、共享次数 |

图40-2-3 延长石油数据治理方法论

**404** 数据治理——工业企业数字化转型之道

**40.3** **组织保障**

延长石油成立了信息化工作领导小组，在小组的统一领导下，按照“五统两分”的原则， 建立集团、企业两个层面的信息标准化管理体系，如图40-3-1所示，具体介绍如下。



集团层面数据标准化及治理组织

集团公司信息化工作领导小组

数据标准化办公室(科技与信息化管理部)

**业务组** 技术组

业务流程组 数据指标组 主数据管理组 延长石油业务组 各项目组

企业信息

化领导小组

信息标准化

管理组

企业信息 化领导小组

信息标准化 管理组

企业信息 化领导小组

信息标准化 管理组

**图40-3-1** **延长石油信息标准化管理体系**

集团公司信息化工作领导小组对信息标准化工作进行统一领导，确定指导思想、目标和任 务，协调解决信息标准化相关的重大问题。

数据标准化办公室设在科技与信息化管理部下面。科技与信息化管理部是信息标准化的归 口管理部门，负责信息标准化的统一规划、综合管理。

业务组由相关职能部门组成。各相关职能部门作为某类业务流程模板、数据指标和信息代 码的制定牵头部门，负责该类信息化标准的需求收集及确认、业务审核、培训宣传及应用情况 的监督、检查等管理工作。

技术组主要由项目组组成，负责日常运维和技术支持，与各专家一起提出信息化标准制定 的技术方案，负责相关信息化标准在项目中的实施及补充完善。

企业信息标准化管理组负责对集团统一发布的信息标准在本企业的贯彻落实及应用检查， 负责本企业信息标准的培训等工作。

第40章 能源化工行业：数据治理助百年油企数字化转型 **405**

**40.4** **主要成果**

延长石油通过此次项目主要取得了以下成果。

(1)建立了数据治理体系和标准，确保企业各信息系统所使用的主数据的唯一、完整和统 一；梳理与建立企业的各项统计数据指标，为企业提供及时准确的决策分析数据基础；构建覆 盖整个企业范围的信息化标准，建立信息化标准管理组织和规范制度，保障企业信息化建设的 方向和质量。

(2)编制了集团公司急需的数据标准，实现了通用基础、单位、财务、人事、物料、数据 指标等十大类数据标准的编制及在各单位的贯彻实施。确保各系统中主数据代码的一致性，有 助于保障各信息系统之间信息传输的一致性。为ERP、 物资采购等核心系统提供统一的数据标 准，有利于对相关数据进行统计和分析，为企业的经营决策提供支撑。

(3)确保了企业核心系统(ERP、MES、 电子商务等多个系统)的互联，满足集团的信息

共享的需求，避免形成“信息孤岛”。

(4)提高了工作效率，减少了数据变换、转移所需的成本和时间，降低了数据的冗余度，

提高了信息的有序化程度和存储效率。

(5)提升了信息系统的灵活性，实现了企业范围内数据的平滑性与一致性，极大地提高了

信息系统集成的灵活性，降低了新系统的引入时间及减少报表的生成成本。

**1.** **数据标准化主要成果**

延长石油在数据标准化领域取得以下7个方面的成果。

(1)初步构建了集团数据标准化体系框架。延长石油建立了延长石油集团信息化标准体系， 对集团公用的数据信息实现了统一标准、统一源头、统一管理，解决了数据标准不统一、数据 不一致、分类口径差异较大等问题。

(2)规范了 ERP 业务模板，建立了勘探开发数据模型标准、生产营运层面的 MES (制造 执行系统)工厂模型、炼化生产计划优化模板等。

(3)建立信息代码体系表，统一编制了物料、内/外部单位、财务等共10大类约100万余 条的信息代码。

(4)对集团公司数据指标进行了整体规划，定义了3000多项数据指标，保证了数据指标的

业务含义定义的一致性、应用规则的一致性。

**406** 数据治理 工业企业数字化转型之道

(5)建立了一套完整的技术标准，制定了开发框架功能规范、开发框架技术规范和开发过 程规范。

(6)建立了一套数据标准化制度和规范，制定了《主数据管理办法》《数据指标管理办法》 等，保证数据标准的统一管控和统一应用，形成了集团公司数据管理及运行维护体系，确保集 团数据标准化管理系统的平稳运行，培养了集团公司的运行维护管理队伍，切实将信息化标准 执行贯彻落实到位。

(7)建立了统一的技术平台，搭建了一套统一的信息标准化管理平台，为多个应用系统提 供数据共享服务，实现主数据从发布到应用、维护的全过程管控；将各类数据标准及数据明细 进行了统一管理，为集团公司及下属企业的信息系统提供了公共、开放、统一的数据共享服务， 为与集团公司及下属企业重点系统的集成、业务协同贯通扫清了障碍。

**2.** **数据治理的主要思路和做法**

(1)统一数据治理体系。

延长石油通过统一数据标准，形成了数据资源开放目录，构建了一整套数据管理机制，具 体包括以下4个方面。

一是建全数据治理体系。延长石油借鉴IT 治理国际标准ISO 38500和 DAMA 数据管理理 论，结合其信息化建设成果与经验，制定了延长石油数据资源管理制度；在业务层面明确相关 业务数据拥有者和技术支持者，厘清了各部门的数据管理、维护及共享交换的职责、权利和义 务；对数据的生产者、使用者、拥有者和管理者在使用数据过程中的流程和权责进行界定；结 合数据的采集、存储、处理、共享、销毁的全生命周期管理需求，搭建数据工具，全方面健全 数据治理体系。

二是开展数据资源盘点工作。延长石油成立了工作小组并选取典型的应用场景进行调研， 形成数据盘点方法论和相关执行流程及模板；在此基础上以应用为导向进行数据盘点，明确数 据的责任主体、重要性、使用频率、存储地点等属性并进行编码；同时梳理数据中存在的质量 问题，形成盘点报告及下一步数据治理建议，作为数据资产目录的输入。

三是整合数据，形成数据资源开放目录。延长石油根据应用场景和数据盘点结果，开发基 础信息采集工具，从不同类型的源系统进行数据探查，并自动获取基础信息；建立数据资源分 类标准和分类机制，对采集后的数据资源进行分级、分类；对数据资源属性进行完善，形成数 据管理模型、业务模型、技术模型及权限和安全模型；通过数据接入和传输监管实现数据的血

第40章 能源化工行业：数据治理助百年油企数字化转型 **407**

缘分析和影响分析；提供对外的数据服务、数据订阅服务及数据评价服务。

四是提升数据安全保障能力。针对网络信息安全新形势，延长石油遵照《中华人民共和国 网络安全法》及《延长石油信息安全等级保护》中关于数据安全的有关要求，整合延长石油的 信息安全资源，加强大数据安全技术和产品的应用；强化大数据基础设施安全管理、大数据应 用安全管理和身份鉴别及访问控制；构建符合延长石油业务特色的大数据安全保障体系，实现 数据共享安全。

(2)初步建立统一的大数据服务平台。

延长石油基于前期数据仓库建设成果，初步构建了统一的大数据服务平台，通过整合内/ 外部数据，提供多元化的数据服务，助力数据资产的开放和运营，具体包括以下两个方面。

一是强化技术产品的功能开发。延长石油以应用为导向，在合理引进国内外先进软件的基 础上，对技术服务模式进行创新，形成了技术先进、生态完备的技术产品体系。通过分析自身 业务和数据的特点，并结合大数据技术和行业发展趋势，延长石油对自身的大数据应用现状进 行全方位、多维度的评估和查缺补漏，加快大数据关键技术产品的研发。

二是整合行业内/外部数据。考虑到能源、化工行业数据的特殊性，并结合应用系统的实际 情况及发展趋势，延长石油充分利用内/外部数据，实现了大数据服务平台可以获取来自不同渠 道的数据，同时配合分层存储，为企业提供全面的数据共享。

**3.** **制度、体制及机制**

在管理制度方面，延长石油在集团公司层面制定了《延长石油信息化工作管理办法》《延长 石油主数据管理办法》《延长石油数据指标管理办法》《延长石油内部单位代码维护细则》《延长 石油外部单位代码维护细则》等。其中的管理办法明确了各主要代码的业务牵头部门和职责分 工，进一步规范了代码的申请、审批、制定和修订、审核、发布、维护、复审、废止等管理及 维护流程。其中的细则从编码规则、使用要求及维护细则、代码申请及审核流程、岗位及职责 等方面，对各类代码的具体维护方式和要求进行了详细的说明。

延长石油还形成了“共享服务”机制,其数据标准化工作的总体思路建立在实现标准化管 理与应用的“4 个转变”基础上(包括从以日常运维为主向以工厂化运维为主转变，实现“共 享服务”转变；从以人工线下沟通为主向以全过程在线管理为主转变；从以信息代码建设为主 向以信息代码应用为主转变，实现应用、监控、评价等全面管控转变；从以信息代码的发布为 主向以集成自动的主动分发为主转变，提高数据质量),具体包括如下3点。

**408** 数据治理——工业企业数字化转型之道

(1)强化标准化运维管理。

延长石油加强运维体系建设，从组织、流程、制度、知识库、系统平台等方面加强运维工 作的规范化管理，为“工厂化运维”提供支撑，实现“共享服务”。

(2)实现数据的全生命周期管理。

延长石油从主数据代码的标准化、主数据的规范化、主数据应用的一致化出发，全面推进 主数据的全生命周期管理。

①主数据代码的应用与维护：主数据代码逐步以建设阶段转入应用与维护阶段，并紧紧围 绕信息代码表进行代码管理、监控、在线维护、数据质量日常检查和数据分发等；建立数据推 送和提醒机制，保证各信息系统中的代码使用的一致性。

②分发监控与应用检查：通过开发外挂式监控窗口，对主数据代码的分发情况进行监控； 通过开发数据比对工具，定期对目标系统(如 ERP 、HR 系统)中的数据与标准化数据进行比 对，形成各类主数据应用情况差异表。

③主数据代码应用情况分析：通过对主数据应用情况差异进行分析，为相关职能部门对违 规行为和不规范操作提供检查和考核的依据。

(3)实现标准化应用监控与评价。

延长石油加强信息标准化应用的监督、检查力度，健全应用考核制度；加强对新建项目的 贯彻标准管理，规范立项和验收的标准化应用模板；通过技术手段，检查各系统中主数据的应 用情况并分析差异，对各应用系统(从主数据需求、应用问题等维度)进行全面管控。

**40.5** **炼化公司智能工厂数据治理实践案例**

**1.** **背景介绍**

炼化公司是延长石油下属公司之一，主要负责石油炼化生产、天然气液化及相关工作。为 了提升精细化管理水平，炼化公司先后建设了财务管理、物资管理等业务系统。这些信息系统。 在公司内部的信息管理中发挥了积极的作用。但是，由于这些信息系统之间缺乏统一的数据标 准，造成系统之间的数据难以共享和集成应用。同时，由于数据难以共享，在一定程度上制约 了公司业务的发展。为了实现对资源的合理利用及集中管理，做到系统之间的数据有效交互， 炼化公司急需在公司层面建立统一的数据标准，为炼化公司的信息化发展奠定基础。

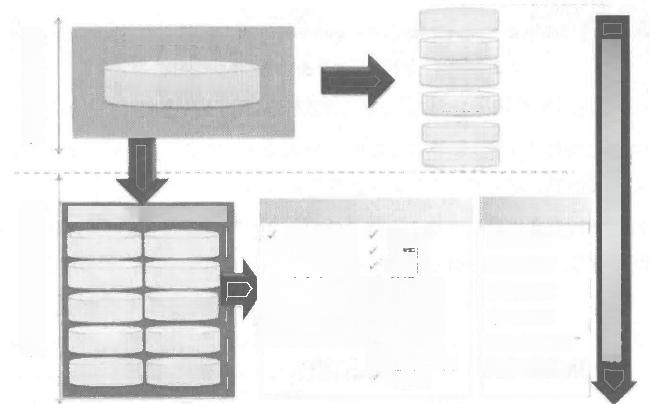
类 类 类

**2.** **建设目标及内容**

本着“统筹规划，标准先行”的原则，并紧密结合公司的实际需求，炼化公司建立了数据

标准化体系，实现对炼化生产数据标准的集中、统一、规范管理。

炼化公司通过制定统一的数据标准，建立统一的数据标准代码库和规范的运维体系，实现 了数据全生命周期管理并提供了标准数据共享服务，为建设智能工厂奠定了基础。炼化公司的 数据标准项目建设内容如图40-5-1所示。



**集团数据治理项目**

**集团通用类主数据**

**炼化公司数据标准化项目**

|  |  |
| --- | --- |
| 装置类  罐区类  装卸台  仓库类  数据相标类 | 侧线类  储罐类  进出厂点类  库位类  设备类 |

**通用基础类**

**人事类**

**单位类**

**物料类**

**财务类**

**指标类**

**组织范围**

·炼化公司 延安炼油厂 永平炼油厂

榆林炼油厂

延安石油化工厂 天然气公司



数据范围

通用基础类 √外部单位类 √内部单位类

集团数据标准 炼化公司数据标准

装卸台类

√进出厂点类 √ 设备类

统 一 信 息 代 鸦 规 则 规 范 ： 信 息 代 弱 库 ， 管 理 机 制

装置岁 侧 现5 料仓料仓

图40-5- 1 炼化公司的数据标准项目建设内容

**40.6** **建设主要内容**

炼化公司智能工厂的全局数据架构主要分为3 部分：业务域应用数据库、企业数据仓库 (ODS) 和大数据应用平台。

(1)业务域应用数据库用于实现交易数据的存储。

(2)企业数据仓库用于实现数据标准化、数据共享及数据分析。

( 3 ) 大 数 据 应 用 平 台 通 过 分 布 式 存 储 用 于 实 现 结 构 化 、 半 结 构 化 和 非 结 构 化 数 据 的 挖 掘 和 分 析 。

炼化公司智能工厂的数据标准化体系框架包括标准化规范、标准化模板、标准化管控3部 分，如图40-6-1所示。

星 兴 面

猛 美

类分爱对

坏排疆故

微指啼推

年曲帝推

区藏

高章照工

排交的能生业退

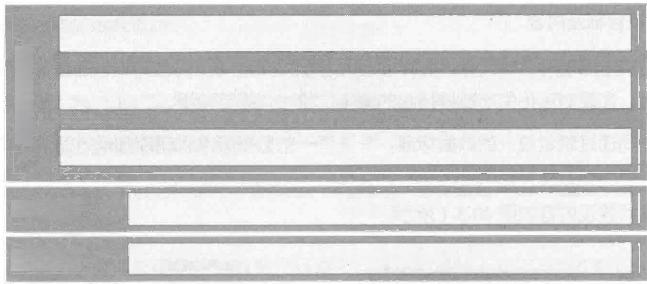
区理管盛

推 繁

播路帷治彩污

标 准 化 规 范

**410 数据治理——工业企业数字化转型之道**

应用标准

数据标准

技术标准

建设模板

**标准化模板**

主数据管理功能

**标准化情控**

**图40-6-1** **智能工厂的数据标准化体系框架图**

(1)标准化规范涵盖应用标准、数据标准和技术标准，用于规范智能工厂的设计与开发过程。

(2)标准化模板覆盖智能工厂各建设阶段，涵盖设计、开发、实施共3类模板，用于沉淀

智能工厂项目建设经验，助力智能工厂的快速推广和建设。

(3)标准化管控聚焦主数据管理，涵盖“总部一企业”两级的主数据管理功能。图40-6-2

为延长石油炼油化工生产层面主数据标准框架体系。



运 方 式



籍

计 单 位

能遥碧点

药测福能

应急物器存故占

空发事件耕脚

游 献 色

消 防 鬣

应意以恬

职 黄

职业离言晶囊

职出健师商重段

明弱张

清账糖

率 霸 重

角感代学点

作业莲可塞商

安金表查高题类胜

安 食 区

分 期 分

除 篇 推

采程息

单无能书占

三和节点

进 出 点



装置

故 隆 困

·设备警理

故监瓶象

精性选的值

对 象 性

生 物 科

生指料

生 物 料

其他单位

外部单位

内瓣单位

延长石油炼油化工生产层面 主数掘标准

工

辐码规则

中

盘

术语与定义

区 烟

主数据属性

安全环保类

环请蛋属点

 环集

安 全

质量类

装面台

首 法

鑫 分

能盖

应意

物料

念

线

盘

\*

**图40-6-2** **延长石油炼油化工生产层面主数据标准框架体系**

炼化公司智能工厂的信息模型的设计及应用，可作为智能工厂开发和实施的参照标准，指

导延长石油智能工厂的建设，如图40-6-3所示。

延长石油的信息标准化管理平台建设实现了对炼油化工生产数据标准、流程及相关制度的

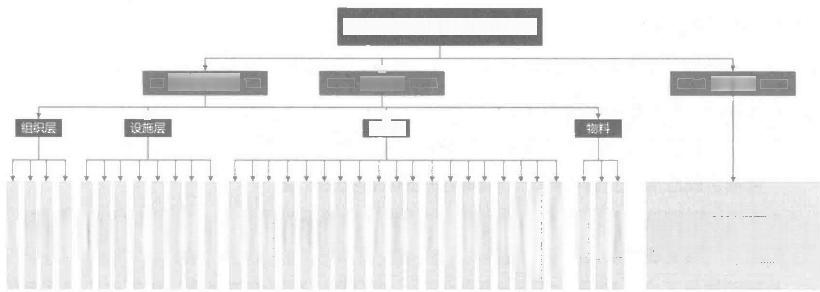
风 睡

器 储

线嘴喻籍

第40章 能源化工行业：数据治理助百年油企数字化转型 **411**

落地执行，为炼化公司的信息化建设提供了基础数据支撑，促进了炼化公司的快速发展。



化工三剂

能源个质

生产物料

应急物资库存点

危险化学品莲存点

职 业 店 点 排故点

密封点

安 圣 惠

安金风脸点 设备

巡检点

无 点

采 样 点

进 画 点

库 位

料 仓

三辆节点

能酒节点

物班测线

件区路施

办公俊施

HSE

系统管量

装卸合料

仓崖

生 装 置

生产车间

生 班 帽

工

企业

87类字典数据

工厂类型、设施类型、装置业务类

型，装置工艺类型，罐区工艺类型

、罐区业务类型、仓库业旁类型、

仓库工艺类型…

延长石油智能工厂信息模型标准

术语与定义

工 节点层

元数据

元模型

工



**图40-6-3** **延长石油智能工厂信息模型标准**

**40.7** **总结与展望**

**1.** **完善组织机制**

根据延长石油数字化转型发展的相关政策，延长石油建立了总部和下层企业联动的数据治 理发展协调机制，明确责任分工，加强资源共享和沟通协作，制定政策措施和行动计划，并加 强数据治理相关应用的组织保障，具体包括以下6个方面。

(1)加强规划统筹，对于涉及面宽的重点大数据应用项目，牵头部门要统筹考虑，加强

顶层设计。

(2)在充分与相关部门协商研究、沟通的基础上，制定综合规划和方案，确定重大项目推 进的工作任务、政策措施等。

(3)加强工作的衔接性，对涉及多个领域的大数据应用项目，根据具体情况，在综合规划 的框架下，由各相关数据管理组织制定分领域或分专题的专项规划。

(4)建立规划及实施保障机制，根据规划确定的工作任务，各相关数据管理组织制定具体 实施方案并精心组织实施，使各项工作任务落到实处。

(5)各数据管理组织密切合作，共同推进规划的实施，并分阶段对规划实施情况进行跟进 评估、调整完善。

(6)牵头部门要加强跟踪各部门的协调联动事项办理情况，定期对协调联动工作会议确定

**41 2** 数据治理—— 工业企业数字化转型之道

的工作任务完成情况进行通报，形成有效的跟踪及督办机制。

**2.** **健全规章制度**

延长石油制定了数据资源保护和开放的制度性文件及信息资源管理办法，加强对大数据知 识产权的保护，具体包括以下内容。

●制定延长石油数据流通共享规则，推进流通环节的风险评估，支持通过第三方机构进行 数据合规应用的监督和审计，保障相关主体的合法权益。

●强化对关键信息基础设施的安全保护，推动建立关于数据流动的规章制度和管理机制， 加强对重要敏感数据流动的管理。

●推动对员工个人数据、企业数据的保护，健全网络数据的防泄露、防篡改和数据备份 等安全防护措施及相关管理机制，加大对数据滥用、侵犯个人隐私等行为的管理和惩 戒力度。

**3.** **加大支持力度**

根据延长石油的数字化转型战略实施和数据应用发展的相关政策，延长石油在财政、技术、 人员保障、基础建设等方面加大扶持力度，具体包括以下内容。

●制定大数据资金扶持政策，优先支持重点大数据应用示范类和创新研发类项目，优先支 持相关企业参与申报国家专项资金项目。

●引进云计算、人工智能及区块链等信息技术服务，为总部和下属企业的大数据应用创新 提供技术支持。

●推动大数据服务基础设施建设，为总部和企业的数据分析应用提供硬件环境和平台基础。

**4.** **加快人才培养**

延长石油先后组织了上百次的业务培训，以及面向全集团100多家单位(涉及上千余名工 作人员)的数据治理培训，提高了员工对于大数据的整体认知和应用水平，建立了多层次、多 类型的大数据人才培养体系。

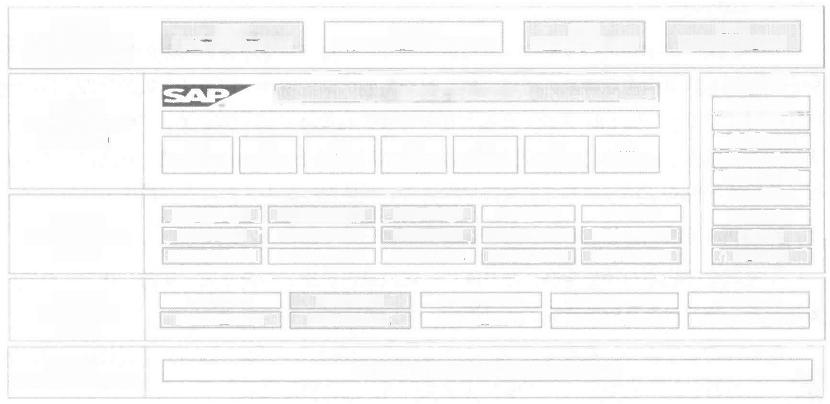
第41章

钢铁行业：酒钢集团数据治理实践

**41.1** **背景介绍**

**1.** **公司介绍**

酒泉钢铁(集团)有限责任公司(以下简称“酒钢集团”)始建于1958年，是黑色金属与 有色金属并举的多元化现代企业集团。经过60多年的建设发展，酒钢集团已初步形成钢铁、有 色金属、电力能源、装备制造、生产性服务业、现代农业六大产业板块协同发展的格局。图41-1-1 为酒钢集团公司信息化五级系统架构应用全景图。



CRM

客户关系管理系统

其他应用

协同办公系统 企业门户系统

视的

质量管理

智能卡管理系统

电子交易系统

资金管理系统 电子商务 企业私有云

发电控制系统

电解铝模糊控制系统

(PCS)

一级系统

基础自动化系统(PLC)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 选绕焦MES系统  炼铁MES 系统  碳钢炼钢MES系统 | 碳钢热书MES 系统  玻间冷轧MES所统 碳钢罐择MES 系统 | 炼轧MES 系统  榆用MES系统 不锈钢MES 系统 | 物流管理系统  物资动力计量系统  实验室管理系统 | 电厂SIS系统  电解铝MES系统 铁合金MES系统 |

三级系统

生产制造执行系统

(MES)

二级系统

过程控制系统

五级系统

企业间管理系统及 商务智能系统(BI)

四级系统

企业资源计划管理 系 统(ERP)

BW/BO

数据仓库/企业决策支持系统

实现了钢铁、电解铝、电厂、铁合金基础自动化全覆盖

不锈构体轧控制系统 不佛胸冷轧控制系统

选矿控制系统

烧结集化控制系统

SRM

供应商关系管理系统

碳阿热轧控制系统 碳阴冷轧控制系统

SEM

企业战略管控系统

炼铁控制系统 碳钢炼钢控制系统

PLM 产品生命周期管理

MM

采购、库存

SD

销售管理

HR

人力资源

PM

设备管理

PP

生产计划

PS

项目管理

Fl/CO 财务/成本

QM

**图41-1-1** **酒钢集团公司信息化五级系统架构应用全景图**

**414** 数据治理 工 业 企 业 数字化转型之道

**2.** **信息化现状**

随着酒钢集团信息系统建设步伐的不断加快，其下属公司及部门的信息系统应用数量正在 不断增多，系统之间的数据横向共享、纵向交互的需求也在逐渐增加。酒钢集团迫切需要加强 数据标准化管理工作，通过数据集成接口，实现对各系统之间的公用数据进行统一、集中管理。

**41.2** **项目目标**

酒钢集团的数据治理项目的目标是建立酒钢集团主数据标准体系、主数据管理平台，实现 对主数据标准的统一、集中、规范管理，最终实现“数据同源、规范共享、应用统一、服务集

中”,具体包括如下内容。

(1)建立适用于酒钢集团的主数据(物料、客户、供应商)标准、物料主数据代码库。

(2)建立主数据标准体系，包括运维组织机构、流程、制度等内容。

(3)建立适用于酒钢集团业务需求的主数据管理平台。

(4)在酒钢集团范围内针对新物料、客户、供应商推行主数据标准。

**41.3** **项目实施**

**1.** **存在问题**

长期以来，酒钢集团缺少专门从事数据治理及主数据标准体系建设方面的管理机构，这就 导致物料、客户及供应商等基础数据未能在整个集团内得到统一、规范和有效的治理。酒钢集 团的主要信息系统(如 ERP 系统)普遍存在一物多码和描述不统一的现象，这给酒钢集团及各 单位的经营数据的统计和分析造成了障碍，严重影响了统计分析、财务核算、财务合并的准确 性，也极大地制约了各信息系统之间的互联互通。具体包括如下3个方面问题。

(1)物料主数据面临的问题。

① 酒钢集团的物料种类复杂，涉及面广，包括黑色金属、有色金属、医药、化工、食品等 多种行业，物料分类及标准建立工作难度大。

② 物料数据量庞大，现有 SAP 系统中的数据总量达到67万条，同时，无效数据(重复数 据、不完整数据等)数量庞大，需要对数据进行科学的分析及清洗。

第41章 钢铁行业：酒钢集团数据治理实践 **41 5**

③酒钢集团在物料主数据管理上采用品名规范的管理方式，其中存在品名交叉重叠、无法 层级化管理和分析、无法控制品名数量快速增长等问题。这导致其自主开发的 MDM(主数据 管理)系统从2014年上线至2019年，已产生约3万多条品名规范，物料主数据增量达到约7 万条/年。

④现有 MDM系统中的查询功能欠缺，缺少数据重复性校验、数据质量管理等关键功能， 这些都有待于在新系统中提升。

(2)客户主数据面临的问题。

①现有系统中的客户数据量约为1.6 万条，其中存在重复的客户数据，客户主数据中的某 些关键字段值存在空值或错误的情况，这些数据都需要进行清洗或标准化。

②客户数据的维护方式是，由工厂的业务人员将需要创建的数据发送给信息中心的运维人 员，运维人员通过SAP系统进行维护。

③客户账户组数量多，且各组之间的界限不清晰，需要整理并调整。

(3)供应商主数据面临的问题。

①供应商主数据主要分为两类：采购类供应商主数据和财务类供应商主数据，这两类主数 据存在以下问题。

采购类供应商主数据：从电子交易平台录入，供应商在投标前自行注册，填写关键信息并 提交资质文件照片，经过相关人员审核通过后生成供应商编号；在交易平台中，标的供应商通 过手工方式将数据录入SAP 系统。

财务类供应商主数据：由工厂的业务人员通过 Excel 将新增的供应商需求提交至信息中心 的运维人员，运维人员通过手工方式录入 SAP 系统。

②系统中的供应商数量约为1.7 万条，其中存在部分重复的供应商数据，供应商主数据中 的某些关键字段值存在空值或错误的情况，这些数据都需要进行清洗或标准化。

③供应商账户组数量多，且各组之间的界限不清晰，需要整理并调整。

**2.** **解决方案**

面对以上问题，酒钢集团的具体解决方案有以下5个。

(1)建设一个成熟的主数据管理平台(见图41-3-1)。通过引入与应用推广平台，替代酒钢 集团原有自主开发的 MDM 系统，对主数据进行规范、高效、集中的管理；结合酒钢集团的主

**416** 数据治理——工业企业数字化转型之道

数据管理现状及需求，对新平台进行部分优化及改造， 建立适应酒钢集团实际业务应用的主数 据管理平台，提升酒钢集团主数据管理平台的功能。



门户管理 标准文档 待办事项 相关链接

动态模型 主数据维护管理 系统管理

主数据模型 数据表 申请 审核 查询 用户管理 角色管理

视图设置 属性列表 冻结 解冻 修改 权限管理 资源管理

字典表模型 字典表 归档 校验 分发 日志管理 参数管理

通用基础类 物料类 外部单位类 银行类

PI

SAP OA 电子交易平台 门户

**图41-3-1** **主数据管理平台**

(2 )编制一套成熟的主数据标准及建立主数据运维体系。通过引入主数据标准及运维体系， 酒钢集团建立了一套适合钢铁冶金行业的主数据标准及运维体系，以及一套适合酒钢集团的标 准化管理组织及管理制度，为持续提升主数据管理做好制度保障。

(3)形成一套标准代码库。酒钢集团引入和借鉴了标准代码库(以物料主数据为核心),减 小了大规模组织专业人员集中进行数据治理工作的实施难度，缩短了实施周期，提高了实施效 率；结合引入的标准代码库，通过对主数据编码进行分析、规划、改进和完善，形成了适合酒 钢集团业务需求的标准代码库。

(4)实现与业务系统的接口集成。酒钢集团在主数据标准及主数据管理平台实施的过程中， 做好接口集成工作，实现了物料主数据、客商主数据通过PI系统到业务系统(SAP 系统、电子 交易平台)的贯通及无缝衔接，以及物料主数据管理平台和协同办公平台的集成(待办推送和 返回)。

(5)培养一批主数据专业人才。在项目建设与实施过程中，酒钢集团通过各种方式让关键

用 户 和 内 部IT 人员学习技术、管理经验和实施经验，以及贯彻标准体系。

第41章 钢铁行业：酒钢集团数据治理实践 **41 7**

**3.** **实施效果**

酒钢集团的数据治理项目已取得了如下成效。

(1)建立了主数据标准。酒钢集团建立了通用基础类主数据标准(中华人民共和国行政区 划、世界国家及地区、运输方式、币种、计量单位)及物料主数据标准、外部单位主数据标准， 使各类主数据有标准可依，提高了主数据的规范性。

(2)实现了对重要主数据的集中管理。酒钢集团对通用基础类主数据、物料主数据、外部 单位主数据及其标准实现了集中管理，提高了主数据的真实性、准确性。

(3)实现了数据同源。酒钢集团建立了主数据管理平台，取消了ERP 系统、电子交易平台、 MES(制造执行)系统中的主数据创建功能，使主数据管理平台成为酒钢集团中唯一的主数据 来源系统，保证了主数据在各系统中的唯一、共享。

**41.4** **项目总结**

**1.** **项目成果**

酒钢集团的数据治理项目主要达成了以下目标。

(1)建立主数据标准及代码库。

①建立酒钢集团4类主数据标准，包括通用基础类主数据标准(中华人民共和国行政区划 主数据代码标准、世界国家及地区主数据标准、运输方式主数据标准、币种主数据标准、计量 单位主数据标准)、物料分类主数据标准、物料描述主数据标准、外部单位主数据标准(客户、 供应商主数据标准)。

②建立酒钢集团主数据代码库，包括通用基础类标准代码3797 条、物料分类主数据代码

5408条、物料描述主数据代码9158条、客商主数据代码337条、银行主数据代码14万余条。

(2)建立主数据管理体系。

①建立覆盖集团及成员单位的数据标准化工作机构，明确各成员单位的总负责人、机构负

责人、专业分组负责人及分组成员(共40个机构，138位负责人，360余位分组成员)。

②梳理并制定5个主数据标准维护流程，包括通用基础类主数据维护流程、物料主数据新

增流程、物料主数据修改流程、外部单位主数据新增流程、外部单位主数据修改流程。

③编制及发布酒钢集团主数据管理制度，以及编制物料主数据、外部单位主数据的提报指

**418** 数据治理 工业企业数字化转型之道

南、操作手册。

(3)建成了适用于酒钢集团业务需求的主数据管理平台。

(4)在酒钢集团范围内推行对新物料主数据、外部单位(客户、供应商)主数据标准。自 2019年9月20日起，酒钢集团已全面推行新的物料主数据标准。

(5)完成2017—2018年创建的84450条旧物料数据的治理工作。截至2019年9月20日， 共完成5123条物料数据治理，冻结10584条物料数据，其他物料数据按照《酒钢集团公司信息 中心关于在集团范围内全面推行新物料标准和禁止使用旧物料编码的通知》要求治理：自9月 20 日起，禁止在电子交易平台中使用旧物料编码，所有采购业务使用新物料编码；自12月1 日起，所有生产下线的产品、半成品、中间产品使用新物料编码；自2020年1月1日起，所有 物流业务禁止使用旧物料编码，旧物料编码统一被冻结。

**2.** **技术创新点**

酒钢集团的数据治理项目具有以下技术创新点。

(1)建立了适用于钢铁行业的主数据标准及管理体系。

此次数据治理项目建立了适合钢铁行业的物料代码、客户及供应商代码、世界各国和地区 名称代码、货币代码，以及计量单位、运输方式、银行联行号等标准，并且通过建立酒钢集团 的主数据管理制度、流程及主数据标准化工作机构，为主数据标准在酒钢集团范围内的推行和 落地提供了保障。

(2)实现了主数据相关的多系统集成和改造优化

此次数据治理项目实现了MDM 系统与PI、ERP、MES、OA、 电子交易平台等系统的集成， 解决了多个关联信息系统之间的数据传输与分发问题；通过对不锈钢 MES 系统、碳钢冷轧 MES 系统、电子交易平台、运输计划系统、线棒仓储系统、劳保超市系统等进行适应性改造，使之 适应新的主数据标准。

(3)规范并统一了信息系统的物料解析模式。

此次数据治理项目使得各物料编码应用系统均通过解析物料描述特征量进行物料特性的逻 辑判断，废除了以往有意义的编码方式，全部启用无意义的流水码编码方式，提升了物料编码 在集团范围内的通用性和适应性。

第41章 钢铁行业：酒钢集团数据治理实践 **41 9**

**3.** **经济效益**

酒钢集团的数据治理项目主要取得了以下经济效益。

(1)清库存，降低成本。通过规范物料主数据，能够有效反映库存量，为制订物资需求计 划提供准确的数据支撑；通过与各业务需求无缝集成，及时提供所需物资，有效减少各环节的 库存，达到清库存，降低成本的目的； 按照纳入集团公司ERP 系统核算的单位2018年年末辅 助材料、备件库存余额计算，在实施主数据标准化后，每年可降低库存余额1%,每年可减少辅 助材料、备件库存占用资金327.52万元。

(2)降低采购成本。在采购过程中，酒钢集团以统一、标准、规范的物料主数据作为基础， 规范招/投标流程、提高采购过程的透明度，减少采购过程中的舞弊行为，发挥集中采购优势， 从而降低了采购成本；按照纳入集团公司 ERP 系统核算的单位2018年年末辅助材料、备件库 存采购金额计算，在实施主数据标准化后，每年可降低采购金额0.5%,每年可减少辅助材料、 备件采购金额1300万元。

**41.5** **未来展望**

酒钢集团的数据治理项目在取得以上成果的同时，在未来仍需面临着一系列的挑战。

(1)主数据标准需要不断完善。

酒钢集团的经营业务覆盖钢铁冶金、铝业、装备制造、电力能源、酿酒、建筑、医疗等多 个产业，因此，主数据标准覆盖面广、类型多，在应用的过程中需要不断修正和完善，逐步形 成各个产业的标准化分类标准和物料描述标准。

(2)物料主数据治理工作长期而艰巨。

在本次项目实施前，原有的 SAP 、ERP 系统中已产生70万条旧编码数据，需要根据业务 的迭代过程不断对旧编码有计划地开展治理工作，最终实现新旧编码的替换。

**第42章**

**汽车行业：数据驱动长安汽车数字化转型**

**42.1** **背景介绍**

**1.** **公司介绍**

重庆长安汽车股份有限公司(以下简称“长安汽车”)在全球拥有16个生产基地、35个整 车及发动机工厂，并与41个国家和地区有合作关系。长安汽车以数字化、信息化、自动化为基 础，以平台化、轻量化、精益化为抓手，集成大数据、云计算、人工智能、物联网技术，以实 现高质量、柔性制造，快速满足客户个性化定制需求。

**2.长安汽车数字化发展历程**

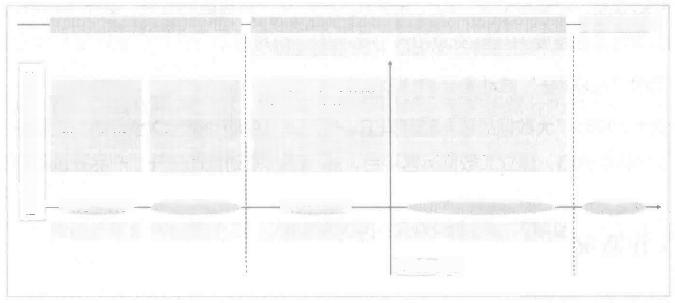
汽车行业的数字化建设历程分为信息化、数字化、智能化3个阶段，其中数字化建设阶段 可被细分为数字化转型、数字化重塑两个阶段。长安汽车从2001年开始信息化建设，通过数字 化战略尝试进行数字化转型，在2010年开始进入数字化转型阶段，在2019年开始进入数字化 重塑阶段。图42-1-1所示的是长安汽车数字化建设历程。

2000年是长安汽车信息化建设的一个分水岭。因为业务战略的转变，长安汽车此前建立的 分布在各部门、各领域的信息系统已经无法满足新的业务发展的需要。为了长远发展，长安汽 车随后成立了专门的信息公司，开始全面推进信息化建设。

长安汽车的信息化建设经过十余年的发展，基础数据建设依靠内部系统已建设完成，但各 系统中的数据如何保持统一的源头，并成为公司的数据资产，成为当务之急。2011年，长安汽 车开始建立汽车超级 BOM 平台，实现了产品策划部、产品部、制造中心、生产工厂、销售公

第42章汽车行业：数据驱动长安汽车数字化转型 **421**

司、售后服务部等全环节覆盖，以及汽车的产、供、销全链条的数据化连通，并以此展开主数 据管理建设。2016年，为实现价值链升级，以及从制造型企业向制造服务型企业转型，长安汽 车规划了“以大数据推动第三次创业”的宏伟目标。



信息化(T)

数字化转型

·基本完成各业务领域数

·流程

与出数、化监业控务、

信息安全平台

2010—2018-

创新协同

Here we are

·数据驱动管理

·客户体验数字化支撑 ·创新业务孵化

·生态合作、数字化生态圈

·初步开展研发、 制造、供应链、

·务

管理等数字化

系统

·进一步拓展并 推广研发、制

造、供应链数

·新增管

理、营销数字 化工具

2001—2005-

综合集成

2019—2025

数字化长安

2006—2009

管理融合

长 安 汽 车 数 字 化 建 设 历 程

智能化(T)→

数字化(DT)

数字化重塑

20XX—

**图42** **-** **1** **-** **1** **长** **安** **汽** **车** **数** **字** **化** **建** **设** **历** **程**

**3.** **数据治理背景**

(1)汽车消费市场基于环境发生了巨大的变化。

如今，汽车消费市场已进入中低速增长阶段，并且逐步进入刚性消费与消费升级并行发展 阶段，汽车消费者更加注重品牌、舒适性和操纵便利性。在这样的背景下，汽车消费市场的几 个主要特征发生了明显变化，主要表现在以下3个方面。

一是客户变了。“80后”“90后”用户逐渐成为中国汽车消费市场的主力军，他们更关注产 品的个性化和产品的使用体验。

二是产品变了。电动化、智能化、互联化、共享化逐渐成为汽车产业技术和商业模式进化 的新方向。

三是技术变了。人工智能、3D 打印、物联网、区块链等新技术正在颠覆一切，包括汽车 产业。

因此，作为汽车制造厂商，想要在日益激烈的市场竞争中占得一席之地，就必须得提高对 外部市场的反应能力，提高对消费者的服务能力，而其根本则是提高企业内部的运营效率。

(2)传统汽车行业迎来由IT 向 DT 转型的机遇。

自进入21世纪以来，以云计算、大数据、人工智能等为代表的前沿新技术掀起了一股技术

**422** 数据治理——工业企业数字化转型之道

浪潮，并在各大领域中得到了广泛的应用。在此背景下，传统制造行业的“以客户为中心”的 新营销模式优势日益突出，“以产品为中心”的营销模式已不再具备竞争力。此时，企业迎来由 IT 时代向 DT 时代转型的机遇。企业的“以流程为核心”也开始向“以数据为核心”转变，以 及“从业务到数据，功能是价值”的传统理念也开始向“从数据到业务，数据是价值”的理念转 变。数据已成为企业重要的战略资源和核心竞争力，也是驱动行业和企业转型升级的重要引擎。

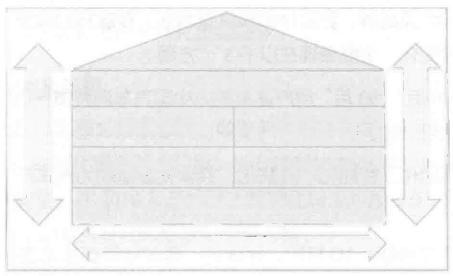
(3)借助“互联网+”推动第三次创新创业。

在国家大力推动“大数据战略”等背景下，长安汽车启动“第三次创业——创新创业计划”, 并制定数字化转型战略，建立大数据运营平台，通过数据驱动管理提升、产品升级和构建生态圈。

**42.2** **工作概况**

**1.** **建立数据治理架构**

2017年6月，长安汽车启动数据治理项目，成立CA-DDM (长安一数据驱动管理)项目团 队，并确定长安汽车的数据治理框架(见图42-2-1),以及数据治理目标，以保障公司数据的安 全和质量，提高数据的使用价值，真正实现数据是企业的核心资产。长安汽车数据治理的总体 目标如下。



数据治理

数据架构管理

数据质量管理 数据安全管理

元数据管理 主数据管理

数据生命周期管理

统一平台

统 一 运 营

统 一 数 据

**图42-2-1** **长安汽车的数据治理框架**

(1)满足并超越公司和所有利益相关者的信息需求，提高数据的使用价值。

(2)确保数据资产的安全性。

(3)持续提高数据和信息的质量(准确、及时、 一致、完整、关联、实用)。

(4)对数据资产价值有更广泛和更深入的理解，保持信息管理的一致性。

第42章 汽车行业：数据驱动长安汽车数字化转型 **423**

**2.** **明确运营工作思路**

长安汽车通过统一平台、统一数据、统一运营的数字化管理手段，将数据治理的成果应用 于生产实践中，实现基础管理标准化、业务数据财务化、经营成果指标化，具体介绍如下。

通过统一平台，打通数据孤岛，实现数据的全面、正式、透明、共享。

通过统一数据，统一数据标准，建立指标体系，规范业务行为。

通过统一运营，强化数据应用、创新业务运营模式，支撑运营分析科学化，业务改善持

续化。

(1)统一平台。

长安汽车建立了大数据分析平台，为公司决策提供数据支撑。大数据分析平台的价值具体

体现在以下两个方面。

融合外部数据，精准营销：对市场进行全局掌控，把握客户、产品及市场变化，实时采集 主流汽车网站、论坛等数据，及时掌控行业市场格局，为改进产品、优化服务和提升销量提供 全方位的决策支持；整合用户数据进行用户画像，构建用户全景视图，形成识别客户的量化指 标，为营销活动提供精准支持。

运营车联网数据，提升用户体验：实时采集车联网数据，通过分析用户评价等，促进产品 改进；通过分析用户操作数据、车辆故障信息、车辆运行状态，为用户提供贴心的服务。

(2)统一数据。

统一数据：旨在整合并完善公司数据体系，即对标最佳实践，建立数据管理体系，从而推 进企业价值链全面数字化，实现数据资产的有效利用和价值挖掘。

建立数据指标体系，即结合市场洞察和创新需要，对公司战略进行拆解，构建数据指标体 系，促进以客户为中心的数据驱动管理；从服务、质量、效率、效益、成长5个方面建立响应 指标，促进企业成长。

开展数据治理：统一数据标准、口径和计算逻辑，重点对数据指标和主数据开展治理和重 新审视并发布相关标准，以规范业务行为，提升数据质量。

(3)统一运营。

实现数据运营，用数据说话，建立实体+虚拟的运营组织，制定运营机制，开展全价值链的 数据运营：通过数据运营发现问题，持续推动管理提升；逐步将运营结构化、自动化、智能化，

沟通与推广

提高整体运营效率。

开展大数据场景建设，为精准营销、产品策划、电商应用、客户运营等应用场景，提供精 准“炮弹”:以数据中台为核心架构，对数据进行组件化、服务化，推进自助分析能力建设，全 面提升员工的数据分析能力，为业务赋能。

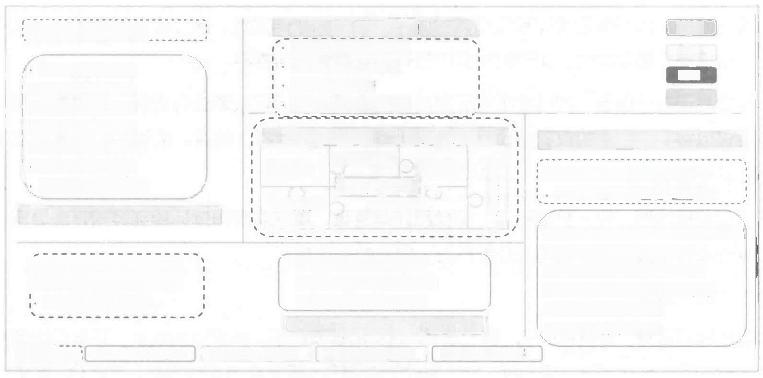
按“看数据→信数据→用数据→数据文化”的思路，建立数据文化氛围：在全公司范围内 推广通过大数据平台进行用户运营；综合运用培训、宣传、竞赛、抽查、指标评价等多种方式， 建立数据文化氛围，推动以数据分析手段改善公司业务运营质量、洞察客户及市场机会，鼓励 员工掌握数据分析技能。

**3.** **建立数据管理标准体系**

长安汽车通过建立数据管理体系框架，确定了数据治理的范围，规划了长安汽车数据治理 的提升路径；通过建立数据管理标准，把数据治理工作落到实处。

(1)建立数据管理体系框架。

建立数据管理体系框架，以数据治理为核心展开各项工作，将数据治理作为规划、控制和 提高数据及信息资产的一个业务职能，是有效利用数据开展运营的基础保障。为推动基于大数 据平台开展业务运营，长安汽车创建了数据管理体系框架，其中涵盖了数据架构管理、数据安 全管理、数据质量管理、主数据管理、元数据管理、数据生命周期管理等领域，如图42-2-2所 示。



DA.数据架构管理

DAI 建立企业数据模型

DA2. 信息价值链分析

DA3. 数据流分析

DA4 维护数据技术架构

DS.数据安全管理 DQ 数据质量管理

项目和服务

数据估值

DG. 数据治理

组织和角色

· ML4. 元数据创建和维护

·ML5. 元数据采集存储管理

·ML6. 元数据应用及分析

MT, 主数据管理

2—行业一般水平 3—行业先进水平 4—行业最佳水平

计划

建设

控制

运营

MD.主数据管理

·MD1. 定义和维护主数据架构

MD2. 定义和维护主数据匹配规则

·MD3. 定义和维护主数据层次及关联关系 ·MD4. 理解主数据需求及分析

·MD5. 计划和实施数据源的整合

·MD6. 设计及实施主数据解决方案

·MD7. 主数据发布管理

· MD8.主数据变更管理

DL!. 数据规范和定义.

