

数据资产管理实践白皮书 (2.0版)

中国信息通信研究院云计算与大数据研究所
2018年4月

版权声明

本白皮书版权属于中国信息通信研究院云计算与大数据研究所，并受法律保护。转载、摘编或利用其它方式使用本白皮书文字或者观点的，应注明“来源：中国信息通信研究院云计算与大数据研究所”。违反上述声明者，本院将追究其相关法律责任。

牵头编写单位：中国信息通信研究院云计算与大数据研究所

参与编写单位：中国电信股份有限公司云计算分公司、中软国际有限公司、中国移动苏州研发中心、中国移动通信研究院、联想（北京）有限公司、成都四方伟业软件股份有限公司、北京东方金信科技有限公司、北京国脉互联信息顾问有限公司、北京供销大数据集团股份有限公司、烽火通信科技股份有限公司、上海新炬网络技术有限公司、中国建设银行、石化盈科信息技术有限责任公司、北京数语科技有限公司

编写组成员：

中国信通院云计算与大数据研究所：魏凯、姜春宇、刘成成、闫树、马鹏玮

中国电信股份有限公司云计算分公司：蓝曾威、杨杨、李慧

中软国际有限公司：贾丕星、符山、陈涛、周庭宇

中国移动通信集团公司：钱岭、齐骥、侯志强、石在辉、刘童桐、吴嘉

联想（北京）有限公司：于辰涛、王晟、陈志凌

成都四方伟业软件股份有限公司：刘俊良、周刚、王军、李今朝、李岳璘

北京东方金信科技有限公司：王伟哲、骆阳、石棋玲、戴桂春、张义水、李弘思

北京国脉互联信息顾问有限公司：孙泽红、邹惠

北京供销大数据集团股份有限公司：熊杰、杨慧

烽火通信科技股份有限公司：蓝海、邹素雯、董喆、熊威

中国建设银行：车春雷

石化盈科信息技术有限责任公司：蔡春久

北京数语科技有限公司：王瑋

前 言

党的十九大报告提出要“推动互联网、大数据、人工智能和实体经济深度融合”，进一步突出了大数据作为国家基础性战略性资源的重要地位，掌握丰富的高价值数据资源日益成为抢占未来发展主动权的前提和保障。

经过了 3 到 5 年的发展，大数据技术逐渐渗透到各行各业，以 Hadoop、Spark 为代表的大数据技术也慢慢为大家所熟悉，产业界的注意力回归到数据价值本身。数据成为资产，已经是行业共识。然而现实中，数据资产的管理和应用还处于摸索阶段，数据资产管理面临诸多挑战。一是大部分企业和政府部门的数据基础还很薄弱，存在数据标准混乱，数据质量层次不齐，各条块之间数据孤岛化严重等现象，阻碍了数据的共享应用。二是数据的应用空间亟待开拓，受限于数据规模和数据源种类的丰富程度，多数企业的数据应用刚刚起步，主要集中在精准营销，舆情感知和风险控制等有限场景，应用深度不够。三是数据的价值很难评估，数据资产的不确定性，导致了无法对数据的成本以及对业务的贡献进行评估，也就无法像运营有形资产一样管理数据资产。

国外的组织和机构长期在数据管理方面有着丰富的经验，形成了以国际数据管理协会（DAMA，Data Management Association International）、能力成熟度模型集成（CMMI，Capability Maturity Model Integration）为首的几大流派，但在数据资产管理方面也处于起步阶段，而且国外的数据管理的理论方法与国内实际情况有着很大鸿沟，不能完全直接套用，必须要进行专门的本地化。我国的金融和电信等行业，由于较早地开展信息化和大数据工作，已经积累了一定的数据

管理、治理乃至资产运营的经验，这些经验的整理和输出对于促进国内数据资产管理的发展有着重要意义。

为了促进数据资产管理的研究，中国信息通信研究院云计算与大数据研究所大数据部组织编写了《数据资产管理实践白皮书》。本白皮书分为四大部分，第一部分介绍了数据资产管理的内涵及大数据背景下的新特征，第二部分从实践的角度出发阐述了数据资产管理的主要内容。第三部分重点介绍了数据资产管理的实施步骤、工具平台和成功要素。最后结合实践经验，介绍了电信、金融、政务、医疗和工业等相关领域的数据资产管理案例。

本白皮书可以为政府和企业开展数据资产管理工作提供参考，也可以作为相关产品和服务提供商的参考依据。

由于时间仓促，水平所限，我们的工作还有很多不足。下一步，我们还将广泛采纳各方面意见建议，进一步深化相关研究，持续完善白皮书内容，在已有版本的基础上，适时修订发布新版。我们诚邀各界专家学者参与我们的研究工作，积极献言献策，共同完善国内数据资产管理理论和方法论体系，为促进大数据与实体经济深度融合做出积极贡献。如您有意愿，请联系我们：liuchengcheng@caict.ac.cn。

目 录

版权声明.....	1
前 言	2
图 表 目 录.....	6
一、 数据资产管理概述.....	1
(一) 数据资产管理的定义与内涵.....	1
1. 定义数据资产管理.....	1
2. 数据资产管理的内涵.....	2
3. 数据资产管理：数据管理 2.0.....	3
(二) 数据资产管理的重要性.....	4
(三) 数据资产管理的新特征.....	7
1. 数据对象纷繁复杂.....	8
2. 处理架构更新换代.....	8
3. 组织职能升级变迁.....	9
4. 管理手段自动智能.....	9
5. 应用范围不断扩大.....	9
二、 数据资产管理的主要内容.....	10
(一) 活动职能.....	10
1. 数据标准管理.....	11
2. 数据模型管理.....	12
3. 元数据管理.....	13
4. 主数据管理.....	15
5. 数据质量管理.....	16
6. 数据安全的管理.....	17
7. 数据价值管理.....	18
8. 数据共享管理.....	19
(二) 保障措施.....	20
1. 组织架构.....	20
2. 制度体系.....	22
三、 数据资产管理的实施要点.....	23
(一) 实施步骤.....	24
1. 第一阶段：统筹规划.....	24
2. 第二阶段：管理实施.....	25
3. 第三阶段：稽核检查.....	26
4. 第四阶段：资产运营.....	27
(二) 软件工具.....	28
1. 数据标准管理工具.....	28
2. 数据模型管理工具.....	29
3. 元数据管理工具.....	31
4. 主数据管理工具.....	32
5. 数据质量管理工具.....	33
6. 数据安全的管理工具.....	33
7. 数据生命周期管理工具.....	34

(三) 成功要素.....	36
1.明确责权利标, 有效推进管理.....	36
2.合理引进技术, 提升治理能力.....	36
3.着眼业务应用, 释放数据价值.....	37
4.加强数据合规, 注重风险风控.....	37
四、 总结与展望.....	39
附录: 数据资产管理的实践案例.....	41
(一) 中国电信集团公司大数据资产管理案例.....	41
(二) 中国移动省公司大数据资产管理案例.....	49
(三) 中国保信保险业务数据资产管理案例.....	55
(四) 基于数据模型的证券期货行业数据资产管理案例.....	60
(五) “云上贵州”政务数据资产管理案例.....	64
(六) 某省政务大数据资产管理案例.....	68
(七) 某大型医药公司数据资产管理案例.....	76
(八) 某大型钢铁集团公司数据资产管理案例.....	80
(九) 湖北省物价数据资产管理案例.....	84

图 表 目 录

表 1 数据资产管理组织架构角色职责.....	21
图 1 数据资产管理在大数据体系中的定位.....	2
图 2 数据资产管理四阶段实施步骤.....	3
图 3 大数据背景下的数据资产管理特点特征.....	7
图 4 数据资产管理体系架构.....	10
图 5 数据资产管理保障措施组织架构.....	20
图 6 数据认责机制.....	22
图 7 一种典型的制度体系架构.....	23
图 8 数据资产管理实施方法论.....	24
图 9 三段八步法.....	51
图 10 某省政务大数据资产汇聚架构图.....	70
图 11 某省政务数据汇聚主题模型.....	70
图 12 某省政务数据资产治理架构图.....	71
图 13 某省政务数据资产应用架构图.....	72
图 14 数据资产管理流程图.....	81
图 15 多元化数据集成.....	81
图 16 原始检测数据实时采集.....	82
图 17 钢材合格预测模型.....	82
图 18 物价大数据平台总体架构示意图.....	85

一、 数据资产管理概述

数据成为资产，已经是行业共识，甚至有人建议将数据计入资产负债表。但如果对比实物资产，对数据资产的管理，还处于非常原始的阶段。往往一个机构针对其数据资产类别和数量都缺乏全面了解，数据质量、数据安全、资产评估、资产交换交易等精细管理、价值挖掘和持续运营则更为薄弱。

数据资产管理是现阶段推动大数据与实体经济深度融合、新旧动能转换、经济转向高质量发展阶段的重要工作内容。本章将阐述数据资产管理的定义与内涵，分析数据资产管理在大数据中的重要意义，并对数据资产管理在大数据发展中的趋势进行分析。

(一) 数据资产管理的定义与内涵

1. 定义数据资产管理

数据资产(Data Asset)是指由企业拥有或者控制的，能够为企业带来未来经济利益的，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。

数据资产管理(DAM, Data Asset Management)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。数据资产管理是需要充分融合业务、技术和管理，来确保数据资产保值增值。

2. 数据资产管理的内涵

数据资产管理在大数据体系中的定位如图 1 所示，它位于应用和底层平台中间。数据资产管理包括两个方面，一是数据资产管理的核心活动职能，二是确保这些活动职能落地实施的保障措施，包括组织架构、制度体系。

数据资产管理在大数据应用体系中，处于承上启下的重要地位。对上支持以价值创造为导向的数据应用开发，对下依托大数据平台实现数据全生命周期的管理。



图 1 数据资产管理在大数据体系中的定位

数据资产管理贯穿数据采集、应用和价值实现等整个生命周期全过程。企业管理数据资产就是通过对数据的生命周期的管理，提高数据资产质量，促进数据在“内增值，外增效”两方面的价值变现。数据先被规范性定义、创建或获得，然后存储、维护和使用，最终被销毁。数据的生命周期开始于数据获取之前，企业先期制定数据规划、定义数据规范，以期获得实现数据采集、交付、存储和控制所需的技术能力。一般来说包括统筹规划、管理实施、稽核检查和资产运营四个主要阶段。



图 2 数据资产管理四阶段实施步骤

3. 数据资产管理：数据管理 2.0

数据管理的概念是伴随上世纪八十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。国际数据管理协会（DAMA，Data Management Association International）在 2009 年发布的数据管理知识体系 DMBOK1.0^①中定义为规划、控制和提供数据资产，发挥数据资产的价值。DAMA 数据管理体系将数据管理划分为 10 个领域，分别是数据治理、数据架构管理、数据开发、数据操作管理、数据安全、参考数据和主数据管理、数据仓库和商务智能管理、文档和内容管理、元数据管理和数据质量管理。2015 年，DAMA 在 DBMOK2.0 知识领域将其扩展为 11 个活动职能，分别是数据架构、数据模型与设计、数据存储与操作、数据安全、数据集成与互操作性、文件和内容、参考数据和主数据、数据仓库和商务智能（BI，Business Intelligence）、元数据、数据质量等。

在数据资产化背景下，数据资产管理是在数据管理基础上的进一步发展，可以视作数据管理的升级版。主要区别可以从三方面看：一是从数据管理变成数据资产管理，在数据资产管理的概念下，强调的是紧紧围绕着把数据作为一种资产，基于数据资产的价值、成本、收益开展全生命周期的管理。二是活动职能有所调整，和 2015 年 DAMA

^① The DAMA Guide to the Data Management Body of Knowledge

的活动职能相比，数据资产管理沿用数据模型、元数据、数据质量、参考数据和主数据、数据安全等内容，整合数据架构、数据存储与操作等内容，将数据标准管理、数据生命周期管理纳入活动职能，还针对当下应用场景、平台建设情况，将传统数据管理活动职能的具体内容进行了升级，增加了数据资产价值评估、数据资产运营流通两个活动职能。关于活动职能的描述详见第二章。三是管理要求有所升级，在“数据资源管理转向数据资产管理”的理念影响下，管理制度和组织架构也要有相应的变化，需要有更细致的管理制度和更专业的管理队伍来确保数据资产管理的流程性、严谨性和安全性。

此外，业界也经常使用“数据治理”、“数据管控”等说法。

（二）数据资产管理的重要性

提升数据质量、降低成本已经成为行业企业热点关注话题。如果不能对数据进行有效梳理及精细化管理，其价值就得不到很好体现，严重影响数据价值发挥，甚至会给运营管理带来负面作用。此外，日益全面、严格的监管措施和信息披露要求，也对企业数据提出了前所未有的挑战。这些痛点反映了数据资产管理的重要性，主要体现在以下几个方面：

痛点之一是缺乏统一数据标准。数据登记盘点流程缺乏统一的数据标准，无法有效避免数据混乱冲突、一数多源、多样多类等问题。统一标准是解决数据的关联能力，保障信息交互、数据流通、系统访问功能顺畅的必要前提。

痛点之二是数据周期规划混乱。对于部分企业来说，其内部数据的采集、传输、存储、应用、开放共享等全生命周期流程的各个环节的规划存在不合理现象。如收集数据时数据源用户处于不知情/非同意状态、违约超范围加工或未做到加工信息隔离等。

痛点之三是难以统筹业务管理。数据的增删、修改、使用等权限管理混乱，难以建立全面、准确、完整地反映企业运营状况的单一数据视图。数据需求、数据质量、数据应用等问题的管理和解决分散在不同业务和技术部门，没有一个清晰的协调机制和统一的数据管理渠道，业务不能及时、按需获得数据支持。

痛点之四是数据处理效率低下。数据采集、预处理等工作的周期较长，方法不够便捷，处理效率低下，无法快速挖掘整理出完善优质的数据属性供分析应用，需要提升开发及治理效率。

痛点之五是数据质量参差不齐。数据冗余、数据缺值、数据冲突等数据质量问题不能被及时发现和有效解决。需要建立规范的数据治理流程和考核机制等途径加以完善。

痛点之六是数据垃圾亟待解决。大量的历史留存冷数据无法被有效识别及处理，形成数据“包袱”。这些数据“包袱”很难变成数据“金矿”，又占用存储空间，浪费成本，造成损失。

痛点之七是安全监管势在必行。缺乏有效的数据安全管理机制，对敏感信息、隐私信息、保密信息的访问缺乏有效控制使其脱敏脱密合规，甚至对企业形成潜在的声誉和法律风险等。建立一个可靠的“数据加密保险箱”势在必行。

痛点之八是数据价值难以评估。数据评价体系以及数据资产化目前处于初级阶段，数据增值保值以及数据估值衡量问题亟待解决，可以说数据资产变现任重而道远。

数据作为越来越重要的生产要素，将成为比土地、石油、煤矿、劳动力等更为核心的生产材料，但是，实现数据资源向数据资本的转变还需要面对一系列的问题和挑战，数据资产管理正在成为企业赋能商业创新的具有影响力和战斗力的核心竞争领域。

数据资产管理除了对企业比较重要外，对于政府乃至国际层面都有重要的意义。可以说，构建科学的数据资产管理体系是个人层面、企业层面、社会层面、国家层面乃至国际层面都关注的热点话题之一。

个人层面，个人数据利用与数据安全保护之间需要合理的平衡，降低组织运营与合规方面的风险。大数据商业化应用中涉及的用户数据处理需要对用户隐私进行脱敏加密，以实现可控的隐私保护目标。

企业层面，数据资产价值的估算可以帮助企业更准确的掌握信息化投资收益，也是数据交易流通的前提之一。好的数据资产管理策略能有效规避风险，节约投入成本。

社会层面，受限于数据汇聚程度、数据规模和数据源种类的丰富程度，社会能够感知的数据应用场景较为单一，主要集中在精准营销，舆情感知和风险控制等有限场景，应用深度不够，应用空间，尤其是能够惠及大众的应用空间亟待开发。

国家层面，数据资产的运营流通需要国家层面的监管治理，合法合规性是数据运营流通的首要前提，是国家促进大数据发展，保障人民群众权益的关键纽带。

国际层面，数据资产管理知识体系涉及管理、技术等多个学

科，是一个非常复杂的系统工程，涉及很多技术难点和管理内容需要结合多方力量达成一致标准，国际化的共识机制是数据资产管理有效执行的重要前提和保障。国与国之间除了可以探讨标准内涵之外，还可以在合法合规的情况下有效进行数据传输，数据出入境问题需要法律、政策和规章指南的约束

(三) 数据资产管理的新特征

随着大数据的迅猛发展，数据资产管理相对传统的数据管理体现出明显的新特征，可以从数据对象、处理架构、组织职能、管理手段、应用范围五个方面来认识，如图 3 所示。

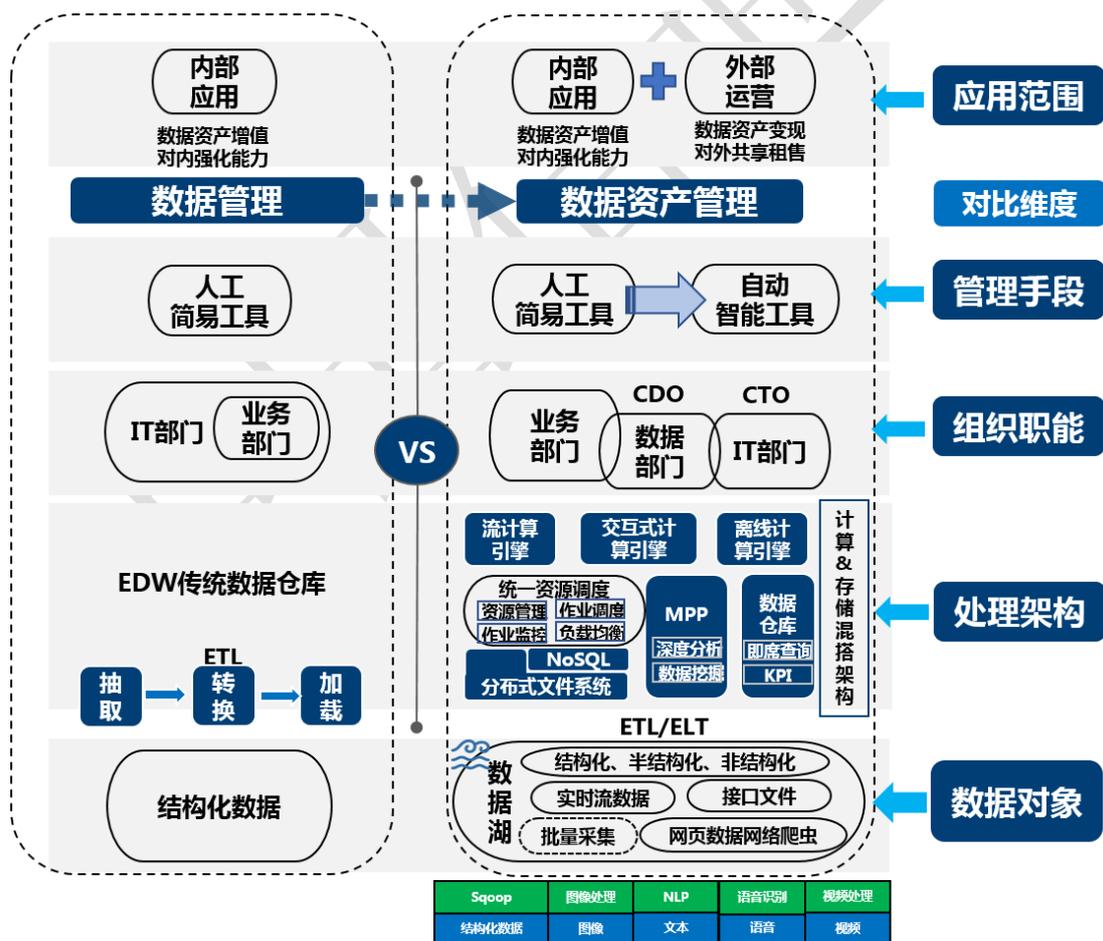


图 3 大数据背景下的数据资产管理特点特征

1. 数据对象纷繁复杂

数据作为数据资产管理的对象，体现出数据规模海量庞大、数据格式种类繁杂以及数据来源各式各样等特征。在数据量方面，单一机构的数据规模由以前的 GB 级上升到 TB 级，甚至 PB 级、EB 级，数据增速快。在数据格式种类方面，除传统的结构化数据之外，文本数据、图像数据、语音数据、视频数据等半结构化数据或非结构化数据占比越来越大，种类日益丰富。在数据来源方面，数据既包括内部数据，也包括来自第三方的外部数据，既包括传统业务处理采集的业务数据，也包括手机终端、传感器、机器设备、网站网络、日志等技术产生的数据。数据资产呈现数据对象海量、多样、多元化等特点。

2. 处理架构更新换代

处理架构的更新换代体现在两个方面。一方面是底层架构，数据处理的底层架构快速的向分布式系统迁移。以 Hadoop、Spark 等分布式技术和组件为核心的“计算&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以及灵活的业务需求。另一方面是数据的预处理流程正在从传统的 ETL 结构向 ELT 转变。传统的数据集成处理架构是 ETL 结构，这是构建数据仓库的重要一环，即用户从数据源抽取所需的数据，经过数据清洗，最终按照预先定义好的数据仓库模型，将数据加载到数据仓库中去。而大数据背景下的架构体系是 ELT 结构，其根据上层的应用需求，随时从数据湖^②中抽取数据建模分析。

^② 数据湖（Data Lake）：数据湖被定义为一种区别于原始数据仓库的数据存储架构，以原生格式保存大量的原始数据，可以包含结构化、非结构化数据，直到数据被使用的时候再进行处理。来源：《Big Data Requires a big new Architecture》,2011.

3. 组织职能升级变迁

传统的管理制度体系中，数据管理职能主要由 IT 部门来负责，是 IT 部门的一项工作，业务部门配合 IT 部门执行数据管理，提出需求。随着数据分析与业务融合越来越深入，业务部门逐步成为大数据应用的主角，因而数据资产管理在企业中扮演越来越重要的角色。出现了越来越多的企业设置专门的“数据管理”职能部门或首席数据官（CDO，Chief Data Officer）岗位。在这种变迁背景下，数据管理的组织架构也面临革新的需求。

4. 管理手段自动智能

依靠“手工人力”的电子表格数据治理模式即将被“自动智能”的“专业工具”取代，越来越多的数据管理员、业务分析师和数据领导者采用“平台工具”来梳理元数据、主数据，构建模型和管控质量。随着机器学习、深度学习技术的成熟，相关专项解决方案和平台工具系统的技术局限性如效率低、差错率高、扩展性差等将被一一攻破，能够有效地解放人力，提高效率和精度。

5. 应用范围不断扩大

数据的应用范围在不断扩大，由传统的支持管理需要的战略决策分析为主，发展为支撑业务一线业务场景的战术性决策。数据资产的意义价值也从对内强化能力扩展到了对外合作开放上，从而实现数据资产保值到增值的跨越。**战略决策分析**一般包括管理优化、研判决策、风险合规、业务拓展、管控成本等。由原来的只应用于领导决策

场景扩展到部门级业务分析使用。**战术性决策**包括智能推荐、精准营销、分析报告以及风险防范等。在跨企业的业务合作中，数据不可避免的也会流通到上下游的合作伙伴。从使用对象来看，数据资产的使用者不仅包括企业决策人员，还包括运维用户、业务管理人员、数据分析人员、数据科学家等各种角色。数据应用范围越广，就越需要智能化的数据资产管理技术及完备的管理制度作为支撑，从而实现数据精细化管控和数据价值最大化。

二、 数据资产管理的主要内容

数据资产管理框架如图 4 所示，包含 8 个活动职能和 2 个保障措施。活动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活动职能实现的一些辅助的组织架构和制度体系。本章主要描述具体的活动职能和保障措施的详细内容。



图 4 数据资产管理体系架构

（一）活动职能

数据资产管理的活动职能包括数据标准管理、数据模型管理、元数据管理、主数据管理、数据质量管理、数据安全管理和数据价值管理以及数据共享管理等 8 个方面，详细阐述如下。

1. 数据标准管理

数据标准是指保障数据的内外部使用和交换的一致性和准确性的规范性约束。**数据标准**一般包括三个要素：标准分类、标准信息项（标准内容）和相关公共代码（如国别代码、邮政编码）。数据标准通常可分为基础类数据标准和指标类数据标准。

基础类数据标准一般包括数据维度标准、主数据标准、逻辑数据模型标准、物理数据模型标准、元数据标准、公共代码标准等。**指标类数据标准**一般分为基础指标标准和计算指标（又称组合指标）标准。基础指标一般不含维度信息，且具有特定业务和经济含义，计算指标通常由两个以上基础指标计算得出。

数据标准管理是指数据标准的制定和实施的一系列活动，关键活动包括：

- 理解数据标准化需求；
- 构建数据标准体系和规范；
- 规划制定数据标准化的实施路线和方案；
- 制定数据标准管理办法和实施流程要求；
- 建设数据标准管理工具，推动数据标准的执行落地。
- 评估数据标准化工作的开展情况

数据标准管理的目标是通过统一的数据标准制定和发布，结合制度约束、系统控制等手段，实现企业大数据平台数据的完整性、有效性、一致性、规范性、开放性和共享性管理，为数据资产管理活动提供参考依据。

2. 数据模型管理

数据模型是现实世界数据特征的抽象，用于描述一组数据的概念和定义。数据模型从抽象层次上描述了数据的静态特征、动态行为和约束条件。数据模型所描述的内容有三部分：数据结构、数据操作（其中 ER 图数据模型中无数据操作）和数据约束，形成数据结构的基本蓝图，也是企业数据资产的战略地图。数据模型按不同的应用层次分成概念数据模型、逻辑数据模型、物理数据模型三种类型。

- **概念模型**：是一种面向用户、面向客观世界的模型，主要用来描述现实世界的概念化结构，与具体的数据库管理系统（DBMS，Database Management System）无关；
- **逻辑模型**：是一种以概念模型的框架为基础，根据业务条线、业务事项、业务流程、业务场景的需要，设计的面向业务实现的数据模型。逻辑模型可用于指导在不同的 DBMS 系统中实现。逻辑数据模型包括网状数据模型、层次数据模型等；
- **物理模型**：是一种面向计算机物理表示的模型，描述了数据在储存介质上的组织结构。物理模型的设计应基于逻辑模型的成果，以保证实现业务需求。它不但与具体的 DBMS 有关，而且还与操作系统和硬件有关，同时考虑系统性能的相关要求。

数据模型管理是指在信息系统设计时，参考业务模型，使用标准化用语、单词等数据要素来设计企业数据模型，并在信息系统建设和

运行维护过程中,严格按照数据模型管理制度,审核和管理新建数据模型,数据模型的标准化管理和统一管控,有利于指导企业数据整合,提高信息系统数据质量。数据模型管理包括对数据模型的设计、数据模型和数据标准词典的同步、数据模型审核发布、数据模型差异对比、版本管理等。数据模型管理的关键活动包括:

- 定义和分析企业数据需求;
- 定义标准化的业务用语、单词、域、编码等;
- 设计标准化数据模型;
- 制定数据模型管理办法和实施流程要求;
- 建设数据模型管理工具,统一管控企业数据模型。

数据模型是数据资产管理的基础,一个完整、可扩展、稳定的数据模型对于数据资产管理的成功起着重要的作用。通过数据模型管理可以清楚地表达企业内部各种业务主体之间的数据相关性,使不同部门的业务人员、应用开发人员和系统管理人员获得关于企业内部业务数据的统一完整视图。

3. 元数据管理

元数据 (Metadata) 是描述数据的数据。元数据按用途不同分为技术元数据、业务元数据和管理元数据。

- **技术元数据 (Technical Metadata):** 描述数据系统中技术领域相关概念、关系和规则的数据;包括数据平台内对象和数据结构的定义、源数据到目的数据的映射、数据转换的描述等;

- **业务元数据（Business Metadata）**：描述数据系统中业务领域相关概念、关系和规则的数据；包括业务术语、信息分类、指标、统计口径等；
- **管理元数据（Management Metadata）**：描述数据系统中管理领域相关概念、关系、规则的数据，主要包括人员角色、岗位职责、管理流程等信息。

元数据管理（Meta Data Management）是数据资产管理的重要基础，是为获得高质量的、整合的元数据而进行的规划、实施与控制行为。元数据管理的内容可以从以下六个角度进行概括，即“向前看”：“我”是谁加工出来的；“向后看”：“我”又支持了谁的加工；“看历史”：过去的“我”长什么样子；“看本体”：“我”的定义和格式是什么；“向上看”：“我”的父节点是谁；“向下看”：“我”的子节点是谁。元数据管理的关键活动包括：

- 理解企业元数据管理需求；
- 开发和维护元数据标准；
- 建设元数据管理工具；
- 创建、采集、整合元数据；
- 管理元数据存储库；
- 分发和使用元数据。
- 元数据分析（血缘分析、影响分析、数据地图等）

通过元数据管理活动，可以使企业数据信息的描述和分类实现格式统一，有助于理解数据的真实含义，为数据资源的管理和数据应用

奠定了基础。

4. 主数据管理

主数据 (Master Data) 是指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、共享应用于多个业务流程的、跨越各个业务部门、各个系统之间共享的、高价值的基础数据，是各业务应用和各系统之间进行信息交互的基础。从业务角度，主数据是相对“固定”的，变化缓慢。主数据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。例如供应商、客户、企业组织机构和员工、产品、渠道、科目 COA、BOM 等。

主数据管理 (MDM , Master Data Management) 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。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数据同步

- 方便修改、监控、更新关联系统主数据变化

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的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、跨系统数据融合应用。

5. 数据质量管理

数据质量是保证数据应用的基础。衡量数据质量的指标体系有很多，几个典型的指标有：完整性（数据是否缺失）、规范性（数据是否按照要求的规则存储）、一致性（数据的值是否存在信息含义上的冲突）、准确性（数据是否错误）、唯一性（数据是否是重复的）、时效性（数据是否按照时间的要求进行上传）。数据质量是描述数据价值含量的指标，就像铁矿石的质量，矿石的质量高，则炼出来的钢材就会多；反之，矿石的质量低，不但练出来的钢材少了，同时也增加了提炼的成本。

数据质量管理是指运用相关技术来衡量、提高和确保数据质量的规划、实施与控制等一系列活动。数据质量管理工作中的关键活动包括：

- 开发和提升数据质量意识；
- 定义数据质量需求；
- 剖析、分析和评估数据质量；
- 定义数据质量测量指标；
- 定义数据质量业务规则；
- 测试和验证数据质量需求；

- 确定与评估数据质量服务水平；
- 持续测量和监控数据质量；
- 管理数据质量问题；
- 分析产生数据质量问题的根本原因
- 制定数据质量改善方案
- 清洗和纠正数据质量缺陷；
- 设计并实施数据质量管理工具；
- 监控数据质量管理操作程序和绩效。

通过开展数据质量管理工作，企业可以获得干净、结构清晰的数据，是企业开发大数据产品、提供对外数据服务、发挥大数据价值的必要前提，也是企业开展数据资产管理的重要目标。

6. 数据安全 管理

数据安全 管理是指对数据设定安全等级，保证其被适当地使用。企业通过数据安全 管理，规划、开发和执行安全政策与措施，提供适当的身份以确认、授权、访问与审计等功能。

数据安全 管理的关键活动包括：

- 理解数据安全需求及监管要求；
- 定义业务敏感数据对象
- 定义数据安全策略；
- 定义数据安全标准
- 定义数据安全控制及措施；
- 管理用户、密码和用户组成员；