DL2. 数据采集管理

·DL3. 数据储存管理

·DL4. 数据整合管理

·DL5, 数据展现管理

·DL6. 数据分析与应用

·DL7. 数据运营管理

DLB.数据归档与销毁

DL 数据生命周期管理

{MT1. 定义和维护元数据架构 |:MT2. 定义和维护元数据标准

MT3.元数据系统管理

0二( 1—行业基础水平

政策和标准

数据战略

**图42-2-2** **长安汽车数据管理体系框架**

第42章 汽车行业：数据驱动长安汽车数字化转型 **425**

(2)建立数据管理标准。

①基于长安汽车的数据治理项目的总体规划与体系框架，长安汽车累计制定并发布了25 项数据管理标准，包括《数据质量管理规范》《数据展现标准管理程序》《指标通用技术规范》 《指标评价管理程序》《指标建设流程规范》等，有效地指导长安数据管理体系建设、数据治理 及后续推广等工作的开展。

②长安汽车根据数据管理标准，完善了数据管理成熟度评估标准(见图42-2-3),持续按季 度对各业务领域开展数据管理成熟度评估，以提升数据管理成熟度。

③长安汽车依据数据管理成熟度评估结果，推动各业务部门开展主数据治理，明确数据责 任部门。

数据管理成熟度

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 评估  等级 | 初始级 受管理级 稳健级 量化管理级 优化级 | | | | |
| 分值  区间 | 1.00~1.99 | 2.00~2.99 | 3.00~3.99 | 4.00~4.99 | 5 |
| 等级  特征 | 以项目集体现 缺乏统一的被 动式管理 | 意识到数据的 重要性，要求 制定相关流程 | 数据反映组织  绩效目标，制  订管理体系 | 数据作为竞争 优势的来源  量化分析、监控 | 数据作为竞争 生成的基础，  持续改进和提升 |

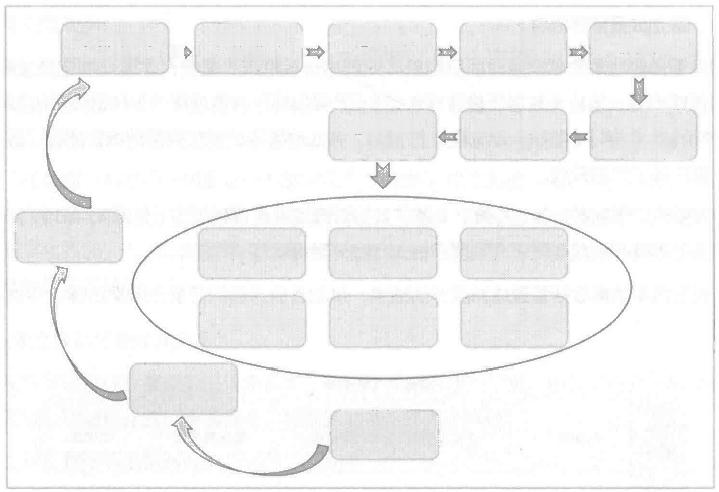
**图42-2-3** **长安汽车数据管理成熟度评估标准**

④长安汽车建立了数据运营机制，定期发布运营分析报告。

**4.** **数据管理实施方法和路径**

长安汽车以数据架构管理、数据安全管理、数据质量管理、主数据管理、元数据管理和数 据生命周期管理为中心，并结合业务应用的具体情况，采用 PDCA 循环(计划、执行、检查、 处理)模式进行测量和改进，如图42-2-4所示。

**426** 数据治理— — 工业企业数字化转型之道



定义数据业务

范围 

制定数据治理 标准

持续改进

数据安全管理

元数据管理 主数据管理 数据架构管理

评估与监控

运行与测量

数据生命周期 管理

建立元数据

仓库

开展成熟度 评估

搭建数据 管理框架

建立组织

蓝图

建立实施 路线图

数据质量管理

数据调研

**图42-** **2-4** **长** **安** **汽** **车** **数** **据** **管** **理** **实** **施** **方** **法** **和** **路** **径**

**42.3** **组织保障**

大数据运营体系建设工作是长安汽车第三次创业的重点举措之一，此项工作自启动以来， 一直受到高层领导的重视和关注。为有效推进与持续开展大数据运营体系建设工作，长安汽车 组建了由高层领导挂帅、业务“一把手”牵头、核心骨干员工参与的矩阵式团队，其中包括领 导小组、总体推进组和13个业务领域团队，有来自30多个业务部门共300多名专/兼职人员参 与，其中管理层人员占比接近50%。

(1)领导小组：负责确定大数据运营体系建设方向，制定总体规划及推进策略，对项目进

行评价并提出建议，解决在项目推进中存在的重大问题，保障项目的各项资源配置。

(2)总体推进组：负责统一管理大数据运营体系建设、确定子项目推进目标、解决项目相 关问题，同时负责相关培训和辅导，对各子团队的项目推进情况进行评价与激励。

(3)各业务领域团队：负责各领域团队的搭建、推进目标的制定和分解、推进各领域的项 目实施，以及开展领域运营管理并持续改善。

第42章 汽车行业：数据驱动长安汽车数字化转型 **427**

**42.4** **项目成果**

长安汽车的数据治理工作经过持续不断的沉淀和积累，在数据驱动企业数字化转型升级方 面取得了众多成果，具体有以下几个方面。

**1.** **建立统一的数据管理体系**

(1)建立《CA-DDM 运营评价标准》:长安汽车制定并发布企业数据治理标准及管理办法 等相关文件和规范共计24篇。

(2)初步建成企业内部指标体系：长安汽车基于《指标评价管理程序》《指标编码规范》文 件和规范，进行指标体系建设，累计实现用户大数据、集团运营等板块中的448个一级指标建 设。

(3)初步建成主数据管理体系：长安汽车制定并发布《产品主数据管理》《经销商主数据管 理》《生产型供应商主数据管理》等文件，并通过开展产品主数据质量运营，每周发布产品主数 据质量周报，通报产品主数据质量(在首批试点部门中，产品主数据领域有3 个事业部和15 个业务单元被纳入数据质量通报范畴),实现对产品主数据的系统化管理，使产品主数据质量的 完整率从93%提升到 99.7%。

(4)构建数据管理能力成熟度评估模型：长安汽车根据国家标准，完成了数据管理能力成 熟度评估标准及细则细化，以及8大能力域的28项能力评估细则的编制；构建了初始级、受管 理级、稳健级评估模型，确保数据管理落到实地。

**2.** **有效支撑数据治理相关工作**

在建立数据管理标准的基础上，长安汽车基于“绝情抓数据，诚信不妥协”原则，发布《数 据治理周报》《数据质量红黑榜》,有效地支撑了数据治理的相关工作，具体包括以下内容。

(1)指标治理持续开展：长安汽车以《指标及数据项检核评价》为标准，在2019年累计组 织291场业务项目指标评审，有1249人次参与评审，319个指标通过检验与审核。

(2)产品主数据治理初见成效：长安汽车对3年内产生的3.1 万余条生产、销售、服务方 面的产品主数据的41项属性进行数据收集、整理、清洗、入库、发布，确保属性值的完整、准 确、规范，满足各业务单位对产品主数据的使用需求。

(3)其他基础治理工作并行推进：长安汽车统一了经销商主数据共61项维度属性的定义及 9大属性项的逻辑关系，确定了44项供应商主数据属性项，并明确各属性项的定义，统一了属性

**428** 数据治理 教字化转型之道

项列表、属性值，有效支撑了产品研发、销售、采购、财务等业务的开展。

**3.** **整合企业数据，建立大数据平台**

长安汽车整合企业内部核心业务系统，通过融合互联网数据、统一客户数据并集成产品数 据，构建了集数据管理、分析、应用为一体的大数据平台，服务公司经营，以实现“现状可见、 问题可查、风险可辨、未来可测”,促进数据的全面、真实、透明、共享。具体包括以下内容。

(1)完成数据整合。

①整合企业数据：长安汽车将经营结果指标化、业务数据财务化。其全面整合了100 多个 内部核心业务系统、300多项结果指标、3000多个关键数据项及2000多个业务报表，准确计量 和评估业务经营活动的投入、产出，实现“业财融合”及一体化运作。长安汽车从集团战略出 发，将集团战略指标分为3层(决策层、管理层、操作层)并层层深入分析，直击产生问题的 根本原因：

决策层通过看行业，找差距，辅助领导找到经营决策的关键核心指标，解决经营痛点；

管理层通过看目标，找问题，帮助决策层找到发现问题、定位问题的核心指标，找到问题 的改善点；

操作层通过看问题，找办法，从专业层面解决具体业务的问题。

②融合互联网数据：长安汽车通过互联网集成技术，结合自然语言处理技术，并融合10000 多个外部网站数据，打造了集客户洞察和产品分析(产品剖析+质量口碑)、新技术趋势分析等 于一体的在线分析平台。

③统一客户数据：长安汽车打通了18个与客户触点相关的内部 IT 系统并形成客户个体画 像，以及融合外部数据形成客户群体画像，通过线上、线下活动打通个体画像和群体画像，形 成统一的客户画像。长安汽车还融合2800万条长安汽车用户数据，构建了客户数据管理、客户 数据服务、客户数据分析三大能力，实现客户数据管理的“干人千面”。

④集成产品数据：长安汽车统一采集产品信息，掌握产品的动态，指导产品研发。长安汽 车还围绕人、车、路开展全方位的信息采集，进行“三位一体”的数据分析，并及时反馈用户 评价、产品状态和整改效果，促进产品改进。

(2)形成企业级平台。

长安汽车构建了集数据管理、分析、应用于一体的CA-DDM (长安一数据驱动管理)大数

第42章 汽车行业：数据驱动长安汽车数字化转型 **42 9**

据平台(以下简称“大数据平台”),对公司数据资产进行管理，服务公司经营，并初步实现了 公司数据管理的全面、真实、透明、共享。大数据平台的总注册用户人数达7200多人，月均访 问量近30万人次，广泛应用于12大领域，推动了60多项管理制度的改善。具体介绍如下。

①完善数据门户，随时随地获取可信任的数据，实现平台的数据化、系统化、透明化管理， 面向全员开放，及时掌握各领域及各部门的大数据运营状态，以及共同建立和持续改善数据文 化氛围。

②构建数据分析工作台，让业务人员可在线进行数据分析，实现效益提升了40%。

③建立数据实验室，研究数据分析模型等，将销售预测、客户洞察、市场洞察等核心模型

投入业务应用中。

**4.** **开展数字化运营，驱动数字化转型**

长安汽车通过建立运营组织，完善运营机制，以数据为驱动，开展内容运营、产品运营、 用户运营，逐步深化数据在线、报表在线、分析在线、决策在线，为业务赋能，促进了业务的 持续创新。

(1)实现报表自动化。

为摆脱手工操作报表，由“报表流程化”向“报表自动化”转变，提升业务效率和数据质 量，长安汽车在大数据平台上建设了15大领域共500多个关键报表，通过系统抓取数据自动生 成报表。各业务部门可按权限访问报表，大幅提升了业务效率。

长安汽车在各业务领域推广“自助分析”,为业务赋能。各业务部门的业务分析人员经过培 训后，可通过大数据平台自行从系统中抓取数据，自定义生成个性化报表，充分满足业务各方 面的需要。目前，长安汽车累计在35个单位中开展了800多人次的培训，为公司输送了大量的 数据分析人才。

(2)用量化指标驱动公司数字化转型。

2017—2019年，长安汽车为了推动公司数字化转型，坚持以客户为中心，将公司战略层层 分解，构建了涵盖产品研发、生产、销售等各业务板块的各层级指标1800多个，并解决了指标 及属性定义、数据来源、数据口径不统一的问题，初步实现了业务工作指标化、共享化。

(3)为业务赋能，助推业务创新。

数据经过治理和整合后，通过运用大数据技术，可突破传统业务模式促进业务创新，例如

深度解析制造过程，提高生产效率；实现多维用户画像，开展用户个性化服务；发现潜在人群

**430** 数据治理——工业企业数字化转型之道

和需求，实现精准营销，深入挖掘潜客资源，具体介绍如下。

①用户360°全景视图。

用户360°全景视图提供了360°全景视图的数据服务，实现了用户标签向客服座席人员的 实时在线输出，帮助长安汽车发现新的商机，提升用户体验。用户360°全景视图通过整合用 户触点、打通数据系统，形成了统一的用户画像。其通过对接客服呼叫中心，根据来电号码智 能生成用户全景视图，对用户的特征及诉求进行预判，使得客服座席人员咨询业务处理时长缩 短8%,提升了客服与用户的沟通效率和用户体验。

②新车改款用户评价分析。

新车改款用户评价分析可以动态跟踪媒体报道热度、报道关注点，智能分析用户对改款车 型的喜恶程度，提取优化建议。整个新车改款用户评价分析过程为在线分析，节省了人力，缩 短了时耗，提升了效率。例如某改款车型发布后，收集到改款报道1000多篇、报道评论20000 多条、个人舆情评论30000多条。如果按照传统的线下调研+手工统计+报告撰写的模式，则需 要1个月以上才能输出结论和建议。现在采用线上大数据+实时统计+智能分析的模式，则只需 要1天即可输出结论和建议。

③新竞品上市跟踪对标分析。

通过实时分析新竞品上市信息，可以捕捉新竞品的讨论热度，智能计算新竞品上市后的市 场格局。

④有效促进工厂之间的“比、齐、追、赶、超”。

通过整合7大基地工厂的生产制造数据，统一标准、统一口径，可以在相同标准下进行对 比分析。这样做一方面可支撑业务分析，供车间、工厂每日早会及公司的每月分析会议使用； 另一方面可促进业务改善，规范工厂业务(例如规范工厂定编管理、停线管理、生产线管理、 班次切换管理、复产开班管理等业务问题)。

⑤车联网数据分析。

通过实时采集车联网客户数据，对客户评价、车机功能的使用情况等的分析，可以促进产 品改进。还可以实时采集客户车联网设备的使用情况、功能应用程度，并收集客户评价意见， 对问题整改进行跟进监控。

第42章 汽车行业：数据驱动长安汽车数字化转型 **431**

**42.5** **工作价值**

长安汽车通过开展大数据运营体系的建设与应用，产生了以下工作价值。

(1)填补了企业在大数据运营领域的空白，并且可推广和复制，产生了良好的综合效益。

长安汽车通过开展大数据运营体系的建设与应用，统一了平台、数据、运营，实现了以数 据为核心开展企业生产经营， 一改以往的“以经验为导向解决问题”“解决的问题重复发生”“因 数据不统一导致的混乱”等局面。长安汽车在真正意义上实现了数字化运营“零”的突破，同 时，通过在内部、外部的推广应用，均取得了直接和间接效益。

通过不断应用和优化大数据运营体系，累计有50多个二级部门运用大数据开展业务运营， 有效地推动了企业数据文化氛围的形成，同时，企业运营效率提升了30%以上，逐步降低了企 业的运营成本。

(2)提升了企业运营决策能力和水平，奠定了企业核心竞争力基础。

长安汽车的大数据运营体系作为企业的一项战略，是一场通过数据驱动管理的变革，最终 在分析、决策、行动和监控等方面形成共同语言和数据文化。对管理者而言，它通过一套完整 的大数据运营管理平台，提高了分析和决策的准确性和效率，让管理和业务的持续改善变得容 易且有效；对员工而言，它能提高员工的工作效率，减少重复劳动。

大数据运营体系主要赋予业务管理者以“数字化思维”“数字化理念”“数字化经营”的模 式，通过参与数据在“现状可见”“问题可查”“风险可辨”“未来可测”各阶段的运营，可以有 效地对企业在市场预测、产品投放、舆情跟踪、质量改进、产品创新、管理创新及市场创新等 方面提供分析、决策，有效地促进了企业运营决策能力和水平的提升，奠定了企业在市场、行 业中的地位。

(3)建立了一套运营体系，支撑企业数字化战略落地，推进了传统制造企业的转型升级。

大数据运营体系的建设与应用能够从真正意义上将互联网数据、企业数据应用于企业运营 分析决策中，并且取得了显著的成效。具体包括以下内容。

首先，以技术为后盾，实现企业数据的统一。大数据运营体系的建设与应用，打破了以往 数百个系统被零散分布为“数据孤岛”的局面，确保了十多个业务领域的数据集成、共享；对 企业内部、互联网、客户、车联网等的有效融合，超越了其他同行业企业的“分散式”“单领域 的建设模式，使企业在行业中处于领先地位，为大数据运营体系平台的有效应用打下坚实的基础。

其次，以“统一运营”为核心，推动长安汽车在营销、财务、人力等方面的内部运营分析，

**432** 数据治理——工业企业数字化转型之道

有效地促进了企业在管理、市场及产品上的创新和持续改善。 一是实现企业内、外部数据的融合， 从经验驱动的管理模式向高效、精准的数据驱动的管理模式转变；二是利用互联网大数据，直达 用户，倾听用户之声，可以更全面、更直接地了解用户对产品的需求，指导企业的产品策划及 产品改进；三是融合内、外部数据，对用户进行画像、洞察，通过自动化分析，可以更加精准 地进行营销服务，提升效率，降低成本；四是依托大数据平台，形成“发现问题一剖析原因一 制定措施一执行监控—效果评估”的闭环运营机制，通过数据驱动业务和管理的持续改善。

长安汽车大数据运营体系的建设与应用，促进了研发、制造、营销、质量、采购、用户大 数据等13个业务领域的工作开展。通过持续开展大数据运营，有效地推动了长安汽车的管理变 革，挖掘了更多的市场机会和商业价值，有效地推进企业的转型升级。

**42.6** **交流分享**

长安汽车在全公司内持续开展业务培训，提升了业务人员在数据管理方面的能力。仅2019 年，长安汽车就开展业务培训25场，覆盖公司35个单位，参加人员达832人。CA-DDM 大数 据平台不仅被应用于自主板块，也逐渐被应用于合资板块。

长安汽车以开放、包容、学习的心态，向先进企业学习，沉淀了大数据最佳实践经验。他 们通过“管创研习社”、案例分享会、大数据研讨会、产品培训等方式，组织了50多起与行业 知名企业、大学、研究机构等关于大数据的交流会、研讨会(其中包括密歇根大学、重庆大学、 北京理工大学等),并参与政府机构及行业组织的各类大数据论坛、交流会(其中包括百度、阿 里巴巴、腾讯、华为、美的等企业),有效地学习和沉淀了先进企业的经验，为长安汽车大数据 运营提供有效的参考依据，同时也加强了长安汽车的品牌影响力。

**第43章**

**核工业：物料主数据治理助力核电智慧运营**

**43.1** **背景介绍**

**1.** **公司介绍**

中国核工业集团有限公司(以下简称“中核集团”)是中央直接管理的国有重要骨干企业之 一，拥有完整的核科技工业体系。

中国核能电力股份有限公司(以下简称“中国核电”)由中核集团作为控股股东，主要经 营范围涵盖核电项目的开发、投资、建设、运营与管理，以及清洁能源项目的投资、开发等领域。

中国核电的 ERP 系统和相关基础工作建设主要依托秦山核电业务团队开展。建设数字化、 智能化核电站是秦山核电当前及未来一段时间努力的方向和目标。而大数据应用更是秦山核电 重点突破的方向，其对于提高核电运行效率，保障安全、经济、高效、清洁的电力供应具有重 要的意义。秦山核电作为中国核电的物料主数据运维管理部门，在数据质量方面秉持着不断改 进的理念陆续开展着各项治理工作。

**2.** **物料主数据管理现状**

自2015年秦山核电的 ERP 系统上线以来，秦山核电的信息化管理水平在不断提升， 曾数次在集团信息化测评中获得A 级水平。回顾秦山核电30多年的物料主数据建设历程， 主要分为3 个阶段：分散管理阶段、整合优化阶段、集中发展阶段。按照“标准先行、流 程支持、优化推广和迭代更新”的建设思路，秦山核电的物料主数据管理已延伸到公司管理的 各个领域，有力地促进了公司管理水平的提高。同时，由秦山核电负责的中核集团和中国核电

**434** 数据治理——工业企业数字化转型之道

的物料主数据标准的建立和优化工作也在不断发展中，并且还存在一些有待优化的领域，具体 包括以下3个方面。

(1)物料主数据管理整体缺乏完善的标准和架构体系(包括物料主数据相关标准和架构的 战略、组织、流程、体系及相关技术手段),不利于集团统筹物料主数据的规划和建设。

(2)秦山核电的数据质量建设虽然历经多年，但数据质量还不能完全满足业务管理的需 求，还未完全融入数据全生命周期管理的各个阶段。

(3)秦山核电的物料主数据应用效果不佳，没有从客户端考虑用户体验，客户接受和使用 程度不高，信息化能力成熟度不足。

为夯实大数据发展战略基础，秦山核电从2015年开始展开物料主数据专项清理工作，并陆 续制定了《物料主数据实物清理专项工作计划》和各类专项工作细则和样例，明确了数据质量 治理的具体实施方法，确定了数据质量建设的内容、目标和规划，以指导数据治理工作。秦山 核电以“钉钉子”的精神，夯实基础，把数据质量治理工作落到实处并取得了实效。

**3.** **数据质量治理背景**

中核集团、中国核电和秦山核电等各级核工业管理公司在开展数据治理工作时，始终把数 据质量和保障能力放在第一要务。不同于其他消费类行业，所有涉及核工业生产的数据必须要 具备高度的准确性和可靠性，以保证相关信息的准确、完整及与实物高度一致，确保生产所使 用的信息万无一失。

秦山核电的 ERP 系统于2015年年初上线。在系统上线前，秦山核电对原有的多个系统内 的各类主数据和业务数据进行了清理并传输到 ERP 系统中，但是因ERP 系统的建设周期限制， 在主数据清理的过程中，损失了质量监督环节，致使ERP系统中的主数据质量不能很好地满足 后续生产及经营管理的需求。为此，秦山核电开展数据质量治理工作。

从2015年7月开始，秦山核电启动物料主数据专项清理工作，成立了公司级专项组，并投 入了大量的人力、物力，对 ERP 系统中的几十万条物料主数据进行专项清理。该专项工作主要 完成两个基本目标：

(1)针对 ERP 系统内的主数据的格式优化，主要内容是格式整理和信息完善。

(2)在物料主数据信息已经比较完善的前提下，开展实物核查工作，保证物料主数据和实 物信息的相符性，并通过系统展示成果，以便多方位保障现场生产所需。

第43章 核工业：物料主数据治理助力核电智慧运营 **435**

**43.2** **工作概况**

秦山核电的物料主数据专项清理工作于2015年7月正式启动，经过充分的调研和分析，确 立了“全面覆盖、统一规划、分步实施、注重应用”的总体原则。

(1)全面覆盖。

本次专项工作针对秦山核电的所有物料主数据，开展全样本的数据规范和影像留存，涉及 数据规范20多万条，影像资料约100万份。

(2)统一规划。

因为涉及的业务部门和人员比较多，还要搭建信息平台和外部支持团队，故需要成立公司 级专项组，按照项目管理模式，统一规划、组织和推动。

(3)分步实施。

因为涉及几百名工程师、几十万条物料主数据，同时考虑到平台和外部团队的准备情况， 以及在保障质量的前提下，充分考虑业务排程的最优化，故该专项工作总体分为两阶段实施。

(4)注重应用。

数据本身没有价值，只有在业务流程中通过信息流来实现价值，故如何把数据清理成果嵌 入业务流中，也是项目组重点关注和必须实现的目标。

秦山核电的项目管理延续数据治理项目的总体思路：“标准体系建立，搭建管理平台，试点 固化推广，监督评估迭代”,即通过制定标准、搭建平台来理顺业务流程，通过试点项目来验证 成果，然后通过日常监督和评估来不断完善管理，真正实现对物料主数据全生命周期的统一和 规范管理，提高公司对数据的管控能力，加强数据对公司经营决策的支撑力度。

**1.** **数据清理专项工作历程**

在分析秦山核电物料主数据质量现状后，2015年，秦山核电通过搭建组织机构、建设制度 规范、开发信息平台和分步实施清理4个环节陆续开展相关工作。具体介绍如下。

(1)搭建组织机构。

该专项工作涉及的数据量大、部门多，为了有效推动各项工作，秦山核电根据秦山核电物 料主数据的管理现状，首先确定了秦山核电物料主数据清理专项组织机构。该专项组织机构的 领导组由公司主管领导组成，工作组由物料主数据管理部门负责人组成，具体工作成员包括来 自技术、维修、化学、储运等各部门的代表。

**436** 数据治理 字化转型之道

(2)建设制度规范。

秦山核电成立物料主数据清理专项组成立后，即制定了各类专项工作所需的标准(包括制 定工作方案、技术规范等),以确保该专项工作在统一的高质量要求下开展。专项组陆续发布了 《ERP 物料主数据第一阶段清理专项工作计划》《ERP 物料主数据第一阶段清理管理指令》《ERP 物料主数据第一阶段清理模板(实施细则)》《ERP物料主数据第二阶段清理专项工作方案》《ERP 物料主数据第二阶段清理管理指令》《ERP 物料主数据第二阶段清理技术规范细则》《ERP 物料 主数据第二阶段实物清理分类物资照片样例》和《ERP 物料主数据第二阶段清理技术规格书》 等文件，为后续数据清理工作提供了业务标准和流程支持，也为后续工作的开展奠定了基础。

(3)开发信息平台。

为了能够及时跟踪和处理相关数据清理业务，秦山核电以现有物料主数据平台为基础，开 发了实物清理专用平台。最终实现照片和主数据——对应和挂接，为备件管理、仓库管理和库 存核库等工作提供实物照片数据。

(4)分步实施清理。

整个专项工作采取分步实施的策略。第一阶段主要是数据规范清理，确保 ERP 系统内的所 有数据符合系统和业务的格式规范；第二阶段主要是实物信息清理，即通过实物影像资料收集、 与系统内的信息比对、最终数据发布等工作环节，确保实物信息、系统信息的一致性和可视化。

**2.** **管理模式**

整个专项工作历时4年，涉及几十个部门、400多人和几十万条数据，其中，如何按工作 计划高效推动专项工作是专项组面临的一个挑战。专项组通过不断摸索，比较迅速地建立了符 合项目需求的管理模式，全面实施项目标准化管理，即利用专业清理模块和质量控制工具，以 及应用标准化管理流程来有效地标准化结构数据，确保最终为企业的各业务系统提供高效的数 据，如图43-2-1所示。

秦山核电的数据清理专项管理具体包括以下4个方面。

(1)结合项目特点开发专业清理模块，并进行适应功能优化工作。

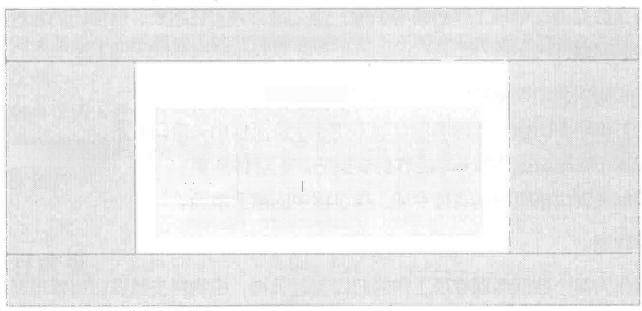
(2)引进PDCA 循环等质量控制工具，对数据清理工作进行质量控制。

(3)根据数据清理需求和各部门之间的业务关系，明确数据清理的业务步骤和职责。

(4)核电行业对数据质量的要求较高，因此其只针对标准结构化数据开展清理，其他非结

第43章 核工业：物料主数据治理助力核电智慧运营 **437**

构化数据不纳入专项清理范围中。



**《项目标准化管理》**

**专业清理模块**

**流程模块**

线上数据 线上澄清 标准数据

**数据识别** **清理答复** 数据输出

主数据管理平台 ERP

**质量控制**

**实物照片** **上传工具**

**标准化结构数据**

**标准化管理流程**

**清理系统**

图43-2-1 秦山核电数据清理专项管理模式

**3.** **工作目标**

本次专项工作围绕公司各项业务，分析当前与未来的数据需求，以“钉钉子”的精神落实 工作，做好基础数据的整理和核实，不追求短期绩效，而是建立将数据应用于业务的长效机制， 具体工作目标介绍如下。

(1)规范性：在清理完所有数据后，数据必须符合系统和业务总体要求。

(2)准确性：清理后的系统内的信息应与库存实物信息一致，确保财务账与实务账相符。

(3)完整性：清理后的物料主数据应尽可能地包括本次清理后补充的实物关键参数和影音 文件资料，如实物图片、技术文件、铭牌信息和钢印信息等。

(4)统一性：在进行账实相符性处理后，通过数据分析进行部分秦山地区物资的排重工作， 即确保秦山地区的同一种物资采用唯一码管理。

(5)过程可追溯性：实物清理工作要借助物料主数据平台对清理过程，以及已清理、已合 并的物资进行监控。

**43.3** **组织保障**

在秦山核电总经理部的统一领导下，按照“归口管理，分工负责”的原则，秦山核电建立

了数据清理专项领导组、管理组、工作组3个层面的数据清理管理体系。

**438** 数据治理 工业企业数字化转型之道

(1)领导组。

领导组是数据清理专项工作的领导机构，由公司领导担任组长，各部门的管理人员作为组 员，其主要职责是：

●领导和组织秦山核电物料主数据清理专项工作；

●协调并确定秦山核电数据清理专项工作体系及流程相关接口关系；

●审议秦山核电数据清理专项工作的实施方案及总体规划；

●协调解决秦山核电数据清理专项工作过程中的重要事项。 (2)管理组。

管理组作为整个数据清理专项工作的归口管理机构，由物料主数据归口管理部门负责人担 任组长，各部门归口协调人员作为组员，其主要职责是：

●定期修改数据清理专项工作的总体工作计划；

●制定数据清理专项工作的实施方案，包括管理细则和技术细则；

●负责制定、修订和收集与本部门业务相关的实物清理执行的技术标准、管理标准、工作 标准或规范制度；

●建立和完善数据清理专项工作的体系及流程接口关系；

●负责数据清理过程中的数据导出及上载；

●根据方案，定期组织例会以收集并反馈各工作组的工作成果；

●负责采集后信息的重码识别比对工作；

●协调解决日常工作中的问题。 (3)工作组。

工作组下设7个小组，分别为：工作一组、工作二组、工作三组、化学组、工器具组、储 运组、信息采集组，各工作小组由专业部门人员组成，其主要职责是：

●负责对采集的实物信息进行维护及确认(维护或重码合并);

●负责参与数据澄清讨论会；

●负责对实物清理结果的最终确认(签字);

●负责对实物清理后的数据进行技术文件补充工作；

●负责日常到库物资与实物清理后的数据再确认工作；

●负责提供实物清理所需要的工作场地、人员和工具的支持工作；

●负责实物影像采集清单内的物资下、上货架及清点确认工作；

第43章 核工业：物料主数据治理助力核电智慧运营 **439**

●负责对实物清理人员的在库劳动保护及相关安防监督管理工作；

●负责实物清理过程中影像信息的采集、信息数据核查及维护工作；

●负责对原系统中的拍照物资的整理和信息采集(不符合采集要求的重新拍照并按要求整 理)工作；

●组织备件负责人最终签字确认实物清理结果；

●负责对实物清理期间及清理结束后到库物资的实物影像采集工作(根据实物清理影像采 集要求执行)。

**43.4** **项目成果**

随着数据清理专项工作的稳步开展，清理专项组从数据清理的过程中不断总结技术和管理 经验，通过不断完善项目管理，逐步建立了统一的数据清理管控组织、制度、规范及数据质量 问题处理流程，并通过统一的数据清理平台进行落地，满足了秦山核电对高质量数据的需求， 为秦山核电的经营和管理工作提供了全面、可靠、准确、及时的数据服务，促进了后续数据深 度应用和数据挖掘工作的有序开展。

**1.** **探索数据清理机制**

通过物料主数据清理专项工作，秦山核电逐步理顺了数据清理工作的治理机制，为后续其 他类型主数据的清理提供了可借鉴的经验，并形成了一套可提供服务的数据支持机制，具体包 括如下内容。

(1)建立健全的数据管理组织。

通过此次数据清理专项工作，秦山核电明确了数据清理组织、人员、职责，建立了专业的 数据清理支持团队，以及各业务部门和数据清理组织之间的数据沟通机制。

(2)统一数据清理标准。

秦山核电基于数据现状和业务需求，统一了各类数据清理标准，为数据清理打下坚实的基 础，通过建立数据清理质量管理机制，保障了数据质量。

(3)统一数据清理工具。

秦山核电结合现有的主数据管理平台和所需的数据架构，统一规划了符合数据运维和数据 清理的集成数据管理平台。

**440** 数据治理——工业企业数字化转型之道

(4)实现数据清理过程的管控。

秦山核电通过分析数据质量需求、数据风险，确定了高效、顺畅的沟通和协调机制，以便 快速、高效地解决数据清理专项工作的瓶颈，并不断优化项目管理。

**2.** **提升物料数据质量**

秦山核电为了保障核电安全，用“笨办法”和“苦功夫”来实施数据清理。

“苦功夫”是指投入几百名设备管理工程师，对所有信息系统内的数据逐条、逐字段地核对技

术文件和图纸，做好基础数据的整理和确认， 一劳永逸地解决了基础数据“常治常乱”的情况。

“笨办法”是指为了避免小概率事件(即领用物资的系统数据与实物信息不一致的事件), 秦山核电对所有有库存的物资进行包括实物拍照、信息抽取和数据比对的全过程质量提升工作。

通过几年的数据清理，无论是在数据结构上，还是在物账相符性上，都得到了切实的管理优化。

**3.** **数据清理专项工作的两个阶段**

整个数据清理专项工作历时4年多，分两个阶段实施，各阶段的工作重点不同，故投入的 人力和资源也不相同，具体介绍如下。

(1)第一阶段：即数据规范性整理阶段。在此阶段清理数据总计204717项，投入人数总计 425人。

(2)第二阶段：即实物清理阶段。物料主数据平台共计挂接16.3万条物料的实物照片，约 97万张照片，数据识别组共计对15.9万条主数据与实物进行了信息比对，在线上发出11.89万 条技术澄清，收到物资负责人11.86万条技术答复。

**43.5** **项目总结**

在中核集团和中国核电的数据化战略基础上，秦山核电紧紧围绕“核安全第一”的理念，夯 实数据基础，为后续核电数据化、智慧化打下坚实的基础。秦山核电通过不断开展数据治理项目 并积累项目经验，在数据治理领域建立了一套科学、适用的数据治理体系和数据清理机制，其通 过不断地完善和优化，实现对数据治理全生命周期管理，为中核集团的信息化发展做出了贡献。

**第44章**

航空行业：军工企业的“三位

一体”数据治理体系建设实践

**44.1** **背景介绍**

**1.** **公司介绍**

中国航天科工集团有限公司(以下简称“航天科工”)是战略性、高科技、创新型企业，涉 足航天、建筑、医疗、汽车、食品、化工、石油装备、软件开发等多个行业。

中国航天科工防御技术研究院物资供应站隶属于中国航天科工集团有限公司，承担航天科 工集团的主数据管理与运维、云平台协作与物品编码业务的运行支撑。

**2.** **信息化现状**

(1)数据管理机制尚未建立。

2015年，航天科工提出了“建设智慧企业”的长远发展目标，加快建设三大平台(智慧企 业运行平台、专有云运行平台、航天云网运行平台),努力打造“信息互通、资源共享、能力协 同、开放合作、互利共赢”的环境。但是在推进平台建设的过程中，其首要面临的就是数据管 理机制尚为空白、数据(编码、属性)定义不统一、各层级标准自成体系、顶层标准执行不彻 底、系统平台不统一等问题。因此，加快推动航天科工的数据规范化、信息代码化、数据资产 化，以及统一组织开展数据治理工作，从根本上提升航天科工的主数据管理水平和信息化条件 下的管控能力迫在眉睫。

**442** 数据治理——工业企业数字化转型之道

(2)信息系统孤岛众多。

近年来，多品种、小批量、短周期、高质量成为航天科工产品研制生产的典型特征，这对 企业的生产能力提出了更高的要求。

2017 年，航天科工制定了数字化科研生产体系建设三年专项行动，以构建“模型、数据、 流程驱动”的科研生产新体系。但其面临的难题是信息系统孤岛多，数据交互断点多，设计、 试验、生产等信息尚未完全贯通。主数据是信息化建设的重要内容，也是在建设数字化科研生 产体系过程中实现信息交换和集成的基础。可以说，没有良好的数据管理能力，数字化、信息 化工作将无法有序推进。因此，想要实现对型号产品全生命周期及各阶段的有效管理，以及对 型号产品各层级信息的有效关联和追溯，提升科研生产的管理能力，必须从数据基础抓起。

(3)数据质量亟待提升。

为提升全级次经营管控和科研生产管理水平，从2016年开始，航天科工大力推进 ERP 系 统的建设及应用。

统一全级次各类主数据及编码，是确保 ERP 等系统有效运行，实现系统之间集成应用的前 提条件，也是通过 ERP 系统开展自动化及智能化制造、采购、结算管理的基础，还是实现供应 链管理、合同管理、计划管理等信息化系统互联互通、信息共享的基础。

航天科工是集团型企业，由于具有企业规模大、经营范围广、管理链条长、层级主体多等 特点，其在主数据管理上还存在集中管控难、精益管控难、横向协同弱、管理水平不均衡等问 题。从2016—2017年航天科工的 ERP系统一期建设经验来看，由于其数据标准化体系不健全、 主数据描述缺乏细化的标准、主数据管理系统功能不完善、主数据管理员责任人体系不健全 等问题，造成了基础数据质量不高、编码利用率低等现象，因此各系统之间不能顺利切换，无 法有效运行。

因此，确定适合航天科工自身特点的主数据管理体系建设和实施策略，成为尤为重要的一

项基础工作。

**3.** **数据治理背景**

为了解决上述问题，航天科工自2017年开始，逐步开展数据治理工作，以通过数据资源创 造数据价值为长远发展目标，深挖业务和管理需求，从集团层面自顶向下统筹规划布局，制定 数据治理体系总体规划，构建 “三位一体”的数据治理体系格局：制定一套统一的主数据技术 标准体系和运维管理体系，搭建一个集中的主数据管理平台，建立一支专业的技术团队，实现

第44章 航空行业：军工企业的“三位一体”数据治理体系建设实践 **443**

对主数据的统一管理。

按照航天科工的主数据两级管理架构的要求，各二级单位全部完成了主数据系统建设，打 通了与集团级主数据系统的传输链路。航天科工通过深入开展以“数据集成”为核心的信息化 深度建设工作，实现了航天科工各类信息系统中数据的互联互通，打造了“信息互通、资源共 享、能力协同、开放合作、互利共赢”的环境，为建立具有航天科工特色的“模型、数据、流 程驱动”的科研生产体系提供了基础保障，逐步实现航天科工的精细化管理和现代化管理。

**44.2** **工作历程**

**1.** **构建主数据标准体系，实现流程驱动和数据管控**

主数据具有特征一致性、识别唯一性、长期有效性和交易稳定性等特征，这就要求主数据 的分类和编码应遵循统一的数据标准。建立主数据标准是数据标准化建设的核心工作，具体包 括以下内容。

(1)制定标准，让主数据管理有章可循。

航天科工集团公司总部负责主数据管理系统运行各项工作的组织和协调，各业务归口部门 负责其业务范围内主数据结构和管理规范的编制、修订和废止；各二级单位可选派本单位具有 数据专业能力的人员与集团公司共同编制标准，并结合本单位实际工作，对标准进行补充或细 化。航天科工打破原有“各自为政”的数据管理模式，制定了统一的分类标准、描述模板、编 码规则、数据模型等，让主数据管理有章可循，确保企业编码达到“一物一码”的效果。

结合《中国航天科工集团公司企业架构 V4.0》对主数据的要求，航天科工通过召开多次专 题工作会议、在全集团范围内征求意见等方式，制定并发布了机构、供应方、客户、项目、物 品(物资和产品)、会计科目、人员、合同、信息系统组织机构、信息系统用户共11 类主数据 规范及57类代码主数据标准。其中物资数据分类涵盖元器件、原材料、标准件等产品700余类， 以及属性特征80余个，基本满足航天科工各单位的科研生产、日常经营管理等业务需求。

(2)清理数据，保证数据的完整性、准确性和唯一性。

在制定好主数据标准后，航天科工根据主数据标准规范，对历史数据进行清洗、排重、编 码，保证数据的完整性、准确性和唯一性，并且对已上线、在建等业务系统，根据系统所处阶 段及重要性的不同，采用“完全、映射、择机”不同的策略导入标准数据。

航天科工结合 ERP 系统建设需求，按照 ERP 系统建设与数据治理同步并行建设的方式推

**444** 数据治理——工业企业数字化转型之道

进集团主数据清理工作。各单位在ERP 系统的蓝图设计中期，即启动本单位的主数据清理工作。

航天科工的主数据运维机构梳理了数据准备及数据清理过程中的关键步骤及节点，在全集 团范围内对二级单位的数据清理团队开展数据规范、数据清理方法、数据填报模板的培训工作， 确保了各数据清理团队对规范要求和数据质量要求的理解一致。截至2018年，航天科工累计清 理数据50余万条，有力保障了航天科工的308家单位的ERP 系统集中上线任务的圆满完成。

**2.** **建立高效的主数据管理体系，保障对主数据全生命周期的动态管理**

航天科工主数据管理要实现“统一领导，职责清晰，制度规范，流程优化”的目标。而实 施统一编码管理是一项系统性极强的基础性工作，其中涉及企业经营管理和科研生产管理的方 方面面，具体包括以下工作。

(1)建立主数据管理制度。

集团主数据建设和应用是一种业务管理活动，通过执行统一的数据标准和管理制度，可以 将规范、流程和技术平台三者有机地结合，以保证主数据的一致性、完整性、及时性。主数据 管理制度用于规定主数据管理工作的内容、程序、章程及方法，是主数据管理人员的行为规范 和准则。

航天科工制定了《中国航天科工集团公司主数据管理系统运行管理要求》《中国航天科工集 团公司物品编码管理要求》《ERP物资数据上线检查规范》《物品编码提报审核工作规范》等规 章制度，明确了主数据管理机构、管理人员及其工作职责，规范了主数据管理系统日常运维和 问题处理机制，提高了主数据质量。

(2)建立主数据管理流程。

建立主数据管理流程是保障主数据落地和提升数据质量的一个非常重要的环节。通过梳理 数据维护及管理流程，建立符合航天科工实际应用情况的主数据管理流程，可以保证主数据标 准得到有效的执行，对主数据应用过程中发生的各类问题进行及时处理，实现主数据的持续性 长效治理。

航天科工还制定了《Q/QJB213B-2018 主数据管理规范》,规定了集团公司各级次数据的主 数据管理岗位的职责、主数据管理状态、主数据管理流程(各单位申请、修改、传递、使用集 团主数据的流程)。

(3)建立主数据管理评价体系。

航天科工梳理并制定了集团公司主数据 KPI 指标，明确考核范围、考核规则、考核形式，

第44章 航空行业：军工企业的“三位一体”数据治理体系建设实践 **445**

形成主数据考核评价体系，从而可以动态反映主数据管理工作、主数据应用情况及系统建设情 况。航天科工从数据管控过程、数据质量和数据标准执行情况3个方面形成28项指标，并纳入 《中国航天科工集团有限公司 ERP 建设与应用评价方案V1.0》中。航天科工通过将指标分解并 落实到二三级单位及主数据管理员，以及强化考核和细化评价，使各单位逐步加强主数据管理 及应用意识，进而加强责任制的落实。

**3.** **统一主数据应用架构，实现各业务系统和决策系统的集成应用**

航天科工通过主数据管理平台，固化了主数据全生命周期管理流程，主数据的新增、变更 和废止统一由业务组提出申请，根据不同类别的主数据管理模式，执行相应的审核制度，并通 过系统设置的数据校验公式和人工审核双重数据确认方式，严格控制数据质量。本着“业务谁 主管，数据谁负责”的原则，航天科工强化主数据管理员的数据管理责任，并且通过系统实时 监控数据的审核时效性、提报准确性，规范数据管理流程。

(1)搭建集团级主数据管理平台。

航天科工的主数据管理系统是由集团内部单位自行开发、建设及运维的系统，管理各类主 数据的增、审、改、停、用、控等全生命周期业务，以实现数据的统一、集中、规范管理，提 高数据质量和增强信息共享。

航天科工的主数据管理系统主要包括主数据建模、主数据整合、主数据管理、主数据服务、 基础管理、标准管理等功能模块，如图44-2-1所示。

(2)制定外围系统应用主数据要求。

航天科工集团两网主数据管理系统采用两种部署模式， 一种是“二级部署，多级应用”的 架构模式，即集团本部、各二级单位各部署一套主数据管理系统，两级集成应用。

另一种是采用“一级部署，多级应用”的架构模式，即集团本部部署一套主数据管理系统， 下级单位应用集团主数据。 一级架构系统支持多组织机构模式，集团所属单位的应用系统(包 括 MDM 系统)可以直接与集团级主数据管理系统集成，统一应用集团级主数据。

航天科工运维团队梳理了用户单位对集团主数据的贯彻情况及系统应用情况，明确了各类 主数据的数据流向、管理流程、订阅分发流程，制定了用户单位的 MDM系统与业务应用系统 的集成路线。各二级单位接收并分发集团级主数据，形成自顶向下的数据流转链路，实现主数 据的采集、分发等交互操作，保证数据源头唯一。集团各单位 ERP 等外围系统应用主数据的架 构如图44-2-2所示。

盐 特 齿

日 常 管 理

字 典 盛 理

数 据 审 核

酸 据 的 建

数 据 完 善

数 据 清 洗

数 据 合 井

数 識 汇 律

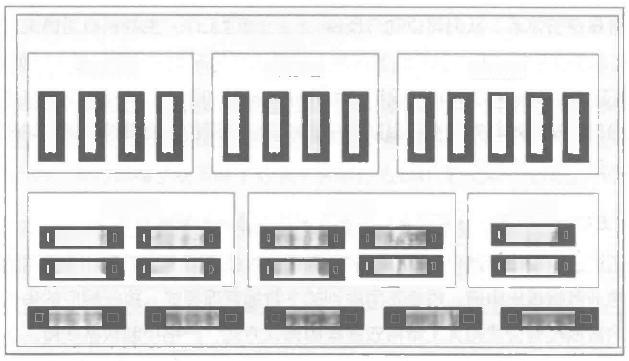
流 程 定 义

属 性 定 义

编 码 规 则

对 象 定 义

**446** 数据治理——工业企业数字化转型之道

主数据管理系统

(MDM)

**主数据建模** 主数据整合

**主数据管理**

**主数据服务**

|  |  |
| --- | --- |
| 数据分发  查询服务  排重引举 | 订阅服务  异常处理  流程引擎 |

基础管理

用户管理

日志管理

系练管理

权深管理

检索引擎 分发引擎

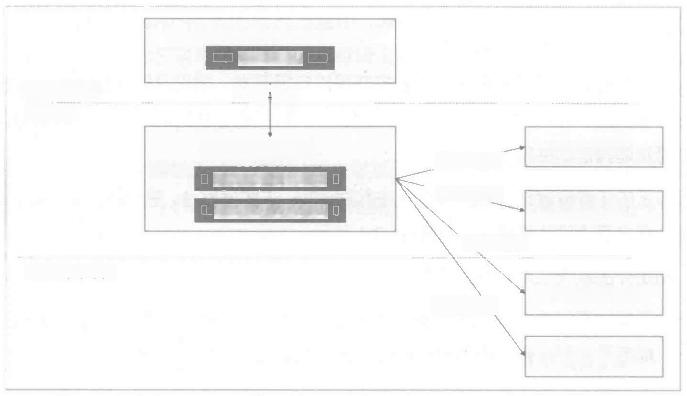
标准管理

标准维护

标准查询

高速缓存

图44-2-1 主数据管理系统功能架构



集团级主数据管理系统

通用主数据

分发主数据

|  |  |  |
| --- | --- | --- |
| 院级主数据管理系统  集团下发通用主数据  下级自建个性主数据 | 分发主数据  分发主数据  分发主数据 | 院级ERP系统  院级应用系统 |

三级及以下单位

厂所级ERP系统

分发主数据

厂所级应用系统

集团公司总部 二级单位

图44-2-2 外围系统应用主数据的架构

**4.** **开展全级次主数据贯彻应用工作**

航天科工开展的全级次主数据贯彻应用工作包括以下几项。

(1)制订主数据应用推广计划，在推行建立全集团ERP 系统和财务共享中心的过程中，在

全集团范围内各级单位和各级信息系统中全面的应用和推广主数据。

(2)制定《集团公司主数据建设及应用情况调研工作方案》,定期走访航天科工下属单位，

实地调研21家二级单位主数据建设及应用过程中遇到的需要集团协调解决的问题等，深入了解

第44章 航空行业：军工企业的“三位一体”数据治理体系建设实践 **447**

并协助各单位做好主数据建设与应用工作。

**44.3** **组织保障**

航天科工建立了统一的主数据管理专业团队，实现业务常态化运行：建立了明确的主数据 管理机构和组织保障体系，落实各级部门的职责，具体介绍如下。

**1.** **建立主数据责任人体系**

航天科工集团公司顶层明确了主数据管理的归口负责部门，并将有关二级单位的主要领导 纳入主数据管理责任人体系；明确了业务流程和主数据管理部门，并已建立了一支包括2300 余人的主数据责任人体系。航天科工的主数据管理责任人体系主要包含以下3层组织架构。

(1)决策层。

航天科工设立了主数据领导小组，由公司信息化领导小组成员组成，对主数据标准化工作 进行统一领导，确定指导思想、目标和任务，协调解决主数据管理中相关的重大问题。

(2)管理层。

按照“业务部门牵头，信息部门统筹，归口管理，分工负责”的原则，航天科工信息化部 门设立了主数据归口技术管理组，负责主数据管理系统中各项运行工作的组织和协调。业务组 由职能部门中的相关业务专业人员组成，负责管理对应类别的主数据结构和数据质量，以及编 制管理规范等。

(3)执行层。

执行层由航天科工集团总部和下属单位中的主数据管理员专职或兼职组成。他们负责主数 据在本单位的贯彻落实、应用检查工作，本单位主数据需求的收集、申请提报、审核工作，以 及本单位主数据的培训、宣传贯彻和日常维护等工作。

建立主数据管理责任人体系，使得业务部门与IT部门之间合作密切，各业务部门及人员深 度参与。航天科工的主数据管理由企业的业务需求驱动，业务部门的人员负责推进，每类主数 据由相关的管理部门牵头制定并发布数据标准，即实现了业务牵引，信息化驱动，协同推进。

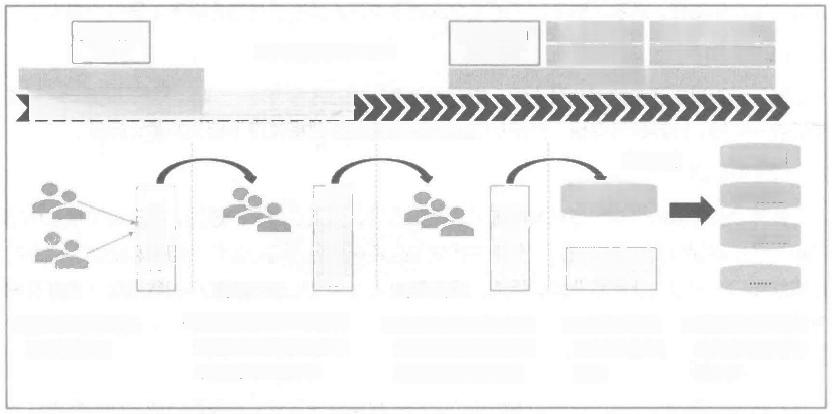
**2.** **建立主数据专业化管理团队**

航天科工成立了两个专业化机构(二院基础数据与编码管理中心和三院三O 四所大数据产 品中心),分别负责主数据管理和主数据系统的技术运维，高效地支撑主数据在各类信息化系统

和业务应用系统中的应用，实现了集团主数据的精准、有效、协同、共享。

二院基础数据与编码管理中心是航天科工集团主数据管理及日常运维工作抓总的支撑机构， 拥有30多位主数据与编码管理专业人才，面向全集团提供主数据建设和应用服务。其负责制定 及修订主数据结构标准和管理规范，审核和发布委托管理的主数据，定期检查数据质量，推进 各单位应用集团主数据。

三院三O 四所大数据产品中心承担主数据管理系统的日常维护工作，具体包括编写主数据 管理系统用户操作手册及建设指导方案，及时解决主数据管理系统的集成及应用等相关问题； 主数据管理系统与相关系统的集成联调、测试等，全面保障主数据管理系统的安全、稳定及可 靠运行，如图44-3-1所示。



实施/技术

支持

维护组

>>>>>>>>>>>>

信息代码日常维护流程

自动分发

代码车/标准 库

MES 系统

代码及信息 发布

代码维护组

·审核通过的信·标准代码由系统自 息由系统配码、动分发到目标信息

系统中

专业审核组

·专业审核组负责用户申 请的最终审核，确保信 息的准确性和权威性

·信息标准化维护组负责代 码的初审，以及代码的日 常监控、检查和维护

用户

·用户到主数据管理 系统申请代码

物品管理组

项目管理组

各专业审核组

供方客户管理组 会计科目管理组

财务系统

HR 系统

信息标准确 认/审核

支撑运维部门统筹管理

合法性检查

代码申请

专业审核

初审

发布

**图44-3-1** **主数据管理系统的日常维护工作**

**3.** **全方位组织贯标培训**

为了加强基础数据管理，保证集团数据质量，2017年航天科工下发文件，要求各单位建立 物品、客户、供应方、项目主数据管理员团队，并施行“持证上岗”机制。另外，航天科工通 过每年定期举办“中国航天科工集团有限公司数据资源治理”培训班，让这些主数据管理员学 习主数据管理的基础知识、基本理念，并结合上机操作掌握系统的基本操作方法，具备系统运 维能力，从而培养了一支常态化的专业实施和运维团队(涵盖财务、科研生产和信息化等主要 业务领域的人员),实现全级次主数据管理工作的常态化、规范化、流程化。

第44章 航空行业：军工企业的“三位一体”数据治理体系建设实践 **449**

**4.** **全天候在线运维服务**

为提高技术支持的工作效率，航天科工面向全集团推出主数据“在线答疑管理”模块，集 中面向集团公司总部、各单位提供技术咨询、故障申报、问题处理及系统管理和维护等方面的 服务。

为加快沟通效率，航天科工开通了“二院基础数据与编码管理中心”微信公众号，系统展 现集团主数据知识体系，深入报道数据治理行业案例。其通过微信公众号搭建了主数据运维客 服平台，并设置10位不同专业类别的人工客服提供在线答疑服务，解决各单位在主数据运维管 理过程中遇到的各类问题。

**44.4** **实施效果**

经过3年持续不断的努力，航天科工的数据治理工作取得了以下效果。

“数据强基”的思想不断增强，有3000余人接受培训，在广泛普及基础知识的同时，也让 员工的规范意识生根发芽。

主数据标准体系进入常态化良性循环建设模式并不断被完善，确保业务数据管理有据可依。

主数据在 ERP 、PDM 系统、型号 BOM系统、设计基础数据资源等方面被广泛应用，形式 多样，为数字化科研生产体系铺建“高速公路”。

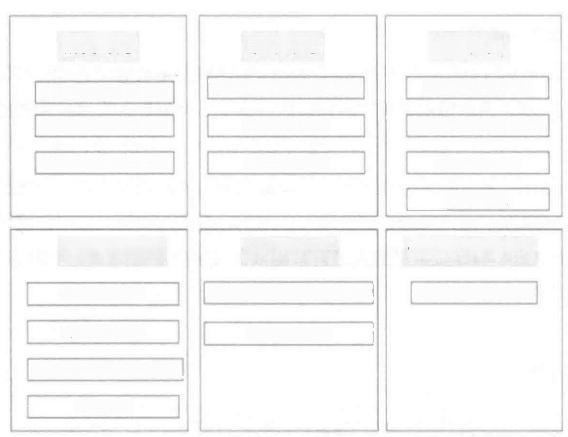
建设7大类共覆盖200万项物资的数据字典库内容，从源头解决数据标准化问题，确保数 据质量得到优化及提升。

主数据应用监管力度加大，通过现场走访、调研各级单位、诊断及解决百余项问题，确保 主数据管理机制完善、数据链路畅通，有力地支撑各单位系统上线及业务精准化。

主数据的数据量快速增长，数据资产初具规模，各类数据总量从2017年年初的50万条， 增长至2019年的近100万条，增长率达100%。

另外，航天科工基于主数据开展了各类数据治理工作，全面打通了各职能部门、业务部门 及业务系统之间的数据传输链路。各类主数据与业务深度融合，在销售、采购、供应链管理、 高层决策、合同与风险管控等多个业务领域中发挥了作用，保证了管理经营和决策支持中数据 源的准确性。图44-4-1是主数据为各管理领域带来的价值全景图。

**450** 数据治理——工业企业数字化转型之道



|  |  |  |
| --- | --- | --- |
| **销售部门** | **采购部门** | **IT部门** |
| 降低产品发布的成本 | 统一采购，降低采购成本 | 降低数据清洗成本 |
| 提升销售和市场效率 | 提升采购效率 | 降低信息系统维护成本 |
| 增加交叉销售的机会 | 优化资产配置 | 减少错误数据  优化IT架构 |
| **供应链管理** | **高层决策** | **合同与风险管控** |
| 优化需求预测 | 更准确、更全面的数据支持 | 降低法规遵循的成本 |
| 降低物流成本 | 更实时的反馈 |  |
| 降低仓库接收和处理成本 |  |  |
| 降低库存 |  |  |

**图44-4-1** **主数据为各管理领域带来的价值全景图**

**44.5** **项目总结**

**1.** **基于供应商主数据实现产品供应商名录管理**

基于供应商主数据，航天科工建立了产品采购统一供应商名录，二级单位在集团名录内建 立本单位名录。

通过供应商主数据关联 ERP 系统中的项目及采购业务，有效地控制了供应商采购不合格产 品的现象，促进了产品采购业务的精细化管控。

供应商主数据作为合同签订、付款的重要标的信息，进入ERP、 财务共享中心、专有云等 系统中，从而打通企业各层级信息，严格审计及风险管控，有效杜绝了企业自行创建不规范信 息带来的数据统计难、业务规范性差等问题。

主数据是管控黑、灰名单的有效手段：在黑名单内的供应商禁止使用，在灰名单内的供应 商予以提醒。航天科工基于客商主数据对 CRM 系统中的客户数据进行清洗，数据量压缩了约 20%,有效降低了管理成本。

**2.** **基于产品主数据的数字化科研生产管理**

产品主数据作为科研生产活动中不可或缺的一类数据，在多个应用场景中发挥着重要的

第44章 航空行业：军工企业的“三位一体”数据治理体系建设实践 **451**

作用，具体介绍如下。

一是作为产品的代号，在多网络环境及系统中应用，可以降低管理成本，为产品数据管理 提供便捷。

二是作为项目、物资、供应商等多类业务数据的共用属性项，有利于快速抽取数据和统计 分析数据，并基于数据在 MPM系统、财务系统、ERP 系统中的流转识别业务情况，例如在物 资采购与送筛过程中，可根据产品区分筛选标准并发放装机证，提升采购与筛选效率。

**3.** **基于项目、会计科目主数据，提升经营成本管控能力**

项目主数据在项目计划闭环管理、项目成本统计、经营绩效核算等方面发挥了较大的作用： 一方面作为 MPM 系统中项目管理计划分解的源头，在拓展属性后被推送至 ERP 系统中，与物 资采购需求进行关联；另一方面基于ERP 系统中从项目到核算的全成本管控，有效提升经营 绩效。

会计科目主数据被全面应用于 ERP 系统、财务共享中心中，从而达到“六集中”的目的： 核算处理集中、数据存储集中、资金收付集中、规范制度集中、监管审计集中、管控治理集中， 实现了应收款报表的实时展示，持续提升了财务事项管控治理。

**4.** **基于物品主数据的供应链管理、设计选用优化**

依托物品主数据，研发生产类单位重点构建了设计基础资源管理平台，逐步向数据化科研 生产、协同设计迈进，具体包括以下几个阶段。

(1)2010年，集团协同供应链系统及集团物资编码开始应用，并落实了物资统一采购业务。

(2)2012年，随着业务及数据的深入应用及管理精细化程度的提高，部分二级单位提出了 应在集团物品编码基础上建立满足各下属单位业务管控需求的二级物资编码数据库。

(3)2017年，各单位的 ERP系统中陆续上线了采购模块，对二级物资编码应用的需求愈加

迫切。

(4)2018年，研发生产类单位依托物资编码，重点开展了设计基础资源库建设，目的是解 决多年未实现的物资数据管控问题。

设计基础资源库作为设计生产单位的物品主数据的主要管理系统，在承接集团公司所有物 品主数据的同时，按照二级单位的要求，自动甄别出自用物资，并可按院级、厂所级管控范围 分别推送至下游业务系统中(如 ERP 系统、PDM 系统等)。该系统通过限定合格供应商及其供

**452** 数据治理——工业企业数字化转型之道

货范围，实现了科研生产管控物资采购范围的系统化，极大提高了数据流转效率，节省了人工 维护成本，降低了人为判断出错率，具体介绍如下。

(1)旧流程费时费力，管控颗粒度粗。

2017年年初，某二级单位的 ERP 系统一期上线。因未打通集团公司主数据到该单位 ERP 系统的应用链路，新增物品主数据的应用流程较为烦琐，需要手工操作将数据导入ERP系统中， 每天需投入0.5人进行处理，不仅费时且易出现错误。2018年，各单位的 ERP系统二期全面上 线，如果不解决这个问题，则数据质量难以保证，人工运维工作量也将成倍增长，影响将越来 越大。

(2)新流程省时省效，管理精细化程度高。

2018年11月，设计基础资源库正式启用，3家单位申请的89条元器件物品主数据作为第 一批数据，从集团主数据系统中被自动分发至该单位的 ERP 系统中并应用于业务，改变了原有 人工操作的模式。

用户在集团公司主数据系统中申请物品编码后，数据通过院级 MDM 系统进入设计基础资 源库，接着系统自动甄别出申请单位所用物资，并实时推送至ERP 系统中。另外，对于未能直 接推送至 ERP系统的特殊数据，系统提供人工选用功能，通过设置对申请、审批环节予以管控。

系统能够自动甄别的核心关键在于，其依据该单位合格供应商名录及供货范围，与集团公 司物品主数据中的分类索引码进行全面对照，梳理出一套根据物资编码分类码即可判断是否为 物资管控范围的“对应表”,系统根据“对应表”实现了自动甄别。

(3)基于统一物料编码设计基础资源。

物品主数据的价值不仅在分类索引对照，其更大的价值还在于它承载了基于业务管控需求 的扩展属性信息。在集团物资固有属性基础上，航天科工还扩展了采购周期、筛选周期、价格 信息、质量信息、元器件对可获得性、供应商等级等信息，同时整合各所现有设计资源，以主 数据为主线，建立三维模型、EDA 模型、产品图片、产品资料等信息资源数据库，并打通与 PDM系统的链路，实现将主数据和结构模型库从设计基础资源库向院、所两级 PDM系统自动 推送，为全院设计师提供统一、丰富的设计数据资源。

(4)基于物品主数据的产品 EBOM 建设。

航天科工的物品主数据已作为制造类单位产品建立 EBOM 的数据基准，成功与PDM 系统 挂接，其主要做法就是要求所有研制设计类产品 EBOM 结构树展开到元器件级，并强制应用集

第44章 航空行业：军工企业的“三位一体”数据治理体系建设实践 **453**

团统一物资主数据，否则 PDM 系统无法接收。基于 EBOM, 航天科工已实现快速统计出各类 产品、各层级产品应用元器件的品种、规格、国别、价格情况，并开展物资采购需求提报，极 大压缩物资采购投产通知下达后的物料清单梳理、库存梳理的周期，实现“设计一采购—试验 一生产”一体化的物资保障新模式。据初步统计，备料周期压缩至少1个月，且统一的编码可 实现各单位合并采购，降低了议价成本。

(5)基于物品主数据的元器件数字化选用目录建设。

航天科工首先基于完善的物资分类体系及带有元器件类别、单价等信息的物资编码数据库， 完成产品设计用元器件数字化选用目录建设，对设计师选用进行数字化管控，有效防止超目录 选用。

“三分技术、七分管理、十二分数据”,航天科工“三位一体”的数据治理体系以智能制造 类流程和智慧企业管控类流程为复杂协同业务主线，围绕经营管理构建了一套覆盖集团公司全 业务领域的主数据管理体系，支撑了跨单位、跨系统、跨网络的协同设计、协同制造、协同采 购，这是开展数据价值资产评估和数据资产内部运营流通的必要条件，也是实现航天科工数据 战略目标的基石。

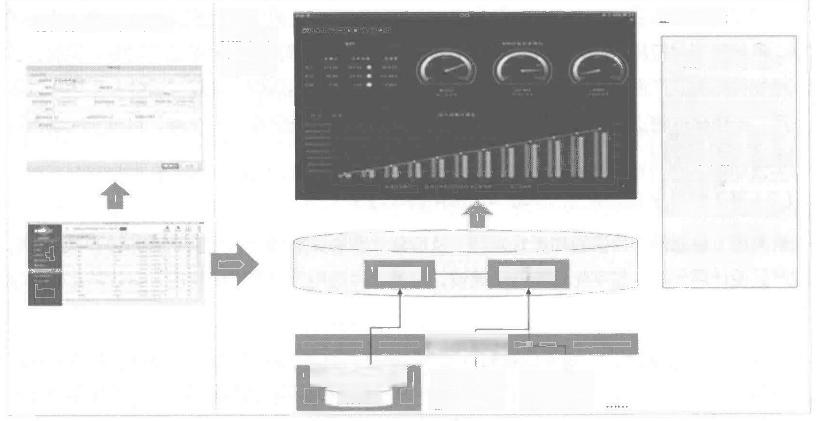
**44.6** **未来展望**

数据标准化与数据建设会促进企业业务的精细化管理，为企业提供全方位的可视化平台、 标准化的 KPI分析体系管理。

因此，在未来一段时间内，编码中心将以智慧企业建设、ERP 系统全面建设、数字化科研 生产体系建设对于数据标准化的需求为出发点，解决当前各项工作中的突出矛盾，理顺各层级 问题，保证各系统之间数据标准统一，实现快速、高效的数据交互，进而实现各业务领域的大 数据分析，并以数据为创新载体作为长期目标。

**1.** **数据标准化支撑企业决策**

航天科工在此次数据标准化工作中建立了丰富、准确的主数据仓库，为企业全景化数据融 合提供了一致性、高质量的数据标准化支持，如图44-6-1所示。



企业驾驶舱

主数据对数据仓库 的支撑作用：

·为指标管理、决 策支持提供一致 性、高质量数据 的支持

·为企业全景化数 据融合提供数据 标准化支持

业务数据

统单评台

**人力资源**

**经管系统**

目标：

数据自动 采集、汇 总、统计

主数罐管理平台

MDM

**现阶段：手工填报数据**

**财务系统** **集采系统**

**数据仓库**

主数据

**图44-6-1** **主数据仓库**

**2.** **数据标准化支撑财务业务一体化**

航天科工通过主数据推进财务业务一体化，建立了包含人、财、物、生产、采购、设备管 理高度贯通的信息流，梳理、优化并规范了业务流程，实现了各级次、各部门之间的高效协同。

**3.** **数据标准化实现“三流合—”**

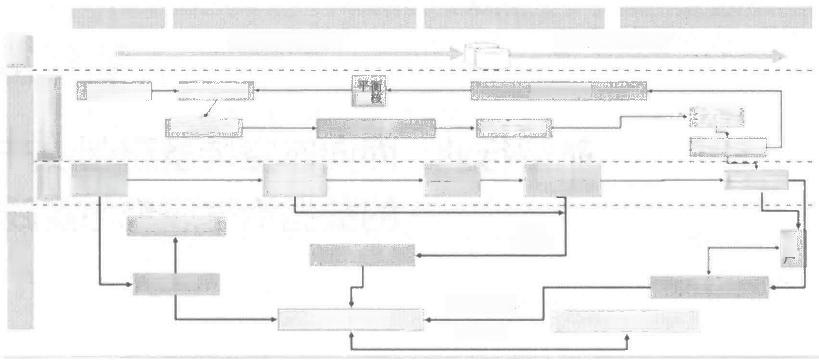
航天科工以各类主数据编码为引导，实现数据全生命周期综合运维，以及全型号、全过程 有效管控数据。其通过提高项目过程管理、物流和物资管理等方面的精细度，在交易数据层面 做到无缝衔接，实现物流、业务流、信息流的高效协同(即“三流合一”),如图44-6-2所示。

在当前以数字化、网络化、智能化、云化为特征的信息文明时代，数据资产已经成为比石

油和矿产价值更高的战略资源。

航天科工实现了各业务系统中所需的主数据编码的统一应用及全覆盖，逐步建立了集团编 码技术服务产业化体系：对内满足企业能力建设和实现要求，对外加快基于数据的企业形态与 商业模式创新，逐步形成数据产业基础支撑能力，打造基于数据服务的新型核心竞争能力和新 业态。

第44章 航空行业：军工企业的“三位一体”数据治理体系建设实践 **455**



成员企业

一 级物流(陆运)

需求计划

二级配送计划

工程建设现场

发票/发货

凭证

内部

结算单

外部结算

发票/出库单

共享服务中心

资金信息

商运公司

一 级物流(海运)

物 流 信息流 业务流

采购计划

一级配送计划

发票

发货凭证”

审

一级配送

计划与执行 神 单 流

发票校检 /物流费用

收货

匹配量价

向供应商付款

级配

供应计划

最终客户

收款/发票

结算中心

内部结算

外部结算

供应商

客

**图44-6-2** **实现“三流合一”**

**第45章**

航空行业：面向航空装备研制生产

的数据治理研究与实践

**45.1** **背景介绍**

**1.** **公司简介**

中国航空工业集团有限公司(下文简称“航空工业”)是由中央管理的国有特大型企业，是 国家授权的投资机构。集团公司设有航空装备、通用航空、航空研究、飞行试验、资产管理、 金融、工程建设等产业。

**2.** **航空工业集团统一编码**

(1)开展集团统一编码工作的背景。

随着航空信息化建设的深入推进，许多航空企业加大了信息化的投入，相继实施 PDM、ERP、 MES 等业务信息系统。在以往，建设或者更新一次系统，都需进行一次基础数据整理工作，不 仅浪费了人力和物力，而且各个信息系统之间的基础数据不一致，各自为政，难以实现信息的 交互与共享。为了打破基础数据不统一的问题造成的信息化建设的瓶颈，各单位需要开展信息 编码工作。

各单位在开展信息编码工作时，没有一套统一、科学的理论作为指导，没有对关键信息的 命名、描述、分类和编码等开展标准化工作，仍按照各单位或各职能部门的需求和习惯进行， 很容易出现对同一个信息对象的标识不统一、信息分类方法不统一、名称不统一等问题，难以

第45章 航空行业：面向航空装备研制生产的数据治理研究与实践 **457**

满足现代化企业信息管理的需求。为了推进航空企业信息统一交换与共享，为航空装备研制生 产提供共享规范的基础数据，集团公司制定了统一的信息编码标准化体系和管理策略，组织开 展集团统一编码工作。

(2)集团统一信息编码工作的指导方针和工作策略。

① 指导方针——共建共享

航空工业的信息编码工作的重点在于将信息代码应用于企业数字化建设中，解决集团及各 成员单位对于信息编码的各项需求，使集团的各单位都能共享集团统一的代码数据库，从而以 集团统一代码为基础开展企业的信息编码工作，支持企业的信息化建设及航空产品的协同研制。

②工作推动策略

●统筹规划，建立体系

航空工业对信息编码涉及的一些工作(工作机制建立、编码标准体系框架建设、编码标准 编制、编码数据库构建),实行全行业统筹规划，明确顶层要求，加强培训宣贯并加强应用统一 代码的技术指导，强化信息编码工作执行力的检查考核；通过完善制度、健全组织机构、明确 责任、梳理流程、强化服务等措施，建立了规范统一、高效运行的信息编码管理体系；同时， 在航空企业内建立领导重视、全员参与、重点突破、持续推进的企业级信息编码工作机制。

●急用先行，重点突破

信息编码工作是一项系统工程，也是一项长期的贯穿于企业数字化建设全过程的基础工作， 不是通过编制几项编码标准就能完成的，而是需要多方协同、持续推进、不断完善，才能适应 日益增长的信息化建设需求。因此，航空工业的信息编码标准编制工作坚持“急用先行”的原 则，重点解决在行业和企业数字化建设过程中急需的信息编码标准的编制问题，并构建统一的 代码数据库，重点突破制约企业数字化建设的数据标准化问题。

●标准引导，数据支撑

航空工业以编制统一信息编码标准为引导，规范了全行业统一代码应用标准，建立了基础 编码数据的规范要求，从数据源头管控数据质量。航空工业将重点放在建立集团统一的信息编 码数据库上，以行业统一的编码数据库支撑信息编码标准的贯彻及应用，同时也重点解决了企 业信息化和数字化建设对信息编码的需求。

●服务产品，逐步推广

在开展统一信息编码工作的过程中，航空工业以重点航空产品为突破点，在产品中应用集

**458** 数据治理——工业企业数字化转型之道

团统一的单位及供应商代码、物品码，将统一的信息编码作为信息传递和交换的桥梁，贯穿于

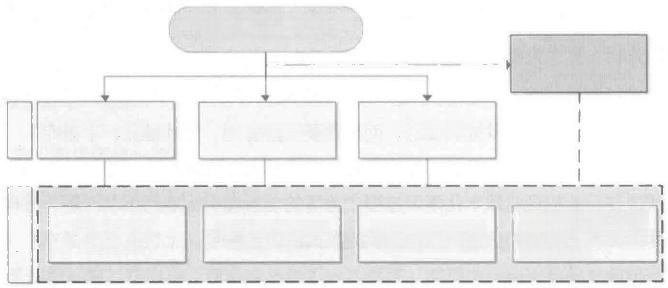
航空产品的研制、生产、采购、使用、维修等全过程。

●重在应用，服务企业

航空工业在统一信息编码工作中始终坚持以应用为主的原则，以解决信息系统中数据的规 范应用为出发点，对于集团级的信息系统(例如集团统一认证系统、电子采购平台、协同数字 平台等),都要求应用集团统一编码；对于企业级的信息系统(例如 ERP 、MES 、PDM等),则 将集团统一编码作为数据基础直接在信息系统中应用，实现了数据来源统一，减少数据的重复 定义和多源头输出，进而保证数据的一致性。

**3.** **编码中心及其发展历程**

为了加强集团公司的信息代码管理工作，加快统一信息代码的推广和应用，在航空工业集 团公司各级领导的大力支持下，航空工业于2009年成立了“中国航空工业集团有限公司信息编 码管理中心”(以下简称“编码中心”),负责集团统一代码的推广及应用工作。航空工业各所属 单位也根据集团公司的要求建立了企业级编码中心，明确了企业代码管理工作的主管领导、归 口管理部门和责任人员等。航空工业编码中心的组织架构如图45-1-1所示。



中国航空工业集团有限公司

信息编码管理中心

外部专家组

技术保障组

负责各类对象数据分 类、描述模板、填写 规范的制定

负责代码管理系统 日常运行和维护

负责提供编码技术 咨询和服务

负责集团公司 代码审核

业务咨询组

代码审核组

组织

职责

图45-1-1 航空工业编码中心的组织架构

为了更好地推进各单位开展统一信息编码工作，航空工业在2009年发布了《中国航空工业 集团有限公司信息代码管理办法》及《中国航空工业集团有限公司信息代码管理实施细则》,明 确了集团代码管理工作的主管机构、编码中心的机构设置和职责，完善了统一代码的应用范围， 以及代码的申请、审核、赋码、变更、注销的管理流程。

从2015年之后，航空工业建立了统一的信息编码工作季度通报机制，每个季度都会对集团

第45章 航空行业：面向航空装备研制生产的数据治理研究与实践 **459**

统一代码的推广及应用、各单位的信息编码工作开展情况予以通报，组织成功实施案例的宣传 和推广。

为了满足航空工业在科研、生产、经营管理等相关领域中对信息分类与编码标准化的需求， 支撑集团公司及各单位的信息化建设和长久发展，航空工业发布了信息分类与编码标准体系框 架。此体系框架符合航空产品研制生产的特点，将信息按照科学的原则和方法进行分类并加以 编码，作为在一定范围内进行信息处理与交换时共同遵守的准则。体系构架主要包括基础标准， 产品研制、生产、服务和管理的信息分类与编码标准等分支。

根据“急用先行”的策略，航空工业在信息分类与编码标准体系框架下发布了《单位及供 应商代码编码规则》《人员代码编制规则》《物品码编制规则》《产品条码标印要求》等顶层编码 标准，明确了集团公司的人员、单位、供应商、物品统一代码的适用范围、编码原则、代码结 构及赋码方法。

在发布以上集团公司顶层标准后，编码中心编制各项标准的贯彻及宣传材料，通过培训、 宣讲等形式提高各单位对于标准的理解，指导各单位开展人员、单位、供应商、物品统一代码 的申报工作。

**45.2** **两级数据管控模式**

航空工业的信息编码工作采用两级管控的模式：集团级——集团制定整体战略，并发布标 准体系；组织制定核心标准，并建立统一的代码数据库及管理系统；企业级——各单位申请代 码，并在产品研制过程中应用代码。

自2012年开始，航空工业建立了完备的代码申请、审核与赋码的动态管理机制，下属的各 单位开始在集团公司的代码管理系统中申请各类代码。

例如，对物品码按照“申请→审核→赋码→发布→使用”的流程进行管理。首先用户单位 向编码中心提交代码申请，编码中心在接到申请后进行审核和赋码，赋码后编码中心将新物品 码在代码管理系统中发布并反馈给用户单位。同时，用户单位可以查询、下载相应范围的物品 码及其属性信息并使用。