- 管理数据访问视图与权限；
- 监控用户身份认证和访问行为；
- 定义数据安全强度，划分信息等级；
- 部署数据安全防控系统或工具；
- 审计数据安全。

数据安全管理的目标是建立完善的体系化的安全策略措施，全方位进行安全管控，通过多种手段确保数据资产在“存、管、用”等各个环节中的安全，做到“事前可管、事中可控、事后可查”。

7. 数据价值管理

数据价值管理是对数据内在价值的度量，可以从数据成本和数据应用价值两方面来开展。**数据成本**一般包括采集获取和存储的费用（人工费用、IT设备等直接费用和间接费用等）和运维费用（业务操作费、技术操作费等）。**数据应用价值**主要考虑数据资产的分类、使用频次、使用对象、使用效果和共享流通等因素。

当前，对于数据资产评估的研究还处于早期阶段，评估方法手段还不成熟。可能的方法包括市场法、成本法和收益法三种。以收益法为例，将企业数据资产未来可能产生的收益折现为现金流进行计算。对数据资产价值的估算可以帮助企业更准确的掌握信息化投资收益，也是数据交易流通的前提之一。

进行数据价值管理的关键性活动包括：

- 确定企业数据集成度水平；
- 确定企业数据的应用场景；

- 计算数据在不同应用场景下的收益；
- 计算企业数据资产的总体价值。

8. 数据共享管理

数据共享管理主要是指开展数据共享和交换,实现数据外部价值的一系列活动。数据共享管理是指数据的所有者通过对数据的建模分析挖掘,把隐藏在海量数据中的符合共享开放层级的信息作为应用商品,以合规安全的形式完成共享交换或开放发布,使得数据具有流通属性,能方便供数据消费者使用。目前来看,拥有海量数据是企业开展数据资产运营的前提条件,在数据流通环境下,数据资产运营流通职能的服务对象包括了数据提供者、数据消费者、数据服务者和数据运营者四类角色。

数据共享管理的关键活动包括:

- 定义数据资产运营流通监控指标;
- 设计数据资产运营流通管理方案;
- 制定数据资产运营流通管理办法和实施流程要求;
- 监控数据资产运营指标;
- 监督落实数据流通等合规性管理要求;
- 分析运营流通指标,评价运营效果并改进。

重视数据资产管理、运营、流通可以为企业带来未来经济利益,同时这也是数据保值增值的重要手段。数据资产运营流通是使数据资产流动和发挥价值的核心,它将推动数据价值创造模式的不断创新,

从根本上改变企业管理、社会管理和政府治理的发展趋势。

（二）保障措施

数据资产管理是体系化非常强的工作，需要充分考虑企业内部 IT 系统、数据资源以及业务应用的开展现状，同时也要考虑围绕业务开展所设立的人员和组织机构的情况，在此基础上设计一套有针对性的数据资产管理组织架构、管理流程、管理机制和考核评估办法，通过管理的手段明确“责权利”以保障数据资产管理工作有序开展。保障措施包括组织架构和制度体系，本章将进行详细阐述。

1. 组织架构

典型的组织架构主要由数据资产管理委员会、数据资产管理中心和各业务部门构成。组织架构划分和角色设定如下图所示：

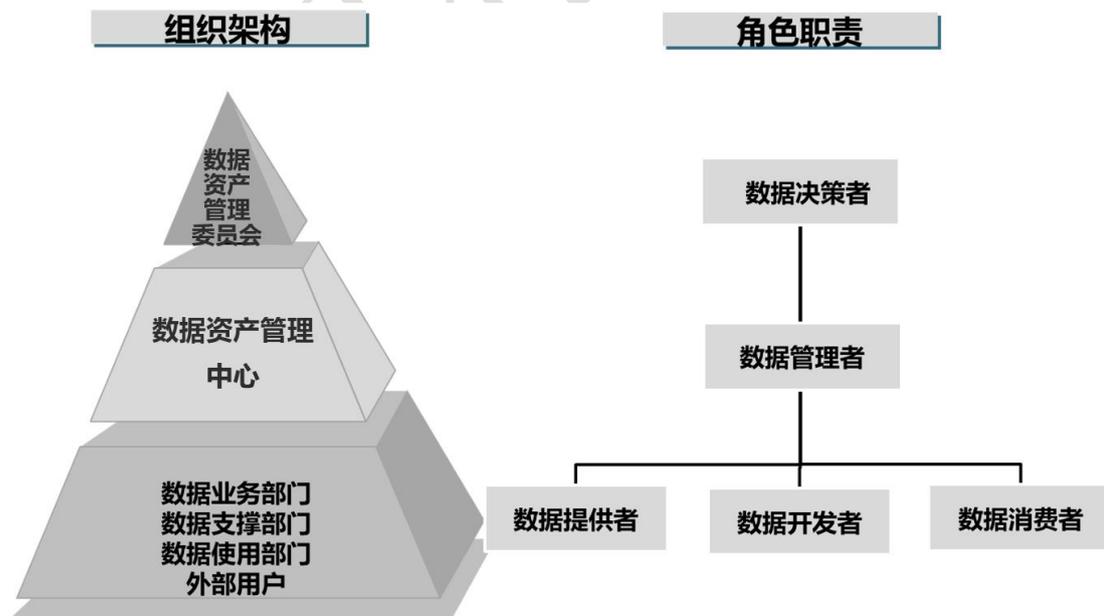


图 5 数据资产管理保障措施组织架构

为了让组织架构中的各个角色相互配合，各司其职，还需要明确

他们相应的职责，让工作职责融入到日常的数据资产管理和使用工作中。与上述典型的数据资产管理架构相适应的角色职责如下表。

表 1 数据资产管理组织架构角色职责

组织结构	角色	角色描述	角色主要职责
数据资产管理委员会	数据决策者	由公司主管领导和各业务部门领导组成	负责领导数据资产管理工作；决策数据资产管理重大工作内容和方向。在数据角色方出现问题时负责仲裁。
数据资产管理中心	数据管理者	数据管理中心机构的平台运营人员	负责牵头制定数据资产管理的政策、标准、规则、流程，协调认责冲突；监督各项数据规则和规范的约束的落实情况；负责数据资产管理平台中整体数据的管控流程制定和平台功能系统支撑的实施；负责数据平台的整体运营、组织、协调。
各业务/技术部门	数据提供者	相关数据所有人和权限管理人员	配合制定相关数据标准、数据制度和规则；遵守和执行数据标准管控相关的流程，根据数据标准要求提供相关数据规范。作为数据出现质量问题时的主要责任者。
	数据开发者	数据开发人员	负责数据开发，有责任执行数据标准和数据质量内容，负责从技术角度解决数据质量问题。作为数据出现质量问题时的次要责任者。
	数据消费者	数据使用人员，包含内部用户和外部用户	作为数据资产管理平台数据的使用者，负责反馈数据效果，作为数据资产管理平台数据闭环流程的发起人。

数据认责是数据资产管理在服务各领域、各环节工作落到实处的有效手段，通过数据角色职责开展数据认责相关工作，其主要认责流程如图 6 所示。具体认责条例、管理办法及相关制度流程由数据资产管理委员会进行制定。

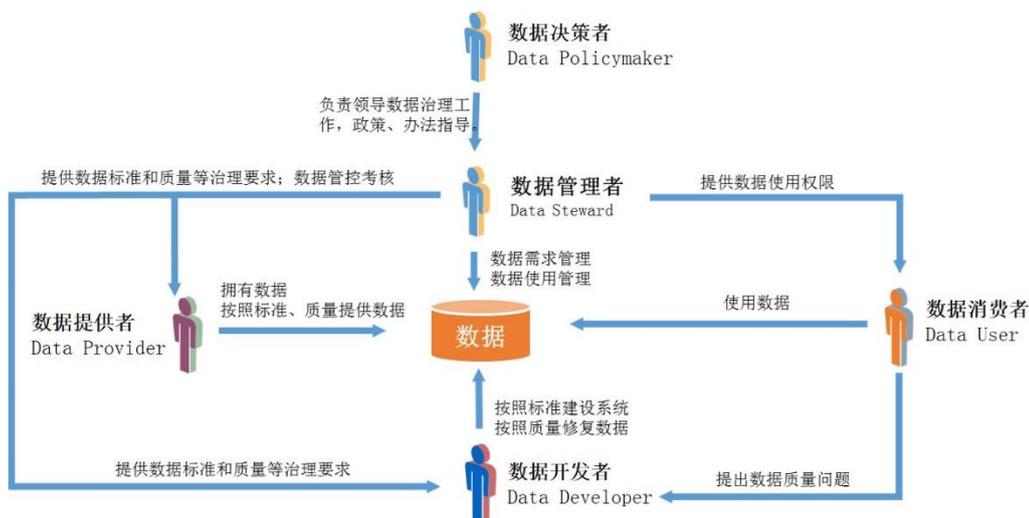


图 6 数据认责机制

2. 制度体系

为了保障活动实施和组织架构正常运转，需要建立一套覆盖数据引入、使用、开放等整个生产运营过程的数据管理规范，从制度上保障数据资产管理工作有据、可行、可控。

数据资产管理规范包括元数据管理规范、生命周期管理规范、数据质量管理规范以及数据安全规范等对应活动职能的具体规范。在此基础上，规范需细化至接口设计、接口开发、模型设计、模型开发、数据开放以及服务封装等内容。规范的标准一般包括基础分类标准、命名规范要求、数据架构划分、存储与数据权限规则、元数据信息完整性要求等。规范和标准在执行的过程中执行监控规定，要求事中检查和事后监控。事中检查指的是在开发和上线时进行控制，包括命名规范，信息完整性，合理性等；事后监控指的是对存储周期，数据安全敏感信息和加密信息，权限赋权常态化检查。图 7 是可参考的一种典型的制度体系架构举例。



图 7 一种典型的制度体系架构

三、 数据资产管理的实施要点

完整的企业或机构大数据能力的构建步骤一般是“建立组织架构→应用需求梳理→数据盘点梳理→引进平台技术→汇聚多源数据→治理数据→数据应用→数据运营”等。数据资产管理以数据价值为导向，分布在大数据能力构建的多个环节。本章将主要围绕数据资产管理，具体阐述实施步骤、主要工具平台的功能，并基于实践经验，提出数据资产管理成功的要素。数据成熟度不同的企业或单位开展数据资产管理的具体步骤和实施内容要根据自身情况制定。

（一）实施步骤

数据资产管理可按照“统筹规划→管理实施→稽核检查→资产运营”四个阶段的方法策略执行，每个阶段对应的活动职能如图 8 所示。以业务应用目标为指引，企业可以按照自身数据及管理情况制定不同的实施步骤顺序。

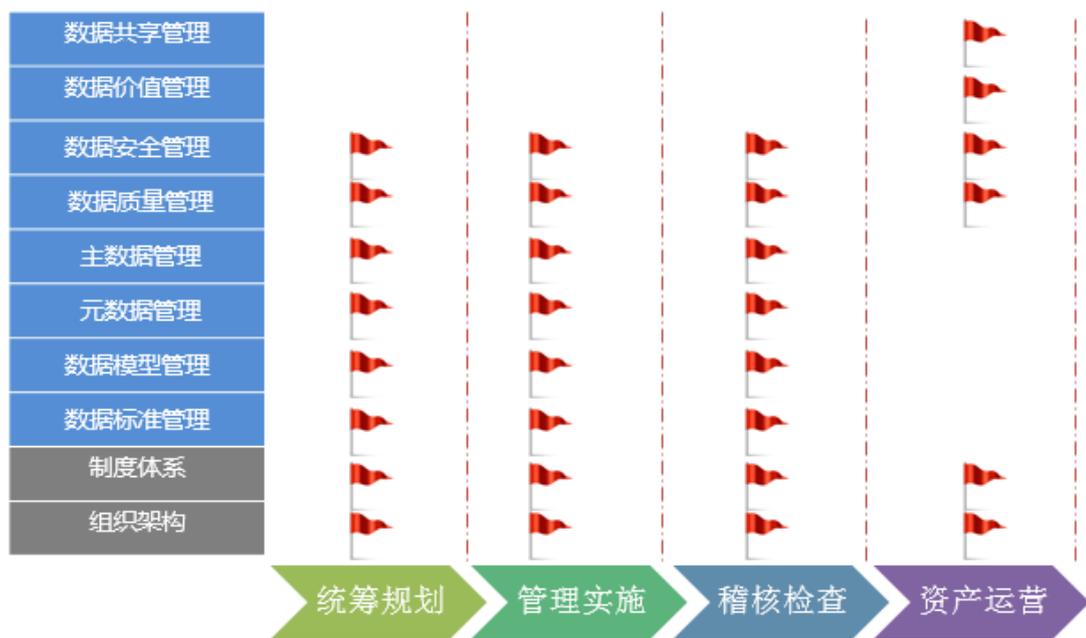


图 8 数据资产管理实施方法论

1. 第一阶段：统筹规划

第一阶段是统筹规划过程，涉及建立数据资产管理组织和制度作为保障措施，盘点数据资产，制定数据资产标准规范等，该阶段成果是后续工作的基础。

一般情况下，数据资产管理的**第一步**是建立组织责任体系，根据自身情况，制定数据资产管理制度规范。需要建立一套独立完整的关于数据资产管理的组织机构，明确各级角色和职责，确定兼职专职人员，保障数据资产管理的各项管理办法、工作流程的实施，推进工作

的有序开展，并逐步打造管理及技术的专业人才团队。

第二步是结合业务盘点数据资产。对基础数据的盘点是开展数据资产管理工作的前提之一，需要分析企业战略及业务现状，结合当前大数据现状及未来发展，盘点企业内外部数据现状，确立数据资产管理的目标，并逐渐实施需求调研、盘点资产、采集汇聚等专题任务。与此同时，了解企业数据来源、数据采集手段和硬件设备情况，以定位自身数据资产管理能力，规划未来数据资产管理成熟度提升方案。

第三步是制定数据资产相关的标准规范。标准规范包括元数据标准、核心业务指标数据标准、业务系统数据模型标准、主数据标准、关键业务稽核规则等，使得数据管理人员在工作中有明确的规则可依，同时，需要对企业数据资产设计信息化管理方式，规划设计企业的数据仓库、元数据库、主数据库等。

2. 第二阶段：管理实施

如果说第一阶段重点还在于对数据资产的定义、规划、梳理，第二阶段就是对第一阶段成果的落地实施。**首先**，在搭建大数据管理平台、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资产管理平台或引入第三方工具以支撑管理工作，切实建立起企业数据资产管理能力。**其次**，要建立安全管理体系，防范数据安全隐患，执行数据安全活动职能。**再次**，还需要制定和管理主数据，以明确企业核心业务实体的数据，如客户、合作伙伴、员工、产品、物料单、账户等，从而自动、准确、及时地分发和分析整个企业中的数据，并对数据进行验证。

在第二阶段里，需要从数据资产管理的相关业务、技术部门日常工作流程入手，切实建立起企业数据资产管控能力，包括从业务角度梳理企业数据质量规则，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。针对关键性数据资产管理工作，可以借助管理工具，建立数据资产的管理流程，保证相关事情都有专人负责。第二阶段的工作目标主要是为企业打造核心的管理数据资产的能力，同时为企业内数据资产管理部门形成数据管理的工作环境，概括起来，就是企业数据资产可管理、可落地。