对于申请的集团统一代码数据，各个单位建立了自己的主数据编码管理系统以便对其进行 管理，并作为单位统一的数据源，为各个业务系统建设和集成提供一致、规范的代码数据。

编码管理工作同时也作为航空工业各单位的常规工作任务之一开展。航空工业针对每一类集

**460** 数据治理——工业企业数字化转型之道

团统一代码(人员代码、供应商代码、物品码)制定了相应的申请审核流程，在单位内部审核通 过后，再提交给集团的编码中心进行审核及赋码。对于单位内部统一编码的对象，航空工业也制 定了代码的申请、审核、维护、注销的工作流程，并明确签审节点的职责，从而保证代码在各单 位内的良好应用。

**45.3** **信息分类与编码标准**

**1.** **标准体系制定**

航空工业的信息分类与编码标准体系是信息化标准体系的重要组成部分，是航空工业各单 位开展信息分类与编码标准的研究与制定、贯彻实施、应用检查等工作的顶层指导文件，也是 各单位开展信息资源规划、数据治理和信息系统集成的基础标准和参考依据。

航空工业的信息分类与编码标准体系包含以下3个方面的内容。

(1)信息分类与编码基础标准。

信息分类与编码基础标准主要包括术语和定义、方法标准、元数据与数据字典标准、信息 分类与编码管理标准等，为信息分类与编码工作在航空工业各业务领域中的开展提供了通用要 求，规范了信息分类与编码标准的规划、编制和贯彻实施。

(2)主数据信息分类与编码标准。

主数据信息分类与编码标准包括产品及零部件分类与编码标准、物资分类与编码标准、文 件(文档、资料)分类与编码标准、主数据技术应用标准与编码数据交换及集成共享标准。

(3)业务数据信息分类与编码标准。

业务数据是指为完成航空产品研制所涉及的各项业务活动而产生的事务性数据，是对业务 活动的动态描述。航空工业按企业、价值链、产品生命周期3个维度对业务数据信息分类与编 码标准开展：规划和管理。

**2.** **标准制定与贯彻**

(1)标准编制的总体原则。

航空工业的标准编制的总体原则包括以下几条。

①在制定集团级信息分类与编码标准时，遵循“顶层、科学、实用”的原则从集团统一赋 码和顶层管理的需求出发，充分兼容和考虑各单位对物品码的使用需求，将编码标准作为物品

第45章 航空行业：面向航空装备研制生产的数据治理研究与实践 **461**

码赋码的依据。

②集团物品码的应用贯穿在产品全生命周期中使用，同时也应用在各单位的 PDM 、ERP、 CAPP、MES 等信息系统中，作为信息传递的纽带。

③从集团公司和各成员单位信息化建设和应用角度出发，要统一协调航空工业各单位对于 物品种类的标识需求，并给出编码规则。

④从物品码的编制要求来看，要求做到科学性与实用性相结合。

(2)信息编码标准的制定。

下面以物品码标准的制定为例，介绍航空工业的信息编码标准的制定。

在集团统一管理的代码中，物品码是保障产品协同研制的核心。物品码是指在航空工业中 用于标识物品种类(指区分不同特征、功能、参数等的物品类型)的代码，其实质是一种物品 种类的标识符，而不是物理个体的标识符。此处所指的特征、功能、参数包括如下。

●特征：包括外形特征、接口特征、物理化学特征等技术特征。

●功能：主要指产品能实现的功效和作用。

●参数：主要指产品的型号规格、功能技术指标值等。

航空工业在确定物品码编码对象范围时，主要从以下两个出发点考虑。

首先，要解决产品协同研制生产中对产品统一编码的问题。因此，先要实现集团层面产品 代码的统一，为协同研制生产提供统一、规范的代码数据，为产品交付后的使用提供信息化服 务。

其次，要解决集团多数成员单位共性的编码问题，统一代码要具有基础性、实用性、共享 性。因此，还要统一集团范围内的基础通用物品码，能为集团信息化建设提供规范的基础通用 数据，提升航空工业信息化建设的整体水平。

根据上述原则，航空优先工业确定了在科研、生产中涉及的下列产品应统一编制物品码：

●基础通用产品：包括标准件、材料、电子元器件、通用工具工装；

●成品及附件；

●飞机、直升机、发动机等主机产品及其保障设备；

●其他在集团公司范围内需要申请物品码的产品。

对物品码标准而言，确定物品码的代码结构是最重要的技术内容之一。航空工业在确定物 品码的代码结构过程中，首先收集了国内外现有相关先进编码标准，研究并分析其代码结构及

**462** 数据治理——工业企业数字化转型之道

特点，并与企业标准化专家、信息化专家就代码结构方案充分交流与探讨。同时，航空工业充 分考虑了集团统一赋码需求和各单位对物品码的应用需求，并结合目前信息编码技术发展的主 流趋势，最终确定采用8位无含义流水码的代码结构。

航空工业的物品码标准已经在集团范围内得到了很好的应用，并广泛应用在各个航空产品 的研制生产全过程中，数据总量和数据质量也处于很好的状态。表45-3-1所示的是集团物品码 数据示例。

**表45-3-1** **集团物品码数据示例**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 物 品 码 | 产品名称 | 一级  分类 | 二级  分类 | 三级  分类 | 标准级别 | 标 准 号 | 规格单位 | 材料牌号 | 表面 处理 |
| 10212943 | 六角头全螺 纹螺栓 | 标准件 | 紧固件 | 螺栓 | 国家标准 | GB 5781-2000 | M20×30 | ML35 | 镀锌 钝化 |
| 10213054 | 半圆头螺栓 | 标准件 | 紧固件 | 螺栓 | 航空行业标准 | HB 1-152-1995 | 4×18 | ML30CrMn  SiA | 氧化 |
| 10213055 | 半圆头螺栓 | 标准件 | 紧固件 | 螺栓 | 航空行业标准 | HB 1-152-1995 | 4×20 | ML30CrMn  SiA | 氧化 |
| 10213317 | 六角头螺栓 全螺纹 | 标准件 | 紧固件 | 螺栓 | 国家标准 | GB 5783-2000 | M6×12 | 1Cr11Ni2W  2MoV |  |
| 10213354 | 六角头螺栓 | 标准件 | 紧固件 | 螺栓 | 国家标准 | GB 5783-2000 | M3×12 | 30CrMnSIiA | 镀镉 钝化 |
| 10213375 | 六角头螺栓 | 标准件 | 紧固件 | 螺栓 | 航空行业标准 | HB 1-101-2002 | 4×18 | ML30CrMn  SiA |  |
| 10213376 | 六角头螺栓 | 标准件 | 紧固件 | 螺栓 | 航空行业标准 | HB 1-101-2002 | 4×19 | 30CrMnSiA |  |
| 10213576 | 螺栓 | 标准件 | 紧固件 | 螺栓 | 航空行业标准 | HB1-101-1983 | 5×16 | 1Cr17Ni2 |  |
| 10213577 | 螺栓 | 标准件 | 紧固件 | 螺栓 | 航空行业标准 | HB1-127-1983 | 4×20 | 1Cr18Ni9Ti | 钝化 |

**45.4** **数据应用场景**

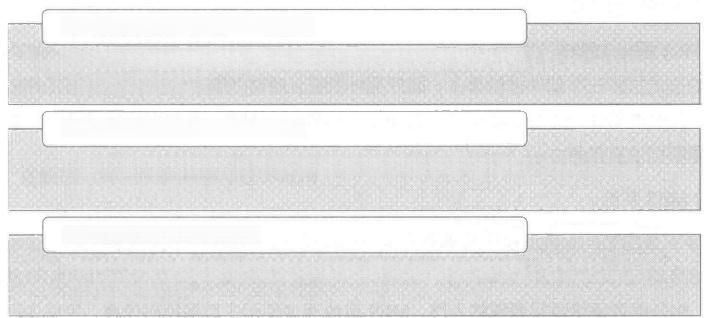
**1.** **在产品设计环节的应用**

在生产制造型企业中，设计部门往往掌握着企业的核心技术。设计部门的主要工作内容是 根据产品需求和技术指标要求开展产品的设计工作，产物为图纸及技术文件，用以指导下游部 门制造产品。其通用的业务流程如图45-4-1 所示。

由于设计部门是产品数据的源头部门，下游的业务部门均会依据设计物料清单(BOM) 等 资料开展工作。例如生产部门将以设计 BOM 为基础指导生产，采购部门以设计部门提供的外

购件清单和设计 BOM 为依据进行采购。因此，设计环节的数据质量及编码应用效果，将直接

影响企业的生产成本及产品质量。



**1.产品总图的设计与产品BOM** **的搭建**

·由该部门产品总设计师或主设计师进行总图的绘制，并制定各个零部件图的图号与编 码，搭建产品BOM。

**2.零部件图纸的绘制，零部件图清单**

·根据总图的绘制，结合技术协议的要求，进行相应的计算校核、材料选择、外购件协 议、零部件图绘制等工作。

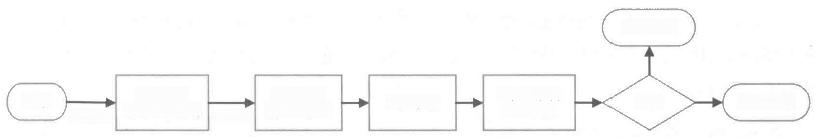
**3.校核图纸等文件，批准发图**

·对设计工程师绘制的零部件图及其他过程文件：计算表、外购件协议等文件进行审核， 审核后核准发布。

**图45-4-** **1** **设计部门通用业务流程**

**2.** **在** **制** **造** **环** **节** **的** **应** **用**

工艺部门在接到设计文件和 EBOM 之后，要设计产品工艺路线、绘制产品加工图等工作， 最终形成 PBOM。PBOM与生产订单一同被下发到生产部门，由生产部门在 MES 系统中进行 排产，并将生产任务下发到工人和设备上。物料和库房管理员进行备料，设备管理员对设备进 行调试。做好生产准备工作后，工人按照工艺路线依次完成每一道加工工序及送检工作，检验 合格的产品即可交付入库。生产制造环节通用的业务流程如图45-4-2所示。



废品处理

开始

EBOM

工艺部门 编制工艺路线

生产部门 排产、派工

人员、设备 生产执行

结束入库

生产准备

PBOM

送检

图45-4-2 生产制造环节通用的业务流程

在生产制造环节中对于数据规范化的要求，主要体现在通过数据的载体在不同业务部门和

环节之间的信息交互。常见的产品生产数据载体如下。

(1)工艺流程卡。

工艺流程卡随待加工产品一同被下发到工人手中，并在工序之间流转，承担着对生产车间

现场进行进度管理、状态管理、检验等重要职能。工艺流程卡中记录着物料编码、产品名称、

**464** 数据治理 工业企业数字化转型之道

产品图号等重要的主数据编码信息。通过工艺流程卡上的物料编码信息，工人可以在 MES 系统 中唯一、准确地确定一条生产任务，方便现场工人查询订单生产进度。

(2)零部件检验单。

在MES 系统中生成的零部件检验单，包括产品物料编码、批次号、生产检验单的编码，并 以条码的形式打印在零部件检验单上，随产品一起在工序中流转。其中记录了各工序的检验结 果信息，同时能够保证合格的零部件检验单与实物的——对应。借助条码技术能够快速地将检 验信息更新到信息系统中。

(3)MES 系统。

MES 系统是生产制造过程中最重要的信息管理系统。物料、人员、工具工装、设备等主数 据编码信息均在MES 系统中流转。生产管理人员根据产品的基本属性、加工属性和财务属性进 行排产，将任务指派给满足要求的人员。MES 系统通过使用主数据编码信息，完成生产制造环 节数据的贯通。

(4)二维码标印技术和 RFID 芯片。

通过二维码(又称二维条码)标印技术和 RFID 芯片，可以将主数据编码信息在实物上体 现，实现虚拟数据信息与物理实体的关联。在车间现场管理中，通过应用二维码标印技术和RFID 芯片，使信息采集的速度更快、准确度更高、操作更加方便，有利于数据的交换与读取，确保 了信息流与实物流的一致。

**3.** **在采供环节的应用**

采供环节是整个生产制造线中最为重要的环节之一。在企业中，这块业务的实现通常需要 多个部门协同开展，是一项涉及频繁协作和信息交互的业务。所以，在采供环节采用统一的数 据语言尤为重要。

采购清单是采购业务的主要依据之一，是根据生产要求采购的原材料、标准件和成套部件 等产生的，对应的文本载体主要包括外协件明细表、自制件明细表和材料明细汇总表。采购清 单信息一般来源于设计图纸和工艺流程卡上的信息汇总，由采购部门或生产准备部门根据采购 清单安排采购计划和生产计划。PDM 系统一般都可以根据图纸和工艺信息汇总相应的采购清单 信息，但是，如果要针对产品批量获得动态的采购清单信息，就必须在ERP系统中完成。表45-4-1 的是采购清单示例。

第45章 航空行业：面向航空装备研制生产的数据治理研究与实践 **465**

**表45-4-1采购清单示例**

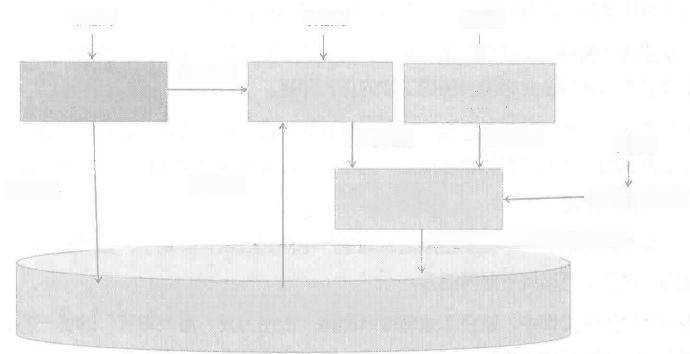
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序 号 | 物料编码 | 材料名称 | 型号规格 | **单位** | 库存量 |
| 1 | 10046090 | 铆钉 | 2×6 | **个** | 0 |

物料编码在采购清单中作为物料唯一的标示信息，可以使采购部门准确地识别物料。同时， 物料编码也是采购部门与库存部门检查库存时使用的唯一标示。所以，统一编码对于采购清单 的准确性有决定性的作用。

**4.** **在航空产品协同研制中的整体应用**

在航空产品研制过程中，航空工业统一代码是飞机整机及机载产品的唯一合法索引，也是 飞机装机数据采集、交付、现场维修等全生命周期管理的信息系统之间流转的主线。图45-4-3

为航空产品研制过程中的统一代码应用场景 。



**物品码**

↓

主机厂

产品标识

单件数据

**主机所构型数据**

装机数据(整机单件交付数据包)

**代码管理系统(编码中心)**

主机厂

装机数据采集软件

主机厂

成品接收软件

配套单位 产品标识软件

单件数据

单件数据

实物条码

单件数据

**物品码**

**物品码**

**物品码**

**图45-4-3** **航空产品研制过程中的统一代码应用场景**

各配套单位负责研制的产品，统一在集团代码管理系统中申请统一代码，各单位将统一代 码在条码标印工具中进行管理 ， 并通过标印工具将代码连同其个体码生成二维条码 ， 标印在产 品的实物表面、外包装和履历本上。航空产品在交付时需要同时交付带有集团统一代码的条码 信息、产品信息、履历信息等。

配套单位将产品的单件数据再次上传到集团代码管理系统中 ， 形成集团统一的航空产品单 件库。编码中心再将数据连同产品接收软件一起发给主机厂。主机厂在接收到来自配套单位交

付的产品时 ， 进行 “ 三码合一 ” 的校验 ， 即验证产品实物 、 外包装 、 履历本与产品接收软件中

**466** 数据治理——工业企业数字化转型之道

的条码的一致性。

当有了上述基础后，飞机产品在装机时，可以利用编码中心开发的装机数据采集软件，对 每一架次飞机的物理构型和计时构型进行采集和管理，作为综合保障系统的初始数据。

在开展此项工作之前，工业部门之间、用户与工业部门之间的编码数据规则、规范不一致， 产品标识各不相同，实物和信息难以关联，导致航空产品在不同研制单位之间的协同程度低， 研制成本高，并且产品备件一致性、互换性差，维修的难度和成本也很高。集团统一代码在产 品研制生产中贯彻应用，有效解决了上述问题。

**45.5** **总结与展望**

在航空工业集团公司的统一领导和持续推进下，以及在各相关单位的积极配合及支持下， 航空工业的统一编码工作顺利开展，并取得了丰硕的成果，主要包括：

(1)集团统一编码工作制度完善、组织机构健全，集团统一信息编码管理体系有效运行， 有效地支撑了集团级和企业级信息编码工作的正常开展。

(2)在航空行业内，对于信息编码的重要性认识到位、执行力强，基于集团统一代码及企 业编码数据标准的数据治理工作稳步推进，为集团公司与企业的信息系统的集成和业务数据的 贯通起到了较强的促进作用。

(3)条码标印与识别技术在航空产品研制生产中得到应用，实现了产品质量信息全生命周 期的可追溯，提升了产品服务保障能力。

在下一个阶段的工作中，航空工业将继续坚持“应用为先，落地做实”的统一编码工作原 则，拓展统一编码和单件采集等工作应用范围，持续支撑航空产品研制生产数智化的提升和转 型。

首先，坚持贯彻集团标准化顶层战略，做好信息分类与编码标准化专业的顶层规划，完善 标准体系，编制一批急需的信息分类与编码标准，在集团范围内开展贯彻应用工作，支撑各单 位的信息系统建设与集成。

其次，丰富标准体系中的条码及 RFID 等自动识别技术的应用标准，积极推进条码等自动 识别技术在产品研制生产中的进一步深度应用，重点解决条码标印与识别设备、信息系统的应 用集成。

第45章 航空行业：面向航空装备研制生产的数据治理研究与实践 **467**

最后，基于航空产品单件数据交付工作，完善产品单件数据交付标准及数据管理模式，重 点推进产品单件状态数据在交付、用户使用等环节的应用，实现单件平台汇总及传递设计、制 造、交付、备件管理、使用、大修、报废等全生命周期环节的状态数据，为产品全生命周期大 数据分析提供支撑，进而实现航空产品全生命周期数据状态可视化。

**第46章**

重型装备制造行业：数据标准，装备中国——

中国一重的数据标准化管理项目

**46.1** **背景介绍**

**1.** **公司介绍**

中国一重集团有限公司(以下简称“中国一重”)主要为钢铁、有色金属、电力、能源、汽 车、矿山、石油、化工、交通运输等行业提供重大成套技术装备、高新技术产品和服务，以及 开展相关的国际贸易。

中国一重正加快推动传统产品的优化升级，在做好装备制造板块的同时，大力发展新能源、 节能环保、新材料、农业机械、金融等新业务板块，努力形成优势突出、结构合理、创新驱动、 开放协同的发展新格局，以及成为多元化发展、多级支撑的现代化企业集团。

**2.** **行业特点及信息化建设背景**

中国一重属于离散型装备制造业中的重型机械行业，其特点是产品类型多，产品结构复杂 (少则由几千个零件组成，多则由几万个零件组成),覆盖的工艺类型多，生产工艺流程长及制 造周期长。由于中国一重的客户对产品的需求各异，在经营过程中，经营人员、产品和工艺设 计人员共同进行经营活动，并需要与用户进行长时间的技术交流。虽然产品复杂，但市场要求 产品设计周期要越短越好。

中国一重的生产形式决定了其信息化实施工作与其他企业有很大的差异，其产品和生产特 点如下。

第46章 重型装备制造行业：数据标准，装备中国——中国一重的数据标准化管理项目 **469**

(1)产品品种多、生产数量少、规格多变，按客户订单进行设计、生产，工作性质依客户 要求的品种、规格、交货期、价格而定；产品各部件之间的时序约束关系和成套性要求严格， 关键设备的能力平衡和利用率是生产与控制的关键环节。

(2)产品结构复杂、生产周期长、重复作业比率低，较难采用流水线或专用工装设备生产。 为了保证按期交货，企业一般难以在技术资料全部准备齐全后才开始生产，而是边设计，边生 产，边修改。

(3)产品质量难以控制。虽然企业会采用全面的质量管理体系，但生产中存在的质量问题 常常会造成工期延误，打乱生产计划，致使制造成本上升，不能及时交货。

(4)成本难以动态监控。大型单件产品是根据订单组织生产的，产品工艺成本及构成需要 在完成产品设计、工艺编制，以及形成物料清单、工艺线路、工时定额后才能确定；产品定额 成本的控制视工程项目而不同，并且数据量大，从而导致成本控制困难。

(5)对交货期要求严格，且每次订单上所需要的产品同以往生产的产品都有区别，虽非全 新产品，但可能在设计、尺寸、形状上有新的变化。企业根据交货期，以订单中的独立需求为 对象下达工令号组织生产。

(6)生产组织困难。大型单件产品在完成产品设计和工艺编制后，才能制订出较为合理可 行的生产计划和作业计划，因此生产准备周期较长。在生产组织中也不可避免地存在如下问题： 生产资源冲突难以避免，产品总装不能有效配套，物资供应难以保证，等等。

因为以上这些特点，导致离散型装备制造业的信息化建设和整合难度大。因此，离散型装 备制造业的信息化建设应以提高企业竞争力为宗旨，以实现产品设计信息化、生产过程自动化、 管理现代化、决策科学化为目标，以企业的信息流、资源流、业务流融合统一为途径，以提高 产品性能、产品质量、缩短产品研制和生产周期，以及提供优质和及时的服务等为落脚点，分 层次、分类型推进，分阶段、分目标落实。

**3.** **数据治理背景及必要性**

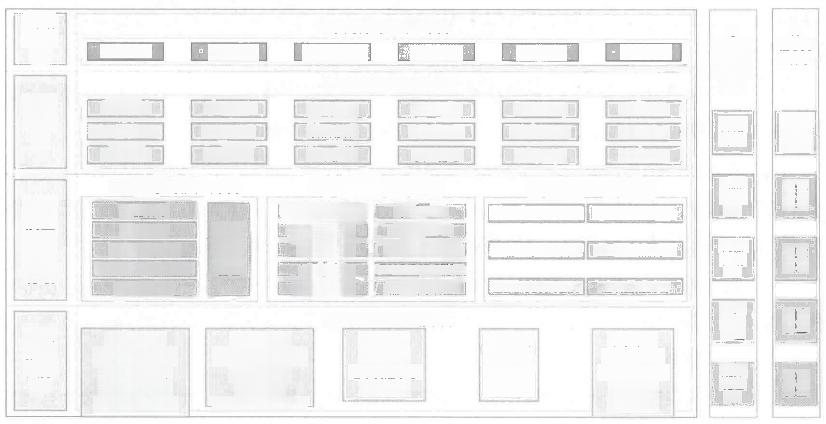
中国一重的信息化建设的重点是开展平台化整合和系统集成。做好此项工作的前提是统一 数据标准，提高数据质量，尤其是主数据质量，其直接关系到平台化整合的成败和效果。中国 一重的信息化规划架构图如图46-1-1所示。

(1)数据治理是提高信息化标准管理水平的迫切需要。中国一重的信息化建设正在逐步展 开，众多已建系统、在建系统、未来规划建设的系统之间使用着多种基础信息数据，如物料、

安 腐

组织机构、员工、客户、供应商等数据。这些数量众多的基础信息数据，缺乏统一的标准、数 据规范、管理手段，导致各信息系统之间的数据不一致、不正确，不能集成应用。中国一重迫 切需要通过数据标准化管理平台加强数据治理，有效缓解集团数据代码标准化管理工作的压力，

切实提高公司的信息化标准管理能力。



G

信息化

治理与

标准化

体系

数据 标准

电子商务

主数 据管 理

元数 据管 理

业务

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **工业物联网**  **标识与识别系统** **数据采集系统** **智能检测设备** **装备网络化集成** | **企业私有云**  **私有云模式与架构**  **服务器系统**  **存储系统**  **网络系统终端系统** | **F** **集成化基础应用平台** **数据中心**  **机房设施**  **运维管控中心**  **容灾备份** | 融合通信  **视频会议**  基础链路  安防集成 | 展示中心  管理驾驶舱  生产调度指挥  能源管控  运营状态 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **D** **智能化制造执行平台**   |  |  | | --- | --- | | 生产作业指导书  生产专业化 生产协作 质量管理 能源管控 | 军工MES  洁净钢MES  高效物流配送 DNC/MDC管理 生产现场监控 | | **E 协同化现代服务平台**   |  |  | | --- | --- | | 产品故障远程诊断  产品运行状态管理  工程协同设计服务 | 再制造应用服务  工程项目群管理  后端营销管理 | |

|  |  |  |
| --- | --- | --- |
| **财务核算**  售后服务  基建技改 | 费用报销 集中采购 资产管理 | 资金管理  供应链管理  成本核算 |

**A** **敏捷化辅助决策平台**

管理驾驶舱 国资监管 知识管理

|  |  |
| --- | --- |
| 人力资源  **客户关系管理**  **科研管理** | **全面预算** **销售管理**  **投资项目** |

产品设计

工程设计

产品物料清单

计算机辅助工艺设计

仿真与优化

**研** **发** **及生**

**翻** **行层**

**基** **础** **技** **术** **支撑**

层

战略 决策 层

**运营** **管** **理**

层

企业门户 协同办公

数非

管

规范 化运

维

财务报表

费易管理

**C** **数字化产品开发平台**

**B 精益化运营管控平台**

H

信息安 全体系

戆

据 金 数 安

综合报表

品命期理

产生周管

用 全

应 安

据 理

端 全

网安

络全

**图46** **-** **1** **-** **1** **中** **国** **一** **重** **的** **信** **息** **化** **规** **划** **架** **构** **图**

(2)信息孤岛和信息化重复建设是信息化工作中始终要解决的问题。数据治理有助于减少 信息孤岛、避免信息化重复建设，具体主要体现在以下几点。

①保证系统数据统一标准、统一管理。中国一重的信息系统呈“竖井式”,往往一个产生数 据有多个源头，从而造成了数据不一致、数据不统一、数据不准确等现象。通过数据治理可以 实现各类数据统一标准、统一管理。从数据创建、修改、应用等多个关键环节进行控制，可以

确保数据“一数一源”及“—物一码”。

②提升数据质量，为系统集成提供基础保障。中国一重将一致的、唯一的、标准的数据发 送给各系统应用，确保异构的信息系统中的数据的正确性和标准性，以及各信息系统应用的数 据标准一致，并在此基础上实现系统的集成工作。

③提高数据的复用性，降低信息化建设成本，缩短建设工期。中国一重形成全集团的标准

数据资源库，标准数据资源库中保存着各类标准的数据，当企业新建系统时，可以快速应用已

有数据，无须重新收集、分类、整理数据，降低信息化建设的成本和缩短信息化建设的工期。

第46章 重型装备制造行业：数据标准，装备中国——中国一重的数据标准化管理项目 **471**

中国一重于2017年开始建设数据标准化平台，进行数据治理。

**4.** **数据治理难点分析**

质量、成本、交货期和新产品开发速度决定了企业的竞争力，离散型装备制造业对设计数 据、产品数据、营销数据及制造数据依赖程度比较高。中国一重在其信息系统建设过程中，主 要面临以下4个方面的困难。

(1)编码体系、产品结构、生产模式和服务过程复杂，业务多变性强。离散型装备制造业 生产具有典型的多规格、多品种、小批量，甚至单件定制或按订单定制的特点，并且零部件品 种及规格多，产品结构复杂，从而使得生产计划调度十分困难。

(2)不同形态的生产车间数据管理需求不同。离散型装备制造业的每个生产车间的生产对 象和加工工序都不同(如中国一重的装备制造板块的生产过程包括炼钢、模型、热处理、焊接、 机加、装配、包装、运输等),因此具有以下特点：

●生产过程经常处于边设计、边生产、边修改的状态；

●对 BOM 数据的分段发布、并行管理和汇总能力的要求强；

●生产过程难以达到稳定状态，需要较强的生产应变能力；

●客户需求变化大，动态变更频繁，对响应速度的要求高；

●工艺计划与生产过程密切关联，需要实时跟踪。

(3)数据采集自动化程度低。中国一重各生产单位的自动化程度都不同，目前其大部分生 产车间都还没有引入自动数据采集技术，只有部分数字化车间的数控机床联网系统可以直接从 数控系统中获取生产信息和机床运行状态信息。

(4)系统集成非常困难。中国一重自大规模启动信息系统项目建设以来，相继实施了设计、 合同、生产、物流、财务、设备等多个信息系统建设。而贯穿于多个系统的基础数据入口没有 固化。有的是在某一个系统中产生并维护部分属性后被传递到其他系统中使用，在使用过程中 会扩展属性，甚至会改变原来已有的属性值，造成各系统数据不一致，以及系统集成困难。随 着集团信息化进程的不断推进，信息化程度的不断深入，各业务信息系统的持续运行，产生了 大量的数据。分散的业务数据让决策者无法总揽全局，传统的查询统计功能限制性又强，复杂 的数据环境及数据不规范、不完整、不一致等造成了数据质量差。

**472** 数据治理 工 业 企 业 数字化转型之道

**46.2** **数据治理概况**

**1.** **数据治理目标**

中国一重建立数据标准化管理平台，旨在实现对公司各类数据的统一标准、统一管理，实 现在整个集团的数据资源共享服务。建立统一、集中、规范的数据标准化管理平台，可以提高 数据质量及数据决策分析的效率，实现各单位、各系统之间的互通互联，切实加强集团信息的 共享服务能力及标准化水平。

中国一重具体要实现如下目标。

(1)建立数据标准化体系，制定主数据标准、数据标准化制度、流程等规范性文件。

(2)收集现有的数据信息，组织专家完成通用基础、财务、人员、内部单位、外部单位、 物料、生产共7大类主数据的清洗工作，并形成标准代码库，提升数据质量。

(3)建设数据标准化管理平台，并完成相关业务系统的改造，实现数据标准化平台与各业 务系统的集成，达到统一标准、统一编码、统一管理、统一运维的目标，以及实现集团统一管 控主数据。

(4)组建数据标准化运维团队，建立数据标准化运维体系，规范数据新增、修改、审核、 冻结、解冻、查询等管理流程，实现对主数据的全生命周期管理。

(5)规范各业务系统数据，打通从设计到制造的全流程，消除信息孤岛，为数据统计、分 析及商业智能提供支撑。

**2.** **数据治理过程**

数据治理要从基础的主数据治理着手。经调研分析，中国一重的物料主数据管理主要存在 以下问题：

●数据分类不科学，描述不规范，存在大量重码、错码、废码。

●计量单位不准确，同类物资计量单位不唯一，数据质量不高，编码利用率低等。

●其他(通用基础、财务、人员、内部单位、外部单位、生产)6类主数据存在编码不统 一、描述不规范、入口不唯一、流程不规范等问题，以及各业务系统各自为政。

2017年6月，中国一重启动数据标准化管理平台建设项目。通过此项目，围绕中国一重各 项业务和现有信息系统，在集团范围内建立一套科学、适用的主数据标准和管理流程，为中国 一重的信息共享和管理奠定信息标准基础。



第46章 重型装备制造行业：数据标准，装备中国——中国一重的数据标准化管理项目 **473**

物料主数据标准化是数据标准化项目中实施的重点与难点。中国一重的物料涉及范围广， 纵向细分深。项目组经过研究，决定从下属各单位中抽调50名专业人员组成中国一重物料主数 据标准化组，下设9个专业小组(配套件组、金属材料组、专项产品组、设备备件组、工具组、 冶金炉料组、化工材料组、电器仪表组、产品组),经过1个月的集中办公，共同制定了中国一 重的物料主数据标准，如图46-2-1 所示。

中国一重物料主数据标准化组

组长：XXX

副组长：XXX、XXX、XXX

技术支持：XXX、XXX、 XXX

亚

平

平

业

亚

工

工

工

工

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 配套件组  组 长 ： X X X 副组长：XXX  组 员 ：  XXX  XXX  XXX  XXX  XXX  XXXXXX  XXX | 金属材料组  组 长 ： X X X 副组长：XXX  组 员 ：  XXX  XXX  XXX  XXX  XXX  XXX  XXX | 专项产品组  组 长 ： X X X 副组长：XXX  组 员 ：  XXX  XXX  XXX  XXX  XXX  XXX | 设备备件组  组 长 ： X X X 副组长：XXX  组 员 ：  XXX  XXX  XXX  XXX  XXX  XXX | 工具组  组 长 ： X X X 副组长：XXX  组 员 ：  XXX  XXX  XXX  XXX | 冶金炉料组  组 长 ： X X X 副组长：XXX  组 员 ：  XXX  XXX  XXX  XXX | 化工材料组  组 长 ： X X X 副组长：XXX 林  组 员 ：  XXX  XXX  XXX  XXX | 电器仪表组  组长：XXX 副组长：XXX  组 员 ：  XXX  XXX  XXX  XXX | 产品组  组长：XXX  组 员 ：  XXX  XXX  XXX  XXX |
| 10人  要求熟恶液压  缸、减速机 接轴、联轴器 气缸、自润滑  青 假 和 。律 簧、碟簧.紧 固件、管路射  件等 | 11人  要求熟恶金属  材料如型材，  板材、管材  焊材、有色金  属及金属制品  等物质 | 6人  要求熟恶专项  产品，专用配  套件与专用化  工材料等物资 | 5人  要求熟恶轴承，  液压件、管件  制动器、高合  器、机加件等  物资 | 7人  要求熟恶进口  刀具、量具、  刃具、五金工  具、磨具磨料，  焊接工具等物  资 | 7人  要求熟恶合金、  生铁、废钢、  耐火材料、冶  金辅料、煤炭  等物资 | 7人  要求熟恶油品  化工、劳保  杂品、木材.  气体、轮胎、  油漆、办公用  品等物资 | 6人  要求熟悉电器、 仪表、电偶、 电机、风机  电脑耗材等物  资 | 8人  要求熟悉产品  等物资 |

**图46-2-1** **物料主数据标准化组及业务分工**

接下来是物料数据清洗，其中涉及集团的各项业务，技术性强、数据量大。物料数据清洗 历时3个多月，共清洗库存物料数据6万余条。为了让数据标准化成果能够用得起来、用得方 便、用得好，2018年年初，中国一重采用一种全新的方式对将要设计的产品进行预先清洗，这 样既不影响设计部门的工作效率，又能让数据标准化成果得到深入和快速应用。据统计，在数 据清洗阶段中国一重共计清洗了12万余条物料数据。同时，在此阶段中，中国一重结合业务特 点和数据标准化应用体会，将大家遇到的问题进行整理和汇总，创新性地编制出了《物料分类 说明及提报指南》和《数据标准化管理平台应用100问》,极大地提高了工作效率。

数据标准化平台的建设和应用，需要由专业、稳定的运维队伍来完成。中国一重的领导高 度重视，提前组织、抽调中国一重物资仓储部领导全职担任数据标准化运维组组长，并组建了 由8人组成的专职物料编码审核小组，分别负责物料编码审核。

在项目实施阶段，中国一重开发了物料类别快速检索功能，从物料申请人员的角度看，这 方便了物料申请人员快速找到对应分类，降低了对申请人员的专业性要求，以及申请人员对系

**474** 数据治理——工业企业数字化转型之道

统的熟悉度；从物料运维人员的角度看，这方便了物料运维人员对模板合理性的检验，以及对

同类物料的科学归类。

**46.3** **数据治理成果**

中国一重的数据治理项目主要取得了以下成果。

**1.** **建立数据标准**

中国一重建立了以下数据标准。

(1)建立《主数据标准》,涵盖通用基础、财务、人员、内部单位、外部单位、物料、生产 共7大类主数据，以及规范编码规则、分类规则、描述规则等，对数据的新增、修改、审核、 冻结、解冻等流程进行规范化管理。

(2)建立《数据标准化平台接口规范》,规范各业务系统与主数据系统的接口规则。

(3)建立《数据标准化运行管理办法》,规范数据运维的职责划分、管理流程和考核。

(4)编制了《数据标准化管理平台操作手册》《数据标准化管理平台应用100问》《物料分 类说明及提报指南》等辅助文档，有利于数据标准化平台的推广及应用。

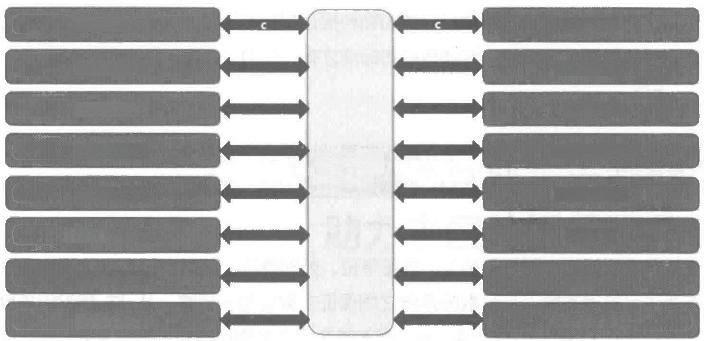
**2.** **形成标准数据资源库**

通过近一年的努力，数据标准化项目形成通用基础主数据4000条，会计科目主数据2000 条，银行主数据15万条，内部单位主数据800条，外部单位主数据1.6万条，银行账户主数据 1万余条，人员主数据1万条，物料代码12万条，生产主数据2万条，共计33万余条主数据， 培训1600余人次，重要交付物达到24项。

**3.** **实现业务系统集成**

中国一重以数据标准化管理平台为核心，完成主数据管理系统与集中采购系统、合同管理 系统、人力资源系统、财务核算系统、资金集中管理系统、费用报销系统、全面预算系统、协 同办公系统等系统的集成，实现了数据源头唯一，数据标准一致，数据规范共享。图46-3-1为 中国一重主数据管理系统集成对接情况。

第46章 重型装备制造行业：数据标准，装备中国——中国一重的数据标准化管理项目 **475**



**协同办公系统**

**人力资源系统**

**设备管理系统**

**全面预算系统**

**财务核算系统**

**资金集中管理系统**

**费用报销系统**

**综合报表系统**

MES系统

**报产发运系统**

数据标准 化平台

集中采购系统

合同管理系统

CAPP工艺系统

CRM系统

PDM 系统

物流系统

**图46-3-1** **中国一重主数据管理系统集成对接情况**

通过主数据与各业务系统的深度集成，中国一重实现了物料主数据的全流程、全生命周期 管理，有效降低了库存、采购成本与管理成本，提高了工作效率。

**4.** **搭建运维体系**

数据治理并不能通过企业的单一部门来完成，需要 IT 部门与业务部门协作，而且必须始 终如一地进行协作，以改善数据质量，从而为关键业务提供支持。数据标准化管理平台的建成 只是数据治理的开始，还需要对数据质量进行持续管控，因此，中国一重组建了数据标准化运 维组，由15名业务专家组成的数据标准化运维组按数据类别进行分工，其中，8名物料专家专 职负责物料主数据的分类和审核工作。运维组每个月在集团范围内对主数据的运行情况进行通 报，主要反馈数据总量及变动情况、各单位主数据的提报通过率、存在或需要注意的问题等， 督促各单位认识到数据的重要性，不断提高数据质量。

**46.4** **总结与成效**

**1.** **项目总结**

(1)要从思想上真正理解主数据的长远价值，并使各级领导和员工引起足够重视，特别是 集团管理层的直接参与和支持，将对数据治理项目的成败起到关键作用。

(2)数据标准化管理项目是标准化与信息化高度集成的项目，特别是在标准和规则制定前

期，既要考虑规则的前瞻性和规范性，又要考虑信息化落地的可行性和方便性。

**476** 数据治理——工业企业数字化转型之道

(3)制定的物料标准要包括物料分类和物料描述模板，不能只考虑各自部门的岗位需求，

要站在集团整体层面统筹把控，这样建设的标准才有生命力。

**2.** **主要成效及展望**

数据标准化管理平台是企业信息化建设的基础性平台，对企业信息化的建设质量、应用效 果和数据分析起到支撑和决定性作用，能够规范企业编码、消除信息孤岛、提高数据质量，这 也是系统集成的前提。

通过规范通用基础、财务、员工、内部单位、外部单位、物料、生产这7类主数据，使得 集团能够集中化管理数据，在分散的系统之间保证主数据的一致性。从 IT 建设的角度看，主 数据管理可以增强IT 架构的灵活性，构建覆盖整个企业范围内的数据管理基础和相应规范，以 更灵活地适应业务需求的变化。

通过建立标准化管理平台，中国一重实现了集团各类信息数据的统一标准、统一管理，实 现了在整个集团的数据资源共享服务。通过统一、集中、规范的数据标准化管理平台，中国一 重提高了数据质量及数据决策分析的效率，实现了各单位、各系统之间的互通互联，切实加强 信息的共享服务能力及标准化水平。

通过建立数据标准化管理平台，中国一重主数据治理已经达到行业领先水平。在大数据时 代，数据已经成为企业的核心资产，主数据治理的成功只是数据治理的开端。下一步，中国一 重还将持续推进数据治理工作：

一是充分利用已有的数据治理成果，助力集团数字化转型和智能制造升级，提升集团的信 息化水平，促进集团高质量发展。

二是继续深化应用，整合业务系统，消除业务系统之间的壁垒，打通业务流程。

三是加强对数据的综合利用，加快推进工业大数据分析平台的建立，使业务数据化、数据 资产化、资产价值化、价值最大化。

四是加强对数据质量的管控，在制度、组织、管理流程、人力资源等各个方面给予支持， 不断提升数据质量。

**第47章**

交通物流行业：主数据治理

助力中国外运数字化转型

**47.1** **背景介绍**

**1.** **公司简介**

中国外运股份有限公司(以下简称“中国外运”)是综合物流服务供货商，服务网络覆盖全 国并遍及全球主要经济带。中国外运的主营业务包括货运代理、专业物流、仓储、码头服务、 物流设备租赁、船舶承运、汽车运输和快递服务等，形成了以专业物流、代理及相关业务、电 子商务三大板块为主的综合物流服务体系，为国内外客户提供端到端、高品质的全程物流服务。

**2.** **企业现状**

作为世界先进的物流行业整合商和服务商之一，中国外运一直非常重视信息技术的运用和 提升，通过不断加大投入，持续推进信息化建设，取得了丰硕的成果：各级信息化组织逐步健 全，基础设施不断加强；信息技术应用已经渗透到海运、空运、陆运等业务，以及人力资源、 财务、客户、运营管理等领域。这些信息化建设成果在各自的业务领域中发挥了重要的作用， 也为中国外运的数字化转型升级提供了很好的基础。

近些年来，数据作为企业的重要资产越来越被重视，成为驱动传统企业数字化转型的新动 力。如何实现数据融合，以及如何实现企业内部数据协同和企业外部数据开放和共享成为关键 课题。为此，中国外运明确提出数字化战略，并构建了企业级数据仓库，先后部署了业务综合 分析平台(BI)、 业务可视化平台、大数据平台和统一业务结算管理系统，目的是通过整合和开

**47 8** 数据治理——工业企业数字化转型之道

发利用数据，实现业务运营数字化和管理决策数字化，从而提升企业的运营效率和市场竞争力。

然而，中国外运在实践中遇到了很大的困难，主要是因为存在严重的数据问题，导致数据 难整合、难利用，无法满足企业数字化转型的要求，难以发挥数据价值。这些数据问题主要包 括以下3个方面。

(1)数据标准不统一。

在以往的信息化工作中，信息系统大都是在不同阶段针对不同业务需求建设起来的，虽然 一些业务系统内部已经形成了一些局部、独立的数据标准，但是不同业务系统之间未能形成统 一的数据标准，各系统都是基于自己的需要定义数据代码、数据规格和语义的，致使不同系统 之间的数据标准存在很大的差异，难以实现数据融合和数据共享。比如港口数据，有的系统用 UN/code标准，有的系统用自定义标准，这就造成了数据缺乏统一标准而无法融合。

(2)数据不一致。

企业各系统中的数据都是自成体系、分散管理的，公用的主数据也不例外，这就造成了不 同系统对同一个业务实体的数据描述不一致、数据代码各异等情况，此时无法对不同源头的数 据进行归并，无法消除数据冗余，从而给数据融合和数据共享带来难以解决的困难。比如客户 数据，每个系统都有不同的客户代码、客户名称和描述信息，当将数据集中到CRM 系统中后， 则无法对来自不同系统的客户数据进行合并，无法消除一个客户有多条重复记录所带来的数据 冗余，无法满足客户授信管理和客户分析的要求。

(3)数据质量不高。

企业缺乏有效的数据质量管理，对数据采集、数据传输、数据存储、数据转换处理等缺乏 质量控制，导致数据质量参差不齐，数据的完整性、准确性和有效性低，存在诸多数据垃圾， 难以发挥数据的价值。

**3.** **探索解决之道**

产生以上问题的根源在于企业缺乏统一的数据标准和有效的数据治理，因此，要解决这些 问题还需要从这两个方面着手。中国外运根据自身的实际情况选择了主数据驱动的数据治理模 式，并于2016年启动了主数据管理项目，于2019年年底完成了整个项目的实施工作。

通过这个项目的成功实施，中国外运构建了完整的主数据管理体系，搭建了主数据管理平台， 实现了对主数据的统一管理，很好地解决了数据标准不统一、数据不一致等问题，有效提升了数据 质量，并取得了明显的数据治理效果。本案例主要介绍中国外运主数据管理项目实施过程和成果。

**47.2** **项目实施**

中国外运主数据管理项目于2016年启动，分为规划设计和建设实施两个阶段进行。其中规 划设计阶段主要是梳理需求、设计实施方案、编制管理体系框架；建设实施阶段主要是“定标 准、搭平台、整数据、推应用”。经过3年多的时间，此项目已于2019年年底完成了全部实施 工作，具体介绍如下。

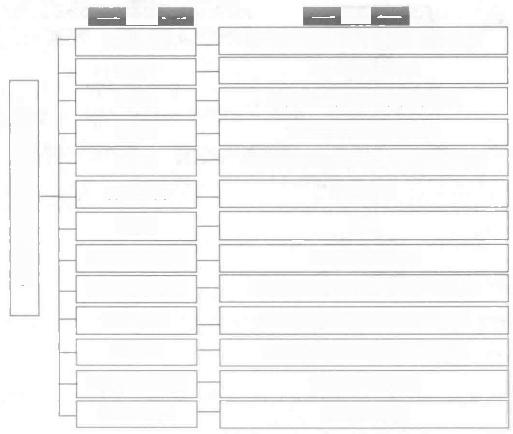
**1.** **总体规划设计**

中国外运主数据管理项目的总体规划设计于2016年9月启动，于2016年年底完成，该阶 段围绕主数据标准体系、主数据管理平台、主数据管理和服务体系进行了详细的调研和规划。 规划内容主要包括以下5个方面。

(1)分析企业业务经营和管理现状，梳理并规划数据管理范围及业务边界，识别数据源头，

编 制 现 状 调 研 报 告 。

(2)从特征一致性、识别唯一性、长期有效性、交易稳定性这4个方面进行识别，确定主 数据范围，对现有主数据进行梳理和清理，形成统一、规范的主数据标准体系表(见图47-2-1)。



中类

行政区划计量方式交通运输行业及其他

内部单位 外部单位

基本信息 岗位信息资质发展信息

科目财务基础类固定资产

业务基础类 业务属性规范值 业务标准数据

物料 设备

项目

合同管理法律业务

印章管理 文档管理 档案管理综合事务

业务类指标财务类指标人事类指标

安全 职业健康 环境保护 应急管理

审计业务风险管理

信息类科技类

大类

通用基础类 单位类

人事类

财务类

业务类

设备物料类

项目类

法律与合规事务类

办公与行政事务类

数据指标类

安健环类

风险管理与内控类

信息与科技类

中国外运主数据标准体系

**图47-2-** **1** **中国外运主数据标准体系表**

》目的：

建立统一的信息

分类代码标准，在信

息化系统建设过程中

采用统一的信息分类

代码标准，作为中国

外运信息代码查询和

应用依据，同时作为

信息代码的全局性和

指导性文件。

》体系表编制原则

1)全面完整

2)分类明确

3)层次恰当

■大类：13个大类

■中类：37个中类

■小类：65个小类

■ 279项主数据

(3)基于数据治理和数据资产管理知识体系与业界最佳实践，设计主数据管理平台建设总 体方案，包括主数据管理功能框架、功能模块、功能点、功能机制及数据交换与集成等。

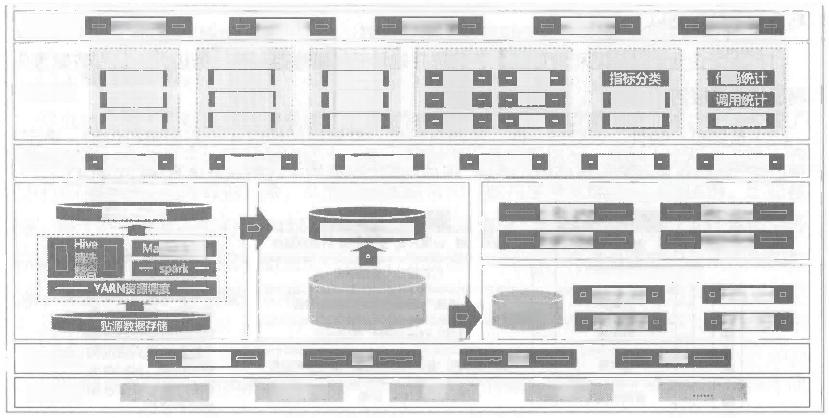
**480** 数据治理- 业 企 业 数字化转型之道

(4)规划实施策略和方法，包括主数据实施的方法论、分期实施范围、组织计划、阶段步 骤、阶段成果及风险控制等。

(5)根据主数据管理要求，从组织、流程和管理规范等方面对主数据管理和服务体系的构 建提出建议。

**2.** **主数据管理平台建设实施**

中国外运按照总体规划设计要求，以业界成熟的产品为基础，并根据自身需求进行扩展， 搭建了集数据采集加工、数据存储、主数据全生命周期管理、主数据服务等为一体的主数据管 理平台。主数据管理平台的功能架构如图47-2-2所示。



单点登录

**体系管理**

体系查询

体系维护

体系发布

用户管理

数据采集加工

主顾效据存确

共享数据资源池

即时服务

**数据查询引擎(ES)**

箱确查询

分词查询

OGG

企查查

下载中心

监控评价 代码

审批效率

任务管理

接口调用

订阅分发

模期查询

全文索引

**展现交互**

**数据管理**

**系统管理**

**服务总线**

**业务系统**

|  |  |
| --- | --- |
| 查询  申请  审核 | 维护  版本  归档 |

**标准管理**

标准维护

标准检素

标准发布

**模型管理**

模型维护

组件管理 校验规则]

**指标管理**

指标维护

指标发布

工作流管理

数据存储

日虑管理 参数管理

**全生命周期管理**

主数据代码库

**数据服务服务**



标准文档

权限管理

页面查询

海运系统

通知公告

待办事项

财务系统

API网关

HR系统

ES存储

EDI

EAI

图47 - 2 - 2 主 数 据 管 理 平 台 的 功 能 架 构 图

主数据管理平台建设包括以下实施要点。

(1)采用业界成熟的主数据管理平台产品，开发并部署主数据管理平台的基础功能，包括 展现和交互、数据管理和系统管理等。

(2)利用数据采集工具(OGG、Sqoop 等)和大数据平台，构建集数据采集、数据清洗、 数据重构、数据存储和数据服务为一体的数据采集处理通道。

(3)完成主数据代码库、只读库、ES 存储的搭建，完成主数据的查询、申请、调用、同步 和ES 智能检索5种服务的开发及部署。

第47章 交通物流行业：主数据治理助力中国外运数字化转型 **481**

**3.** **主数据实施**

主数据实施是构建主数据代码库、实施主数据管理的重要环节，每类主数据实施都必须按 照实施流程进行，包括构建数据管理模型、清洗存量数据、数据加载入库、配置数据接口和服 务方式等。其中具体工作介绍如下。

(1)制定和完善各类主数据代码标准，包括主数据业务标准、主数据管理标准和主数据技

术标准。

(2)选择外部通用标准(ISO 、Uncode 、IATA 、GB/T 等),寻找各类通用基础数据，并配 置数据采集方式。

(3)开发及实施各类主数据的数据模型，配置各类主数据管理流程，以及开发各类主数据 共享服务接口。

(4)实施数据清洗和主数据代码入库，清洗历史数据60多万条，查找和整理标准数据100 多万条，加载入库主数据代码150多万条。

**4.** **主数据的应用推广**

主数据是数据标准的载体，是统一数据标准、确保核心业务数据一致性的重要保障。主数 据只有被使用，标准才能被执行，数据才能被管控。因此，做好主数据的应用推广，对于体现 主数据管理的价值至关重要。主数据的应用推广工作分为总部试点和全面推广两个阶段进行。

(1)总部试点。

总部试点以总部统一建设的集中管理系统和统一业务平台为主，率先按照主数据管理要求 对接及引用主数据。在总部试点应用中要检查主数据质量和应用效果，验证主数据各项服务接 口的功能和性能，检验主数据管理与业务系统的协调性。通过总部试点应用可以不断优化主数 据服务功能，提高主数据的完整性和数据质量，优化主数据运维管理和服务体系，为主数据的 全面应用推广做好充分准备。

(2)全面推广。

全面推广针对整个企业各层级展开。 一般集团化企业层级多，下属单位多，业务系统也多， 这给主数据的全面应用推广带来很大的挑战。为了确保主数据的应用推广成功，全面推广分为 以下4个步骤有序进行。

第一步是前期准备，包括成立应用推广工作组，负责组织培训和宣讲，对重点部门进行需 求调研，制订详细的应用推广实施计划。

**482** 数据治理 工 业 企 业数字化转型之道

第二步是确定应用推广方案，包括需求搜集及整理，确定每个系统的主数据范围和对接方 式，编制应用推广的总体方案。

第三步是应用推广实施，包括业务系统数据对接开发，数据清洗和初始化、联调和切换等。

第四步是应用推广验收，包括验收系统对接功能、数据落地和应用效果。

**5.** **构建主数据管理运营体系**

构建主数据管理运营体系包括建立组织、配备人员，制定管理制度、规范管理流程、明确 管理职责、完善管理机制，形成主数据负责部门牵头、主数据运维团队密切配合的主数据管理 和服务保障体系。具体包括以下内容。

(1)制定并发布主数据管理办法和主数据运维管理细则，明确主数据管理要求，规范各类 主数据的申请、审核、入库的流程。

(2)组建主数据管理和服务组织，并配置相关人员，企业总部成立主数据管理团队，根据 需要配备主数据维护人员、技术服务人员和各部门主数据审核人员，各下属单位指定主数据管 理和服务专员。

(3)建立主数据维护、考核机制，从时效性、准确性和完整性等方面进行考核，确保主数 据维护和管理的质量和效率。

**47.3** **主要成果**

中国外运的主数据管理项目的实施，主要取得了以下成果。

**1.** **解决了主数据标准、主数据统一的问题**

(1)统一主数据标准。

在实施主数据管理项目之前，中国外运的主数据没有统一的标准，数据代码和数据字段定 义只要满足系统的需求即可，各系统之间的差异很大且相互独立。自主数据管理项目实施之后， 中国外运统一了主数据标准，并按照13类主数据构建了主数据标准体系。其中，主数据标准包 括数据代码和属性标准、主数据管理标准和主数据应用标准。数据代码和属性标准对数据代码、 唯一标示、数据项、数据关联、数据来源等方面进行规范；主数据管理标准对管理职责、管理 流程和管理质量等方面进行规范；主数据应用标准对应用范围、应用方式、应用质量要求等方 面进行规范。

第47章 交通物流行业：主数据治理助力中国外运数字化转型 **483**

(2)统一主数据。

在实施主数据管理项目之前，各类基础数据和主数据都是由每个系统自行管理的，没有严 格的管理流程和统一的管理要求，主数据质量参差不齐、可用性差，难以形成统一的主数据代 码库以支持主数据的一致性。主数据管理项目实施后，中国外运搭建了统一的主数据管理平台， 将所有(共13类)主数据由分散管理变为集中管理，统一标准、统一流程、统一质量要求，并 通过集中的主数据代码库统一提供主数据共享服务，彻底消除各系统之间主数据不统一的问题。

**2.** **实现了主数据管理一体化协同**

针对主数据的不同管理模式，中国外运将主数据管理平台与业务系统及数据来源系统完美 集成，对主数据的申请、创建、审核、入库、起用、修改、停用、归档等实现全生命周期一体 化协同，具体包括以下几个方面。

(1)将主数据申请前移至业务系统，实现业务需求与后端管理的协同。

中国外运在主数据平台中对主数据进行集中管理，但需求往往来自业务，因为业务是主数 据的消费端，能及时感知主数据是否满足业务。中国外运将主数据申请前移至业务系统中，并 通过主数据申请服务接口连通主数据平台，让业务人员可以在业务系统操作界面中按需提交主 数据创建和修改申请。

主数据平台接受业务系统的申请并生成任务队列，优先由系统进行自动处理(包括初审、 配置参考数据、创建主数据),系统处理不了的任务(参考数据无法匹配)被转到人工处理队列， 由主数据管理人员进行处理。系统自动处理时效小于3秒，人工处理时效要求不超过两小时， 这样很好地满足了业务需求： 一方面很好地解决了业务人员在面对业务操作和主数据申请时来 回切换系统带来的不便，大幅度提升了用户体验；另一方面将主数据申请和应用在业务端形成 闭环，对于大量不需要人工审核的主数据，实现对业务系统的无感管理。

(2)打通参考数据源，实现主数据自动匹配和采集。

对于依赖参考数据的主数据(比如客户主数据，需要引用企业工商、征信等数据),如何引 用参考数据会直接影响主数据管理的质量和效率。通过人工查找及录入数据不但效率低下，还 容易因误录、错录带来数据质量隐患。通过接入第三方提供的可靠的数据源(如企业征信数据), 系统可自动在线匹配和引用数据源中的数据，实现数据处理自动化，不但简化了人工操作，提 高了效率，还避免了数据质量隐患，并且在实践中的应用效果非常好。

(3)专职专责，实现数据管理质量和业务管理的协同。

**484** 数据治理——工业企业数字化转型之道

主数据管理并不是简单的数据管理，往往还受业务管理的约束。如何协调好两者之间的关 系，是在进行主数据管理时必须要处理好的问题，否则就会产生管理冲突。中国外运通过对主 数据管理设定了两个重要角色：主数据维护人员和主数据审核人员，很好地解决了这个问题。

主数据维护人员是专职数据操作人员，并对数据质量负全责。所有主数据都必须通过主数 据维护人员进行采集、核对、录入，其他人员无权操作数据。

主数据审核人员是代表数据管理责任单位行使数据审核批准权的人员，对入库数据是否适 用于业务负全责。所有入库前需要审核的主数据，必须由主数据审核人员确认通过后才可入库。 这样既确保了主数据管理的质量，又兼顾了业务对主数据的反向约束，责任清晰、分工明确。

(4)主数据自动更新和人工更新协同，确保主数据常用常新。

对主数据进行更新及修改是维持数据有效的重要手段。主数据更新及修改包括被动响应修 改和日常主动更新。

●被动响应修改是指接收到修改需求并响应需求而进行的修改。

●日常主动更新是指为了维持主数据的完整性和有效性而进行的主动修改。

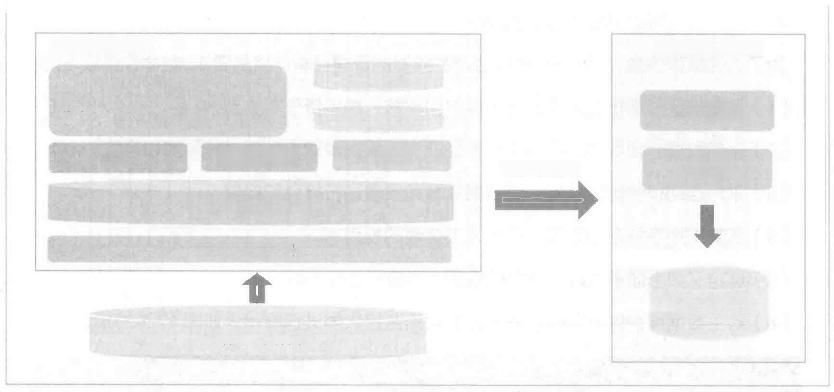
为了提高主数据更新效率，中国外运充分利用技术手段，实现主数据自动更新，如每天客 户主数据的每条记录在第一次被调用时，系统会自动触发更新，以确保主数据的时效性为T+1。 主数据更新及修改由主数据管理员或系统自动完成，不允许业务人员操作。

(5)主数据自主采集，实现源头管理和数据汇集协同。

对于源头在业务系统中的主数据(如员工、内部组织主数据等),中国外运实际上是将此类 主数据管理前移至业务系统中，通过系统改造使这些业务系统既能满足业务需求，又能满足主 数据管理的要求，从源头上做好主数据管理。

考虑到业务系统中的数据结构与主数据的存储结构存在很大的差异，需要经过复杂的整理、 加工和裁剪处理才能满足主数据的存储要求，中国外运采用主数据预处理装置来完成数据采集、 加工的工作，很好地解决了此问题。这种方式充分解耦，摆脱了数据接口方式在接收数据时受 制于数据源系统的业务逻辑，也消除了数据交互异常带来的影响，极大地提高了数据管理的独 立性、灵活性和需求扩展性。同时，由于实现了数据源之间的关联，数据异常排查更便捷，解 决问题的效率也更高。其中，主数据汇集功能架构如图47-3-1所示。

第47章 交通物流行业：主数据治理助力中国外运数字化转型 **485**



**大数据平台TBDS**

整合

HIVE

清洗

MapReduce

WebService

分布式文件系统HDFS

YARN资源调度

主数据代码库

**业务系统只读库**

组织主题存储

员工主题存储

Spark

数据接收

数据校验

主数据平台MDM

TEZ

**图47-3-1** **主数据汇集功能架构图**

**3.** **完成了主数据清洗，彻底清除数据垃圾**

在实施主数据管理项目以前，中国外运的各类主数据都在各业务系统中自行管理，产生了 大量的数据冗余和数据垃圾，严重影响了主数据的质量和使用效率。主数据管理项目通过大量 的数据清洗工作，将各业务系统中积累的数据垃圾彻底清除，对数据之间的关联数据和缺失数 据进行补全，并将清洗后的数据进行加工、整理和完善，汇集到主数据代码库中。具体包括以 下工作。

(1)彻底清理业务关联方数据，解决了以前不知道客户有多少、不知道客户是谁、不知道 客户是否存在等问题，保证了客户数据唯一、真实、完整、有效。

(2)彻底清理内部的组织数据，规范各类组织的代码标准，厘清不同组织树之间的关联， 解决了以前组织口径无法统一、不同组织树之间没有关联的问题。

(3)彻底清理与员工相关的基础数据，将原来分散管理的 HR 员工数据、邮件系统员工数 据和统一用户员工数据进行集中清理，用员工ID 建立强关联，实现了与员工相关的基础数据的 统一。

(4)对各业务系统中的业务基础数据进行清理，彻底清除了非标准数据，实现了业务基础 数据的统一。

**486** 数据治理——工业企业数字化转型之道

**4.** **建立主数据质量管理和监控机制**

为了确保数据质量，主数据管理项目对主数据的管理有着严格的要求，具体包括以下内容。

(1)每类主数据都指定由专人负责维护和审核，确保责任落实到人。

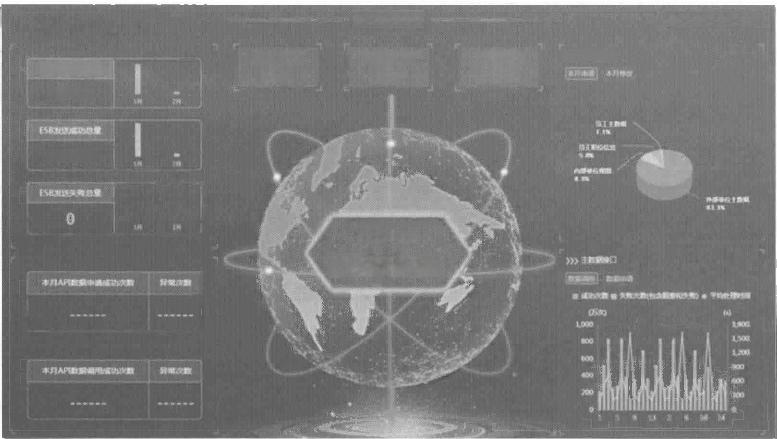
(2)采用数据必须可信，不允许有未经证实的数据源对主数据进行添加和修改。

(3)采用自动校验管理机制，对数据标准和质量进行检查和把控。

(4)尽量采用系统自动处理，减少人工处理的差错率。

(5)制定了数据质量 KPI, 对数据质量达标情况进行考核。

(6)对主数据维护进行监控，并通过主数据监控大屏进行展示，如图47-3-2所示。

主数据监控

2 0 2 0 年 0 2 月 1 7 日

中 初 \*

》

快

3.563

本月RR 834

617

1.488

**12.953**

**主数规总量**

**1,584,982**

**图47-3-2** **主数据监控大屏**

**5.** **实现主数据应用全覆盖**

主数据管理是手段，主数据应用落地才是目的。中国外运的主数据管理项目经过总部试点 和全面推广两个阶段后，完成了各层级主数据的应用推广，接入60多个系统，实现了主数据应

用全覆盖，具体如图47-3-3所示。

第47章 交通物流行业：主数据治理助力中国外运数字化转型 **48 7**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **总部统建系统** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 全国海运主系统 | 订单综合管理系统 | 运输管理系统 | 公共船期服务 | 智慧工程物流 | 外运物流结算平台 | 货代管理系统 | 外部用户统 认证平台 | 化工运输管理系统 | 外运新快线 | 会展物流管理系统 | 跨境电商物流业务系统 | 船代：船代可视化系统 | 云资源系统(云仓 | 人力资源信息管理系统 | 统一结算系统 | 客户信息管理平台 | 股份公司项目管理系统 | 客户关系管理系统 | 股份公司供应商管理系统 | 招商法务系统 | 培训系统 | 11运维服务平台 | 电子邮件系统 | 统一用户认证管理平台 | 单证智能识别平台 | 中外运物联网平台 | FROS | 统一登录系统  LaAP | uws前置系细 | 企 业微信 | 外运大厦门禁系统 | oA办公室系统 |
|  | 2 | 3 | 4 | 5 | 6 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 区域/子公司自有系统 | | | | | | | | | | | | | | | | | | | | | | | |
| 跨境电商物流 | 华北区域 | 华东区域 | | | | | | | 华中区域 | | | | | 华南区域 | | | | | | | | | |
| 空运系统 | 华北区域主业务服务系统 | 浙江货运系统、浙江 | 信控及应收帐款系统 | FMs货代管理系统 | 外运长江scp网上服务系统 | Eoc财务结算系统 | 华东公共微服务 | 长江驳运信息系 | 场站系统 | 汽运系统 | 大商系统 | 威海空运电商系统 | 多式联运平台 | 运易通 | 华南多式联运系统 | 华南基础数据服务模块 | 华南区域结算中心系统 | 海运货代系统 | 海运船代系统 | 内贸系统 | 结算系统 | 新驳运系统 | 新码头系统 |
| 1 |  | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 5 | 16 | 7 | 18 | 19 | 0 | 21 | 22 | 23 | 24 | 25 |

**图47-3-3** **主数据接入的业务系统**

**6.** **主数据驱动数字化应用**

主数据不但可以为各业务系统提供基础数据，还可以作为企业的数据资产。利用主数据可 以直接驱动企业业务的数字化和自动化。

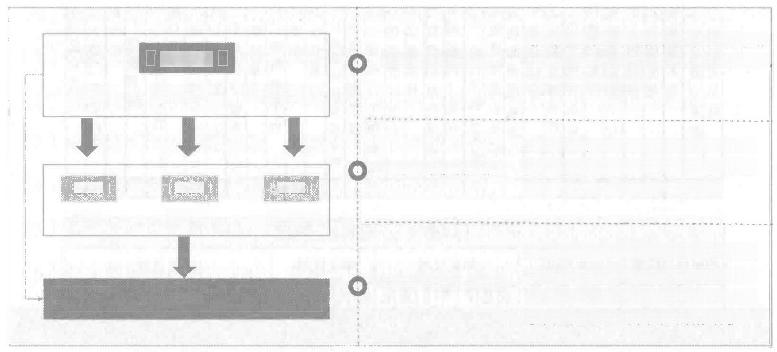
(1)助力提升业务数据质量。

通过对主数据的有效治理和广泛应用，中国外运的业务数据的标准化程度有了显著提高， 为实现数据的互联、互通提供了有力的支持，也为改善数据质量、提高数据可用性提供了有益 的帮助。正是因为有了主数据管理和应用的基础，业务结算系统(BMS) 的基础数据校验和拦 截行动功能才得以实现，数据的合规性和一致性才得到保障。

(2)驱动客户、供应商、结算对象管理数字化。

中国外运通过主数据管理项目，实现了客户、供应商、结算对象管理的数字化协同，确保 了这些业务关联方身份的真实及有效、数据的准确及完整，有效提升了中国外运的运营管理和

风险管控能力，如图47-3-4所示。



利用外部工商数据对历史数据进行了深度的清 洗，将干净数据补充完整，将脏数据剔除干净 彻底解决历史遗留问题。

■供应商

所有客户、供应商和结算对象的基础数据必 须在主数据平台上统一管理，其他系统只能 引用，确保了数据的唯一性和一致性。

结算对象

按照数据维护和审核流程，对数据进行全生 命周期管理，为公司各方面的应用提供准确 可靠的客户供应商数据，彻底杜绝因客户信 息不准确导致的一个客户多个名称等问题。

■建立主数据运营机制

■统一主数据管理规范

SRM

供应商

■统一客商基础数据

CRM

客户

■结算对象

业务系统

■客户

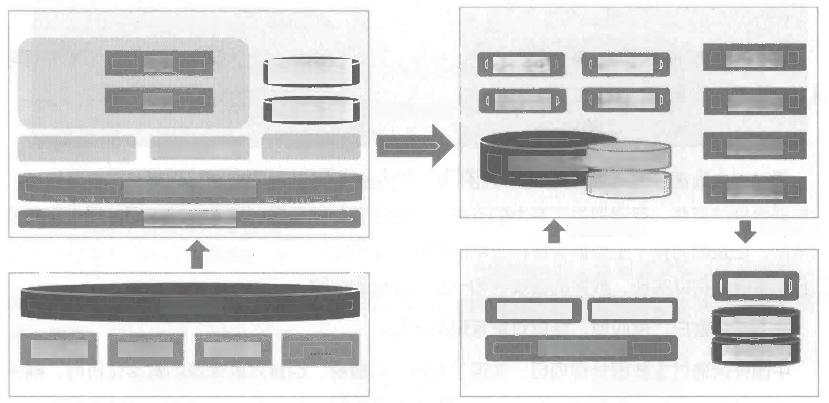
MDM

BMS

图47 - 3 -4 客 户 、 供 应 商 、 结 算 对 象 数 字 化 管 理 示 意 图

(3)实现统一用户、邮箱等事务的自动化管理。

中国外运基于员工主数据，构建了完整的数字化管控场景，实现了邮箱自动化管理，不但 替代了20多个AD 域和邮箱管理员的人工管理，而且大大提高了管理的效率和质量，完美实现 AD 域账号、员工邮箱与HR 员工管理实时联动，再也不用担心出现邮箱账号被离职员工占用的 情况，如图47-3-5所示。



大数据平台TBDS

HIVE

整合

清洗

员工主要存储

MapReduce

组织主数据

分布式文件系统HDFS

YARN 资源调度

WebService

Sqoop

自动化管理

HR 只读库

组织主数据

员工管理 职位管理

人力资源管理系统PS

AD 用户管理 员工邮箱管理

AD 组织树管理

接口调用

分发服务

页面查询

即时服务

主数据平台 MDM

数据裁剪

数据入库

数据接收

数据校验

AD及邮件系统

组织王颗存例

主数据代码库

员工主数据

员工主数据

数据接收

WebService

组织管理

HttpPost

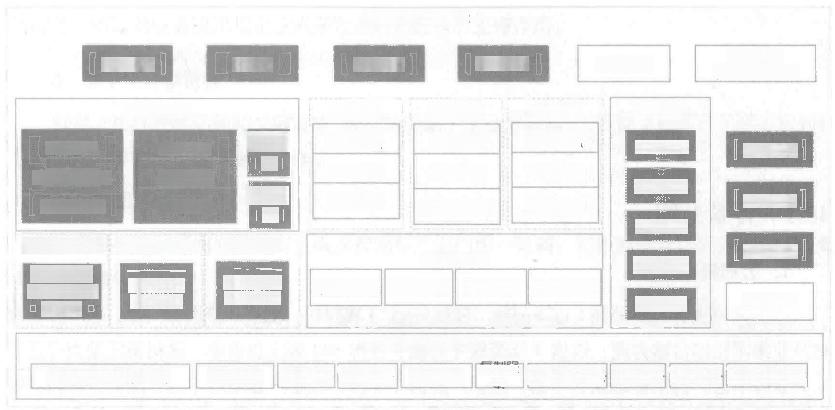
Spark

TEZ

图47 - 3 -5 员 工 邮 箱 自 动 化 管 理 场 景 图

**47.4** **未来展望**

主数据管理是企业数据治理的重要内容和核心基础，经过3年的建设及实施，中国外运的 主数据管理项目实施已基本完成，主数据管理和应用已经达到较高的水平。但是，主数据管理 仅仅是中国外运实施全面数据治理的第一步，接下来中国外运将按照企业数字化、平台化战 略落地的要求，构建统一的数据治理平台(总体架构见图47-4-1),全面推进数据治理工作。



数据治理平台功能

审批待办 移动应用 个性设置

数据质量

质量标准

审核|统计

数据流程

质量问题报告

部署调度

模型评审

元数据管理

源系统配置 元数据采集 血缘关系

自动检测

数据交换与服务工具

服务组件 交换服务 整合服务 传输服务

标准管理

动态建模组件 代码体系表

数据资产管理

资产横型/ 数据目录/数 据地图

数据安全

分级隐私

审查|评价

加密脱敏

数据清洗组件

词库管理/ 横糊匹配

数据指标

指标定义

主题分类

指标数据

数据共享

数据发布/ 管理/服务

数据申请

数据审核

数据变更

数据冻

结

数据归 档

模型标准

模型设计

务 比对验证服务集成服务

测试验证

流程管理

数据开发

服务建模

部署管理

(管理员功能)

共享服务数据加工组件

数据服务总线(ESB)

元数据查询

主数据管理

身份验证

多语言

复制服

图47-4-1 数据治理平台总体架构

为了系统化推进数据治理工作，中国外运已完成了数字化规划一数据治理专项规划工作： 从数据规划、数据标准管理、数据质量管理、数据隐私和安全管理4个领域，确定中国外运全

面数据治理的总体目标和任务；从数据治理价值链和数据治理能力两个维度，规划数据治理实 施 路 线 图 和 推 进 计 划 ， 为 企 业 实 施 全 面 数 据 治 理 绘 制 好 蓝 图 。 中 国 外 运 还 规 划 了 数 据 治 理 支 撑 体系建设方案，包括组织建设、制度流程建设、人员能力建设和技术支撑4个方面，为全面数 据治理提供保障。

未来，中国外运计划用3年左右完成数据治理体系的建设，全面提升企业的数据治理的能 力和水平，为实现数据资产管理和运营、打造“数字外运”、助力平台化战略落地提供强有力 的支撑。

**第48章**

**建材行业：中国建材集团工业大数据应用实践**

**48.1** **背景介绍**

**1.** **公司简介**

中国建材集团有限公司(以下简称“建材集团”)集科研、制造、流通为一体，是综合性建 材产业集团和综合服务商，连续9年荣登《财富》世界500强企业榜单。建材集团重点打造了 6 大业务平台，包括基础建材平台、国际产能合作平台、三新产业发展平台、国家级材料科研 平台、国家级矿山资源平台和金融投资运营平台。

**2.** **信息化现状**

经过多年的开发应用和升级改造，建材集团在信息化建设方面已经具备了一定的基础。集 团总部目前已规划并建设了中心机房、视频会议、集团财务、报表、安全生产、人事、办公等 管理系统。各水泥企业已经基本形成生产厂级的营销、采购、发运、仓储、生产、安全、资金、 项目、成本核算、财务管理等经营管控系统，并通过集成电子招标采购平台、生产调度系统、 磅秤无人值守系统、安全生产系统、电子商城等，为企业决策分析提供了初步的数据基础。

目前，从管理层面和数据层面来说，建材集团急需解决以下两个问题。

(1)集团级的信息管理平台尚未建立，现有系统大多局限于板块级、企业级、部门级应用， 缺乏集团总部与二级公司、三级公司业务系统之间的数据和信息集成；尚未建立流程化运作机 制，集团各业务部门各自制定相关制度、规则，缺乏整体联通，且以管控为主要目标，没有站 在集团整体层面进行考虑。

第48章 建材行业：中国建材集团工业大数据应用实践 **491**

(2)在生产经营过程中，各系统产生了大量的中间业务过程数据和结果数据，并且大部分 分散于各系统中。由于各系统缺乏统一的横向数据交换机制，无法将各业务数据集中进行综合 利用，各系统形成了一个个信息孤岛(甚至同一个指标在不同系统中出现口径不一致的现象), 从而给信息资源的开发及利用带来了较大的困难，影响信息资源价值的充分发挥。同时，建材 集团没有明确集团统一的主数据，员工没有意识到数据即资产，数据在管理、运营中没有起到 应有的作用(目前只是对数据进行简单的统计及分析),且数据还存在不及时、不准确、不全面 等问题，难以对业务决策和企业未来发展起到应有的支撑作用。

**3.** **数字化转型目标**

针对上述存在的问题和关键需求，建材集团基于工业互联网+大数据平台进行了建设规划和 初期应用，其中包括以下重点工作。

(1)集团层面。

①提供高效的数据服务，对全级次数据资产进行统一管理，对数据进行对标分析，实现数 据资产价值化。

②推动基于大数据的创新发展，其中主要是基于大数据的技术创新、模式创新、服务创新。

(2)企业层面。

①基于大数据技术，提升企业自动化水平，稳定生产过程，提升产品质量及降低综合成本； 提升人均效率，降低人员劳动强度及减少人员需求。

②通过基于大数据的预测性维护技术，提升设备的运行管理水平，实现设备的健康、高效 运行，降低维修成本，对设备进行预测性维护，减少故障的发生。

③提升能效、节约能源，建立关键能耗设备的分析模型，减少人为浪费，实现综合节能。

④提高生产过程的精细化管理水平，利用大数据建立标准化、规范化的管理模型，建成科 学化、数字化的管理与考核体系，提高企业的管理效益。

⑤建成实时的安全环保平台，实现符合国家安全环保规范的绿色、安全的生产监控体系。

**48.2** **工作实施**

**1.** **战略规划**

实现集团数字化转型，不仅仅是简单的信息系统或者设备的堆叠建设，而是要基于对业务

**492** 数据治理——工业企业数字化转型之道

流程端到端的梳理与优化，实现全业务流程的数字化运作、数字化管理和数字化决策。其中涉 及流程、组织、IT 架构的改变，以及人的行为和意识的改变，需要全面分析、统筹谋划、分步

开展，更需要一套行之有效的管理机制为这场长期的变革提供持续保障。

**2.** **规划方案**

在理解集团管理现状的基础上，建材集团根据集团战略和调研现状，对核心能力、业务流 程进行评估，全面规划集团的数字化转型战略工作。建材集团的数字化转型规划方案分为以下

4个部分。

(1)以财务为中心的集团智能监管体系规划。

●通过对集团管控模式及信息系统的调研和分析诊断，提出完善和优化建议；

●提出信息化支撑集团监管业务需要具备的能力和变革方向；

●基于财务管控模式，提出集团总部财务平台的总体建设方案。

(2)主数据管控体系规划。

●对集团主数据管理现况及信息系统进行调研和分析诊断；

●提出建立集团主数据管理的方法及流程；

●提出集团数据治理组织建议方案。 (3)IT 及云平台建设规划。

●对集团IT 基础架构进行调研和分析诊断；

●提出集团总部级的云/数据中心建设方案；

●提出集团总部级的IT 开发平台技术方案；

●提出集团IT 治理建议方案。

(4)水泥板块的智能制造体系规划。

●通过对水泥板块集团/区域公司/工厂管理现状及信息系统的调研和分析诊断，提出业务流 程、应用架构、数据架构未来蓝图，并提出未来实施规划线路。

●以标准化、数字化为核心，进行智能制造整合规划，建立集团总部级的标准化架构规划 体系。

●提出集团总部级的智能制造平台的建设方案。



工业 PaaS平台

**3.** **系统架构**

建材集团的工业互联网平台(见图48-2-1)能够根据实际需要，支持集团化或单一水泥企

业的部署和应用。

|  |  |  |
| --- | --- | --- |
| 工业APP | 传统软件云化  (决策分析、生产管控、质量管理、设备管理、供应链管理) | 新型工业APP  (设备监控、设备预测性维护、能耗分析、专家优化、视频巡检) |

(基础服务组件、数据驱动模型、机理模型、业务模型、微服务管理)

工业微服务组件

工业应用开发工具

(微服务框架、开发工具)

通用PaaS平台(开发环境、运行环境、运营环境)

预测分析与优化平台

工业大数据平台

工业物联网平台

数据处理

数据管理

协议转换

规则库

算法平台

设备数据采集

样本库/模型库

数据分析

时序数据库

控制集成

数据可视化

数据集成

laaS层

边缘层

云基础设施(服务器、存储、网络、虚拟化)

协议解析

设备接入

边缘智能



传感器

SCADA PLC 监测/计量 UMS 维修手持终端 视频 环境智能仪表

企业在用系统

次 +

一卡通

■ 显 舍

ERP QMS WMS

前

设备管理

能源管理 安全管理

生产管控

设 备

**图48-2-1** **工业互联网平台架构图**

(1)工业 APP(SaaS 层)即软件应用层，提供包括生产、设备、质量、能源等各种专业应

用系统。

(2)工业 PaaS 平台(技术平台),包含了工业物联网平台、工业大数据平台、预测分析与 优化平台等。其中，工业物联网平台负责与设备进行协同交互；工业大数据平台提供了数据存 储和数据计算服务；预测分析与优化平台提供了专家知识库和控制集成策略。

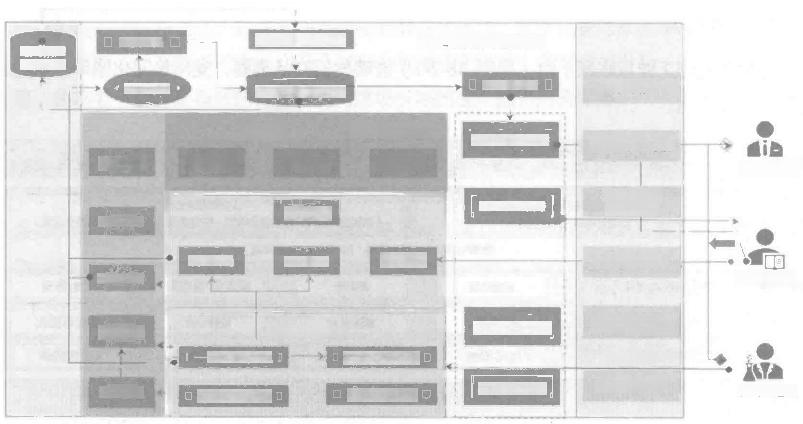
(3)laaS 层(基础设施层),支持混合云或本地等部署模式。

(4)边缘层(设备层),通过工业智能网关实现对运行数据的实时采集。边缘网关通过内置 协议库的方式实现对各种设备通信协议的适配。

**48.3** **应用框架与技术路线**

图48-3-1为面向水泥企业的应用框架和主要技术路线，主要包括以下3个层面。

**494** 数据治理——工业企业数字化转型之道



数据平集 主效最采统一管理 工业互联网及大数据平台

建材集团

数据色库

区域水泥数据色库

集应管控 P 受考管理平台

审批与统程管控

风岭管控 投资管理 HR 繁理

区域管控

统一普销 集中采同 运#管经 在线指挥

区域管理者

跨系统执行银踪

下属企业业务系统

卡通智能物流 生产/能耗管理

行业专用模块

设备/质量管理 产线运他管理

在线查询与 对标管理

经基分析与管控

设备预测性维护

工艺过程优化

生产运慧监控

企业管理者

大数地分析

统一门户

财劳核填

接入适配

财偶合并

成本管理

仓储管理

焦团管理者

财集管授

业务反错

企业

区城

集团

图48-3-1 面向水泥企业的应用架构和技术路线图

(1)集团管控层面：建立集团层面 ERP 经营管理平台，实现集团的 HR 管理、风险管控、 投资管理等；建立集团主数据管理平台，实现集团数据统一管理、统一分发，为集团大数据分 析做好支撑。