3. 第三阶段：稽核检查

稽核检查阶段是保障数据资产管理实施阶段涉及各活动职能有效落地执行的重要一环。这个阶段包括检查数据标准执行情况、稽核数据质量、监管数据生命周期等具体任务。

这个阶段需要抓好三个“常态化”。

一是数据标准执行情况检查的常态化。数据标准管理是企业数据资产管理的基础性工作，通过数据标准管理的实施，企业可实现对大数据平台全网数据的统一运营管理。

二是数据质量稽核的常态化。应对数据质量问题，首先要提升数据质量意识，数据质量意识包括能够将数据质量问题与其实质影响联系起来，同时传达一种“数据质量问题不能仅仅依靠技术手段解决”的理念。其次为数据质量建立一系列流程和程序。最后，应当数据质量管理是循环管理过程，其终极目标是通过可靠的数据提升数据在使用中的价值，提高工作效率，并最终为企业赢得经济效益。

三是灵活配置数据存储策略的常态化。数据生命周期管理，其目标是以完全支持企业业务目标和服务水平的需求，根据数据对企业的价值进行分类分级，形成数据资产目录，然后制定相应的策略，确定最优服务水平和最低成本，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级别的保护、复制和恢复。借助数据生命周期管理，企业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。

4. 第四阶段：资产运营

通过前三个阶段，企业已经能够建立基本的数据资产管理能力，在此基础上，还需要具备以实现业务价值为导向，以用户为中心，为企业内外部不同层面用户提供数据价值的的能力。资产运营阶段是数据资产管理实现价值的最终阶段，该阶段包括开展数据资产价值评估、数据资产运营流通等。

数据资产价值评估能够以合理的方式管理内部数据和提供对外服务。在大数据时代，数据运营企业关于数据价值的实现是体现在数据分析、数据交易层面。数据资产作为一种无形资产，其公允价值的计量应当考虑市场参与者通过最佳使用资产或将其出售给最佳使用该项资产的其他市场参与者而创造经济利益的能力。只有对数据资产价值进行合理的评估，才能以更合理的方式管理内部数据和提供数据对外服务。

数据资产运营流通需要加强管理运营手段和方式方法，促进数据

资产对内支撑业务应用，对外形成数据服务能力，打造数据资产综合运营能力。数据资产运营流通主要是实现数据资产价值的社会化，需要从数据安全及管理合规性、数据资产成本及价值创造、组织结构优化、数据质量提升等方面进行规划并不断迭代，持续优化数据资产管理能力。

（二）软件工具

数据资产管理实践实施过程中，需要依托具体的软件工具来执行。而且随着技术的发展，软件工具的自动化、智能化程度不断地提高，在数据资产管理中的作用越来越大。目前针对上述活动职能，业界很多厂商都开发了相关软件工具，其中，相对比较成熟的工具有数据标准管理工具、数据模型管理工具、元数据管理工具、主数据管理工具、数据质量管理工具、数据安全工具和数据生命周期管理工具等七类工具，这七类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软件平台，其具体意义和主要功能将展开具体阐述。除“数据资产管理”相关的工具之外，在大数据能力构建中，一般还要利用“数据集成工具”、“数据共享交换平台”等，通过传统数据仓库或大数据平台等媒介将数据集成交换到一起，从而为应用分析或开放做准备，涉及工具如“商务智能（BI）分析工具”、“报表工具”、“数据挖掘平台”、“用户行为分析平台”、“数据开放平台”等。

1. 数据标准管理工具

数据标准制定及维护工具可以规范数据资产格式、命名的准确性

和口径的一致性,该工具针对数据标准管理职能而开发,需具备以下基础功能:

- 标准生成:可按照业务领域、业务主题、信息分类、信息项等生成标准细则;
- 标准映射:可以将制定的标准与实际数据进行关联映射,即实现数据标准的落地执行,维护标准与元数据之间的落地映射关系,包括元数据与数据标准的映射、元数据与数据质量的映射,以及数据标准和数据质量的映射,能提供在线的手工映射配置功能,并能对映射结果做页面展示;
- 变更查询:是查询发布或废止的标准的变更轨迹;
- 映射查询:是查询标准项与元数据之间的落地情况并提供下载功能;
- 维护标准:是指对标准状态进行管理,包括增删改、审核、定版、发布、废止等;
- 标准版本查询:是指对发布状态的标准进行版本管理;
- 标准导出:是指按照当前系统中发布的最新标准或者选择版本来下载标准信息。

2. 数据模型管理工具

针对企业在不同业务发展阶段建设的一个个竖井式系统,最大的挑战莫过于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对企业数据模型的管理、比对、分析、展示提供技术支撑，需要提供统一、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。该工具针对数据模型管理职能而开发，需具备以下基础功能：

- 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，能够将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；
- 模型差异稽核：提供数据模型与应用数据库之间自动数据模型审核、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质量；
- 数据模型变更管控：支持数据模型变更管控过程，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、流程化变更管理。同时，实现各系统数据模型版本化管理，自动生成版本号、版本变更明细信息，可以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以及数据模型设计变更的需求来由，实现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；

- 模型可视化：支持将管理的数据模型 E-R 图（实体关系图）转换为图片、数据建模脚本（DDL）等可视化展示形式，方便数据模型管理人员以全局视角监控系统中各类数据实体结构及实体间关系。

3. 元数据管理工具

元数据管理工具可以了解数据资产分布及产生过程，该工具针对元数据管理活动职能而开发，需具备以下基础功能：

- 元数据采集：能够适应异构环境，支持从传统关系型数据库和大数据平台中采集从数据产生系统到数据加工处理系统到数据应用报表系统的全量元数据，包括过程中的数据实体（系统、库、表、字段的描述）以及数据实体加工处理过程中的逻辑；
- 元数据展示：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，展示数据实体间的组合、依赖关系，以及数据实体加工处理上下游的逻辑关系；
- 元数据应用：元数据的应用一般包括数据地图，数据的血缘、影响分析，全链分析等；
- 元数据搜索：可根据数据源库、类型等搜索元数据信息。

4. 主数据管理工具

数据管理工具用来定义、管理和共享企业主数据信息，可通过数据整合工具（如 ETL）或专门的主数据管理工具来实施主数据管理，具有企业级主数据存储、整合、清洗、监管以及分发等五大功能，并保证这些主数据在各个信息系统间的准确性、一致性、完整性。简单说来，存储、整合是数据的“入口”，分发给数据的“出口”，而中间的清洗与监管将担负起数据质量提升的重要任务。该工具针对主数据管理职能而开发，需具备以下基础功能：

- 主数据存储、整合：实现主数据整合、清洗、校验、合并等功能，根据企业业务规则和企业数据质量标准对收集到的主数据进行加工和处理，用于提取分散在各个支撑系统中的主数据集中到主数据存储库，合并和维护唯一、完整、准确的主数据信息；
- 主数据管理：支持对企业主数据的操作维护，包括主数据申请与校验、审批、变更、冻结/解冻、发布、归档等全生命周期管理；
- 主数据分析：实现对主数据的变更情况监控，为主数据系统管理员提供对主数据进行分析、优化、统计、比较等功能；
- 主数据分发与共享：实现主数据对外查询和分发服务，前者用于在其它系统发出针对主数据实时响应类查询请

求时，返回所需数据，后者则用于提供批量数据分发服务，一般采用企业服务总线（ESB 工具）实现方式。

5. 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，针对数据质量管理职能而开发，需具备以下基础功能：

- 质量需求管理：对数据使用过程中产生的问题进行收集、存储、分类并提供查询检索功能，为质量规则的制定提供依据；
- 规则设置：能够提供稽核规则设置功能，用于设置一个稽核规则应用于哪类数据；
- 规则校验：能够对所关注的的数据执行数据质量规则的校验任务；
- 任务管理：能够提供稽核任务调度功能，指定稽核任务周期执行；
- 报告生成：能够对校验结果的质量问题进行记录，积累形成问题知识库，并生成报告，在此基础上，能够根据检核结果，生成对问题数据的质量提高建议，并可直接操作修改数据。

6. 数据安全管理工作具

数据安全管理工作具是结合信息安全的技术手段保证数据资产使用和交换共享过程中的安全。数据管理人员开展数据安全管理，是指

执行数据安全政策和措施，为数据和信息提供适当的认证、授权、访问和审计，以防范可能的数据安全隐患。需具备以下基础功能：

- 数据获取安全：能够支持数据获取需要经过申请与审批流程，保障数据获取安全；
- 数据脱敏：能够支持数据脱敏规则、脱敏算法及脱敏任务的管理及应用，一般情况下，脱敏方式有动态脱敏和静态脱敏两种；
- 统一认证：定义数据安全策略，定义用户组设立和密码标准等；
- 租户隔离：管理用户，密码，用户组和权限；
- 角色授权：划分信息等级，使用密级分类模式，对企业数据和信息产品进行分类；
- 日志审计：审计数据安全，监控用户身份认证和访问行为，支持经常性分析。

7. 数据生命周期管理工具

此外，业界还设计了数据生命周期管理工具，它将协助数据管理组织和人员，在数据从产生、应用、归档到销毁的整个生命周期全过程进行有效的自动化管理，针对不同生命周期阶段的数据建立多层次的数据自动存储体系，既有效管控系统整体在线数据规模，降低系统运营成本，又满足最终用户的数据访问和应用的需求。一般需具备以下基础功能：

- 数据需求分析：通过数据库或者数据平台的各种数据分布分析和访问状态分析，协助数据管理人员对数据生命周期管理策略，有效发现和挖掘当前数据平台或者数据库中历史数据增长最快的关键数据，同时，为管理业务部门需求，满足业务部门对数据使用的要求提供有效的数据化支撑；
- 策略管理：通过可视化的方式，由数据管理员在线建立清晰、合理、完整，可操作性强的数据生命周期管理基线和规范，并且以此为基准实现全企业的数据生命周期管理策略发布和需求管理，为企业建立一个行之有效的数据生命周期变更管理和变更审查机制；
- 自动化数据管理：根据在工具中建立的数据生命周期管理策略，实现数据在异构平台（例如从 oracle -> mySQL, HBase 等）中的全自动化迁移、归档、清理，建立一个多层次的自动化数据分级存储技术体系，实现数据生命周期管理策略实施的自动化、标准化和规范化；
- 其他辅助能力：工具需要支持数据校验能力，保持迁移和清理前后的数据一致性和有效性；迁移前后，支持详细的数据生命周期管理报告，让数据管理员清晰了解企业中各种历史数据的状态；平台支持水平扩展，数据处理能力可以通过增加集群结点方式进行水平扩展。

（三）成功要素

1.明确责权利标，有效推进管理

数据资产管理最重要的成功要素之一就是重视组织管理的作用，将责权利清晰化，逐步建立健全包括管理型人才和技术性人才的适应数据发展的人才结构，减少工作推进阻碍。并注重数据标准化环节以保障信息体系不发生混乱，确保数据规范一致性。数据标准是数据资产管理的基础，是对数据资产进行准确定义的过程。对于一个拥有大量数据资产的企业，或者是要实现数据资产交易的企业而言，构建数据标准是一件必须要做的事情。标准化是解决数据的关联能力，保障信息的交互、流动、系统可访问，提高数据活化能力。保障信息体系不发生混乱，确保数据规范一致性——避免数据混乱、冲突、多样、一数多源。数据资产管理的核心目的是有效综合运营数据以服务企业，让数据成为利润中心的一部分，这离不开管理，更离不开技术。

2.合理引进技术，提升治理能力

人工智能、物联网、新一代移动通信、智能制造、空天一体化网络、量子计算、机器学习、深度学习、图像处理、自然语言处理、4k高清、知识图谱、类脑计算、区块链、虚拟现实、增强现实等前沿技术正在大数据的推动下蓬勃发展。然而，在实现数据资产管理的过程中，应根据自身实际情况，避免盲从，合理引进创新技术以提高数据挖掘准确性和挖掘效率，节省人力成本。信息时代万物数化，企业拥有数据的规模、活性以及收集、运用数据的能力，决定其核心竞争力。

掌控数据，就可以支配市场，意味着巨大的投资回报，数据是企业的核心资产。数据在实现价值的过程中需要充分依托技术，但更离不开结合自身业务与应用，合理规划。大数据和云计算的建立与开放至关重要，可以帮助企业梳理数据内容，高效检索展示，最终给企业带来一定的经济收益和社会效应。但其应用的成功与否还是要取决于企业自身商业模式的建立，以数据融合技术为战略资产的商业模式，可以决定企业未来。

3.着眼业务应用，释放数据价值

数据资产化进程给各类企业带来重生、颠覆和创新，企业应重点关注、顺势而为，建立起符合自身业务和数据特点的数据资产化体系和能力，数据资产管理人员不能只陷于数据资产管理工作，还应紧密联系业务，只有明确了前端业务需求，才能做到数据资产管理过程中的有的放矢，张弛有度。数据的价值体现在决策精准、敏锐洞察，数据资产管理能够使管理具流程化、规范化，结合业务应用的数据资产管理不仅使数据保值增值，还将会给企业带来更加巨大的经济效益和社会效益。

4.加强数据合规，注重风险风控

在数据资产管理的过程中，综合考虑困难及挑战，并全面管控风险，要基于行业模型、行业标准等积累完整、准确的内外部数据以保证数据合规性，进而规避风险。数据资产管理是一项持之以恒的工作，不可能一蹴而就，需要一个循序渐进的过程分阶段进行。要做好充分

地长期作战准备，就一定要加强数据合规操作，避免安全漏洞，及时风险风控。

中国信息通信研究院

四、 总结与展望

“数据之于本世纪，就像石油之于上世纪：它是发展和改变的动力。数据已经产生了新的基础设施、商业领域、垄断机构、政治理论，最关键的是，还产生了一种新经济。数据信息不像过去的其他资源，它采用不同的方式提取、加工、估值和交易。它改变了市场规则，要求使用新的管理方式。”^③

数据作为日益重要的战略资源，需要完善的管理体系。为此，中国信息通信研究院云计算与大数据研究所组织成员单位编写了《数据资产管理实践白皮书》，阐述了大数据背景下的数据资产管理概念和特征，描述了数据资产管理的主要活动职能和保障措施，提炼了实施数据资产管理的主要步骤。白皮书的目的是澄清概念，梳理经验形成方法论，希望引起业界对数据资产管理工作的重视，为数据资产管理研究和实践提供参考。

我们也认识到，数据资产管理知识体系涉及管理、技术等多个学科，是一个非常复杂的系统工程，相关工作在国内刚刚起步，理论还不完善，也缺乏广泛的实践基础。虽然国外 DAMA、CMMI 等机构近几年也提出了数据管理的理论框架，对我们开展数据资产管理研究有一定的借鉴意义。但数据资产管理，除了借鉴国际经验之外，还需要深入结合信息化基础、大数据应用所处阶段、应用和组织架构特点，在理论和实践上开展创新。

^③ 数据是未来的石油：数据如何推动新经济增长，《经济学人》，2017年5月，<https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy>

由于时间仓促，水平所限，我们的工作还有很多不足。下一步，我们还将广泛采纳各方面意见建议，进一步深化相关研究，持续完善白皮书内容，在已有版本的基础上，适时修订发布新版。我们诚邀各界专家学者参与我们的研究工作，积极献言献策，共同完善国内数据资产管理理论和方法论体系，为促进大数据与实体经济深度融合做出积极贡献。

中国信息通信研究院

附录：数据资产管理的实践案例

数据资产管理知识体系涉及管理、技术、金融等多个学科领域，是一个非常复杂的系统工程，本白皮书涉及的知识内容及理论体系尚不成熟，还有很多问题和缺陷有待完善，目前处于一个非常初级的阶段。

为了促进经验交流，本白皮书还征集了数据中心联盟成员内相关专家企业主导或参与的应用案例供业界参考。案例介绍包括基本情况、痛点难点、实施步骤、应用效果以及洞察启示。联盟还将继续征集更多的典型案例补充到未来的版本中。