(2)区域管控层面：按照运营区域对生产运营进行远程管控、在线查询与对标管理和在线 指挥；实现区域统一营销、集中采购，以及运营管控。

(3) 企业业务经营层面：主要是对传统应用软件与区域公司进行数据集成，如设备/质量管 理、生产/能耗管理等；以及对基于人工智能和大数据技术的设备预测性维护、工艺过程优化等 专业模块的推广和使用。

**48.4** **工作成果**

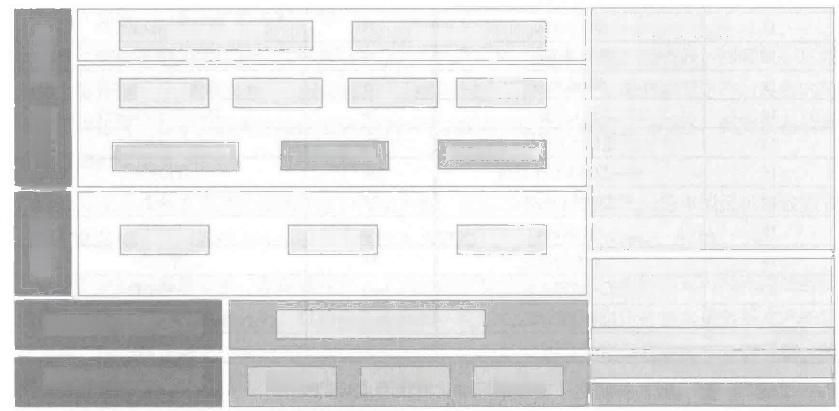
中国建材集团所取得的工作成果包括以下内容。

**1.** **数据治理及主数据管理**

中国建材集团构建了集团及水泥板块对主数据的统一管控框架，以满足各类主数据使用的 规范化需求。其按照“集团总部一板块一企业”三个层级构建主数据管理平台，实现了“数据 同源，规范共享，应用统一，服务集中”的目标。中国建材集团的数据治理及主数据建设的主

第48章 建材行业：中国建材集团工业大数据应用实践 **495**

**要路径如图48-4-1 所示。**



◆一阶段集团总部建设范围

咨询规划

>数据治理咨询规划

>平台建设规划

主数据类

>通用基础类 >财务类

>单位类

>人事类 >指标类

人事指标

◆二阶段水泥板块建设范围 >物资类

主数据管理平台

一期第一阶段建设内容

一期第二阶段建设内容

**数据运维体系**

二期建设内容

数据治理咨询规划 平台建设规划

通用基础 财务类 单位类 人事类

水泥板块专用类物资 新材料板块专用 工程服务板块专用

数据规划 故据治理

**主数据管理平台**

业的指航

管理流程

管理制度

财务指标

管理组织

投资指标

**图48-4-1** **中国建设集团的数据治理及主数据建设的主要路径**

其中主要工作包括以下几个方面。

①基于建材行业协会推出的《水泥工业物资分类与代码》标准文件(由水泥行业协会和建

材集团内部企业编制完成),完成对物资大类的分类与代码定义，如表48-4-1所示。

本标准规定了水泥工业物资的分类与代码。适用于中国水泥行业的物资采购、储存、运输、销售、

等流通全过程，也适用于生产和工程项目的设计、计划、统计、财务核算，以及相关的信息管理。

**表48-4-1** **水泥行业物资大类代码表**

|  |  |  |  |
| --- | --- | --- | --- |
| **大类代码** | **大类名称** | **大类代码** | **大类名称** |
| 01 | 普通钢材 | 23 | 金属加工、维修设备 |
| 02 | 煤炭 | 24 | 交通运输设备 |
| 03 | 水、点 | 25 | 加油加气设备 |
| 04 | 水泥原材料 | 26 | 电气电工设备 |
| 05 | 金属及加工材料 | 27 | 电工材料 |
| 06 | 金属丝、绳 | 28 | 仪器仪表衡器 |
| 07 | 建筑五金 | 29 | 通信设备 |
| 08 | 耐磨/耐热件 | 30 | 工具量具、刃具、磨具 |
| 09 | 建筑材料 | 31 | 消防器材 |
| 10 | 石油及制品 | 32 | 计算机及配件 |

**496** 数据治理——工业企业数字化转型之道

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| **大类代码** | **大类名称** | **大类代码** | **大类名称** |
| 11 | 橡胶及制品 | 33 | 水泥专用设备配件 |
| 12 | 涂料及颜料 | 34 | 通用机械设备配件 |
| 13 | 水泥及制品 | 35 | 交通机械设备配件 |
| 14 | 火工产品 | 36 | 环境治理设备配件 |
| 15 | 通用化工产品 | 37 | 电气设备配件 |
| 16 | 玻璃仪器及化学试剂 | 38 | 焊接材料 |
| 17 | 劳动保护用品 | 39 | 轴承 |
| 18 | 包装物 | 40 | 紧固件 |
| 19 | 水泥专用设备 | 41 | 密封件 |
| 20 | 工程机械设备 | 42 | 日用电器 |
| 21 | 起重输送设备 | 43 | 办公用品及杂品 |
| 22 | 通用机械设备 |  |  |

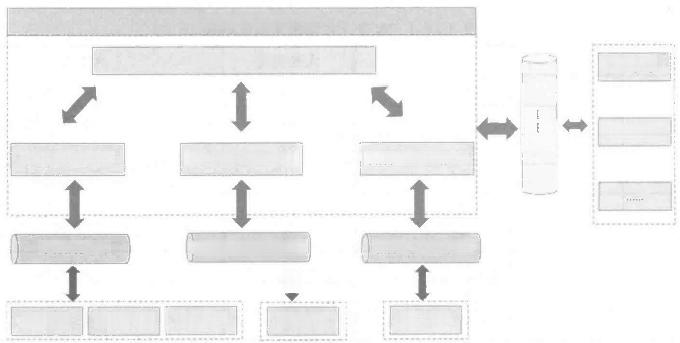
②定义数据模板，并制定《主数据标准化管理标准规范》。

③进行数据清洗，根据《主数据标准化管理标准规范》对历史数据进行清洗、排重、合并、 编码，保证数据的完整性、准确性和唯一性。

④进行数据分发，在数据清洗、编码完成后，需要对已上线、在建等业务系统，根据系统 5个阶段及重要性的不同，采用不同的策略进行标准数据的导入。

⑤实现集团总部及各个业务板块的数据增值与变现。图48-4-2为中国建材集团主数据管理

平台部署模式。



中国建材集团主数据管理平台

集团总部应用实例

团

务 总 线

工程服务板块应用实例

工程服务板块服务总线

个

**ERP系统** 一卡通系统 财务系统 应用系统

新材料板块应用实例

新材料板块服务总线

水泥板块应用实例

水泥服务总线

应用系统

财务系统

HR 系统

集 服

**图48-4-2** **中国建材集团主数据管理平台部署模式**

第48章 建材行业：中国建材集团工业大数据应用实践 **497**

**2.设备预测性维护**

(1)设备预测性维护的主要功能。

在水泥生产中，大部分设备都是以串联的方式运行的(如生料磨、预热器、回转窑、水泥 磨等主要设备),很容易因为一个设备发生故障而造成全工序的停产，因而水泥生产对设备的可 用性要求极高。基于大数据的设备预测性维护可以提高水泥生产设备的可用性，保障设备的持 续、健康和安全生产。

以某水泥基地高温风机预测性维护应用为例，此基地有上百台风机，每年因风机导致的非 计划性年故障停机时间近200 小时，需要投入大量的人力、物力进行点检、巡检。

当涉及监测参数较多、数据量较大、工艺耦合较复杂时，采用一般的方式可能无法涵盖所 有异常情况。通过大数据分析、机器学习的方式，可以让算法模型自行挖掘大量数据之间的耦 合关系，学习在正常工作情况下的数据表现模态，进而更有效地发现异常现象，及时预警，如 图48-4-3 所示。

①系统主要功能

●在线监控：包括设备动态3D 模型、设备运行数据在线监控。

●全方位预警：包括阈值预警、规则预警、综合预警、机器学习预警。

●设备评估：包括设备运行效率评估、设备健康度评估。

●管理：包括设备管理、用户管理、预警记录管理、预警模型管理。 ②系统部署模式

●传感器—SCADA/DCS/PLC 一服务器(OPC/FTP) 一设备预测性检修平台，协议基于

OPC/FTP。

●传感器一网关一服务器一设备预测性检修平台，协议基于Moudbus/TCP/UDP。

●传感器一在线保护器一服务器(消息队列 Kafka/MQTT) 一设备预测性检修平台，协议 基于TCP。

●传感器一在线保护器一服务器(消息队列 Kafka/MQTT)—PaaS 平台一设备预测性检修 平台，协议基于RESTful API。

(2)设备预测性维护的主要内容。

①规则预警

●对每台风机的每个测点的数据分别设置阈值进行预警。

**498** 数据治理——工业企业数字化转型之道

●结合专家经验，实现更全面、更复杂的规则预警系统，多重规则相结合并预警。

●将规则及设备状态数据与工艺数据相结合，综合计算系统各项运行指标并预警。



——支持多种部器模式一 一 应用端 —

惠 湿 风 1

高温风机2\*

高通风机3#

点检/手特设备 数现上传 设备预防性检修平台

本地半台 没备预防性检修平台

设备预防性检修平台

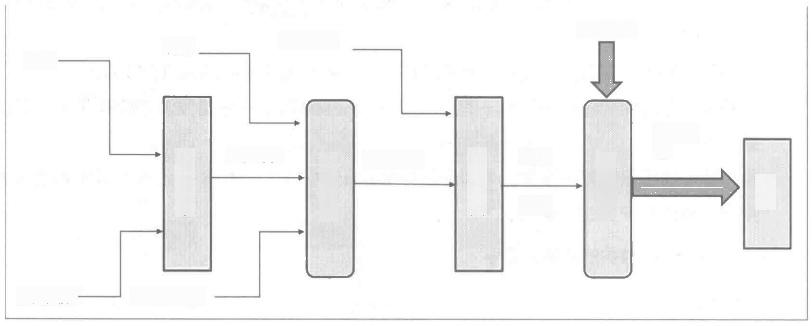
实时上传至

图48-4-3 设备预测性维护部署模式 ②机器学习故障预警(见图48-4-4)

●数据相关度分析：通过大数据分析，挖掘各种信号之间潜在的关系。

●特征建立与提取：基于数据相关性，建立特征单元，对特征单元进行提取。

●机器学习模型：采用机器学习模型(LSTM、XGBoost…), 对数据进行学习和训练、在 线预测，得到较为准确的模型预测结果。利用模型预测结果与在线监测数据的残差进行 预警。



实际数据

转速 实际温度

工况·

残差

聚类

历史数据- 冷却水温度

实际残 差

预测模型

预测温度

聚类模型

聚类算法

特征提取

工况特征

预 警

图48-4-4 机器学习故障预警

第48章 建材行业：中国建材集团工业大数据应用实践 **499**

③实施效果

●实现实时在线监测设备，替代点检员的离线点检。

●引入专家经验与机器学习相结合的算法模型，将点检员的经验转换成算法，利用机器学 习提高故障识别能力，从而提高模型的准确率，替代传统的规则模型。

●将同类设备放在一起监测、预警、分析，便于对该类设备进行总结归纳、深入研究、扩 展管理、综合把控，替代仅按照设备物理位置的分类管理。

●经过训练的机器学习故障预警模型的准确率可达到99%以上，高于传统预警模型的准 确率。

**3.专家优化**

专家优化应用系统是针对水泥生产系统中的大量数据并结合专家的知识及经验开发的应用 系统，以现有硬件及自动控制、数字化、信息化为基础，通过数据分析和算法优化，对运行参 数进行优化和调整，也可以优化和调整系统参数，采用新的控制架构，最终提升生产线的性能 优化。

根据不同的生产流程环节和功能定位，专家优化应用系统将业务划分成不同的场景，包括 窑系统智能控制、磨系统智能控制、工艺故障预警及优化分析。以原料配比优化分析为例，专 家优化应用系统将水泥工艺知识和水泥生产经验相结合，通过数据收集系统采集的实时数据， 对系统的运行状态进行量化评估，实现主动式优化，并将结果以图表的形式反馈给操作员或管 理层。

(1)专家优化应用系统的主要功能。

●依托模型算法和数据处理技术，并结合平台核心算法和专家库知识进行优化。

●基于不同物料的特性，提供非线性模拟掺配比例计算，在保证熟料生产质量和窑炉稳定 燃烧的前提下，显示成本最优、环保最优的掺配比例计算，给出成本优化配比排序表。

(2)专家优化应用系统的主要内容。

●利用物联网、大数据、人工智能技术分析海量的生产数据，在数据中发现原料配比与产 品之间的关系，优化原料配比；在保证产品产量、质量的情况下，降低成本，提高效率。

●以数据驱动替代经验驱动，通过算法优化，得出最合适的可调参数设定值，并将其推送 或直接下达至控制系统，改善操作人员随时关注控制界面并频繁调整的工作状态。

**500** 数据治理——工业企业数字化转型之道

●对生产和能耗数据的采集、完善、存储、平台构建了生产实时数据、企业管理数据集成 应用的数据仓库系统，其由实时数据库和关系数据库组成。生产过程数据通过实时数据 采集平台采集并存储至实时数据库中。

(3)实施效果。

●可对当前生产线进行数据分析，更加符合当前的生产环境和原料条件。

●可根据对当前生产环境变化、原料变化、工况变化的实时响应来优化原料配比。

●人工智能技术在不断完善，机器学习故障预警模型会越来越符合实际的生产环境，效果 会越来越好，越来越逼近理论产量。

●降低了生产成本及人员投入，使产品质量受人员技术水平的影响变小。

**第49章**

制造行业：威孚集团基于

斯欧应用互联平台建设数据通道

**49.1** **背景介绍**

**1.** **公司简介**

无锡威孚高科技集团股份有限公司(以下简称“威孚集团”)是国内著名的汽车零部件生产 厂商之一，目前拥有9家全资和控股子公司。

重庆斯欧信息技术股份有限公司(以下简称“斯欧公司”)成立于2008年3月，其主营业 务是针对各种工业企业信息互联集成场景，提供工业企业软件连接与集成(智能制造互联协同 整体解决方案的咨询、产品、实施及服务，主要客户有吉利汽车、福耀集团、中车青岛四方等 30多家国内特大型和大中型工业企业。

**2.** **数据交换需求及挑战**

(1)数据交换的必要性。

威孚集团从十多年前开始建设企业应用软件系统，并相继建设及实施了 ERP、OA、BI/BO、 BPM、MES、WMS、PLM、HR、CRM、SCM 等30多个企业应用软件系统，用于处理不同业 务领域的业务活动、规则和信息(包含人员、设备、物料、工艺、环境和财务等业务内容)。

这些在不同时期建设及实施的应用软件系统，由不同供应商开发，具有不同的标准、架构

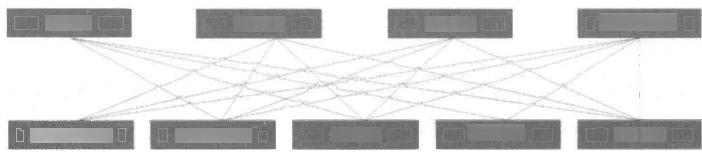
及开发语言，系统之间相对独立、封闭，业务数据不能自动交换和共享。

但是，从企业全局来看，企业不同系统中的业务处理必须相互关联及协同。每个系统在企 业的业务价值链上，既是业务成果的创造者，也是业务数据的使用者；既有业务数据的输入， 也有业务成果的输出。而通过人工进行各个系统的数据传递和业务协同，已不能适应企业信息 化发展的需要。

因此，实现企业业务数据跨系统的自动协同，实现企业跨系统的业务有序对接和数据及时、 可靠、自动地传递，确保企业业务数据的唯一性和准确性，就成了企业持续深入发展数据化、 网络化和智能化所面临的最大挑战。

(2)数据交换技术发展的背景。

为了解决企业应用软件系统之间的数据共享问题，实现企业务数据跨系统的自动交换，企 业在最开始通常会采用端到端定制开发集成接口的技术方式(见图49-1-1),解决局部的数据共 享问题。但是这种技术会造成应用软件系统之间紧密耦合，形成蜘蛛网式的集成架构，如果有 一方增加或调整服务接口，则与之相关的各方都需要随之变动，从而增加了后续扩展升级的难 度和成本，同时给数据的管理和维护带来了极大的困难。

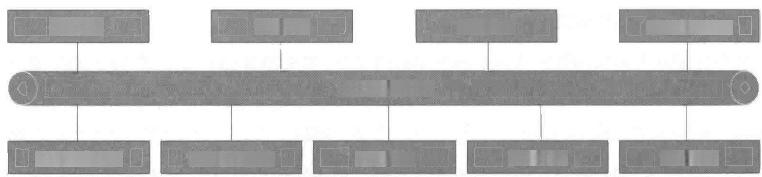


**OA系统** SAP系统 门户系统 人力资源筒理

**BPM** **流程引攀** **设备管理系统** 考勤系统 PLM 系统 财务共享

图49-1-1 端到端的数据交换接口架构

目前，业界比较通用的是基于企业服务总线的数据交换接口架构定制开发集成接口的技术 方式，如图49-1-2所示。企业服务总线的数据交换接口架构虽然比端到端的数据交换接口架构 较为优越，但是定制化开发对数据交换平台开发商依赖严重， 一旦企业新增系统、新增服务或 变更服务，都需要由开发商来完成。



OA 系统 SAP系统 门户系统 人力资源管理

企业服务总线

BPM 流程引擎 设备管理系统 考勤系统 PLM系统 财务共享

图49- 1-2 企业服务总线的数据交换接口架构

第49章 制造行业：威孚集团基于斯欧应用互联平台建设数据通道 **503**

另外，企业应用软件系统之间的信息集成工程的实施工作缺乏统一的标准规范，这就导致 了企业应用软件系统集成工作推进缓慢、接口重复开发、延续性差等诸多问题。

随着市场竞争的加剧和企业业务管理的不断变化，威孚集团各业务单位的数据化需求急剧 增长，应用软件系统逐年增多，数据交换服务持续增加，集成服务变更快速。如果仍采用端到 端的集成技术，或基于企业服务总线的数据交换接口架构定制开发服务集成的技术，则会给企 业应用系统之间的数据交换项目的后续扩展升级和管理运维带来巨大压力，阻碍企业数字化、 网络化和智能化的持续深入发展。

**3.** **项目目标和内容**

(1)项目目标。

为了彻底解决企业跨系统的数据共享问题，实现企业应用软件系统之间的数据信息可靠、 安全、及时、方便地自动交换的业务需求，威孚集团在2013 年启动了数据交换平台项目，至 2019年已经完成三期数据交换项目实施。

本项目按照“统一性、规范性、可管控和易扩展”的原则，基于面向服务架构(SOA) 的 理念，建设企业统一的数据交换平台及相关标准规范，其中具体要实现的目标介绍如下。

①基于斯欧应用互联平台产品和方案，搭建强壮、稳定且性能高效的基于服务总线架构的 标准产品化的数据交换平台。

②基干数据交换平台， 在企业应用软件系统之间自动建立通信连接，实现信息的交换、路 由、分发和转换等功能，同时具有日志、监控、统计、运维等管理功能；实现服务全生命周期 管理，具有服务发布功能(包括服务的注册、测试、授权、发布等),以及服务订阅功能(包括 服务目录、测试、订阅、异常提醒等)。

③通过制定和实施面向数据交换的工程实施标准规范，为后续新建应用软件系统提供集成 的标准规范，规范数据交换工作，提升集成效率和质量，减轻服务接口的维护难度和工作量； 以规范的服务对企业中各种业务的信息化处理能力进行抽象和积累，建设具有支撑性和可复用 性的服务体系，提升企业业务的响应速度，灵活应对市场挑战和业务变革；形成未来重要的可 复用的服务软资产，更好地支撑企业业务战略目标。

(2)项目内容。

威孚集团的数据交换平台项目主要包括以下内容。

①建立标准规范

**504** 数据治理——工业企业数字化转型之道

为了规范数据交换平台项目的建设及实施，提高项目实施的效率和质量，降低实施成本， 方便后续的运维及扩展，本项目制定并落地实施了应用互联(数据交换)工程实施标准，形成 企业实施数据交换项目的标准规范体系(见表49-1-1)。

**表49-1-1** **应用互联工程实施标准各部分简要说明**

|  |  |  |
| --- | --- | --- |
| **序号** | **标准名称** | **标准简要说明** |
| 1 | 应用互联平台 集成技术要求 | 规定了基于企业服务总线架构进行应用软件系统数据交换建设的总体框架和技术要 求，包括集成接口标准、应用集成方法、数据格式标准及集成安全支持方法等 |
| 2 | 应用互联平台 编码及报文规范 | 规定了基于企业服务总线架构进行应用软件系统数据交换建设中将会涉及的系统编 号、服务编号、响应码编号及服务流水号的编码原则，并对接入的报文格式等进行规范 性说明 |
| 3 | 应用互联平台 接入指南 | 规定了基于企业服务总线架构进行应用软件系统数据交换建设中各应用软件系统接入 数据交换平台的技术要求及方法，并作为现有及未来业务系统接入数据交换平台的参考 说明 |
| 4 | 应用互联平台 集成管理规范 | 规定了基于企业服务总线架构进行应用软件系统数据交换建设中各应用软件系统接入 数据交换平台的基本流程和要求，包括新系统接入、服务接口新增、服务接口变更等内 容 |
| 5 | 应用互联平台 集成运维规范 | 规定了基于企业服务总线架构进行应用软件系统数据交换建设中运维阶段的运维总体 要求，包括运维人员角色和职责划分、平台运维检查点、异常问题分类及解决措施、异 常问题报警机制等，以保证数据交换平台的稳定运行 |
| 6 | 应用互联平台 集成实施规范 | 规定了基于企业服务总线架构进行应用软件系统数据交换建设中应用集成的实施规 范，包括资源命名规则、存放路径原则、磁盘分区规则等 |

②实施服务集成

本项目实施主要针对 ERP、OA、BI/BO、BPM、MES、WMS、PLM、HR、CRM、SCM 等企业应用软件系统，梳理业务协同的集成接口，设计并定义交换数据的服务组件，完成了347 个服务接口在注册接入数据交换平台进行服务注册集成(见表49-1-2)。

**表49-1-2** **注册接入数据交换平台的部分服务(实际注册服务347个)**

|  |  |  |  |
| --- | --- | --- | --- |
| 序 号 | 服 务 号 | 系 统 号 | **服** **务** **说** **明** |
| 1 | 2001000000001 | 2001 | 获取OA待办及已办接口：Portal→OA |
| 2 | 2001000000002 | 2001 | OA中的流程启动接口：Portal→OA |
| 3 | 2001000000003 | 2001 | OA中的报表展示接口：Portal→OA |
| 4 | 2001000000004 | 2001 | Portal获取OA中某个文档目录下的文档清单接口：Portal→OA |
| 5 | 2001000000007 | 2001 | PLM触发OA流程接口：PLM→OA |
| 6 | 2002000000001 | 2002 | SAP采购订单接口：PM→SAP |

**续表**

|  |  |  |  |
| --- | --- | --- | --- |
| 序 号 | 服 务 号 | 系 统 号 | **服** **务** **说** **明** |
| 7 | 2002000000002 | 2002 | SAP库存明细接口：PM→SAP |
| 8 | 2002000000003 | 2002 | SAP资产原值净值接口：PM→SAP |
| 9 | 2002000000004 | 2002 | SAP出库领用申请接口：PM→SAP |
| 10 | 2002000000005 | 2002 | SAP中新增和扩展物料主数据接口：CCMS→SAP |
| 11 | 2002000000006 | 2002 | SAP中修改和删除物料主数据接口：CCMS→SAP |
| 12 | 2002000000007 | 2002 | SAP中新增和扩展客户主数据接口：CCMS→SAP |

**49.2** **项目建设技术方案**

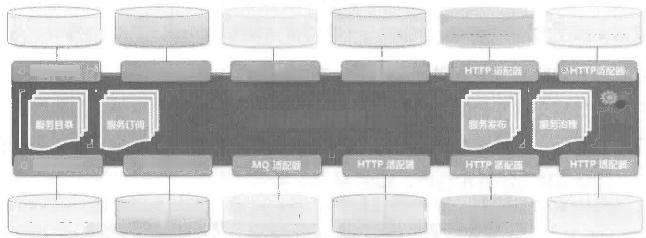
**1.** **数据交换平台技术实现**

(1)数据交换平台的总体架构。

① 总体架构

威孚集团的数据交换平台是基于斯欧应用互联平台产品搭建的。数据交换平台总体架构如

图49-2-1 所示。



OA 系统 门户系统 威孚财务共享 SAP 系统 **人力资源管理** BPM 流程引肇

MQ 通

**威** **孚** **集** **团** **数** **据** **交** **换** **平** **台**

MQ 通配道

SRM 供应商管理

系统

WFLD的PLM 系 统

**威孚税务系统**

**设备管理系统**

OA-Cology

**考勤系统**

Q 配

Me通

MQ

*5AP*

**图49-2-1** **数据交换平台总体架构**

②数据交换场景模式

斯欧应用互联平台产品定义了3种数据交换场景模式，基本能够满足工业企业的应用软件

之间的数据交换需求，具体介绍如下。

●一对一同步/调用(请求/响应)模式，如图49-2-2所示。

●一对多发布/订阅(发布/订阅)模式，如图49-2-3所示。

●同步调用与发布/订阅组合模式，如图49-2-4 所示。

**506** 数 据 治 理 — —工 业 企 业数 字 化 转 型 之 道



请求方发送清求信息，

企业服务总线

数据交换平台

提供方返回处理结果

提供方返回处理结果

请求方发送请求信息

服务提供方

服务请求方

图49-2-2 请求/响应模式



服务订阅方

发布方发布信息

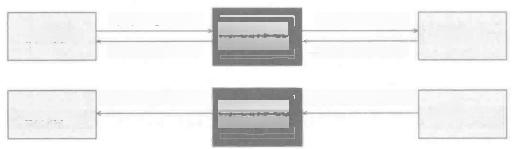
服务订阅方

服务订阅方

企业服务总线 数据交换平台

**服务发布方**

**图49-2-3** **发布/订阅模式**



请求方发送道求信息

**服务请求方**

提供方返回处理结果

提供方分包发送业务信息

**服务请求方**

请求方发送请求信息

提供方返回处理结果

提供方分包发送业务信息[

服务提供方

企业股务总线 数据交换平台

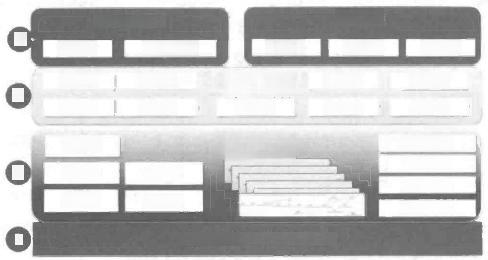
企业服务总线 数据交换平台

服务提供方

图49-2-4 同步调用与发布/订阅组合模式

(2)数据交换平台技术架构。

本项目所基于的斯欧应用互联平台产品版本，采用了IBM Integration Bus(IB) 服务总线 基础中间件作为数据交换平台的运行和部署环境。另外，服务发布平台、服务商店、服务管理 平台、监控分析平台，以及参数化适配器和核心路由消息流等斯欧自主开发的应用互联平台产 品功能组件，均在IB 上部署和运行，如图49-2-5 所示。



**服务发布平台**

服务注册 服务发布测试 服务目录

系统注册 短信告臂 服务管理平台

监控分析平台

服务监控 安全日志 访间日志

参数化适配器

核心踏由消息流

Ws-适配器

服务路由 协议转接

RFC-返配器

OPC-适配器

1 **IBM** **ⅡB消息流部器和运行环境**

服务商店

融务订阅

趋势分析

服务报表

日志记录 植式转换

服务授权

报文数据

服务测试

MQ- 活配器

安全认证

4

2

3

图49-2-5 斯欧应用互联平台技术架构

第49章 制造行业：威孚集团基于斯欧应用互联平台建设数据通道 **507**

(3)参数化适配器技术。

适配器是实现企业应用系统之间服务信息交换的连接接口。标准化的参数化适配器是斯欧 应用互联平台产品标准化的重要组成部分。

参数化适配器是指对于相同数据访问协议下的不同服务，每个服务的数据交换接口，通过 相同适配器的参数定义其实现(见表49-2-1),无须为每个服务单独开发消息流。斯欧公司成功 开发了工业企业典型数据访问协议的参数化适配器，避免了项目实施的开发工作，简化和方便 了项目实施，提高了项目实施的效率和质量。

**表49-2-1** **工业企业典型适配器简要说明**

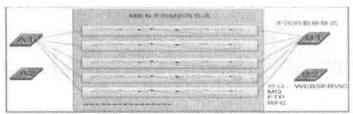
|  |  |  |
| --- | --- | --- |
| 斯欧应用 互联平台产 品提供参数 化可配置适 配器 | 适配器名称 | **适配器简要说明** |
| REST | REST API |
| HTTP | Web Service |
| MQ | IBM MQ协议 |
| SAP | SAP的RFC与IDOC协议 |
| MQTT | 物联网MQTT协议，用于融合OT域数据 |
| FTP | 文件传输协议FTP,支持协议转换 |
| DB | 默认支持MySQL、Oracle、SQL Server,支持其他数据库定制 |
| OPC UA | 工业总线协议OPC UA |
| OPC DC | 物联网协议OPC DC |

(4)消息流通用性技术。

斯欧应用互联平台产品标准化还体现在消息流的开发方面。斯欧应用互联平台产品按照标 准协议进行通用性抽象，其服务总线的消息流是针对不同协议开发不同的适配器，不是针对每 个系统的每个服务，通过系统注册、服务注册及数据参数注册，达到动态路由目的。

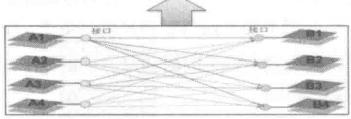
如果采用专用消息流开发技术，则需要进行大量消息流的管理维护工作，这与传统的端到 端方式的数据交换接口架构并无本质区别。因此，通用性消息流技术也是斯欧应用互联平台产 品标准化的重要组成部分，如图49-2-6和图49-2-7所示。

508 数据治理——工业企业数字化转型之道

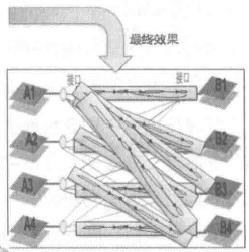


**基于服务总线定制的消息流集成开发，形成大量消息流接口紧耦**

合、逻辑复杂和管理困难



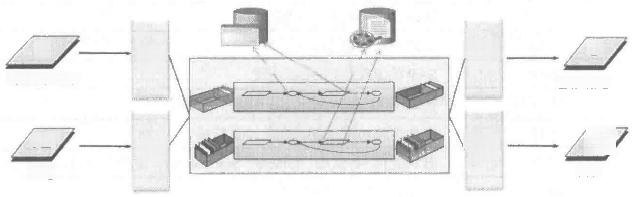
传统的端对端连接接口：网状的架构



形成网状架构

**图49-2-6** **专用消息流的开发效果**

服势目录/路由处理 日志处理

A

**webservice**

协设造

配

**通用核心消息处理罐**

协没造 .

配群

B

**SAP RFC**

A2

MQ

协这适

配器

\协议造

配路

B²

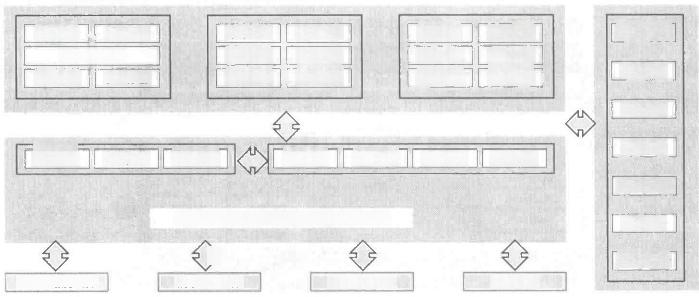
FTP

**图49-2-7** **通用性消息流的开发示意**

**2.数据交换平台功能架构**

威孚集团数据交换平台支持集群、负载均衡和容灾备份。平台的功能架构如图49-2-8所示。

服务发布平台

服务注册 服务测试 服务发布申请

发布订阅 异常报警

身份认证 节流控制

TOKEN服务组件

服务管理平台

|  |  |
| --- | --- |
| 服务发布  服务梳理  服务管理 | 外部发布  平台运维 服务授权 |

消息传输 消息转换 格式转换

基于IBM IB的数据交换运行和部署环境



消息日志组件 服务梳理组件

服务商店

|  |  |
| --- | --- |
| 服务发现  服务订阅  服务调用 | 服务测试  应用管理 异常报警 |

路由分发 消息传输

服务横板

监控分析平台

服务器监控

平台监控

服务监控

服务统计

交易量统计

趋势分析

失败率统计

**图49-2-8** **数据交换平台的功能架构**

第49章 制造行业：威孚集团基于斯欧应用互联平台建设数据通道 509

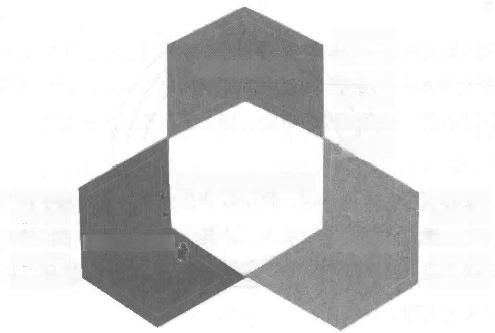
斯欧应用互联平台产品的显著功能是实现了服务的自助管理。其自助管理平台分为服务发 布平台、服务管理平台和服务商店(见图49-2-9);用户角色分别为服务发布者、服务管理者和

服务消费者。

●服务发布平台供服务发布者使用，可进行服务的注册、测试与发布等操作。

●服务商店供服务消费者使用，可进行服务的订阅、测试等操作。

●服务管理平台供服务管理者使用，可进行服务的梳理、管理、监控等操作。



-服务目录 服务管理

-服务测试

-服务梳理 服务监控

服务统计

·服务目录 服务订阅 ·服务测试 ·服务监控 服务统计

·服务注册 -服务测试 -服务发布 服务监控 -服务统计

**自助服务管理平台**

**服务发布平台，**

服务管理平台

服务商店

**图49-2-9** **斯欧应用互联平台产品三大自助管理平台**

**49.3** **项目实施步骤**

威孚集团的数据交换项目实施主要分为3个阶段：

(1)数据交换分析与服务识别；

(2)落地实施数据交换的工程实施标准；

(3)平台部署和服务注册上线。

**1.** **数据交换分析与服务识别**

(1)数据交换分析。

若要实现数据交换的目的，就需要对企业共享数据进行整理及分析。数据交换分析包含以

下工作。

①分析现有企业应用系统的数据交换状况，确定共享数据的服务粒度和交换方式(发布/

**510** 数据治理——工业企业数字化转型之道

订阅或请求/响应等),以及确定数据交换所需的其他特征参数，具体包括：

●数据交互周期(实时或定时);

●数据交换方向(单向或双向);

●数据交换场景(请求方或响应方，批量交换或增量交换，同步或异步， 一对一或一对多);

●数据交换量(数据量大小);

●数据交换协议(文件共享、公共数据表、数据库读取、AP I 、Web Service);

●数据的规范性。

②分析应用系统的集成接口协议和数据访问协议，确定数据交换的可能性。对于个别应用 系统，如果没有标准访问协议、数据字典及数据规范，也不采用标准数据库，或者数据库中的 数据已经加密，并且原供应厂商拒绝提供协助，那么这类系统的数据交换问题就无法解决，应 该放弃。

③分析数据格式和数据定义是否一致，确定数据在交换时是否需要转换；分析数据交换的 响应时间和交换数据量，确定合适的接口接入，还要分析系统之间交换的数据类型，确定是基 础数据(主数据)交换，还是交易数据交换，并制定相应的数据交换规范和协议。

(2)服务发现及服务识别。

服务发现及服务识别是一个梳理和定义需要进行交换与共享的业务信息的过程，即全面地 整理和确定企业应用系统中共同需要的业务信息的过程。可以从企业业务流程和数据共享这两 个维度进行梳理，找出通用的服务信息、设计和配置服务组件，并定义服务级别、服务宿主、 服务提供方式等。

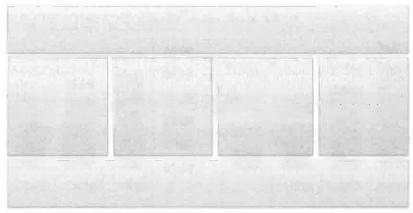
**2.** **落地实施数据交换的工程实施标准**

根据数据交换分析与服务识别成果，要在遵循及继承企业相关标准、规范的基础上，落地 实施数据交换的工程实施标准，如图49-3-1所示。

**3.** **平台部署和服务注册上线**

之后安装并部署企业数据交换平台，完成服务注册、测试、调优及上线。

第49章 制造行业：威孚集团基于斯欧应用互联平台建设数据通道 **511**



第六部分：集成实施规范

第二部分

编码及报文规范

第一部分：集成技术要求

第五部分

集成运维规范

第四部分

集成管理规范

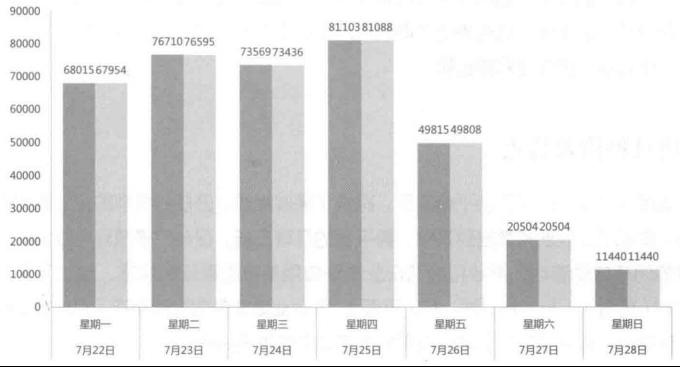
第三部分 接入指南

**图49-3-1** **数据交换的工程实施标准**

**49.4** **项目实施效果**

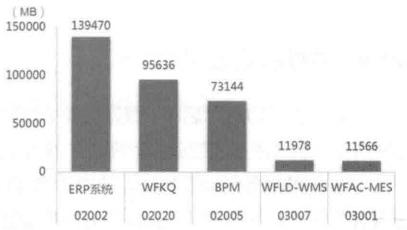
威孚集团在2013年启动了数据交换平台项目，至2019年已经完成三期数据交换项目实施。 数据交换平台注册接入的应用系统共有28个，注册接入并上线运行的接口服务共有347个，其 中SAP 系统的服务接口最多，达到175个。2019年新接入7个系统及70个服务，数据交换场 景模式全部为请求/响应，分别采用HTTP 适配器、MQ 适配器和 SAP 适配器。

数据交换平台集群部署了5台服务器，每周7×24小时运行，当年度故障率为零。数据交换 平台每周(以2019年7月22日至7月28日数据为例)的数据交换请求达38万次，成功响应38 万余次，成功率为99.91%;每周数据交换总流量为1847.80MB, 如图49-4-1和图49-4-2所示。

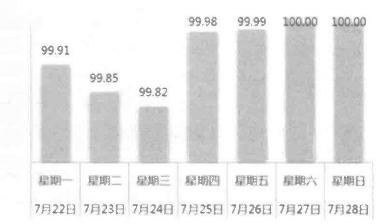


■每日交换请求次数 每日成功响应次数

**图49-4-1** **数据交换平台的每日交换请求次数和每日成功响应次数**



**512** 数据治理——工业企业数字化转型之道

(%)

100.00

99.95

99.90

99.85

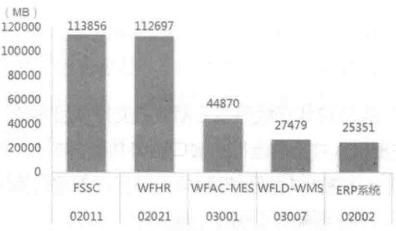
99.80

99.75

99.70

每日数据交换流量

每日交换请求成功率



■(7月22日-7月28日合计次数)交换请求最繁忙的5个系统 ■(7月22日-7月28日合计次数)成功响应最繁忙的5个系统

图49-4-2 平台运行主要指标

威孚集团在基于斯欧应用互联平台产品搭建数据交换平台后，后期新系统和新服务的注册 和集成，原有服务的变更、增减及平台的管理维护，大部分都由威孚集团的员工完成，基本脱 离了对平台供应商的依赖。这也表明了斯欧应用互联平台产品和方案达到了预期目标，让企业 的数据交换更容易、更快速和更省钱。

**49.5** **项目价值及特点**

威孚集团通过实施数据交换平台项目，构建了高度集成、开放和共享的企业数据交换智能 服务平台，实现了企业业务数据跨系统、跨平台的互联互通，促成了多源异构数据的集成、交 换和共享的闭环自动流动，指导和规范企业未来应用系统的建设和实施，建立了企业级层面 的业务状态(包含人、机、料、法、环、财等),为企业级业务向智能制造(包括状态感知、实 时分析、科学决策和精准执行)方向发展奠定了工业软件连接基础。

制造业向智能制造方向发展是一个循序渐进的过程，企业应用系统之间的数据交换是企业 进行智能制造建设的一个重要和必需的环节(或阶段)。威孚集团的数据交换项目的成功落地实

第49章 制造行业：威孚集团基于斯欧应用互联平台建设数据通道 **513**

施和应用，提升了企业的管理效率和竞争力，为企业带来了良好的经济效益。

威孚集团采用的斯欧应用互联平台是基于企业服务总线架构的标准化平台产品，能够确保 数据信息可靠地自动交换，快速变更注册服务信息，快速接入新建系统的服务信息，快速迁移 存量服务，方便了数据交换平台的扩展升级和管理运维，解决了企业应用系统之间数据交换的

工程实施和管理运维难题。

**第** **5** **0** **章**

**战略投资行业：国投集团的数据标准化管理实践**

**50.1** **背景介绍**

**1.** **公司介绍**

国家开发投资集团有限公司(以下简称“国投集团”)成立于1995年5月5日，是中央直 接管理的国有重要骨干企业之一。国投集团逐步形成了“股权投资一股权管理一股权经营”和 “资产经营与资本经营相结合”的独特运作模式。

国投智能科技有限公司是国投集团数字经济产业中的战略投资平台和信息化综合服务平台。 其依托股权投资领域的专业管理经验和国投集团多样化的资源优势，重点投资数字经济产业重 要基础设施、核心技术和高端装备等符合集团战略的重要领域。

**2.** **信息化现状**

国投集团10多年的信息化建设历程主要分为3个阶段：信息化起步阶段(2003—2009年)、 信息化快速发展阶段(2010—2015年)、信息化优化改革阶段(2016年以后)。按照“总部、子 公司、企业分级管理”的信息化建设思路，其信息化系统已覆盖到企业其他管理的各个领域， 有力地促进了企业管理水平的提高，但是也存在一些需要调整的方面和问题，具体包括以下几 个方面。

①国投集团整体缺乏完善的信息化标准和架构体系(包括信息化相关标准和架构的策略、组织、 流程、体系及相关技术手段)不利于集团统筹信息化规划和建设，以及子公司实现应用系统“小集中” 的目标。

第50章 战略投资行业：国投集团的数据标准化管理实践 **515**

②国投集团的整体信息化能力成熟度不足，IT 建设和运维能力相对较弱，IT 服务相关能力 缺失，需要进一步提高相关能力，才能满足未来成员企业的信息化建设和运维需求。

③国投集团总部的管理系统在功能支持、技术等方面已不能匹配于国投集团“成为具有全 球竞争力的世界一流资本投资公司”的定位：横向上存在信息孤岛、业务“竖井”等问题；纵 向上集团总部管理类系统向下推广的力度不足。

④国投集团中没有统一的数据集成平台，缺少相关的数据管理技术和工具；各业务部门从 各自的需求出发，独立建立了一系列专业数据库并积累了大量的历史数据与经营管理资料。

为实现集团的战略目标，国投集团共完成3次集团信息化规划。

2004 年，国投集团组织制定了《国家开发投资公司信息化建设总体规划》,明确了公司信 息化发展战略，确定了公司信息化建设的内容、方法、步骤，规划了八大类应用系统的信息化 建设目标。

2010 年，国投集团制定并印发了《国家开发投资公司(2010—2015)信息化总体规划》, 规划了母/子公司信息化建设的蓝图和计划，并针对各业务板块的信息化建设，设计了《行业信 息化建设指导意见》。

2018 年，国投集团启动《国投集团信息化规划(2018—2021)》编制，理顺了集团信息化 治理体系，明确了集团2018—2021年的信息化建设路径，以及以数据集成为主，向板块“小集 中”发展的思路。

国投集团的信息化工作始终坚持服务于公司发展战略，坚持“总体规划、分步实施、注重 基础、务求实效、稳步推进”的指导思想，并以规划为目标，以测评为抓手，以标准为指引， 以制度为准则。通过持续不断地建设与优化，国投集团的信息化已覆盖到集团管理的各个领域， 有力地促进了集团管理水平的提高。相信随着大数据及云技术的普及，信息化将在集团未来的 发展中发挥更大的作用。

**3.** **数据治理背景**

伴随着大数据技术的发展，数据资产化成为日益明显的趋势。国投集团的数据资产管理仍 然存在数据质量难以保障、数据垃圾难以处理、数据转化效率低下等管理痛点。如何充分挖掘 数据是提升国投集团数据管控能力的关键，也是难点。

随着各类业务系统的上线运行，系统之间的信息交换和实时集成的难度越来越大，集团成 员企业之间的信息交互也越来越频繁。由于各系统之间的数据标准不统一，导致不同系统之间

**516** 数据治理 工业企业数字化转型之道

的数据集成困难，各系统之间的数据无法实现快速、高效地交互，难以满足集团对各层级、各 类型数据的抽取、转换、分析等管控要求；另外，数据规范性较差，数据不一致、不完整的现 象普遍存在，各系统数据分散，数据质量不高，难以满足集团构建综合数据平台的要求。

要及时、有效地沟通信息，建立长效的数据管理机制，实现数据标准化，就需要建立完整 的数据治理管理体系。

**50.2** **工作概况**

国投集团的数据治理项目于2015年正式启动。经过充分的调研和分析，国投集团确立了以 “主数据建设模式”作为国投集团数据标准化及数据治理实践的重要切入方法，搭建了集团统一 的数据标准体系，制定了集团通用类数据标准与各板块专业的数据标准，建立了数据编码规范、 数据管理流程、数据指标和数据管理体系标准。

国投集团通过建设主数据管理平台，固化和落实了数据标准和管理体系，实现了数据全生 命周期的统一和规范管理，提高了集团对数据的管控能力，加强了国投集团对投资企业经营决 策的数据支撑力度。

**1.** **建设历程**

自2015年国投集团启动数据治理及国投交通板块试点工作以来，通过近几年持续不断的推 广与深化工作(2017年启动国投矿业板块的数据治理推广，2018年启动电子工程院板块的数据 治理推广，2019 年启动国投生物板块的数据治理推广等工作),国投集团将全面落实各项数据 标准，在国投贸易、中成集团等板块中贯彻应用。国投集团的数据治理项目主要经历了以下几 个阶段。

(1)体系建立阶段。

国投集团根据管理现状，首先确定了集团总部统一管理平台、统一通用标准，以及板块公 司试点先行，再逐步推广的建设思路，即集团总部搭建了统一的数据治理平台及数据管理体系， 制定了统一的通用类数据标准。板块公司要严格执行集团通用的数据标准，并在集团统一的标 准体系下自建个性化的板块标准体系。

(2)平台实施阶段。

制定完标准体系和通用数据标准后，国投集团梳理了数据治理平台的功能需求，并结合成

第50章 战略投资行业：国投集团的数据标准化管理实践 **517**

熟的数据治理平台产品，搭建了国投集团数据治理平台，实现标准体系的落地和数据全生命周

期的管理。

(3)板块试点阶段。

根据数据治理思路，在建成集团数据标准体系和数据治理平台后，国投集团选择了需求相 对迫切、基础相对成熟的国投交通公司进行板块试点。国投交通公司严格执行集团通用数据标 准，并建立了国投交通公司4类板块专用的标准。经过试点，国投集团后续在矿业、生物、电 子工程院等企业中进行推广实施，为国投集团“统一建平台、集团管通用、板块管专用”的数 据治理模式进行了验证，也为后续推广及实施积累了经验。

(4)优化提升阶段。

国投集团的数据治理平台在集团总部和试点单位中上线运行一年左右，收集到了用户在使 用中提出的新需求，结合数据治理体系和平台在实际运行中存在的问题，集团总部牵头对数据 治理体系和平台进行了优化： 一是完善及修订标准体系和标准；二是完善平台功能和优化服务 接口；三是通过集成对接将集团统管系统都纳入主数据管理；四是通过进一步的数据清洗提升 数据质量。

(5)板块推广阶段。

结合在集团总部及试点单位中得到的经验，国投集团制订了板块推广计划。截至2019年， 其已经先后在国投矿业、国投电子工程院、国投生物板块中推广实施，后续计划在国投贸易、 国投电力、中成集团等板块中进行全面推广。同时在推广的过程中，国投集团也继续对系统进 行优化及完善，集成对接各个业务系统，实现数据的统一集中管理。

**2.** **工作框架**

国投集团的数据标准化工作框架如图50-2-1所示。其中的具体内容包括以下4个方面。

(1)规划并设计国投集团信息标准化体系，制定主数据相关标准。

(2)通过搭建数据标准化管理平台，固化和落实数据标准和管理体系，实现主数据全生命 周期的统一和规范管理。

(3)制定标准化管理组织与规范，实现主数据的统一、集中、规范管理，提高对主数据的 管控能力，为信息系统建设提供标准和规范保障。

**518** 数据治理 — —工业企业数字 化转型之道

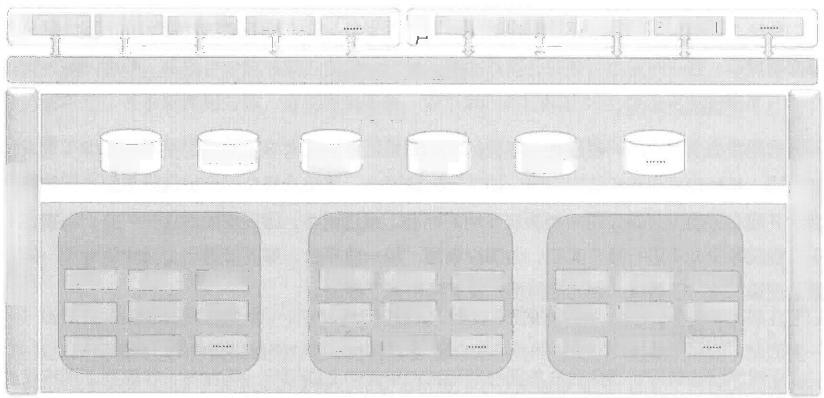


图 货团用户

数据资源“共享服务”

代码库

内部单位

代码库

编码标准

员工

供应商

会计科目

输码标准

员工

供应商

会计科目

员工

供应商

会计科目

子公司 投资企业 子公司部门

ERP **预值系统** **OA** **系统**

**内部单位**

项目

整警理组和種程

内部单位

项目

基础

客户

物料

数 權 复 和 耀

基础

客户

**物料**

基础

客户

物料

内部单位

项目

外部单位 代码库

物资代 码库

员工代 码库

项目代码

电子采购

编码标准

系统

**图50** **-** **2** **-** **1** **国** **投** **集** **团** **数** **据** **标** **准** **化** **工** **作** **框** **架**

④为总部各部门、各企业、各系统提供高质量、高效的信息化标准数据支撑及主数据服务。

**3.** **工作目标**

国投集团的数据标准化及数据治理的工作目标包括以下几个方面。

(1)以主数据建设模式为切入点，建立国投集团统一的数据标准体系；打通各业务链条， 统一数据语言，统一制定数据标准，实现数据共享，使数据资产价值最大化。

(2)统一业务信息定义，消除理解歧义；提升系统开发及实施的效率，实现资源共享，支 撑战略协同。

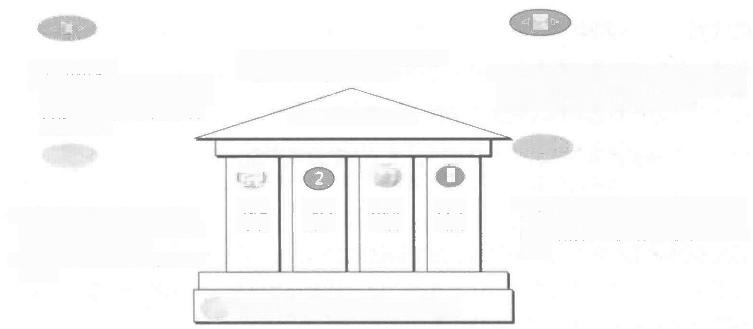
(3)构建准确、唯一、权威的数据来源，统一数据规范，进一步提高企业数据质量和数据 资产价值，全面增强企业的核心竞争力。

(4)为国投集团信息系统建设和深入应用提供标准和规范保障，为各单位、各部门、各系 统提供高质量、高效的信息化标准数据支撑，推动信息系统的深度集成、数据共享和深化应用。

**4.** **实施方法**

如图50-2-2所示，国投集团数据标准化及数据治理方法论框架主要包括制定标准、建代码 库、搭建平台、代码转换、运维体系共5部分。其中制定标准是基础，建代码库是过程，搭建 平台是技术手段，运维体系是保障，代码转换是实现工具。

第50章 战略投资行业：国投集团的数据标准化管理实践 **519**



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ·现状调研、需求分析及标杆企业分析 ·规划和设计主数据管理体系  系统架构规划与设计   |  |  |  |  | | --- | --- | --- | --- | | ①  制定  标准 | 建代  码库 | 3  搭建  平台 | 5  代码 转换 | | 2  ·编制符合数据标准和规范的主数据代码库。包 括按照数据标准进行数据检查、数据排重、数 据编码、数据加载、数据监控策略等  3  ·建设主数据管理平台，为数据的管理提供 技术支持，实现主数据申请、主数据管理 和主数据发布功能、数据清洗，包含系统 接口 |

运维体系(组织与管理流程、考核)

·确定数据范围

。与业务部门共同制订主数据编码 标准。数据标准内容包括确定分 类规范、编码结构、数据粒度、 属性描述等；

4

·建立主数据管理组织和管理流程， 包括建立标准管理和编码管理的运 维组织架构及考核流程；

·建立并完善管理流程、实现知识转 移等；

4

1

图50-2-2 国投集团数据标准化及数据治理方法论框架

**50.3** **组织保障**

在国投集团信息标准化领导小组的统一领导下，国投集团按照“归口管理，分工负责”的 原则，建立总部、子公司、投资企业3个层面的数据标准化管理体系。

**1.** **信息标准化委员会**

信息标准化委员会是国投集团信息标准化工作的领导机构，由集团总部各部/室和重点子公 司的管理层人员组成，主要职责包括以下几项。

(1)组织和协调集团信息标准化工作，对重要事项进行把关和决策。

(2)审定集团信息标准化相关制度、办法及信息化标准规范，经集团总裁办公室批准后发

布实施。

(3)听取信息标准化相关工作汇报，并进行指导。

**2.** **信息标准化办公室**

信息标准化办公室是国投集团信息标准化工作的日常协调和管理机构，由集团总部各部/

室的人员组成，主要职责包括以下几项。

(1)制订信息标准化工作计划，经批准后协调组织实施。

(2)按照集团要求，协调和组织具体事项的落实。

(3)协调和组织各职能部/室做好物流业务板块相关标准制定(或修订)和集团层面相关标

**520** 数据治理——工业企业数字化转型之道

准制定(或修订)的协助工作。

(4)协调和组织各项信息化标准的应用、推广和实施。

(5)督促及检查信息标准化各项工作的落实。

(6)负责与集团信息标准化办公室的日常工作联系和沟通协调。

**3.** **信息标准化工作组**

信息标准化工作组是为编制信息标准而组建的临时工作组织，根据标准化工作性质的不同， 分为信息标准化工作组和技术标准化工作组。信息标准化工作组一般采用项目工作组的形式， 完成信息标准的编制工作。

**50.4** **项目成果**

国投集团的数据标准化及数据治理工作取得了以下成果。

**1.** **数据治理成果**

随着国投集团信息化工作的稳步推进，其数据治理工作也在有序进行。在项目顶层设计上， 集团总部设计了统一的数据蓝图，在各板块及子公司层面还制定了分板块的数据蓝图和标准， 对内向所有板块进行推广，对外向投资企业及其他企业进行交流和推广。在顶层数据蓝图战略 的指引下，国投集团对数据架构和标准进行了详细设计，建立了统一的数据管控组织、制度及 规范，以及数据质量问题处理流程，并通过统一的数据治理平台进行落地，满足了国投集团对 数据集中管控的需求，为国投集团的经营和管理工作提供了全面、可靠、准确、及时的数据服 务，促进了国投集团的精细化管理和长效运营。

(1)数据蓝图。

国投集团设计了统一的数据蓝图，主要目标是建立健全的数据管理体系，统一数据标准和 数据架构，提升集团上下利用数据资产的效率，让企业既能知晓“过去与现在是什么”,又能清 楚“为什么”以及“将来会发生什么”。

数据蓝图主要包括4个部分。

①建立健全的数据管控体系：明确数据治理组织、人员、职责、流程及工具等；建立专业 化的数据集中管理组织，并建立各业务部门、数据管理组织和项目组之间的数据沟通机制。

②统一数据标准：基于业务内容和业务逻辑划分数据主题域，为数据应用打下坚实的基础； 以现有主数据标准为基础，对基础类数据和分析类数据制定标准和规范，在数据集成时发挥纽 带作用；建立数据质量标准化管理机制，保障数据质量。

③统一数据架构：设计集团与各子公司的统一数据架构，以实现数据集成；基于集团业务， 规划企业数据交换平台、数据服务总线等中心化的集成平台，指导数据集成。

④服务决策支持：扩大数据获取范围，提升数据获取的时效性，开发系统的数据统计分析 功能，强化数据深入应用；针对不同层级的管理人员和业务人员，提供有针对性的数据信息和

展现场景以支持决策分析。

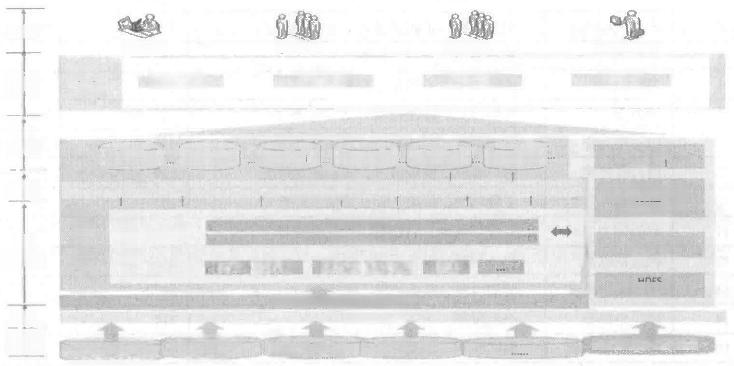
基于以上内容，国投集团规划了一套统一的数据管理体系框架，使企业数据“看得见、说

得清、管得住”,从而可以充分利用企业数字资产，创造和提升企业业务价值。

(2)数据治理架构。

国投集团规划了统一的数据治理架构，包含3个核心部件：主数据管理系统、数据仓库和 数据交换层。建立数据仓库和数据交换总线并应用大数据平台技术，可以满足高层管理者、管

理部门、外部监管及各子公司等不同层面的需求(见图50-4-1)。



用户层

高雇领导

管理驾驶舱

四的础

DM

Hive/Pig (政表分析

MapRaduce/spak

做期计银处理

Hbae/ (数据存)

资 产

皮件得础

操作型数据存储(ODS)

数据交换总线

非结构/半结构化数据

主数解管理系统

深度汇总层 轻度汇总层

投资 风险 财务

数据分析用户

建模分析

汇 基

核心数据区

业务人员

自定义报表

财务综合数据 管理系统

风险集市 资产集市

总 层

础 层

ow

数据仓库



展示 应用

大数据平台

电子采购平台

财务共享平台

数据仓库

数据交换层

多维度查询

投资集市

报表集市

战略集市

管理人员

对标集市

集市屉

交换层

战 略

ODs

**图50-4-** **1** **国投集团的数据治理架构图**

(3)数据标准。

国投集团也建立了统一的数据标准(一整套的数据规范、管控流程和技术工具),用以确保

企业中的各种重要信息(包括战略、投资、资产、项目、人力资源等)在全集团内外的使用和

**522** 数据治理——工业企业数字化转型之道

交换都是一致、准确的。国投集团的数据标准主要包含基础类数据标准和分析类数据标准。

基础类数据指在日常业务过程中所产生的具有共同业务特征的基础性数据，如战略、投资、 项目、人力资源等数据。

分析类数据指为满足国投集团内部管理需要及外部监管要求，在基础类数据的基础上按一 定统计、分析规则加工后的数据。

国投集团建立的数据标准主要包括以下内容。

①建立了100多万条基础类数据标准信息代码库。

国投集团建立了信息代码体系表，统一编制了物料、内/外部单位、财务类、各板块专用等 15个数据类型，截至2019年6月，共形成了100 多万条数据标准信息代码库，为国投集团总 部及下属企业的核心业务系统提供了标准化的数据服务，如表50-4-1所示。

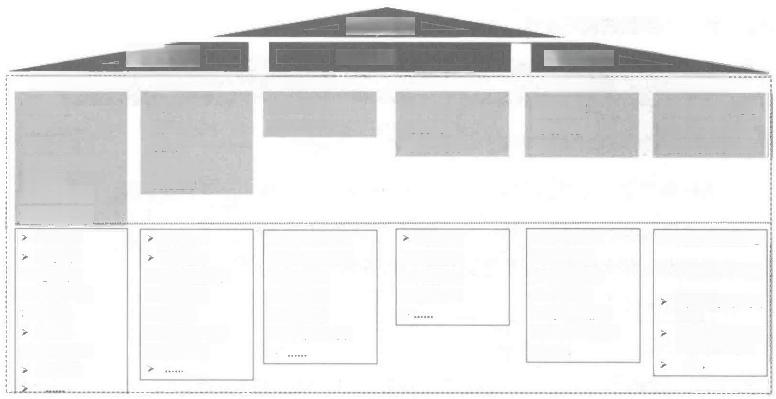
**表50-4-1国投集团信息代码体系表**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序** **号** | **数据名称** | **数据内容** | **代码数量(条)** | **备** **注** |
| 1 | 通用基础类 | 行政区划、计量单位、语种等 | 17684 | 集团通用 |
| 2 | 财务类 | 会计科目、报表项目、部门段、银行等 | 180195 | 集团通用 |
| 3 | 外部单位类 | 客户、供应商 | 46333 | 集团通用 |
| 4 | 内部单位类 | 编码规则、主数据机构、视图信息等 | 5858 | 集团通用 |
| 5 | 项目类 | 编码规则、项目名称等 | 767 | 集团通用 |
| 6 | 员工类 | 编码规则、主数据结构等 | 29022 | 集团通用 |
| 7 | 物料类 | 物料分类、描述模板、取值附表等 | 715199 | 集团通用 |
| 8 | 数据指标类 | 分类标准、指标维度、编码规则等 | 422 | 集团通用 |
| 9 | 港口基础类 | 港口概况、港域情况、港口运营情况 | 679 | 交通专用 |
| 10 | 基建类 | 工程项目、单项工程、单位工程等 | 103 | 交通专用 |
| 11 | 安健环类 | 安全、环保、职业健康、应急管理 | 649 | 交通专用 |
| 12 | 铁路基础类 | 机车车辆、工务工程、通信信号、供电 | 91 | 交通专用 |
| 13 | 设备设施类 | 装卸机械、铁运设备、船舶等 | 3470 | 交通专用 |
| 14 | 矿业专用类 | 产品段等 | 717 | 矿业专用 |
| 15 | 电子院项目类 | 编码规则、电子院项目名称等 | 6473 | 电子院专用 |

②定义了1万多项数据指标。

国投集团对总部的数据指标进行了整体规划，按照自上而下的梳理方式，从财务、人力资 源、生产经营等业务领域中定义了1万多项数据指标，保证了业务定义和概念的一致性、应用 规则的一致性，如图50-4-2所示。

第50章 战略投资行业：国投集团的数据标准化管理实践 **523**



六大主题

82个维度

生产经营

·渔口设施

·港口装备

·港口生产

·港口设施.装备运

用及效率

·港口生产质量

·港口经济评价

泊位个散

库存面积

>港区面积

>货物吞吐量

>操作量

装卸量

>装卸自然吨

工作台时

>科技人员活动数量

> 工程技术人员数量

科技活动经费支出 总额

科技经费总额占主 营业务收入比重

营业成本

经营收入

>技术投入比率 > 利润总额

>净利润

>国定资产原价 >资产总计

>外派劳务人数 >职工平均人数 ▶工资总额

>人均工资

>人均利润

>人均人工成本

>

>伤亡事故数 > 重伤人数

>千人伤亡率

》伤亡人数

>伤亡经济损失

>工业废气排放量

财务管理 ·收益管理

·现金管理

·资产负债管理 ·投资管理

·财务能力

**科技创新**

●科技创新基本情况 ·科技创新费用管理 ·科技创新成果管理

项目性质 项目规模 >项目进度 >项目质量

¥

·项目基本情况 ·项目监控

·项目后评价

388个指标

Hs管 理

●安全管理 ·节能减排 ·职业卫生

·人员基本情况 ·人员运作情况

22个子主题

基证管理

人万资源



**图50-4-2** **国投集团的数据指标体系**

(4)数据管控。

国投集团搭建了统一的数据管理组织，并发布了一套数据管理制度和规范，以及制定了《主 数据管理办法》《主数据维护细则》等，保证对数据标准的统一管理和应用。

数据管理组织是数据管理体系中最重要的因素，它负责制定并管理数据标准，提出各系统 的数据质量要求，监测正在进行的数据管理行动，并从数据的角度对信息系统项目提出指导建 议。数据管理组织是否有完整及合理的角色定义、是否有高层领导的参与，是整个数据管理体 系成败的关键。国投集团建立的数据管理组织包括数据管理指导委员会、数据责任部门、数据 管理部门，具体介绍如下。

①数据管理指导委员会(网信安全与信息化领导小组)由集团公司领导牵头，同时包含各

职能部门领导，是数据管理最高权力机构，负责数据管控战略、数据管理授权、预算审批。

②数据责任部门为虚拟机构，由集团公司、各板块(子公司)、投资企业共三级体系组成，

每级都设立数据管理负责人(可以全职或兼职),主要负责以下工作。

●明确本部门相关数据需求及解释相关数据的业务规则。

●提出本部门相关数据的质量要求。

●提出本部门相关数据标准的变更需求。

●数据管理效果反馈。

③数据管理部门为信息化相关部门(即国投集团运营部和数字国投事业部),职责范围包含

**524** 数据治理——工业企业数字化转型之道

数据标准和数据质量的管理工作。

国投集团运营部主要负责：

●制定并维护标准化体系。

●设计并落实相关制度和流程。

数字国投事业部(由10~15位专业人员负责数据管理工作)主要负责：

●统一收集数据标准需求，统一制定集团数据标准并推动落实。

●收集数据质量问题，牵头制定数据质量提升方案并推动落实。

●日常数据生成及维护。 (5)数据质量。

国投集团制定了统一的数据质量管理要求及流程。只有高质量的数据才有应用的价值。在 数据质量管理中，需要对数据进行持续监控，并在此基础上快速定位数据质量的源头和范围， 采取针对性的解决措施。

①数据质量管理的目标。

国投集团的数据质量管理目标是统一数据质量管理体系，统一数据质量度量、执行的规范 化步骤，为企业数据质量管理的落实提供技术平台支撑。

②数据质量管理工作内容。

数据质量管理以主数据和共享数据为核心，涉及以下4个方面内容。

●数据质量状况分析：引用系统内置度量规则和检核方法，对系统表、字段进行多角度的 数据质量检核及评估，为数据分析、数据状况调研提供数据分析报告，对系统中的关键 业务数据的质量情况进行全方位把握。

●数据质量持续监控：确保为将来提供准确、 一致和及时的数据，使数据始终符合业务规 则，确保企业数据始终保持最高的质量，并防止随时间的推移导致数据质量下降。

●快速定位数据质量问题的影响范围：通过元数据关联关系分析功能，可以定位问题数据 出错的源头和受影响的范围，及时通知关联方，降低由数据问题带来的影响。

●基于企业对数据质量问题处理的要求，灵活制定数据质量问题的处理流程，便于对问题 数据进行处理和清洗。

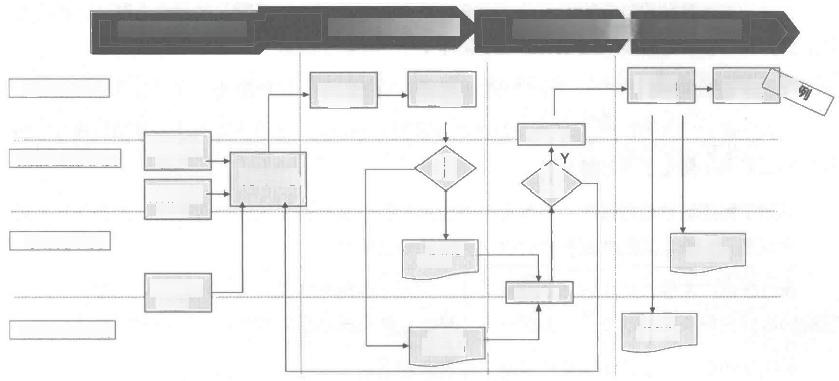
③数据质量问题处理流程。

国投集团在具体的数据质量问题处理过程中，通过发现、分析、验证和解决、评估4个步

朝 重

第50章 战略投资行业：国投集团的数据标准化管理实践 525

骤，使数据质量得到不断提升。国投集团日常的数据质量问题处理流程如图50-4-3所示。



数据质量问题发现 数据质量问题分析 **问题验证和解决，数据质量评估**

数钢因置 分析评估

问题验证

数据质量相关人员

问题 N 解决

系统维护人员

工单处理

源系统

源系统数据人员

数 据 质 量 检查

数据质量 稽核

问题汇总

累统问 题工单

问题工单 生成派单

分析问题 间题展现

系统 评估报告

数据管理人员

数据质量 考核评价

源系统 闻题工单

ETL处 理

过程

评估报告

分

照判

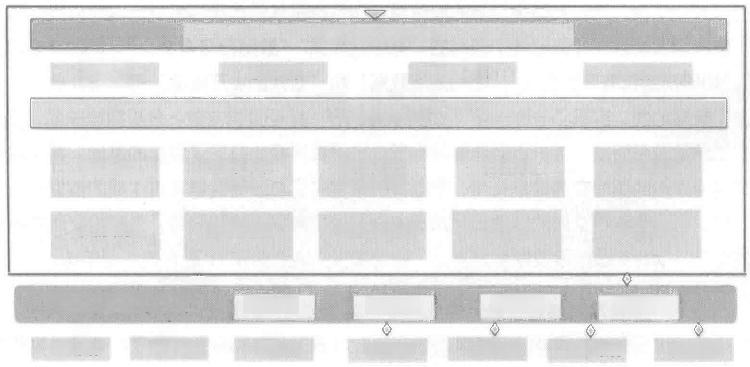
**图50-4-3** **国投集团数据质量问题处理流程图**

(6)管控平台。

国投集团建立了统一的信息标准化平台，为国投集团总部20多个应用系统及各分/子公司

提供了全面、准确、高效的数据共享服务，实现了对基础数据从发布到应用、维护的全生命周 期管控。此平台具有企业级数据存储、整合、清洗、分发及监控共5大功能，具有良好的可扩

展性和灵活性。图50-4-4为国投集团信息标准化平台架构。



**信息标准化综合门户**

移动应用

**核心功能模块**

数据指标 标准化监控

动态建模工具



数据交换

◇

费控系统

服务管理

OA 系统

企业服务总线(ESB)

WMS 系 统 其他系统

代码发布与 查询

主数据全生 命周期管理

工作流工具

MES 系 统

接口管理

数据服务

用户管理

流程监控

系统管理

标准管理

数据清洗

搜东管理

ERP系统

文档管理

HR 系 统



图50-4-4 国投集团信息标准化平台架构

**526** 数据治理——工业企业数字化转型之道

**2.板块推广实施成果**

为了切实有效地贯彻国投集团的数据治理成果，使总部、子公司、控股企业联动，国投集 团启动了在各板块推广及实施数据标准的工作。

通过在各板块中的推广，以及加强数据标准及运维培训，国投集团取得了以下成果。

提高了各分/子公司对数据标准化管理的重视及参与程度，为全面提升国投集团的数据治理 和数据服务水平奠定了良好的基础。

巩固了集团数据治理成果，将数据标准应用到整个集团范围内的各板块中，为各信息系统 的互联互通，系统之间的数据集成提供了更好的专业化服务。

通过在推广实施过程中的数据服务优化，实现业务服务增值，为管理层提供完整、 一致、 准确的数据分析决策辅助支持，为国投集团整体业务向更高层次提升打下了良好的基础。

国投集团在4个板块中的推广实施工作简介如下。

(1)国投交通。

国投交通控股有限公司(以下简称“国投交通”)是国投集团的全资子公司，是对港口、铁

路、油气管道等综合交通基础设施及现代物流项目进行控股、参股经营的投资控股型公司。

在国投集团总部的数据标准体系和数据治理平台建成后，国投集团选择了需求相对迫切、 基础相对成熟的国投交通公司进行板块试点。国投交通严格执行集团通用的数据标准，并建立 了国投交通4类板块专用的数据标准(基建类、安健环类、港口基础类、物料类)。同时其启动 信息资源规划项目，通过对国投交通的业务需求和数据需求的调研与分析，以及开展信息资源 规划，解决国投交通在业务过程中面临的“有哪些数据”“数据存储有多少”“数据流中的血缘 关系”“如何找到我需要的数据”“如何使用数据”和“数据的生产进度”等问题。此项工作为 绘制公司数据地图、实现“一数一源”、统一数据口径、标明数据方位、分析数据关系、打通上 下游数据继承关系断层、维护数据质量和实现业务逻辑可视化打下了坚实的基础。将来以此项 工作为前提建设的国投交通数据仓库，可以有机地实时汇总所有信息系统生成的数据；再通过 数据挖掘，可以为国投交通的管理和经营提供全方位的数据支持和服务，为国投交通的信息化 向集成提升、创新突破阶段发展奠定基础，构建国投交通业务和IT 融合的企业架构。

(2)国投矿业。

国投矿业投资有限公司(以下简称“国投矿业”)是国投集团专业从事矿产资源及其相关产 业投资的全资子公司，经营范围主要包括油气、金属、非金属等矿产资源和可再生资源、危险

第50章 战略投资行业：国投集团的数据标准化管理实践 **527**

废物资源、新能源、建材及相关配套产品设施的投资开发。

基于国投交通板块成熟的推广经验，国投矿业在严格执行总部通用数据标准的基础上，梳 理出矿业6大类的数据标准(设备物资类、安健环类、项目类、数据指标类、矿产资源类、生 产类)。通过编制国投矿业的统一数据标准体系，统一数据标准代码库，实现了对数据“源头” 的集中管理，改变了原有基础数据分散管理的现状，为业务报表编制、数据统计分析提供便利 条件；建立了基础数据共享的“桥梁”,为主营业务系统集中化部署与集约化运营管理奠定了数 据基础。

(3)电子工程院。

中国电子工程设计院有限公司(以下简称“电子工程院”)创建于1953年，2009年被并入 国投集团。电子工程院以设计为龙头，服务范围涵盖前期咨询及规划、环境和节能评价、工程 设计、项目管理、工程监理、工程承包、工程检测评定等全过程，可在工业工程、民用建筑、 节能环保新能源等诸多领域，为国内外客户提供优质、全面的工程服务。

电子工程院的数据标准体系推广策略是严格遵循总部通用类数据标准，其中需求最大的是 单位类数据标准。由于电子工程院是工程服务单位，涉及众多的客户、供应商，客户和供应商 基本信息的准确性、合法性对于后期的工程监理、检测评定结果至关重要。因此，按照总部单 位类数据标准，电子工程院引用外部工商数据源(企查查)清洗所有单位类历史数据，确保客 户和供应商数据的准确、真实、有效、完整；建立了统一的数据运维体系，并配备了信息员和 审核员，利用统一的数据标准化管理平台针对供应商类数据进行全生命周期管理，确保了数据 的质量和时效性，为企业全面风险管控和统计分析奠定了数据基础。

(4)国投生物。

国投生物科技投资有限公司(以下简称“国投生物”)是国投集团的全资子公司，主要负责 统筹国投集团生物能源业务的科研管理、投资经营管理，是国投集团发展燃料乙醇等生物能源 产业的责任主体和市场主体。

国投生物板块的数据标准化项目推广策略是在符合国投集团总部统管要求的基础上，建立 国投生物板块特有的标准体系。从2019年开始启动，其已完成：生物板块特有设备分类，安健 类环、原料类、成品类数据标准的建立，数据清洗及数据初始化，并通过企业服务总线实现与 ERP、MES、OA 等系统的无缝集成。后续国投生物也会逐步建立数据管控组织、流程、规范等， 保证信息系统中数据编码的唯一性与正确性，提高数据质量和增强信息共享，针对生物板块特 有数据，实现全生命周期管理。

**528** 数据治理——工业企业数字化转型之道

**50.5** **工作价值**

在《国家开发投资公司2010—2015信息化总体规划》的指导下，围绕国投集团的各项业务 和现有信息系统，国投集团在集团范围内建立了一套科学、适用的数据标准和管理流程，并通 过主数据管理系统进行落地，实现对数据的全生命周期管理，为国投集团的信息共享和管理奠 定了信息标准基础，具体包含以下5个方面的工作价值。

(1)助力集团实现“建设具有全球竞争力的世界一流资本投资公司”的愿景目标。

国投集团的信息化规划是以“建设具有全球竞争力的世界一流资本投资公司”为目标，集 团信息化管理和建设的重点在于信息化治理、数据治理、IT 基础设施建设，网络与信息安全建 设等工作。其中数据治理是推进集团数字化转型的重要抓手，对于规划企业数据架构、数据标 准，建立集团数据资源管理体系，提升集团化管控能力具有重要的战略意义。

(2)支撑集团加速推进数字化转型，打造国有资本投资企业标杆典范。

国投集团通过统一的数据治理，规划数据架构和数据标准，探索大数据、移动应用、智能 决策等技术在集团信息化建设中的应用，规划了企业IT 基础设施架构及集团云平台，推进了集 团数字化转型。

(3)推动数据治理架构落地，夯实集团数据治理地基。

国投集团的数据治理架构分为治理、管理和应用3个层次。数据治理对集团新的信息化治 理架构的落地具有积极的推动作用，也将对集团的信息化资源优化配置提出建议。

(4)夯实集团信息标准体系，完善集团化管控策略和模式。

实现“数据集成”目标，既是实现集团总部系统之间数据横向集成的需要，也是实现集团 总部、子公司、控股投资企业之间数据纵向集成的需要。国投集团通过建立数据资源管理体系， 实现了数据资源的及时、准确、共享，为国投集团以后利用数据资产提升集团总部掌握全局、 动态分析、风险预控的能力打好基础。

(5)辅助应用优化，助推集团统管系统落地。

数据管理要做的就是从企业的多个业务系统中整合核心、最需要共享的数据，集中进行数 据清洗和数据丰富，并且以服务的方式把统一、完整、准确、具有权威性的数据分发给各个业 务系统、业务流程和决策支持系统等。结合应用生命周期管理过程，国投集团持续不断地进行 数据治理，并辅助进行统管系统和板块专有系统的应用优化。同时，集团的 OA 系统、法律、

第50章 战略投资行业：国投集团的数据标准化管理实践 529

财务共享平台等应用的推广，使得集团能够进行集中化数据管理，构建覆盖整个集团范围内的 数据管理基础和相应规范，以及更灵活地适应各个企业的业务需求变化。

**50.6** **经验分享**

在大数据时代，数据成为资产已经是行业共识，越来越多的企业将“数据价值挖掘”作为 重要的战略方向。从IT 建设的角度来看，通过主数据管理可以增强企业IT 结构的灵活性，构 建覆盖整个企业范围内的数据资产管理基础和相应规范，并且更灵活地适应企业的业务需求变化。

国投集团在数据治理的实践过程中，通过加强内部宣传、贯彻及培训，提升了员工的数据 治理意识；通过与行业内其他企业的交流，总结数据治理经验；通过将数据治理的方案和成果 积极推广至集团内和投资企业中，实现数据治理分析和经营管理过程的联动；通过用数据发现 问题、调整问题、预测问题，让信息化建设呈螺旋式上升，持续支撑业务良好运转，扩大影响 范围，推进业务的不断升级与创新。

**1.** **内部宣传、贯彻及培训**

国投集团分别对管理层人员、关键用户、系统管理员等进行分层级、分角色培训，并将数 据标准进行宣传及贯彻。

**2.** **外部沟通与交流**

(1)行业会议。

作为投资行业的龙头企业之一，国投集团积极参与编制《主数据管理实践白皮书(1.0版)》, 并在行业交流大会上进行经验分享。国投集团的数据治理工作作为实践案例被行业所认可。

国投集团的数据标准化项目所取得的成果丰硕，在2019年大数据产业峰会上，国投智能与 国投交通联合提交的“国投集团数据治理实践案例”荣获2019年大数据“星河奖”——优秀数 据资产管理实践奖，在2019年中国国际大数据产业博览会“第三届数据标准化及数据治理优秀 评选”活动中荣获“数据治理优秀实践奖”。

(2)行业内数据标准化项目分享交流。

国投集团的数据标准化及数据治理的实践经验也不断在投资行业中进行分享与交流，包括 陕西省投资集团有限公司、安徽省投资集团控股有限公司、黑龙江辰能投资集团有限责任公司、 河南省国有资产控股运营集团有限公司、深圳市投资控股有限公司等；同时，国投集团的数据

**530** 数据治理 教字化转型之道

治理经验也不断被投资行业以外的企业学习。

**3.** **经验总结**

通过与行业内外其他企业的分享与交流，国投集团总结了数据治理经验，这也为其他企业 进行数据标准化及数据治理开拓了思路，避免了在实施过程中一些问题。其中以下几点特别值 得其他企业借鉴。

(1)领导重视，全员参与。

数据治理涉及企业管理组织、业务流程、管控流程、人力资源等各个方面，是一项复杂的 系统工程，只有引起企业领导层的高度重视，得到各级员工的认可，才能有效执行，否则会事 倍功半。因此，在数据治理体系建立过程中就应该动员全体员工积极参与，增加员工对数据治 理体系的认知度，并在数据治理体系建立后，通过培训、座谈、考核、推广，扎实分级、分业 务域贯彻标准等多种形式让员工多方面了解及掌握。

(2)循序渐进，自上而下。

由于数据治理涉及面广，因此，想在数据标准化体系建立的过程中一蹴而就是不可能的。 集团管控型企业可以借鉴“集团总部建通用，下属板块建专用”的模式，分级管理，先试点、 后逐步推广。国投集团正是通过在推进总部通用数据标准及板块专用数据标准的过程中逐步规 范各管理线的方式，取得数据标准化管理的成功。