（一）中国电信集团公司大数据资产管理案例

【数据情况】

中国电信集团公司是我国特大型国有通信企业，在全国 31 个省（区、市）和美洲、欧洲、香港、澳门等地设有分支机构，拥有覆盖全国城乡、通达世界各地的通信信息服务网络。中国电信运营着世界上规模最大的固定网、宽带网和 CDMA 网络，骨干网升级流量达到 22T，出口带宽 760G，光网覆盖国内领先并延伸至全球。同时，中国电信拥有过亿固话、宽带和移动用户。

因此，中国电信在进行全集团范围的数据资产管理工作之前，数据分布于全国 31 省 300 余地市之间。数据整体情况具有数据分散、规模大、数据种类繁多等特征。且各省/市之间大多采用 IT 系统自建的方式开展信息化建设工作，所以各平台的数据规范也各不相同。

【痛点难点】

中国电信业务触达全国各地、网元达千万级，且由于各/省市 IT 系统均独立自建，导致业务数据和基础数据种类繁多，各地数据质量参差不齐。再加上当时各级分公司大数据意识淡薄，使得各家数据相互独立，形成数据孤岛，数据融合困难，难以最大限度的发挥数据价值。

1.数据分散、规模大、种类繁多、质量差

中国电信全集团的数据类型包括 M 域、O 域、B 域、网络信息、位置信令、产品、电渠、行业、专业公司、基地等 30 大类 200 子类数据。在进行大数据资产管理之前，经初步统计各省/市、专业公司自建的 IT 系统大约超过 500 个，且各类数据分散在全国 31 个省 300 余地级市、千万级网元之间，导致数据种类繁多，数据规范多样，且各地数据质量参差不齐。

2.缺乏规范管理，各方利益割据，数据变现困难

大数据资产管理初期，集团对数据问题缺乏规范的统一化管理，故存在各省市、平台间的数据规范各不相同，技术标准不统一，管理机制不健全，信息安全和用户隐私保护风险大等诸多问题。再加上当时各级分公司尚未建立完备的大数据意识，导致各方利益割据，数据呈烟囱状形成数据孤岛，很难最大限度的释放数据价值实现数据变现。

【实践路径】

1.设立大数据专业化管理机构，建设两级数据保障责任体系

针对以上问题，中国电信从 2013 年开始部署大数据工作，该项

工作作为一把手工程需要自上而下统一思想，所以我们在数据资产管理前期，首先设立了大数据的专业机构，以及各类保障体系、流程和规章制度，让大数据工作“有人管理，有章可循”。

从整体上来说，数据资产管理工作分为集团和省/专业公司两大数据保障责任体系。在集团方面，为更好的经营大数据，建立政策和环境保障，首先成立了数据管理委员，由委员会来把握全集团大数据发展方向、决策重大事项。并下设专业管理部门，对数据资产管理和大数据运营工作进行统筹管理，其中包括业务产品规划、管理规范制定、需求应用统筹以及对外经营管控等工作。其中，云计算分公司作为大数据集约运营单位，负责整个电信所有对外经营的集约大数据产品的开发、销售以及运营工作。而在二级保障体系中，各省/专业公司主要负责汇聚本省/本公司各类数据以及保障数据质量。

2. 盘点数据资产，统一数据标准，输出数据资产目录

在有了专业机构后，我们对全网数据进行了深度盘点，并统一各类数据标准，其中包括用户基本数据、行为数据、位置数据、终端数据等数据。进而输出了整个中国电信的数据资产目录，对数据资产进行归类、分层管理，使数据更便于沟通、交流、展示和使用。经过3年的数据汇聚工作，目前已汇聚了“云管端”共计18类数据，数据接入总量达30PB，记录总条数达370万亿余条，并且还在以日均200TB的速度增长。

3. 构建“条线汇聚，一点集中”的数据采集汇聚体系

我们将盘点后的数据资产按照数据来源和属性共分为IT数据、

网络数据、增值业务数据和创新数据四大类。如上所述，这些数据分布在各个省公司和专业公司之间，且平台分布也很广泛，所以我们按照数据的属性对不同的数据类型采取了不同的采集策略。最终将采集到的数据集中存储在集团大数据汇聚平台以供后续使用。

4.建设大数据能力平台，整合与提升大数据平台服务能力

根据业务数据量、数据存储周期、数据保存副本、数据存储压缩以及业务处理分析的需求，评估了大数据平台的建设规模。根据评估后的规模进行集群设备选型、网络设计、基础环境准备，完成集群基础环境的部署。最后，合理规划建设了成熟稳定的大数据能力平台。

大数据能力平台采用 2+31+X 的全国资源布局，共计 4000+节点，日数据实时分析处理能力达到 200TB。大数据能力平台的建设帮助实现数据的规范管理、安全存储以及高效便捷的数据分析处理需求，是提升数据价值、实现数据开放共享的基础，更实现了对内支撑全网数据资产管理，对外服务外部客户应用的基本目标。

5.打造完整数据资产管理体系，深挖数据价值，形成数据能力

通过数据预处理、数据能力沉淀和数据封装三项工作的流转配合，打造完整的数据资产管理体系。即对原始数据进行对账、合规评估、数据清洗、数据水印、数据脱敏、数据 ETL 六项工作，使数据成为可用的清单级数据。再将清单级数据进行逻辑关联，最终形成大数据宽表体系，通过数据封装为大数据宽表体系开发服务接口，最后通过各

类融合加工形成数据产品。整个数据生产流程通过前向稽核、后向治理两个策略实现了对数据的治理，使数据具备对外提供服务，对内支撑运营的能力。与此同时，实现数据质量的实时监控与数据服务能力的统一，最终为数据开放共享、最大化数据价值提供基础。

6. 制定开放共享策略，保障数据使用安全，促进行业数据融合

数据能力形成后，根据数据属性以及对数据的需求，制定数据开放共享机制，并围绕共享开放进一步梳理数据资源，制定共享开放数据目录，明确各类数据的共享开放形式和对象，形成体系化、可操作的数据开放共享管理办法。让数据开放共享工作变得有章可依，形成常态化的工作机制保障。

制定“五分一统”的数据开放共享策略，通过分类（明确数据种类，分类开放）、分级（评估数据价值，明确开放级别）、分形（明确数据共享开放形式）、分步（评估数据需求，确定开放序列）、分权（契约开放，信用开放）、一统（统一出口）的战略思想，明确各类数据的共享开放形式和对象，针对数据开放共享涉及到的各个环节，采用相应的安全保障手段，夯实数据开放共享的安全基础。

7. 夯实运营基础，建设综合运营体系

实现大数据体系良性运转，需要强大的运营能力支撑。从而，中国电信围绕客户需求，建设了综合运营体系，以数据运营、平台运营和产品运营为基本核心，不断优化和更新中国电信大数据产品的基础能力以及产品模型，使数据发挥最大的价值，同时实现用户价值最大

化。

其中，数据运营通过建立需求管理、血缘管理、质量监控、价值管理四大工作流程持续建设改进数据综合能力，并在此基础上巩固数据资产目录以及数据血缘管理，使数据有据可依，有源可溯；平台运营则从资源编排、作业调度、质量监控和价值管理四方面对平台运营工作进行整体设计，确保平台稳定、高效地处理业务，进而成为产品运营和数据运营背后最有力的支撑；大数据产品运营以客户需求为核心，将客户服务、产品监控、运营分析和产品优化四项工作不断循环、优化、提升，从对外使用对内运维等方面保障产品的持续迭代和提供更好的服务，协助客户实现数据和产品价值。

【应用效果】

1. 盘点全集团数据资产，形成数据资产目录

中国电信大数据资产涵盖了电信运营服务中所产生的各类基本数据及通信娱乐、交互行为、位置轨迹等数据，数据资产具有中立、横跨移固两网、多屏幕、高价值密度、基础信息完善、天然位置关联等优势。现已汇聚了包含“云管端”共计 4 门 18 类数据，数据日接入量达 200TB，数据累计接入量 30PB。

2. 搭建能力平台，坐拥六大优势

中国电信大数据能力开放平台共有 4000+ 计算节点，日数据处理能力达 200TB。拥有海量数据分析和处理能力、快速开发和部署能力、资源隔离和数据安全管理能力、数据生产过程调度和管控能力、资源动态管控能力和数据及服务开放能力，可有力支撑政府和企业级的数

据进行分析处理及开放共享。

3. 自主研发数据管理工具

中国电信在已有的数据资产管理经验上不断总结提炼,耗费大量时间精力自主研发了 6 大数据资产管理能力和 8 类专业数据资产管理工具。如数据稽核系统对已经接入的数据进行稽核,保持数据的完整性、一致性、准确性、时效性。数据资产运营管理系统对已经接入的数据进行盘点,确认数据的基本信息、归属关系、来源系统、传输方式、生产周期等。

4. 形成 4+1 产品体系, 服务于各行业

已形成风险防控、精准营销、区域洞察、智慧运营、大数据云 PaaS 五个系列 16 个子产品,覆盖金融、旅游、交通、政务、地产、数字营销等多个领域。

【洞察启示】

1. 正确认识大数据

人们总是期望从大数据中挖掘出意想不到的“大价值”,实际上大数据的价值主要体现在它的驱动效应,即带动有关的科研和产业发展,提高各行各业通过数据分析解决困难问题和增值的能力。

大数据可以提高结论的可靠性,但不是 100%的反应结果,正确认识大数据的价值和利益,面对海量、异构、动态变化的数据发起挑战,才能挖掘大数据中潜藏的价值。

2. 建立大数据意识

长久以来,数据意识的缺失导致中国电信的数据资产管理能力十

分孱弱，建立高效、便捷的数据资产管理体系成为提升大数据能力的重中之重。

中国电信通过建立 6 大能力、8 类系统，细化数据采集需求，设计数据采集策略，规范数据采集进度，打造了完整、高效的数据资产管理体系。

数据意识不仅仅是数据价值，更应注意到数据资产管理不当所产生的危害，认识数据的局限性、对数据出处、采集、处理有一定的把握力，建立正确的数据意识，实现对数据正确的使用。

3.把握大数据趋势

大数据技术的发展如火如荼，在各个领域都得到了广泛的应用，而且就其目前的发展情况来看，大数据技术具有十分良好的发展前景。数据分析、数据处理、数据平台等无疑将成为未来大数据发展的核心。

大数据的特点也是有目共睹的，在互联网、云计算发展的推动下，大数据技术未来的发展会朝着更加智能化、先进化、广泛化的方向发展，从而为人们的生活带来更大的便利，为科技的发展注入新的活力。

(二) 中国移动省公司大数据资产管理案例

【数据情况】

中国移动某省大数据整体建设处于发展阶段，现网数据平台存在“竖井式建设”、“数据资产分散管理”、“应用开发成本高”、“安全保障能力分散”等诸多问题。面向未来发展，省公司决定建设全省统一的企业级大数据平台，初期融合 B 域和 O 域数据，并为上层应用提供数据服务、数据存储计算资源、分析挖掘能力等，总共近 300 节点，由中国移动苏州研发中心（简称“苏研”）负责承建，系统采用“两域四层”技术架构，采用 Hadoop、MPP、传统数据库和流处理融合方式支撑上层百花齐放的应用。数据资产管理是其中重要一环，重点从五方面着手，分别为：数据标准化管理、元数据管理、数据质量管理、数据资产管理、数据安全治理。

数据情况如下：

B 域数据包括：经分系统数据，如详单、账单、工单和用户数据；终端库数据，如终端属性数据、IMEI 与终端型号对应关系等。

O 域数据包括：2G/3G 信令、VOLTE 信令、LTE 信令、软采数据、宽带数据、网管数据、测量测试数据、无线数据等

数据规模较大，每天汇总的原始数据达到近 60TB，Hadoop 平台的数据达到近 2PB，MPP 数据达到近 200TB；

【痛点难点】

我省在大数据资产管理工作方面的的痛点难点主要表现为：

- 组织机制：缺乏数据资产管理组织架构，需要明确各级

角色和职责，保障数据资产管理的各项管理办法、工作流程的实施，推送数据资产管理工作的有序开展。

- 数据管控流程：跨域跨专业数据管控流程尚未建立、跨域数据管理技术手段尚不完善，数据质量有时存在“一事一议”现象，效率不高。
- 数据分散：省内数据尚未实现充分整合，无法支撑跨域数据分析应用，比如面向客户体验的4G网络质量保障，需要B域和O域数据融合。
- 信息烟囱：目前数据分散在各域近20多套系统中，未有效整合，数据价值分散。
- 技术标准不统一：各系统建设年代、承建公司不同，存在技术标准不统一问题
- 安全保障：大数据资产管理安全管控和资源监控需要统一纳入到公司级4A系统。并制定数据全周期的安全管控措施和技术保障手段。

【实践路径】

针对现状，苏研按照“三段八步法”制定完整的数据资产管理实施路径。



图 9 三段八步法

第一阶段：规划

第一步：数据资产管理咨询

分析企业战略及业务现状、分析大数据现状及未来发展，分析数据现状，确立数据资产管理目标。

第二步：组织责任体系架构

明确各级角色和职责，保障数据资产管理的各项管理办法、工作流程的实施，推送数据资产管理工作的有序开展。

第二阶段：治理

第三步：梳理数据，建立标准。包括字段标准、元数据模型、数据质量标准、数据隐私字段等。

第四步：搭建平台。实施上线大数据平台（苏研 BC-Hadoop）、ETL（苏研 BC-ETL）、数据资产管理系统（苏研 BC-DM）。

第五步：治理实施。包括数据采集、处理、存储、元数据、规则、监控、安全等各项数据资产管理工作配置及实施。

第六步：数据资产化。包括数据资产生命周期管理、数据资产价

值评估、数据资产分级处置、数据资产视图、资产服务目录等。

第三阶段：长效

第七步：数据资产运营流通。加强管理运营手段和方式方法，促进数据资产对内支撑业务应用，对外形成数据服务能力，打造数据资产综合运营能力。

第八步：能力成熟度评估与优化。从数据风险管理及合规性、数据资产价值创造、组织结构优化、数据质量管理等方面不断优化数据资产管理能力。

【应用效果】

系统上线以来，取得了良好的管理和治理效果，主要表现为：

- 建立了完善的组织架构，工作流程、角色职责清晰，数据资产管理工作井然有序。
- 数据标准化管理方面，制定大数据平台数据标准化管理规范，开发前端界面，建立从标准数据的录入、申请、审批、修改、发布、稽核的闭环管理。截止目前为止已导入和使用近4000个标准字段。
- 元数据管理方面，实现了事前元数据管理；提供血缘/影响分析功能，并在统一运维管理中提供基于元数据的运维故障程序的血缘分析、便于问题定位。支持 Hive, HBase, GBase 等数据仓库，并支持多种业务元数据属性标签，目前已管理元数据模型达到近1500个。
- 数据质量管理方面，实现数据质量规则的增、删、改、

查,并通过规则完成数据质量稽核,生成质量报告。同时支持数据质量分段管控,围绕从源数据接入大数据平台到应用输出的全过程开展数据质量监控,可以实时掌握数据质量情况。

- 数据资产管理方面。实现了资产全生命周期管理、资产地图、资产价值评估及运行审计、数据资产服务接口等。已经形成数据资产综合运营能力。

- 数据安全,全面实现了数据访问安全、数据服务安全、数据存储安全、平台组件安全、业务流程安全。具备完整的 ETL 过程加密、HDFS 透明加密、字段级、表级、用户级、目录和文件级的隔离、数字水印、数据脱敏、安全令牌管理、服务攻击检测等特性。

【洞察启示】

- 重视组织的价值,责权利清晰化,数据资产管理工作才能减少阻碍,推进得当。

- 数据标准管理是数据管理建设中的重要环节,是衡量企业级省大数据平台数据资产运营和管理水平的评估依据。数据标准规划、数据标准化实施、数据标准支撑工作都要到位。

- 通过技术创新,推动自动化、智能化、安全化的大数据资产管理方式,引入机器学习和人工智能提升传统数据资产管理技术能力。

- 数据资产的综合运营是数据资产管理的核心目的,也是让数据资产管理成为利润中心的根本,对数据资产评估以及

数据资产的流通使用需要从管理、技术两方面强化。

- 数据管理产品要具备广泛兼容性。

中国信息通信研究院

(三) 中国保信保险业务数据资产管理案例

【数据情况】

中国保险信息技术管理有限责任公司(简称中国保信)是经国务院批准,由中国保险监督管理委员会管理。公司以服务行业发展、支持保险监管、保护保险消费、辅助社会治理为使命,主要职责是了解保险行业信息交互共享需求,建设和运营集中统一、设计科学、功能完善、安全高效的保险业数据信息共享和对外交互平台。保险业数据信息共享平台是保险业的公共基础设施和综合性服务平台,将涵盖保险主要业务领域的的数据信息,全面建成后能够提供保险生产支持、信息查询、保单登记和认证、结算等综合服务。

【痛点难点】

保险业务要素数据规范是行业信息共享的基础保障,是行业数据资产管理和数据资产管理的关键内容;统一保险数据语义,是整个保险行业大数据工作科学健康开展的先决条件。保险业务开展过程中,保险业务要素数据规范等数据资产管理规范的欠缺,导致关键业务信息产生环节不清晰、范围不明确、内容不规范;制约行业数据质量、安全及共享水平;掣肘保险监管的精细化与科学化;难以支撑巨灾风险及保险信用信息数据库建设;影响保险反欺诈管理、灾害风险防控与疏散;无法客观体现保险在社会治理中的广泛作用。

因此,亟需开展保险业务要素数据规范化工作。在保监会统计信息部直接指导下,中国保信公司负责承建研制保险业务要素数据规范,实现大数据时代保险行业的“书同文、数同轨”的目标。

【实践路径】

作为一个行业级数据资产管理和各机构数据资产管理的基础设施，保险业务要素数据规范化参考行业现行规范，结合保险监管需求、保单登记平台需求及保险公司业务实际，运用 BOR 方法，合理界定规范范围、设计数据模型、制定数据规范，确保规范的可扩展性、可持续性 & 可落地性。具体的工作过程为如下。

1. 全面调研行业现状，力求数据规范研制贴近实务、贴近需求、贴近应用。

- 全面调研现行监管和业务需求
- 全面调阅保险业务流程和数据字典
- 全面梳理业务模型和业务代码

2. 扎实开展研制工作，力求数据规范研制组织专业、定位准确、设计合理。

- 行业专家全面参与
- 反复论证寻找最大公约数
- 合理设计规范架构

3. 认真征求各方意见，不断调试数据规范内容，力求数据规范能用、有用、好用。

- 征求 200 余家保险机构意见
- 面向保监会系统广泛征求意见
- 结合再保应用组织专门测试

其中，数据调研和规范研制的核心技术在于数据资产管理的

BOR 框架体系，该体系确保数据规范体系严谨、与业务活动紧密对接。BOR 方法即借鉴相关行业的成熟实践，结合保险业务实际，通过梳理保险核心业务活动(Behavior)、提炼业务对象(Object)及其属性、总结业务对象的相互关系(Relationship)，从而确定保险行业数据规范的范围、结构和内容。具体的梳理方法包括层层递进和聚类耦合等。

【应用效果】

1. 成果

- 形成首批通用数据规范和专项数据规范，指导保险业实际应用。
- 确定建立“2+N”的规范体系框架，即形成人身险和财产险等 2 个跨险种的基础数据规范，支持跨险种的公共应用。在此基础上，针对各主要险种的特点，扩展形成分险种的专项数据规范。
- 建立主题域、数据实体、数据项三层架构，组织数据规范的有机统一。

2. 应用价值

- 对于监管部门：对保险业务经营评估更全面、对保险行业风险识别更有效、对保险公司服务评价更客观；
- 对于行业平台：从入口把关数据质量、提高数据整合效率、推动数据价值输出。
- 对于保险公司：统一保险行业规范用语、降低行业信息共享成本、支持保险公司系统建设。

- 对于消费者：更准确理解保险条款、更容易比较保险产品。

【洞察启示】

当前，各保险机构积极拥抱大数据，在客户营销和风险控制等领域都取得了企业局部的进展，但在提升行业核心竞争力和抗风险能力、有效整合和利用车联网、信用等线上线下信息、构建完整客户知识图谱实现精准营销和反欺诈等方面一直存在全行业数据共享和协同的瓶颈，保险业的新一轮转型和健康发展亟需在行业级数据基础设施层面破解“数不同轨、用不随心”，大数据知易行难、创新乏力的局面，真正实现大数据在创新、管理和监管方面的大价值。没有全面的行业信息共享不足以谈行业大数据，没有数据资产管理的“规矩”也无法形成数据价值这一“方圆”。数据成为具有系统性价值的企业资产，直到进一步可以形成数据产品对外服务、产生价值，数据资产管理是基础成功因素和必要条件。各保险机构应拥抱行业数据大格局和保监会的这一保险行业大数据的大行动，做三件事：

- 积极参与保险业务要素数据规范的推广与落地，共同推动行业数据协同和安全共享。
- 吸收和学习 BOR 方法论，优化现有数据模型架构规范，审视外部数据合作与内部数据和业务战略的关联，完善自身数据资产管理的组织、流程、平台。

- 充分利用保险业务要素数据规范的成果，加速机构自身数据资产管理和数据资产管理成熟度的提高，积极探索数据资产价值评估和对外变现定价的方法和途径。

中国信息通信研究院

（四）基于数据模型的证券期货行业数据资产管理案例

【数据情况】

现代证券期货行业是依赖信息化程度最高的行业之一，IT 系统运行产生的数据是证券期货行业最核心的资产。中国证券期货行业涉及涉及机构类型多样、数量庞大：参与市场运行的核心组织机构有近 20 家；证券、期货、基金、私募经纪公司等各类注册经营机构有数千家；交易的标的物品种繁多、交易量大，日均交易量在 3.5 亿笔左右；交易方式多样、变化迅速、结构相对复杂，一笔交易需要经过多家主体机构才能完成运行流转。

证券期货行业产生的数据量大（截止到 2016 年底，国内证券期货行业行业结构化数据量已经达到了 PB 级）。但与“通常意义上”的大数据低价值密度、快速处理、非结构化为主相比，证券期货业大数据结构化程度高、数据质量好、附加值高，并且数据需要在需要在多个企业间进行数据交换，形成主营业务闭环，对数据交换要求极高的频率。

证监会作为证券期货行业的监管部门，对证券行业的良性发展起到了不可或缺的作用。在数据方面，证监会一直致力于推进数据标准和治理工作，旨在通过建立统一的框架，指导、规范整个行业数据交换与统计。证券庞大的机构群、巨大的数据量、高结构化、相对复杂的数据交换环境使证券期货行业的数据资产管理工作面临巨大挑战。

【痛点难点】

在目前的国际数据资产管理体系中,各国际标准组织及数据管理组织对数据资产管理的概念、框架、目标要求都已有明确定义,但是在应用与成果理论两个环节,还没有深入和具体的方法论。简而言之,如何将抽象、粗线条的数据资产管理概念和框架具体应用于行业级场景仍缺乏系统性的理论和实践探索。

如前所述,中国证券期货行业机构繁多、交易多变,基础数据定义冲突较多;数据交互相对复杂但跨机构、甚至机构内部的通用程度较差,数据交换接口标准定义各异;在业务协同、监管等领域缺乏清晰可见的行业数据关系图谱。可见,整个行业急需系统化的数据资产管理,以统一明确的数据模型,描述整个证券期货市场业务、数据情况,明确数据定义、规范数据交换、指导行业系统建设。中国证监会在吸收国际先进标准的基础上针对行业应用这一空白研制了中国证券期货行业数据资产管理方法论,建立了以数据模型为核心的行业数据资产管理体系。

【实践路径】

证监会行业数据资产管理以行业数据模型为核心,规范数据定义,服务行业标准化。

证券期货行业数据模型的建设首先是解读行业规则,即以证券期货行业相关法律法规、业务规则、制度及流程等为依据,梳理行业顶层数据流图及机构内数据流图,识别行业数据的现状。

进一步以“交易”、“监管”、“披露”三大业务线条为切入点,分别完成对行业中各种业务的全面遍历,最终形成一系列有关联关系的数

据项和数据表。

以行业现状的识别、行业规则的提取为基础，根据模型的应用深度的不同，形成用于规范行业标准化的抽象模型和指导行业数据应用建设的逻辑模型，最终完成证券期货行业的数据模型建设。

【应用效果】

证券期货行业数据模型已用于规范行业各类标准制定，指导行业各机构应用系统建设，乃至为行业数据大集中项目提供权威数据标准。

1. 规范行业标准化

建立一套基于行业数据模型的行业标准编制、审核机制、关联查询方式，通过行业标准审核，规范、统一在行业中交换数据的含义和口径。

2. 支持核心应用系统建设

运用不同视角的逻辑模型，直接指导行业系统建设，特别是数据仓库类的数据库应用系统建设，规范机构间进而到机构内部系统间的数据定义，提升全行业数据标准化程度。

3. 指导智能数据挖掘方法论

建立智能数据挖掘方法论。通过数据模型中的业务流程和实体关系图，形成数据挖掘地图，开展不依赖于业务需要的工程化数据挖掘，遍历特定业务方向上的数据特征，全面提取业务间的关系列表，提高数据挖掘准确性和挖掘效率。

4. 在行业范围内建立数据资产管理实践

以证券期货行业数据模型为基础，开展行业数据模型治理工作，

梳理证券公司逻辑模型，形成指导证券公司开展数据仓库建设及其他应用系统建设的最佳实践。

5.整合行业数据资源，提升数据利用价值

在建立行业数据资产管理最佳实践的过程中，借助行业力量，整合行业数据资源，使最佳实践更具备行业通用性，更具有行业指导作用，有效提升数据利用价值。

6.加强数据合规、风险和隐私

在数据资产管理过程中，综合考虑全面风险管理，基于行业模型、行业标准等积累完整、准确的内外部数据，更好地服务于风险识别、计量、评估、监测和报告。

【洞察启示】

中国证监会推动的基于数据模型的行业数据资产管理是对DAMA“数据资产行使权力和控制的活动集合”这一数据管理活动抽象定义的具体落地。各行业主管部门、大型公共事业、以及大型企业的管理层应清晰的认识到数据资产管理对于行业数字化发展、行业监管、以及大数据战略落地的意义，高度重视数据资产管理的工作，结合自身业务战略，对照各自行业数据特点、数据管理成熟度、以及未来行业数字化趋势，学习借鉴证监会的最佳实践，以对行业、企业未来和社公众的高度责任心、科学的方法、和严谨务实的态度开展数据资产管理，形成一把手带队、组织机构健全、职责清晰到人、嵌入绩效、自上而下与自下而上结合的数据资产管理文化。

（五）“云上贵州”政务数据资产管理案例

【企业基本情况和数据情况】

2014 年开始，云上贵州大数据产业发展有限公司自主搭建了全城统一的政府平台——云上贵州系统平台，采用国内具有自主知识产权的云操作系统，包括国产的服务器，是全国首个实现全省政府数据统筹存取和共享、统筹标准和统筹安全的云计算系统平台，现在此平台已经被中央网信办确定为全国关键的技术基础设施，并列入中央考核检查范围。平台架构分为三层，在基础设施层主要提供云计算、云存储和云安全服务以及数据服务。在核心层，重点打造全省统一的数据共享交换平台。在应用层主要通过政务网与互联网向全省的公务员、公民和企业提供物联网服务。

【痛点和难点】

APP 上线过多，对公民来说是一个沉重负担，贵州统筹了全省的 APP，不允许各市州自建 APP，分城市频道、部门频道。整合各家政务民生服务，通过统一入口、统一用户认证体系、统一消息推送体系，打造全省统一的政府服务 APP 平台。为公民提供涵盖医疗、教育、交通、生活缴费等多领域的一站式便民服务，聚合了 1976 项服务，148 项服务可网上办结。通过创新电子政务发展方式，整合各类政府应用和数据资源，充分发挥“互联网+政务服务”的价值。

【实践路径】

贵州发展大数据前期在聚、关键在通、目的在用。

1.在“聚”方面

从 2014 年开始，用了四年时间逐渐迁移到云上。针对国家大数据建设要求全省的政府应用系统全部要迁入到云上贵州系统平台上，贵州省大数据中心利用数据架构职能，从 2015 年开始，凡是省级财政新建的信息化系统必须基于同一个云平台进行开发，每个部门的信息化项目在立项之前，方案必须拿到省大数据局做前置审批，不允许单独建服务器、交换器，要基于云的架构、项目，购买云服务来解决信息化建设的问题。截止到 2017 年 5 月底，平台共集聚政府部门 612 个应用系统，220T 的数据。

2.在“通”方面

在今年数博会期间，云上贵州大数据产业发展有限公司与国家信息中心正式签订了协议并进行了两次对接，贵州将率先进入国家重要平台，国家要求贵州先整合省内的共享交换平台，再统一接入到国家平台。

在数据资产管理和数据标准管理方面，在 2015 年开始要求各个部门对数据进行梳理时，却发现缺乏标准。为了数据资产管理的有效性和便捷性，在工信部建交院的指导下，云上贵州大数据产业发展有限公司发布了四个地方标准，即：《政府数据分类分级指南》、《政府数据元数据核心元数据的描述和规范》、《资源目录编制指南》、《政府数据的脱敏指南》要求各级政府部门用这四个标准做数据梳理。数据

标准化管理对数据资产管理中的数据梳理、数据共享和数据开放等都起到了事半功倍的效果，到目前为止省级部门总共梳理 16000 多个数据目录，在平台上共享的数据集有 253 个，共享数据量已经到达 4500000 条。数据开放平台是去年 9 月 30 号上线的，目前已经有 538 项可下载使用、可进行二次开发的政府绿色公共数据资源，API 类型数据达到 129 个。

【应用效果】

1. 全省公共资源交易中心

它把互联互通平台和交易平台全部放置在云上，同时对数据进行建模、加工与分析。过去创标、围标、预约的情况比较多，经过分析后，发现有些标投了很多次，但从来没有中过一次标，而且有的公司会绑定在一起去投标，针对此类问题，用主数据管理和数据模型管理职能，进行数据画像和建模，发现了原来管理上不易发现的问题，最终交易时间能够缩短到五天，效率提高六倍以上。

2. 工商“五证合一”

依托云上贵州数据共享交换实现工商、国税、地税、人社、质监及统计六个部门业务数据实时共享，实现“一照一码”业务数据多跑路，企业少跑腿，为企业注册带来极大的便利。

3. 数据铁笼

数据铁笼即用数据编织规范制约权力的“笼子”，实现数据留痕、权力可视、规范运行，有效防止“权力任性”。利用数据资产流通运营的职能，贵州省各级政府统一采用由省政务中心开发的网上办事大厅，

所有数据都集中在“云上贵州系统云平台”，通过数据汇聚以及数据共享，发挥大数据的应用，取得了很好的效果，并且贵州网上办事大厅被国务院办公厅列为示范。通过采用省公安系统大数据平台，两枪案件破案率同比上升 20.62%，八类案件破案率同比上升 12.99%。贵阳公安块数据指挥中心对全市人、事、物进行实时三维立体画像，实现社会治安立体化管控，改变了传统人工干预的方式。

4.精准扶贫大数据平台

精准扶贫大数据平台利用数据质量管理职能，通过各部门数据共享，统一数据标准，统一数据格式，实现了全省扶贫数据的“通”和“准”。数据共享方面已实现省扶贫办、省公安厅、省卫计委、省水库移民局、省民政厅、省教育厅、省人社厅、省住建厅、省工商局、省水利厅、省国土厅等 11 家省级单位的数据融通。功能方面已完成“扫一扫”、“对象查询”、“数据统计”、“动态异常预警”、“对象关注”等应用。