(3)组织保障，考核激励。

数据治理的复杂性，决定了其实施过程必然是一项耗时、耗力的艰巨任务。这就需要企业 投入必要的资源，包括在企业内部成立管理领导小组、执行小组等组织，并且由企业内的权威 部门来牵头推进，深度参与，负责各自标准的制定和对数据把关，发动业务部门参与数据清洗， 确保数据质量。同时可将这些工作纳入企业绩效考核体系中，对体系建立及实施人员进行充分 的激励。

(4)持之以恒，持续改进。

数据治理需要企业耗费较长的时间，是一项长期工程，要提前做好整体规划，分期实施， 逐步推广。随着企业内/外部环境及战略的变化，还需要对数据标准进行不断调整，以适应企业 的发展要求。在企业发展过程中，企业管理模式没有最好的，只有更适合的。

第50章 战略投资行业：国投集团的数据标准化管理实践 **53** **1**

(5)顶层设计，平台落地。

数据治理涉及业务的梳理、标准的制定、数据的监控管理、数据的集成等工作，复杂度高、 探索性强。因此，在进行数据治理时，要采用系统的方法做好顶层设计工作。国投集团按照“整 体规划，分步实施，先基础，后深化”的总体原则，通过统一的数据治理平台进行落地，并实 现与各应用系统的接口集成，保证数据源头的统一。

**第51章**

**多元化集团：数据治理助力多元化企业集团管控**

**51.1** **背景介绍**

**1.** **公司介绍**

新兴际华集团有限公司(以下简称“新兴际华”)是集资产管理、资本运营和生产经营于一 体的大型国有独资公司，及世界500强企业。其下属的百余家企业遍布于全国30个省(自治区、 直辖市),业务聚焦冶金、轻纺、装备、医药、应急、服务6大板块。

新兴际华采取“战略管控+财务管控”的管控模式，集团被定位于战略投资中心，二级公司 被定位于经营管理中心，三级公司被定位于成本利润中心。

**2.** **信息化现状**

新兴际华是集团型多元化企业，在带领下属企业完成集团的战略目标时，面临着管控难度 大、信息化建设难度大等很多问题，具体介绍如下。

(1)集团管控方面：集团高层领导缺少可以辅助管理决策的信息系统。

(2)数据治理方面：在数据治理工作启动前，缺乏统一的数据标准，数据纵向贯通及横向 共享缓慢、数据质量差；缺乏集团统一管控的数据治理环境，缺乏数据安全保障环境；缺乏统 一规划的数据传递路径，数据的安全性、及时性难以保障；缺乏异构系统之间的数据集成共享 手段，早期各企业建设的应用系统之间相互独立，信息数据相对分散。

如今，不断推由进制造企业走向智能制造企业是企业发展的必然途径。数字化、网络化、 智能化日益成为未来制造业发展的主要趋势，而数字化和网络化是基础，是制造业向智能化发

第51章 多元化集团：数据治理助力多元化企业集团管控 533

展的前提。

新形势带来好机遇，新兴际华积极寻求解决问题的路径，将信息化建设规划上升为集团公 司战略规划之一，用以支撑集团管控及公司决策。

**3.** **数据治理背景**

早期基于业务需求的推动，集团公司、新兴铸管、际华集团均建设了一些业务管理系统。 各系统内部形成了局部、独立的数据标准。但是，不同业务系统之间的基础数据未能形成统一 的数据标准，难以实现各系统之间的数据共享和集成应用，并且存在多头管理、资源分散、编 码不统一、信息整合难度大等问题。经过梳理和总结，新兴际华亟待建设主数据管理系统，主 要用于解决以下问题。

(1)在定义方面，没有统一的标准，没有明确的定义和范围；

(2)在流程方面，数据的创建、维护等管理流程不一致；

(3)在质量方面，缺乏完整性、 一致性、准确性，难以管理；

(4)在共享方面，缺乏源头，缺乏标准。

针对以上问题，在制定信息化总体规划时，新兴际华非常重视主数据的统一管理和建设。 在总体规划中，新兴际华将信息资源体系作为其信息化五大体系之一，并明确提出主数据管理 建设目标：在设计主数据编码体系的基础上，主导建设主数据管理系统，通过主数据管理系统， 统一管理新兴际华集团公司内部主数据，支撑集团公司主数据的编码规范，实现未来各种信息 系统之间数据的统一和连贯，确保重要数据信息在跨板块、跨部门、跨区域、跨业务系统中的 共享和应用，落实及完善主数据的申请、审核、批复、退出等信息化管理流程，保障信息编码 规范和管理流程在集团公司信息化建设中发挥长效作用。

集团公司牵头组织了两次主数据治理项目。

第一次在2013年，启动“新兴际华集团公司主数据系统建设项目”。

第二次在2018年，新兴际华集团公司启动资金统一支付系统、统一核算系统建设，对主数 据中的会计核算科目、物料大类和明细分类、财务组织架构提出进一步深化应用的需求。集团 公司及时牵头启动“新兴际华集团公司主数据深化应用建设项目”,支撑业务系统落地。

**534** 数据治理——工业企业数字化转型之道

**51.2** **数据治理概况**

按照集团公司的战略部署，2012年年底，新兴际华完成了《新兴际华集团公司信息化总体 规划》(以下简称“规划”),并将其作为集团公司战略规划的重要组成部分。规划中明确了信息 化建设要以“打造以有质量、有效益、可持续发展为特征的国际一流强企”为指导思想，以“四 五三四”信息化工程建设(“四”:指管控集中化、平台集成化、应用弹性化、运维标准化；“五”: 指构建应用、IT 技术、信息资源、信息安全、IT 管理五大体系；“三”:指集团公司主导的三大 纵向平台建设；“四”:指推动四大板块横向一体化平台建设)为主要任务，按照“夯实基础， 深化应用，管理创新”三大步骤有序开展，初步建成管控集中、决策智能、产业链协同、服务 敏捷的“智慧新兴际华”。

**1.** **数据治理历程**

新兴际华的数据治理工作曾经历了3个阶段。

(1)第一阶段，电算化代替手工化，实现数据入库。

1994—1999年，新兴铸管(新兴际华冶金板块)启动了以计算机管理代替手工管理的信息 系统建设，相继完成大宗原燃料检验化验系统、磅房称量检斤系统、生产调度管理系统、库存 管理系统。当时，存入数据库中的物料并未进行统一编码，数据标准不统一。

(2)第二阶段，单一企业多系统数据共享，实现成本、利润日核算。

2000年，新兴铸管启动计算机辅助管理信息化建设项目，拟建设日成本、日利润管理信息 系统，采购管理信息系统等。冶金制造企业是长流程企业，成本核算要素涉及指标多、工序多， 其中的数据均取自生产现场源头。而采购成本是产品成本的主要组成部分，因此，采购管理信 息系统要求以工段采购需求计划为源头，以合同为主线，包含采购计划管理、供应商管理、合 同管理、库房管理、大宗原燃料结算等模块。

在项目启动前，考虑到数据编码不统一，无法实现共享的情况，项目组决定优先启动统一 分类编码工作。项目组首先成立编码小组，编制编码手册，形成企业数据标准(以下简称“编 码标准”)从而支持新兴铸管武安工业区的计算机辅助管理信息化建设的顺利开展。

新兴铸管(即冶金板块)有6个大工业区，编码标准在武安工业区成功应用后，在其他工 业区也获得认可并得到应用。虽然编码标准被推广至新兴铸管的各大工业区中，但均在各工业 区局域网内应用，并未实现冶金板块企业统一共享编码资源。

新兴际华的轻纺板块由30多家制造企业组成，包括服装、皮革皮鞋、纺织品、染整、橡胶

第51章 多元化集团：数据治理助力多元化企业集团管控 **535**

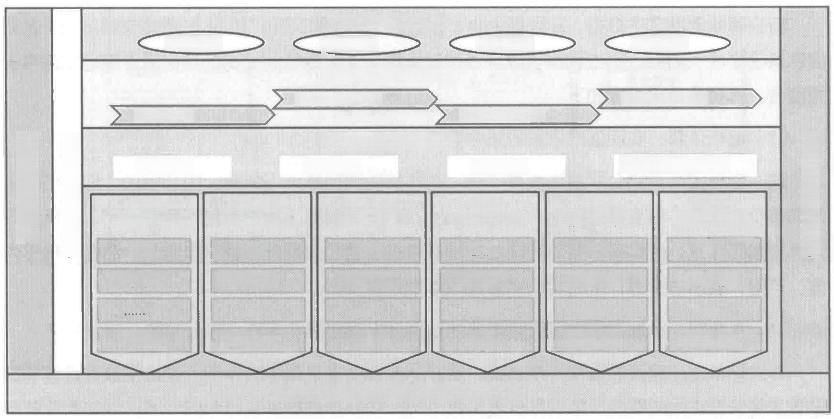
制品、装具等主营业务。其中管理基础好的企业较早就开展了本企业的信息化建设，有直接置 于应用系统的编码，但此时企业之间没有制约管理，未形成统一的数据标准体系。

(3)第三阶段，多板块中的基础数据高质量汇聚应用，支撑集团公司的数据管控。

新兴际华集团公司的成立晚于其旗下的大多数制造企业，信息化建设起步也较晚。集团公 司要实现数据管控目标，需要将多板块中的基础数据(大多来源工业数据)高质量地汇聚应用， 形成可信赖的指导决策的数据。

**2.** **数据治理框架**

2012 年年底，新兴际华集团公司总部牵头制定《新兴际华集团信息化总体规划》(其信息 化建设总体架构如图51-2-1 所示)。此规划经董事会审批，得到了集团公司高层领导的支持与 认可。新兴际华集团公司的信息化建设以规划为指引，实现了数据的高质量采集、传递、存储、 共享、应用，从而提高了数据资产的价值。在规划中提出了“四五三四”信息化工程建设，“五” 是指构建应用、IT 技术、信息资源、信息安全、 IT 管理五大体系，要实现集团信息化建设，这 五大体系建设是重要任务。



决策支持一体化平台

竞争情报

集团管控一体化平台

人力资源管理

财务管理

综合服务一体化平台

党群信访工会纪检

装备制造

板块一体化

ERP

统一研发

……

资源管理

以新一代信息技术为支撑的信息技术体系

应急 服务

板块一体化 板块一体化

ERP ERP

电子商务 电子商务

科技管理

医药

板块一体化

ERP

统一营销

统一门户

冶 金

板块 一 体化

ERP

集中采购

统一的信息资源体系

统一的一管理体系

轻工纺织 板块一体化

ERP

统一营销

应急指挥

战略投资管理

统一的信息安全体系

全面风险管理

物业管理

门店终端

协同办公

知识管理

商业智能

MES

PLM

MES



\*\*\*\*

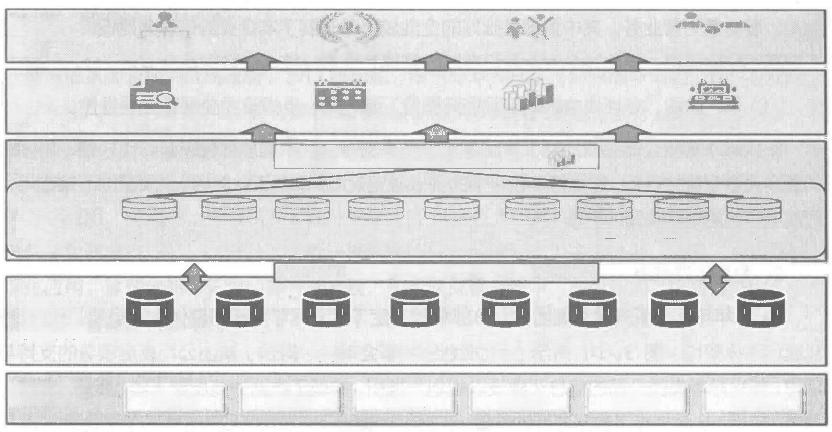
…

**图51-2-1** **新兴际华集团信息化建设总体架构**

其中，集团公司牵头建设三纵平台(集团管控、综合服务、决策支持一体化平台)及数据 来源各板块，组织高质量的数据采集，并在规划中提出了数据资源技术体系建设，如图51-2-2

所示。

**536** 数据治理——工业企业数字化转型之道





各三级公司

多维数据分析 专业数据挖掘方法

 集团级数据仓库

数据资源

集中储存

ODS

ETL过程(抽取、转换、加载)

DB DB

集团办公系统 板块ERP系统

主数据体系 组织类 员工类 设备类

数据消费最终 用户

数据资源

综合利用

DB

应急指挥 系统

科目类

DB

风险管理系统

DB

项目投资 系统

数据基础 资源平台

预算体系 数据

安全质量 数据

人力资源 数据

DB

科技 系统

DB

财务 系统

DB

人力 系统

集团总部处室

报表灵活定制

集团公司领导

**综合数据查询**

生产 数据

计量 数据

市场 数据

工程 数据

销售 数据

财务 数据

二级板块

产品类

……类

图51-2-2 数据资源技术体系框架

**3.** **数据治理目标**

新兴际华启动了主数据、云数据中心、数据通道(集团公司广域网)、数据交换服务(企业 信息系统集成)项目，以主数据管理系统建设为抓手，实现集团公司的数据标准落地和信息资 源整合。其具体建设目标如下。

(1)编码体系：实现信息编码的集中管控。

统一全集团公司的公共基础数据代码，构建覆盖通用基础、物料、内部单位(组织结构)、 外部单位(客户、供应商)、人事(人员、组织机构)、财务(会计科目、固定资产、指标)五 大类的编码体系。此编码体系对集团公司所涉及的主要公共基础数据予以定义、命名，确定内 容、范围、表示方法等，从而实现对编码的集中管控。

(2)主数据：实现集团公司的编码落地和信息资源整合。

新兴际华通过建设主数据管理系统，更好地发挥了信息资源的作用：通过主数据管理系统 打通了各系统之间的壁垒，实现数据连贯和协同，确保重要数据在跨部门、跨区域、跨业务系 统中的一致性和共享应用；建立完善的编码体系和保障机制，实现集团公司编码落地和信息资 源整合。

(3)信息资源：分析企业信息资源，为业务系统整合与集成提供支撑。

第51章 多元化集团：数据治理助力多元化企业集团管控 **537**

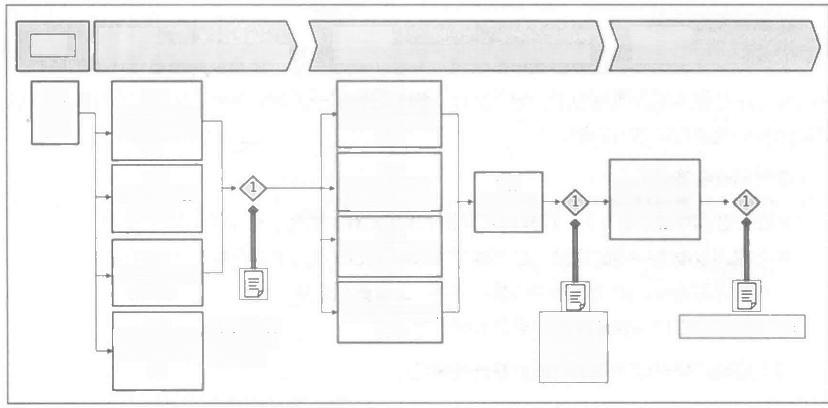
新兴际华全面分析企业在战略管控及财务、生产和业务管理过程中所需的人、财、物等信 息资源，结合集团公司信息化总体规划建设，构建了信息资源部署模式；通过建设与完善人力 资源管理系统、财务管理系统、项目管理系统、ERP 系统等，逐步提高集团公司信息资源的完 整性、真实性、及时性；建立一体化数据资源管理平台，对数据进行统一管理，满足企业横、 纵价值网络中的数据共享、信息交互的要求，为业务系统整合与信息集成应用提供支撑。

**4.** **实施方法**

其实施方法是规范项目组织过程并辅以体系建设，具体介绍如下。

(1)规范组织。

新兴际华的主数据项目建设主要分为以下3个阶段，如图51-2-3所示。



第一阶段

调研访谈、需求分析与标准制定

T06

T02

调研、访谈企

业数据资源现

状。

分析集团主数

据需求。

|  |  |
| --- | --- |
| T04  制定信息代码 体系表。  T05  制定工作范围  内数据标准规 范 | 代码体系表、 各类主数据  编码标准 |

|  |  |
| --- | --- |
| T07  根据业务模型、  进行产品配置  T08  业务培训、系统  培训  T09  数据填报、审核、 发布 | T10  与目标系  统集成  架构设计 实施文档  培训文档  软件产品 |

第二阶段

架构设计、平台搭建、培训填报、系统集成

T11

技术服务、制

定各类主数据

营理规范

构设设计、数据

模型、数据源头

确定

启动

T01 启动 项目

第三阶段

运行维护规范制定

运行维护管理办法

T03

**图51-2-3** **主数据项目建设**

第一阶段：调研访谈、需求分析与标准制定。在本阶段全面调研、访谈各企业的数据资源 现状，分析集团公司的主数据需求，制定信息代码体系表和数据标准规范。

第二阶段：架构设计、平台搭建、培训填报、系统集成。在本阶段开展主数据系统架构设 计，确定数据模型和数据源头，根据业务模型和产品配置，组织业务培训和系统培训。

第三阶段：运行维护规范制定。在本阶段制定和完善主数据运维管理体系，梳理业务流程， 明确工作职责，制定考核标准和制度。

**538** 数据治理——工业企业数字化转型之道

(2)启动云数据中心项目，建立数据存储计算中心。

启动云数据中心项目，建立数据存储计算中心，可以在集团公司(包括集团总部、二级公 司、三级公司)范围内，提高全集团数据资源的完整性、真实性、及时性，实现对人、财、物， 产、供、销等所有经营信息的全面记录和及时汇总、整理，为分析集团公司及下属的二三级公 司开展财务、人力资源、贸易采购等专项业务活动，以及辅助各级领导决策支持，奠定数据资 源基础。

利用数据集成技术，新兴际华完成了统一的信息资源技术框架的构建，支持信息资源的综 合利用与统一管理；在集团公司建立统一的数据平台，可以对集团公司的管理决策、二级板块 主营业务中涉及的业务管理数据进行统一管理，为信息资源的集成与应用提供强力的支持。具 体包括以下两项建设。

①基础设施建设

2015年，新兴际华完成云数据中心建设，实现了数据存储、计算及网络基础环境的搭建任 务。数据中心是实现大数据应用的硬件支撑，建立集团统一的硬件基础设施，可以实现各系统 之间的资源共享和动态调配等。

②数据安全建设

●在云数据中心，服务器、存储、网络等资源虚拟化，增强了机房资源的可靠性、安全性。

●考虑到数据集中度的增加，必须要加强数据安全。在这个项目中，新兴际华同时进行了 数据灾备建设，防止数据中心发生意外，造成数据丢失。

●同时，建立信息安全体系，提高数据安全性。 (3)启动广域网建设项目，建立数据传递通道。

新兴际华基于IPSec VPN 及广域网加速技术，建设了集团公司通信网络体系，覆盖集团总 部、二级板块、三级公司，成为连接集团公司与下属公司的“数据信息数据高速公路”,从而实 现了集团公司的信息数据传输高效、稳定、安全，支撑了集团公司应用系统的正常运行，以及 广域网络的统一建设与集中运维管理。

(4)启动信息集成平台项目，建立数据交互服务。

为解决新兴际华所属的二三级公司信息系统底层技术的异构性，使新兴际华的信息系统互 通，打通集团二级公司与三级公司的信息系统之间的数据交换与共享，实现三级公司的已有系 统、在建系统、待建系统中的信息能够及时、准确地传回二级公司并进行数据的汇总与合并等

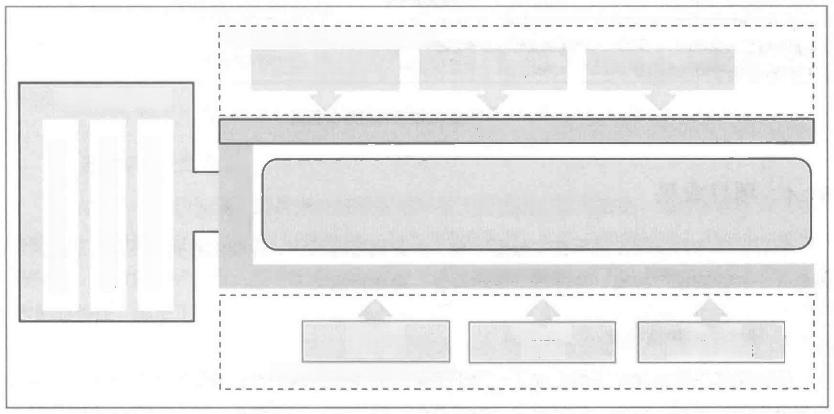
第51章 多元化集团：数据治理助力多元化企业集团管控 **539**

功能需求，新兴际华建立了集团两级信息集成平台体系，实现三级公司信息系统与二级板块和

集团公司之间的纵向数据整合通道 。

新兴际华集团公司的信息集成平台采用“松耦合”的集成模式，引入面向服务框架(SOA)

的设计理念，对集团核心业务系统(包括集团管控平台、综合服务平台、决策支持平台)、二级 板块一体化平台、外部信息系统(包括上级主管企业的业务系统)的相关对外接口模块进行基 于Web Service 的“服务”封装，通过适配器挂接到业务服务总线上，在业务流程引擎的驱动下， 利用消息传输机制，完成系统之间数据的实时(同步)或批处理(异步)传递，实现跨系统业 务处理的自动完成，如图51-2-4所示。



新兴际华集团与二级板块系统集成

集团综合服务 集团管控 集团决策支持一体化

一体化平台 一体化平台 平台

对外数据服务集成

信息集成平台

二级板块一体化平台

信息集成平台

三级企业

专业系统

二级板块与三级企业系统集成

工信部相关信息系统

科技部相关信息系统

国资委相关信息系统

三级企业 原有信息系统

三级企业

…系统

信息集成平台

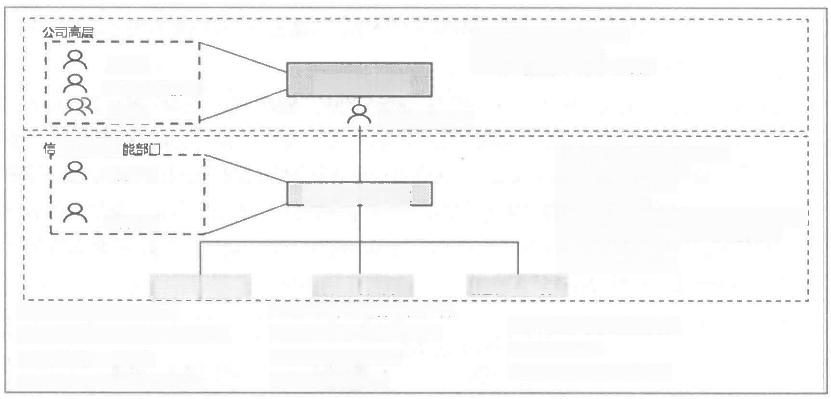
**图51-2-4** **信息集成平台框架**

**51.3** **组织保障**

新兴际华的主数据建设项目构建了覆盖集团公司高层、信息化管理职能部门及项目组(项 目总监与项目经理)3 层组织模式。集团公司高层负责项目战略的制定与监督实施、整体统筹 与协调建设，以及项目建设过程中的重大问题决策；信息化管理职能部门负责项目总体管理工

作 ； 项 目 组 负 责 规 划 与 需 求 管 理 、 平 台 建 设 与 实 施 、 数 据 整 理 与 填 报 ， 如 图 5 1 - 3 - 1 所 示 。

**540** 数据治理——工业企业数字化转型之道



■项目战略制定与监督实施

■整体统筹和协洞建设

■项目建设过程中重大问题决策

信息化领导小蛆

信息化主管领导

|  |  |
| --- | --- |
| 信息化管理办公室  平台建设与实施  ■系统软件、基础设施环境筹备  ■系统新建、更新维护、监控  ■系统应用培训与答疑  ■系统实施文栏编制 | ■项目总体规划的制定和维护  ■项目预算和立项申报  ■项目的招投标管理  ■项目实施监控  ■项目信息化组织架构、标准的制定和落实；  ■确保相关资源包括实施商资源的投入  ■组建项目实施小组  ■项目实施过程管理  数据整理与填报  ■本单位主数据填报工作组织  ■数据填报工作  ■数据更新、质量监控工作 |

■项目调研访谈、分析现状

■规划信息代码标准体系与管理办法

■需求分析、 制定标准

■制定和编制运行维护管理制度

董事长

信息化主管领导I

总经理

项目总监

项目经理

规划与需求管理

息化管理职血

图51-3-1 主数据项目组织及职责

**51.4** **项目成果**

新兴际华主数据建设项目在主数据标准、主数据管理系统、标准化数据、管理办法和运维 制度4个方面取得了成果，并进行了试点验证。具体介绍如下。

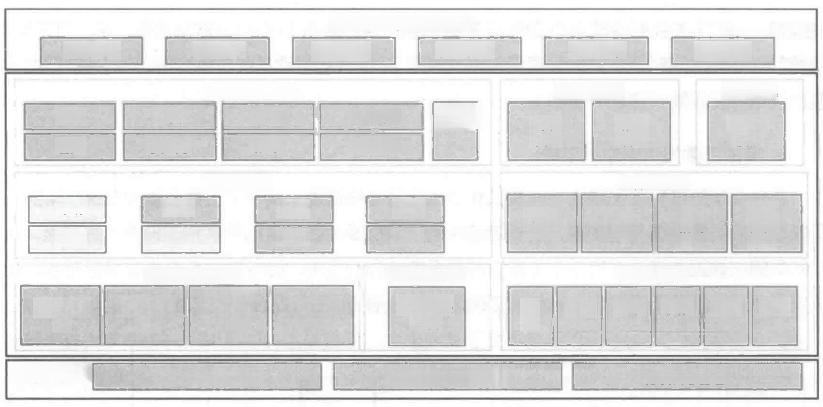
**1.建立了主数据标准**

新兴际华主数据建设项目形成了适用于全集团的统一数据标准，涉及集团及下属公司经营 业务范围内的物料、财务、外部单位、内部单位、员工类、固定资产、数据指标、会计科目、 通用基础共9类主数据，实现了集团公司各类公用数据的统一标准、统一源头、统一管理。其 中，工作量最大的是物料类主数据治理。通过建设统一的数据标准，较好地支撑了新兴际华后 期信息系统的建设与系统整合工作。

**2.** **建成主数据管理系统**

新兴际华集团公司级主数据管理系统总体架构如图51-4-1所示。新兴际华还编制了操作手 册，以方便用户充分了解和熟悉系统的相关业务和技术需求。

第51章 多元化集团：数据治理助力多元化企业集团管控 **541**



新兴际华主数据管理门户

管理社区

实施工具

EXCEL 模板

主数据管理

主数据管理

资源 组织 权限 日志 参数

主数据分发标准服务

分发

分析

内容管理组件

平台模块 整合

审批

冻结

标准化工具组件

新闻资讯

主数据模型

固定资产

会计科目

人员

供应商

逻辑模型 定义

抓取数据 标准化

标准数据 加载

业务系统注册管理

数据源配

置

主数据分类体系

申请

修改

物料

指标

组织

客户

配码 解冻

数据分发接口

分发日志管理

主数据管理

编码 配置

动态 模型

静态 模型

标准管理

公告通知

待办工作

综合查询

通用类

用 户

**图51-4-1** **新兴际华集团公司级主数据管理系统总体架构**

**3.** **形** **成** **标** **准** **化** **数** **据**

新兴际华的主数据建设项目通过梳理数据形成了大量的标准化数据。通过集团公司各部门 的努力，已经形成了10万余条标准数据。在物料数据方面，集团公司累计投入百余人(来自新 兴铸管、际华集团的业务人员占比在80%以上),历时近1年，完成了试点单位中近2万条原始 数据的整理、清洗工作，形成了1万余条标准数据，扭转试点单位原有数据不一致、不标准的 局面。在银行数据方面，集团公司的财务公司提供了标准的银行主数据，通过将其导入主数据 系统中并分发到集团公司的各业务系统中，为资金管理、客户管理、供应商管理提供了标准的 银行信息。在组织人员数据方面，集团公司的人力资源部与各单位归口管理部门给予了大力支 持，仅用不到3个月，就完成了全集团公司的组织机构的数据填报工作，实现组织机构信息覆 盖全面，有效地支撑了全集团公司的人力资源、财务核算、资金管理等业务运行。

4.制定主数据管理办法和运维制度

新兴际华的主数据建设项目制定了主数据管理办法，明确和规范了主数管理组织体系和维 护方法，确立了组织机构，定义了岗位职责，制定了流程与考核办法，规范了主数据的制定、 修订、发布、实施、运维和应用工作。按照“统一规划、统一设计、统一标准、统一管控，以 我为主、外部为辅”的指导方针，新兴际华强化信息化的“统一信息标准”的管理要求，确保 了主数据系统上线后能平稳、有序、顺利地应用。

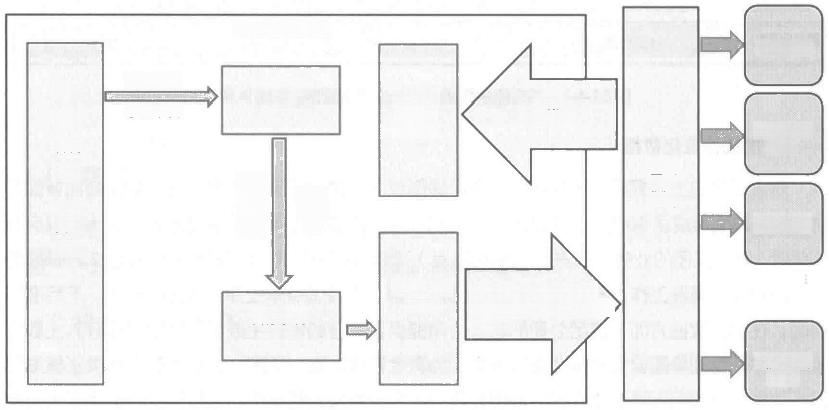
新兴际华的主数据建设项目形成标准规范10个：《主数据管理办法%物资统一分类与代码》

**542** 数据治理——工业企业数字化转型之道

《主数据标准》《主数据管理维护细则》《主数据系统操作手册》《物料主数据管理系统运维方案》 《物料代码提报指南》《主数据管理系统集成规范》《主数据管理系统应用实施方案》《主数据管 理系统用户及权限申请表一模板》。

**5.** **完成主数据试点集成工作**

新兴际华通过标准化集成 API 接口和企业服务总线通道，实现了主数据管理系统与集团公 司各层级应用系统的高效集成，统一管控数据传输权限，达到信息共享和编码统一的目标。新 兴际华的主数据建设项目完成了主数据管理系统与办公系统、财务管理系统、项目管理系统及 试点企业系统等的集成工作，9大类主数据的集成应用得到有效验证(见图51-4-2)。



主数据管理系统

办公系统

目标系统

维护

|  |  |
| --- | --- |
| 数据  关联  目标  系统 | 写入 分发 表 |

XML头信息携带目标

目标系统

维护

试点企业 系统

|  |  |
| --- | --- |
| 企业  服务  总线 | 财务管理 系统  项目系统 |

组织 人员 物料 客商

固定 资产

指标

科目

通用

系统业务类型、数据唯 一标识

审核通过

手工下发

主数据 分发日 志功能

定时推送 功能

ESB响应 推送结果

**图51-4-2** **主数据管理系统应用集成方案**

**6.** **建立了数据治理环境体系**

新兴际华的数据治理为工业大数据全生命周期管理提供环境和服务，具体包括以下内容。

(1)建立云数据中心，为数据存储、计算、网络需求提供服务。

(2)建立集团公司广域网，为数据传递、共享提供服务。

(3)建立同城数据灾备中心，为数据安全提供一份保障。

(4)建立数据信息安全体系，保障网络数据更安全。

(7)提高了全员对数据治理工作的认识程度。

第51章 多元化集团：数据治理助力多元化企业集团管控 **543**

数据标准及环境体系的建设，为新兴际华后期的信息化建设工作提供了方便，也让数据治 理工作不断得到全员的认可，让集团公司营造出良好的数据治理氛围。随着数据质量的提升， 标准数据量的不断增加，使得数据价值逐渐前景可期。

**51.5** **工作价值**

数据是企业信息化的基础，也是支撑建设智慧企业的基础。经过主数据建设项目，新兴际 华的数据治理工作已经从分散管理向集中、标准化管理迈出了跨越式的一步。数据治理是一项 长期、持久并且十分艰苦的工作。在新兴际华这种大型多元化国有企业里推行数据标准化工作， 必须要在政策支持、工作组织、工作方法、技术服务支撑等多方面配合下，才能够持续进行。

未来，新兴际华将继续加大力度推进数据治理工作，健全组织机构，探索 SaaS 化转型，构 建数据资源生态，为集团公司的智能制造、工业互联网、电子商务等创新转型业务提供更高效、 更便捷的主数据服务。

**51.6** **项目总结**

**1.** **宣传及培训**

新兴际华先后组织了5次集团及试点单位培训，其中物料类主数据培训2次，其他类数据 培训3次；并配置239名人员负责填报、审批各类主数据权限，指导各企业按照集团要求，相 互协助、有序建设，共同实现集团公司数据治理的战略规划。

**2.** **经验总结**

集团型多元化企业的管理层级较多，内部工业大数据种类多、来源广，外部相关数据源多， 涉及应用系统多样。在这类企业中进行数据治理是一项复杂的系统工程，需要在管控模式、前 期规划设计、人才培养、资金投入等环节做好充分准备。体系建立得好，数据资产的价值才能 充分发挥出来。

(1)高层领导重视是成功的关键。

新兴际华的集团公司领导高度重视数据治理工作，明确指出集团型企业的信息化建设必须 要建立统一的数据标准，并且在实施的全过程中，精心指导、全力推动，确保所需要的资源调 配到位，并提供政策上的支持。这些都是确保数据治理成功的关键。

**544** 数据治理——工业企业数字化转型之道

(2)顶层设计先行，达成统一共识。

数据治理是一项庞大的系统工程，需要提前做好规划，企业上下要达成共识，才能在规划 指导下有序开展工作。要防止各部门各自为政，这样不仅耗资、耗时，还没有成果。

(3)要重视人才培养与引进。

随着新一代信息技术的爆发式发展，企业对人才的需求非常迫切，掌握和正确应用新型技

术对于工业大数据体系的建设质量至关重要。

(4)持续投入确保迭代推进。

信息化建设不是一蹴而就的工作，每家企业都有各自的特点，支持工业大数据的产品也纷 繁多样，系统建设个性化的需求也会存在，因此，企业必须做好持续投入迭代性优化的准备。

(5)需求导向是项目成功的基础。

工业数据治理涉及企业的多个业务领域，是复杂的系统工程，其中比较困难和重要的是数 据模型的需求定义和架构设计。因此，企业需要认真分析、比较、归纳和提升应用场景，形成 模型定义和架构设计，脱离实际应用只能形成普遍适用的框架体系。在具体应用需求产生时还

需要再进行针对性实施。

(6)要重视数据管理组织体系建设。

数据管理组织体系建设是数据标准能够有效推进、持续应用的保障。新兴际华在信息化领 导小组的统一领导下，按照“归口管理，分工负责”的原则，建立新兴际华和所属企业两个层 面的主数据管理组织体系。其中集团公司层面的主数据管理组织体系由集团信息化领导小组、 信息化管理办公室和相关职能部门、专业代码组、主数据管理组组成；所属企业层面的主数据 管理组织体系由企业公司信息化领导小组和主数据管理组组成；业务人员必须深度参与。

附录 A

工业英文缩写术语表

[1]5WIH:WWWWWH, 原因(Why) 、 对象(What) 、 地点(Where) 、 时间(When)、 人员(Who) 、 方法(How)

[2] ADM:Architecture Development Method,架构开发方法

[3] AGC:Automatic Generation Control,自动发电控制

[4]AHP:Analytic Hierarchy Process,层次分析法

[5] AMP:Advanced Manufacturing Partnership,先进制造伙伴计划

[6]AVC:Automatic Voltage Control,自动电压控制

[7] BCS:Basic Communication Statute,基本通信结构规约

[8] BOM:Bill of Material,物料清单

[9] BPM:Business Process Management,业务流程管理

[10]C2M:Customer-to-Manufacturer, 用户直连制造

[11]CAPP:Computer-Aided Process Planning,计算机辅助生产计划

[12]CDO:Chief Data Officer,首席数据官

[13]CIM:Common Information Model,公共信息模型

[14]CIO:Chief Information Officer,首席信息官

[15]CPPS:Cyber-Physical Production System,信息物理生产系统

[16]CPS:Cyber-Physical Systems, 信息物理系统

**546** 数据治理——工业企业数字化转型之道

[17]CRM:Customer Relationship Management,客户关系管理

[18]CRUD: 增加(Create)、读取(Retrieve)、更新(Update) 删除(Delete)

[19]CTO:Chief Technical Officer,首席技术官

[20]CWM:Common Warehouse MetaModel, 公共仓库元模型

[21]DaaS:Data-as-a-Service,数据即服务

[22]DAMA:Data Management Association International,国际数据管理协会

[23]DCMM:Dada Management Capability Maturiy Model,数据管理能力成熟度评估模型

[24]DCS:Distributed Control System, 集散控制系统/分布式控制系统

[25]DDL:Data Definition Language,数据定义语言

[26]DDM:Dynamic Data Masking,动态数据脱敏

[27]DFFT:Data Free Flow with Trust, 可信赖的数据自由流通

[28]DGI:The Data Governance Institute,数据治理协会

[29]DLP:Data Leakage Prevention,数据防泄密(泄露)防护

[30]DLT:Distributed Ledger Technology,分布式账本技术

[31]DMMA:Data Management Maturity Assessment,数据管理成熟度评估

[32]DMS:Database Management System,数据库管理系统

[33]EBOM:Engineering BOM,工程物料清单

[34]EDW:Enterprise Data Warehouse,企业级数据仓库

[35]ERD:Entity-Relationship Diagram,实体关系图

[36]ERP:Enterprise Resource Planning, 企业资源计划

[37]ESB:Enterprise Service Bus, 企业服务总线

[38]ETL:Extract-Transform-Load, 数据抽取、转换、加载

[39]HSE: 健康 (Health) 、 安全(Safety) 和环境(Environment ) 三位一体的管理体系

[40]laaS:Infrastructure-as-a-Service, 基础设施即服务

[41]ICT:Information And Communications Technology,信息与通信技术

附录 A 工业英文缩写术语表 **547**

[42]IT:Information Technology,信息技术

[43]KKS:Krafwerk-Kennzeichen System,电厂标示系统

[44]MD:Master Data,主数据

[45]MDM:Master Data Management,主数据管理

[46]MES:Manufacturing Execution System,制造执行系统

[47]MQ:Message Queue,消息队列

[48]MQTT:Message Queuing Telemetry Transport, 消息队列遥测传输协议

[49]ODS:Operational Data Store,操作数据存储

[50]OLAP:On-Line Analytic Processing, 联机分析处理

[51]OLTP:On-Line Transaction Processing,联机事务处理

[52]OMS:Order Management System,订单管理系统

[53]OT:Operation Technology,运营技术

[54]PaaS:Platform-as-a-Service,平台即服务

[55]PBOM:Plan BOM,计划物料清单

[56]PDM:Product Data Management, 产品数据管理

[57]PEST:Politics (政治)、Economy(经济)、Society(社会)、Technology(技术)

[58]PLC:Programmable Logic Controller,可编程逻辑控制器

[59]PLM:Product Life-cycle Management,产品生命周期管理

[60]PMS:Production Management System,生产管理系统

[61]RFID:Radio Frequency Identification,射频识别

[62]SaaS:Software-as-a-Service, 软件即服务

[63]SCADA:Supervisory Control And Data Acquisition, 数据采集与监视控制系统

[64]SCM:Supply Chain Management, 供应链管理

[65]SDM:Static Data Masking, 静态数据脱敏

[66]SFC:Shop Floor Control,车间控制系统

**548** 数据治理——工业企业数字化转型之道

[67]SOA:Service-Oriented Architecture,面向服务的架构

[68]SRM:Supplier Relationship Management, 供应商关系管理

[69]SWOT:Strengths (优势)、Weaknesses(劣势)、Threats(威胁)、Opportunities( 机 会 )

[70]TOGAF:The Open Group Architecture Framework,开放组体系结构框架

[71]WMS:Warehouse Management Systems,仓库管理系统

**数据治理**

**联袂荐读**

责任编辑：王 静饮迎投稿wang(@phei.com.cn

封面设计：吴海燕唐 泓梁 琴

工业企业数字化转型之道

**国家信息化专家咨询委员会常务副主任周宏仁作序**

**高金吉**

中国工程院院士、 中国工业互联网研究院技术专家委员会主任

**倪光南**

中国工程院院士、

中国科学院计算技术研究所研究员

**单志广**

国家信息中心信息化和产业发展部主任、 国家大数据发展专家咨询委员会秘书长

**黄澄清**

中国互联网协会副理事长、

国家计算机网络与信息安全管理中心原主任

**余晓晖**

中国信息通信研究院副院长、 工业互联网产业联盟秘书长

**安筱鹏**

阿里研究院副院长、数字化企业研习社副理事长、 中国信息化百人会执行委员

**李向阳**

中国科学技术大学计算机科学与技术学院院长、 IEEE Fellow、ACM Fellow、ACM杰出科学家

**刘云浩**

美国密歇根州立大学计算机系系主任及

大学基金讲席教授、IEEE Fellow、ACM Fellow

**范济安**

中国联通大数据首席科学家

**刘劫**

哈尔滨工业大学人工智能研究院院长、

IEEE Fellow 、ACM杰出科学家

**邱才明**

**汪广盛**

国际数据管理协会(DAMA)

华中科技大学电子信息与通信学院院长、

中国分会主席

EEE Fellow



**上架建议：数据治理**

ISRHLT-T~21+39327-°

9 18?123959 ,

定价：158.00元