【洞察启示】

云上贵州系统平台利用数据资产管理的各个职能，实现全省政府数据统筹存取和共享、统筹标准和统筹安全的云计算系统平台。数据一旦被很好地管理，能够为“我”所用，能让决策力度更强、洞悉发现事物的力度更敏锐以及让信息资产处理更加符合流程化、规范化并其将会带来具大的经济效益和社会价值。

（六）某省政务大数据资产管理案例

【企业基本情况和数据情况】

某省政务大数据资产管理平台通过推进政务数据全面整合、治理和共享，打通信息孤岛，建立基础设施资源池、公共数据资源池、公共服务资源池，打破独立单一系统应用壁垒，重塑数据归集渠道和模式，凝聚纷繁散落的碎片数据，采用国际领先的大数据清洗存储管理核心技术，利用集约化建设策略，结合互联网、云计算和可视化技术，打造一个开放的政务大数据公共服务平台。

【痛点和难点】

政务大数据公共服务平台的建设需要从全省相互独立的多个业务系统中抽取数据，这些业务系统并没有经过顶层设计和统筹规划，他们之间的数据必然存在不标准、不完整、重复、错误、不一致的问题。

当前数据开放的力度还仅限于政府内部机构，尚未形成对社会团体、企业、社会公众的政务数据开放机制。形成政府信息公开领域还存在着两个“不对称”，一是公众对于信息公开的需求和政府公开信息的供给存在明显的不对称，二是政府要求百姓提供的信息和政府自身能向百姓公开的信息不对称。存在数据相对封闭、开放力度不足、开放机制尚无形成的问题。

大数据的发展是支撑电子政务向智慧政务转变的重要基础，但受省内各委办厅局自身数据基础和信息化建设水平现状的影响，由各委办厅局独立的开展本业务内智慧政务的发展建设，一方面存在数据和

底层技术不足无法完全支撑上层应用的问题，一方面也不可避免的形成一定的建设重复情况。独立烟囱式的建设也影响了构建整体性政府、透明化政府和服务型政务的进程。急需通过政务大数据底层支撑平台的建设，为各业务部门行业智慧政务应用的发展夯实基础，提供技术和数据支撑。

【实践路径】

智慧型政府迫切需要从三个方面得到提升：一是充分应用基础设施服务云，二是扎实构建政务大数据资源中心，三是全面打造智慧型政府创新服务，这是一个由基础云、到数据云、到服务云的全面升级的过程，这三朵云聚合在一起，就是政务大数据资产管理体系。

可以从汇聚、治理和应用三个方面统筹考虑、统筹实现。

首先是政务数据资产的汇聚。为了充分实现政务数据的共享与数据资产的管理，通过构建三大资源池，即基础设施资源池、公共数据资源池、公共服务资源池为政府搭建统一的大数据共享平台提供全面支持，实现数据的汇聚。

- 基础设施资源池：充分应用基础设施服务云，搭建大数据平台分布式运行环境。
- 公共数据资源池：整合各部门、各系统、各种类型的数据资产，让数据变成有用的信息。
- 公共服务资源池：基于整合的信息，全面提供各部门各类应用服务。

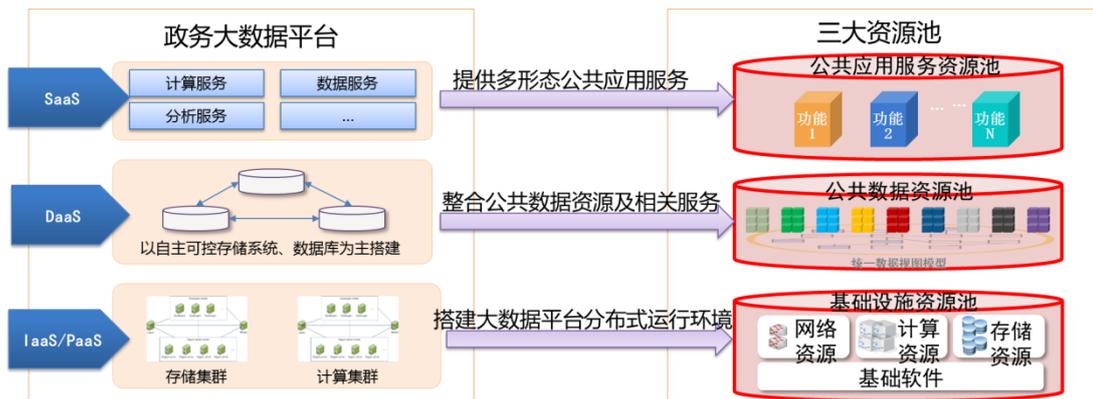


图 10 某省政务大数据资产汇聚架构图

在数据资产汇聚的具体实现上，基于政务数据构建成熟稳定的主题数据模型，涵盖全方位的政务数据，以人口、法人、空间地理、政府机构作为基础主题，涵盖证照、政策法规、宏观经济、监管执法、司法、事件、社会保障、规划、财政、社情民意等社会生活中方方面面的扩展主题。政务数据模型一方面对海量数据进行分类，便于数据存储与管理，另一方面通过技术手段和分析方法构建数据之间的关联性，为数据应用提供更好的服务。

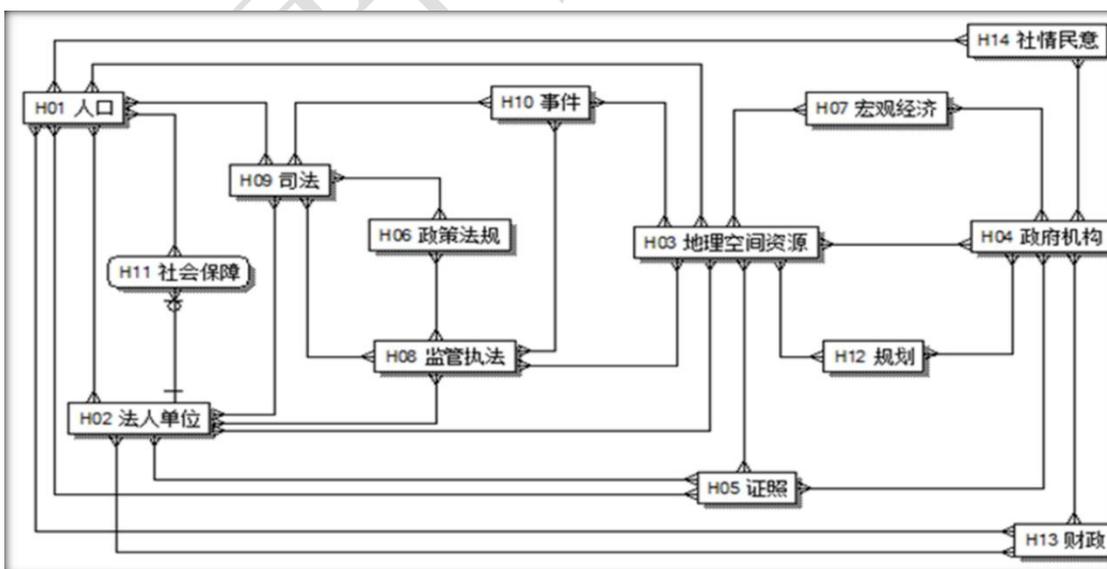


图 11 某省政务数据汇聚主题模型

其次是政务数据资产的治理。就是考虑数据资产如何规范化、数

据质量如何提升、数据安全如何保障等，实现数据资产的可控制、可量化、可变现。数据治理是需要一套体系来保障的，是组织、制度、流程、技术的组合，更重要的是管理、维护、开发内外部数据的活动。该过程一般包括数据架构与模型管理、数据标准管理、元数据管理、数据质量管理、主数据管理、数据生命周期管理、数据安全与隐私管理、数据服务管理 8 个主题，其中数据架构和模型将定义企业数据资产的框架，是数据治理的灵魂；数据标准是经协商一致制定并由相关主管机构批准，共同使用和重复使用的一种规范性文件，是数据治理的核心；元数据是用来定义数据的数据，是企业数据资产的地图，是数据治理的基础；数据质量管理将对数据进行检核，发现数据质量问题，并予以解决，是数据治理的目标。



图 12 某省政务数据资产治理架构图

最后是政务数据资产的应用，这也是政务数据资产的价值所在。政务大数据资产管理平台在数据应用端全面支持惠民、兴业和善政三大方面服务能力，通过数据在民生服务、城市管理、产业经济、生态

宜居等领域提供完善的服务。

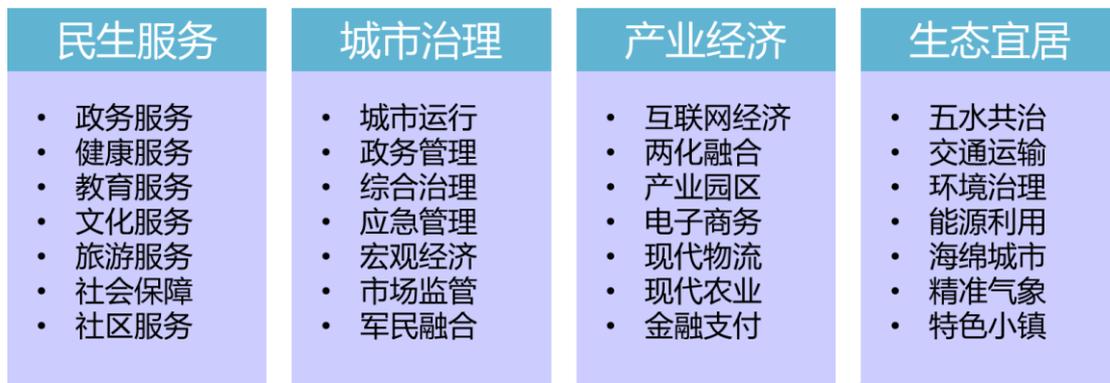


图 13 某省政务数据资产应用架构图

【应用效果】

1. 全省政务大数据资产管理平台

通过政务大数据资产管理平台的建设，将有效整合全省市管理和
服务资源，从而大幅度降低社会管理与服务的成本，有力推进全省大
数据相关应用的建设。

为各委办厅局相关大数据应用建设提供统一的数据存储、计算、
分析、可视化以及技术支撑，降低基础设施建设费用，大大降低大数
据应用建设成本。

重塑数据来源通道，一次采集全省共用，减少工作人员重复采集
的人力成本；各委办厅局可使用该平台进行数据采集，降低大数据采
集的开发及维护成本。

良好的数据质量是大数据平台稳定运行的基础，因此数据质量的
好坏将影响大数据平台的稳定运行；通过该平台，各委办厅局可以及
时的发现大数据应用运行过程中出现的数据问题，从而提升大数据应
用的数据质量。

通过实施公共机构数据开放策略,推进公共机构数据资源统一汇聚和有序开放;面向政府部门、行业事业部门、大中小型企业、广大公众用户等根据开放机制开放平台能力,形成不同资源按需开放,实现平台资源有效利用,促进市区县等地方部门开发地区数据惠民应用。

目前大数据平台服务器全部部署以及大数据组件的安装全部完成,包括数据清洗融合服务器 2 台、数据计算存储服务器 20 台、行业应用服务器 7 台,可支撑各厅局行业大数据的建设,30T 的存储空间,在一年之内使用 15T,具备为其他厅局提供大数据平台服务的基本能力,15T 的空间可以使用,而且可以按需扩展平台节点。

2. 候鸟人才专题分析应用

利用社区网格化工作人员采集结合社会众包采集、互联网爬取人才数据与政务流动人口的数据碰撞,以及通过与教育厅、组织部的人才上报系统的数据交换共享,建设某省高端候鸟人才主题库。进行高端候鸟人才画像、进项候鸟人才统计分析、候鸟政策智能推送服务等,为善待和善用候鸟人才提供技术支撑。

3. 智能办公数据服务示范应用

利用大数据存储和分类建模技术,对各委办厅局日常办公系统中产生的大量文档类数据进行有效的存储和数据挖掘。通过大数据挖掘分析,加强日常办公服务的便利性和智能化,为 OA 系统提高文档快速查询、文档分类检索、文档比对分析的支撑服务能力。

4. 人口民生服务示范应用

通过对某省各厅局各委办局各个部门关于人口和法人数据的汇

总与分析，经过了统一的数据标准化、数据稽查、数据比对、数据关联操作，完善数据结构。以“人口”为核心，在社保、民政、教育、计生、残联等领域，采集户籍人口、长期居住人口、流动人口数据并进行整理、整合，通过人口和社区大数据分析人口的年龄、教育程度、社区分布、从业分布等结构性分析，指导和协助制定人口、教育、医疗、养老以及人才引进等方面的针对性政策，综合运用人口及流动人口、城市规划、土地管理、房产管理、水资源管理等大数据，分析城市的全省户籍、人口总量、人口布局、人口流动、人口迁徙、人口密度以及与城市空间、土地资源、房地产项目、水资源等的承载分析，便于城市管理者从人口角度分析城市的承载能力，实现政府“可持续发展”的目标，为政府政务公开、民生服务、信息惠民提供更好的大数据支撑能力。

【洞察启示】

简单说来，基于大数据的数据资产管理具有全、广、多、快、深的五大特点。

- 全：就是指需要全面支持多种类型数据汇总，包括结构化数据和非结构化数据、实时数据和批量数据、内部和外部数据。
- 广：就是需要数据范围广、应用广、效益广，因此需要打通政府各部门的数据，形成统一视图；同时创造各种惠民、兴业和善政应用，支持政务智慧管理快速、全面扩展。
- 多：就是涉及的部门多，采集的系统多、支持的应用多、

访问的用户多。

- 快：就是要针对一些重要或特殊事件，需要实时监控，实现事中快速响应。
- 深：就是要针对一些重要领域或行业，引入数据挖掘、机器学习、人工智能等前沿技术，实现事前预测，从而掌控未来。

大数据背景下，一方面数据资产越来越多，越来越重要，另一方面数据资产管理将更加困难和复杂，投入也越来越大，要求企业管理人员更加重视企业数据资产管理，更加科学地进行数据资产管理，更加合理地规划和应用数据资产，让数据资产真正成为企业的核心竞争力。为了更好的实现大数据背景下的数据资产管理，认为应该从集成、治理和应用三个方面统筹考虑。数据集成是基础，数据资产管理是手段，数据应用是目标，三者相辅相成，缺一不可。

（七）某大型医药公司数据资产管理案例

【企业基本情况和数据情况】

作为中国第八大药品分销商，分销领域的强大实力使公司服务延伸至全国 322 个城市并覆盖当地医院、零售药店、现代通路、下级商业等各种渠道。除了经营药品，还涉及医疗器械、外科手术器具、特质药品、疫苗、诊断产品和个人保健产品的分销以及配送。其中进口产品和本地产品各占据公司业务量的一半，服务不同跨国企业使该企业成为中国最大的医疗产品进口商。拥有近 2300 名员工，成功搭建了由 15 个物流中心所组成的全国分销平台，在上海、北京、沈阳、大连、无锡、宁波、深圳、成都、重庆、武汉和天津的 11 个主要城市开展本地直销业务。

【痛点和难点】

现有的业务数据系统架构已经近十年，无法适应新时期下的数据增长、管理与使用的需要，管理人员对数据流程监督和管控并不完善，对于现有的业务系统，其扩展性、可靠性、运维、性能都将面临较大的挑战。

现阶段业务处理对于人力的依赖过大，导致团队人员及开支庞大，现有的业务团队已经满负荷运转，已经成为制约业务发展的瓶颈。集团自身还有渠道数据的管理服务，面向自身与分销商。原有 IT 平台数据处理能力无法满足需求，数据质量检查与业务基线查询自动化程度低业务流程中间环节比较机械固化，关键环节数据清洗量大，数据整合的过程中还需要数据清理，ETL 过程占用了大量的人力，在主数

据清洗，质量检验分发，目前 200 多人工团队，每月 500 多万条。数据处理人员工作负荷高，数据处理效率已经成为业务扩展的瓶颈。

【实践路径】

新时期下的新技术、新方式，都无法使用传统的模式进行管控，所以需要使用 New IT 的思路进行重构业务系统架构：

1. 云计算层：

从运行环境、弹性计算、资源扩展等多方面考虑协助企业选型与设计。

2. 数据平台层建设：

采用 LEAP 智能平台，解决数据的采集、存贮、计算与管理：

- 海量数据存储与大规模并行计算；
- 全程的元数据管理与主数据管理；
- 数据集成、处理的自动化、流程化；
- 全程的监控、分析、预警；
- 数据模型重构

3. AI 层建设：

AI 算法+互联网大数据 实现数据管理过程自动化+智能化；

- AI 助力数据处理自动化、机器化；
- 业务流程优化
- 业务运营门户服务建设

4. 业务建设：

业务流程重构，为整个业务流程处理争取最大灵活性和时间缓冲。

更新数据管控理念，在传统管控基础上引入新技术与新方法，使数据管理受惠于技术的发展：

对于数据资产管理职能领域的数据标准管理、主数据管理、数据质量管理，在传统的数据处理基础上引入人工智能的，代替人手，在提升效率，准确度的基础上更节省了人手，大幅降低人力成本：

主数据扩展；定向爬取相关网站数据，形成原始互联网实体库；实体对齐：（存在很多“一个实体有多个别名”、“不同实体名称相同”的情况）；扩展主数据：用互联网实体属性去修正、补充主数据库实体。

数据格式自动识别与规范化；异构数据接入、归一化相似度处理、连接性处理这些都用到“人工智能、机器学习、自然语言处理、语义识别”等算法、技术。

【应用效果】

业务流程明显优化：数据清理环节以减少人工操作环节，工作量下降，业务流程时间缩短了，计算的过程加快手，加速了流程运转，提升了整体效率。

人力成本大幅降低：利用了“人工智能、机器学习”的方法自动维护数据，帮助客户减少 80%人力，每年节省 400 万，实施前大量人工进行数据维护，每月百人团队加班半个月的情况大大改善，人力节省效果显著，整个团队体量从 200 人缩减到 20 人。

数据化转型奠定基础：数据运转进程的加速，数据可依赖性的提升，使得企业在其渠道管理上实现了提升，也使企业对进一步利用数据助力业务发展有了实现的基础。

【洞察启示】

对于需要通过人工规则进行数据清洗、质量提升的工作，通过人工智能进行自动识别与补充，能大大提高工作效率，节省人力。对分销行业，对渠道模式为主的，都会面临渠道数据质量问题，具有很好的参考借鉴意义。

中国信息通信研究院

（八）某大型钢铁集团公司数据资产管理案例

【企业基本情况和数据情况】

该企业是集矿山采掘和钢铁生产、加工、配送、贸易为一体的特大型钢铁联合企业，也是全球不锈钢行业领军企业。集团致力于转型升级，建设资源保障基地和钢材延伸加工基地，构建具有国际竞争力的钢铁上下游产业链；年收入 1500 亿，生产量超过 1000 万吨，为年度中国企业 500 强企业之一。

【痛点和难点】

企业在战略上正在逐步向数据化转型，然而对于在数据的采集、处理过程，以及如何将实时数据处理结果嵌入生产流程中仍然存在种种问题与障碍：

数据采集受限。以往用于传统的分析预测的数据源非常有限，无法有效地支持更复杂的分析方法与技术。大量的非结构化数据的采集处理，原有的集成系统并不能很好地支持实际的需求。

要求实时处理的数据结果无法嵌入生产流程，导致质检仍然只能靠人工抽检，效率低，成本高：由于收集与处理图像等非结构化数据，导致支持质检数据分析处理的结果大大地滞后于生产流程所需时间。企业在钢卷生产过程需要在钢板成卷前 30 分钟内完成质量检测，原有质检系统仍然为半自动化，仍需要大量人工参与检测，而且只能对单个采集点缺陷判断，不能判断整卷钢材是否合格；相机扫描速度快（为 1m/s~5m/s），人工仅能以抽检的方式进行监控；而人工检测准确率低，造成退货成本高。

【实践路径】

1.拓源：针对数据来源进行一系列的改造

第一，扩充基础数据源，从原来业务系统数据包括生产数据如生产任务数据、内部质量检测报告和实时生产设备监控数据等；出货管理数据；销售数据；还扩充了下游汽车销售数据、宏观经济数据、网络数据，聚合内外部多元数据作为数据分析的输入，通过元数据管理将新增数据源纳入到整体数据管控体系中



图 14 数据资产管理流程图

第二，大数据平台的建设，实现了不同数据来源不同结构数据的采集与集成。

(1) 多元数据统一采集



图 15 多元化数据集成

(2) 实时监控数据的采集与处理：和原监控系统在线对接在线对接；流架构实现在线数据，包括图像等非结构化数据的提取计算。

缺陷	数量	面积	左右侧缺陷面积	左右侧缺陷面积	位置	位置
编号	名称	平米	平米	平米	平米	平米
2	裂纹	1.237	42.283	1178.676	9.834	2.643
3	夹杂	1.677	40.962	1179.777	4.215	2.863
4	气孔	3.089	43.144	1177.393	2.811	2.863
5	分层	3.714	151.475	1059.532	5.901	2.863
6	孔洞	5.096	1302.501	105.962	30.396	13.822
10	划伤	5.157	566.311	14.479	415.578	641.473
11	凹坑	5.245	613.917	577.633	41.580	30.713
12	压痕	5.426	1157.349	60.111	4.215	4.826
13	划伤	5.604	140.806	1057.267	18.545	13.654
15	凹坑	5.999	1192.058	13.382	1360.533	45.193

图 16 原始检测数据实时采集

(3) 数据处理：对采集的数据源进行处理，包括数据检查、数据清洗与过滤规则+人工智能化处理

2.治内：针对质量检测数据模型管理：

第一，数据模型的建设：基于整合的数据，通过缺陷聚类与合格专家模型建设，能支持多种型材的合格预测

第二，数据模型的智能管理与优化：数据模型动态管理，质量检测结果支持在线钢卷缺陷人工查看，结果返回系统；机器学习进行模型优化。通过数据补充与检测结果反馈不断进行优化，形成管理闭环。数据质量优化过程嵌入数据消费流程中，提升模型精度。

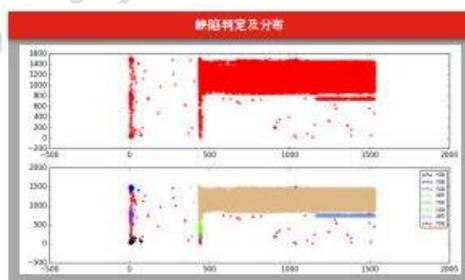


图 17 钢材合格预测模型

第三，在生产流程中嵌入数据实时流处理：建设产线-车间-钢厂 3 层应用联合，预测结果秒级输出，检测覆盖率 98%。

【应用效果】

实现了企业内不同数据源的整合,为数据的分析利用需求打下了基础。数据源的扩充使得企业能从更多维度对客户、对企业身体,对行业进行了解与分析。

以质量数据为突破口,将数据管理环节嵌入到生产流程中,“边管控,边利用”,改变了数据资产管理在业务中价值单一,难以体现,且必须需要较长时间才能显现的固有印象。

引入人工智能的大数据管控,管控手段从传统的手工向自动化与智能化迈进。在项目实施后,数据清洗处理效率,模型优化效率得到了明显提升,而智能化管控也是企业逐步向数据化转型的关键一步,具有里程碑意义。

【洞察启示】

对于大型工业企业来说,数据资产管理问题通常会包括数据源管理,如基础数据源的扩充、采集、清洗、存贮等一系列工作。数据源的扩充是企业能够运用更先进的技术手段使用数据的基础。

传统的数据质量提升价值往往仅体现于侧重 OLAP 的数据消费,而随着业务在实际中对实时分析结果的依赖逐步增强,生产环节的实时分析对基础数据的质量要求也会越来越高。数据管控的切入点需要考虑从传统的集成后再治理往数据流前端推进。

（九）湖北省物价数据资产管理案例

【企业基本情况和数据情况】

湖北省物价局是主管全省价格工作的省政府直属机构，负责贯彻执行国家的价格法律法规、方针政策和管理制度，负责研究提出本省地方性价格调控目标、调控措施和相关地方性法规，并负责全省农业、工业、交通、房地产、公共事业、中介等各方面价格工作的宏观管理、综合平衡、价格和收费的监督检查等工作。

湖北省进行数据资产管理是基于物价大数据平台以及省级政务云平台基础架构开展的，通过打通局办公室、政策法规处、农产品与医药价格处、环境与资源价格处、收费管理处、工业品与服务价格处、人事处、机关党办、监察室、离退休干部处、价格监督检查与反垄断分局、成本调查监审分局、后勤服务中心、价格监测中心、价格认证中心等多个部门相关数据，先后完成决策支持、成本调查、价格监测、移动办公、信用管理、舆情分析、效能监察等应用系统的建设，实现了数据融合、应用的整合、深度数据挖掘分析以及智能化应用，为全省消费者和市场主体提供全面及时的物价公共服务，同时还有效提高物价部门工作效率和水平

【痛点和难点】

随着计算机信息技术和网络技术的快速发展，信息化建设对政府部门的管理水平和运营质量提出了更高的要求。信息化建设从以前的以满足单一部门业务诉求为单位的模式变更为行业融合模式，“平台化、整合化、公开化”等是政府部门电子政务方向发展的趋势。而湖

北省物价局无论从系统建设和愿景规划上，都处于落后状态，主要表现在信息化水平薄弱、纵横向网络不够健全、缺少统一规划、缺少统一标准约束、数据来源单一、数据资源利用率低下等方面，这些问题严重制约了湖北省物价局的公共服务能力，使其无法有效做出科学决策。

【实践路径】

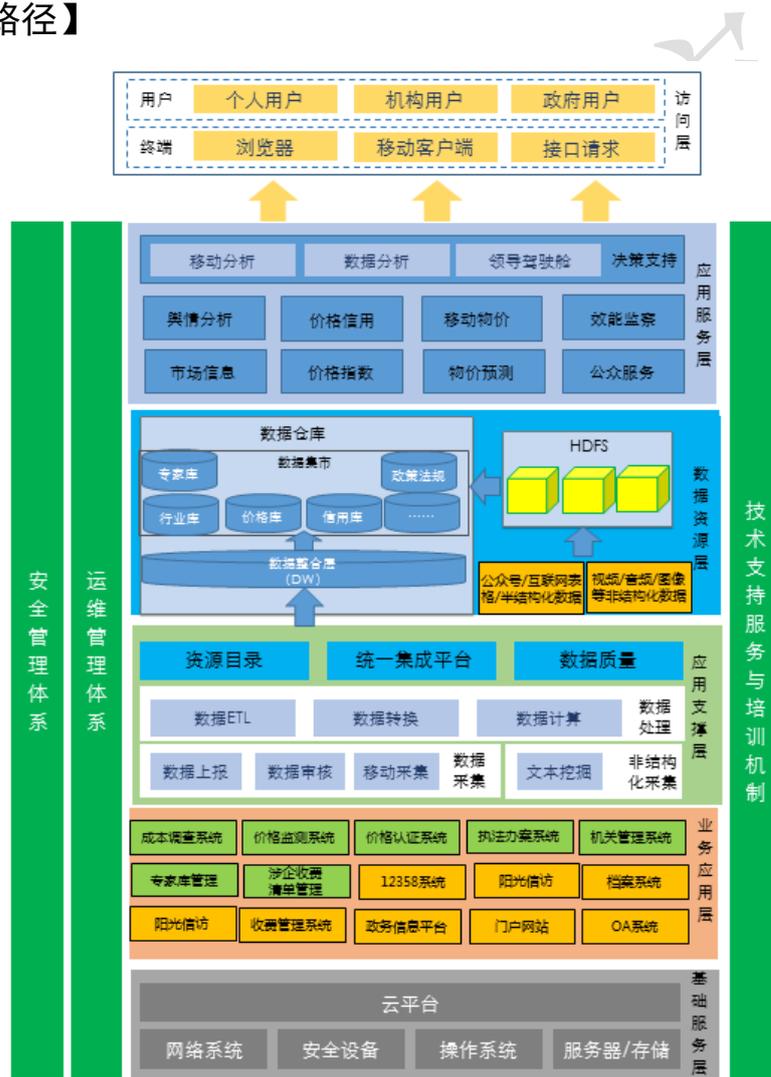


图 18 物价大数据平台总体架构示意图

1.数据融合

加强内部信息资源整合和信息资源的流通，加强数据标准的建设，

同时结合“互联网+大数据”信息化建设新思路，进行结构化和非结构化数据的融合

2.应用融合

从内部业务运用支撑和外部服务应用着手，结合顶层设计思路统一进行大数据应用的规划和设计，全力打造面向管理人员、社会公众、政府机关单位等各方数据服务对象的决策支持系统。

3.深度挖掘与分析

加强数据资源的有效利用，深度挖掘数据的潜在价值，为各项政策的制定和执行推广提供宝贵的数据依据。依托大数据技术和商业智能分析工具，充分利用互联网相关价格数据、社会舆情信息、价格成本数据、交易数据、价格规范标准等信息数据，提供及时可靠的价格预测分析、价格预警分析、舆情热点分析、反欺诈反垄断行为分析等应用分析，将事后监管逐渐延伸的事中监管和事前防范。

4.智能化

通过智能化的应用手段，优化办公效率，服务于公众。

（1）定期发布市民生活必需品、住房、医疗、教育、收费标准等多方面的价格服务，为市民生活消费提供智能化的价格指导。

（2）提供统计分析数据，实现政务决策智能化，建立决策逻辑模型，输出决策建议；实现督查智能化，建立绿灯提醒、黄灯警示、红灯督查的内部管理机制；实现考核智能化，按照各单位、各部门的绩效考核方法建立基于数据的考核体系。

（3）通过移动采集、移动办公等智能终端平台，服务于监测点

数据的上报和查阅,为价格数据的及时性和准确性提供更有效的保障,服务于内部工作人员,满足日常通信、审批公文、待办事项查阅等快捷办公需求。

【应用效果】

1.效率提升

通过大数据平台建立起面向全省范围内的物价信息的共享、交换和应用体系,减少了重复性工作,工作效率提升 70%,总体行政成本降低 50%。

2.智慧生活

通过对各类商品价格信息的统计分析预测,为市民提供衣食住行等多方面的价格服务,帮助市民随时掌握市场行情和价格动态,为生活消费提供智慧支持。同时智能监控市场价格动态,保障价格总水平的基本稳定,营造良好的价格环境。

3.企业发展

通过内部系统与外部发布平台联动实现收费信息的及时公示,畅通企业获取收费信息渠道,杜绝乱收费、乱摊派现象,为反价格垄断和欺诈提供重要的数据支撑,减少价格欺诈和诈骗行为。同时,越来越多的企业谋求自身发展,积极参与价格诚信建设,树立了良好的社会现象。

【洞察启示】

湖北省物价局围绕使市场在资源配置中起决定性作用和更好发挥政府作用,推进简政放权和政府职能转变,以社会信用体系建设和

政府信息公开、数据开放为抓手，充分运用大数据、云计算等现代信息技术，提高政府服务水平，加强事中事后监管，维护市场正常秩序，促进市场公平竞争，释放市场主体活力，进一步优化了湖北经济社会发展环境。

中国信息通信研究院

中国信息通信研究院云计算与大数据研究所

地 址：北京市海淀区花园北路 52 号

邮政编码：100191

联系电话：13920169895

邮 箱：liuchengcheng@caict.ac.cn

网 址：www.caict.ac.cn

数促会网址：www.databench.cn



中国信息通信研究院



大数据发展促进委员会