

# 大数据标准化白皮书

(2020 版)

编写单位：

全国信息技术标准化技术委员会大数据标准工作组

中国电子技术标准化研究院

二零二零年九月

# 《大数据标准化白皮书》（2020 版）

## 专家指导组：

梅 宏 孙文龙 杜小勇

## 编写组（排名不分先后）

周 平	董 建	张 群	尹 卓	刘宇峰	訾津津
闵京华	郭亚茹	张慧敏	刘俊良	刘 定	光 亮
黄先芝	赵春昊	王献芬	戴炳荣	陈敏刚	肖国荣
张 月	李 正	汪先锋	黎俊茂	严 红	朴昕阳
朱 松	刘 莎	王征平	李绪国	张 煜	李晓燕
张 聪	洪 江	梁云丹	杨 钰	康丽丽	曹晓琦
王 枏	王衍强	刘 相	于铁强	尹青云	魏 清
吕令广	王东生	朱艳华	刘 朝	魏丽丽	顾 勤
宾军志					

## 编写单位（排名不分先后）

中国电子技术标准化研究院  
北京华宇软件股份有限公司  
北京亿信华辰软件有限责任公司  
智慧神州（北京）科技有限公司  
成都四方伟业软件股份有限公司  
北京启迪数字科技集团有限公司  
浪潮软件集团有限公司  
上海依图网络科技有限公司  
九次方大数据信息集团有限公司  
中国科学院计算机网络信息中心  
湖南科创信息技术股份有限公司  
陕西省大数据集团有限公司  
交通运输部科学研究院  
山东省生态环境监测中心  
深圳市华傲数据技术有限公司  
中国电子科技集团公司第 28 研究所  
浙江蚂蚁小微金融服务集团股份有限公司  
杭州数梦工场科技有限公司  
成都市大数据股份有限公司  
重庆大数据研究院有限公司  
中通服咨询设计研究院有限公司

阿里云计算有限公司  
上海计算机软件技术开发中心  
成都市标准化研究院  
湖北微安交通安全研究院有限公司  
上海三零卫士信息安全有限公司  
方正国际软件（北京）有限公司  
美林数据技术股份有限公司  
生态环境部信息中心  
成都中科大旗软件股份有限公司  
华为技术有限公司  
北京邮电大学  
中国平安保险（集团）股份有限公司  
普元信息技术股份有限公司  
北京软件和信息服务交易所有限公司  
同方知网数字出版技术股份有限公司  
江苏中堃数据技术有限公司  
深圳中兴网信科技有限公司  
青岛市环境信息中心  
北京东方金信科技有限公司  
中电长城网际系统应用有限公司

# 目 录

1 前言.....	1
1.1 研究背景.....	1
1.2 研究目标及意义.....	2
2 相关政策法规.....	3
2.1 国外政策.....	3
2.1.1 美国大数据政策.....	3
2.1.2 欧盟大数据政策.....	3
2.1.3 英国大数据政策.....	4
2.1.4 法国大数据政策.....	5
2.1.5 日本大数据政策.....	6
2.1.6 俄罗斯大数据政策.....	6
2.2 国内政策.....	7
2.2.1 国家和行业政策.....	7
2.2.2 地方政策.....	8
2.3 安全隐私法律法规.....	10
3 大数据技术发展现状.....	13
3.1 大数据参考架构.....	13
3.2 大数据通用技术概述.....	14
3.3 大数据关键技术概述.....	16
4 大数据产业应用现状.....	19
4.1 大数据产业生态.....	19
4.2 重点领域应用.....	21
4.2.1 社会管理与公共服务.....	21
4.2.2 农业.....	25
4.2.3 制造业.....	26
4.2.4 能源领域.....	27
4.2.5 通信领域.....	29
4.2.6 交通运输领域.....	29
4.2.7 生态环境领域.....	30
4.2.8 司法领域.....	30
4.2.9 市场监管.....	31
4.2.10 金融领域.....	32
4.2.11 科学研究.....	33

4.2.12 卫生健康.....	34
4.2.13 文化、教育和旅游领域.....	35
5 大数据热点领域概述.....	39
5.1 大数据治理.....	39
5.2 大数据与人工智能.....	41
6 大数据标准化现状.....	43
6.1 国际标准化现状.....	43
6.1.1 ISO/IEC JTC 1.....	43
6.1.2 ITU-T.....	46
6.1.3 IEEE BDGMM.....	48
6.1.4 NIST.....	48
6.2 国内标准化现状.....	49
6.2.1 国家标准化现状.....	49
6.2.2 地方标准化现状.....	52
7 大数据标准体系.....	54
7.1 大数据标准体系框架.....	54
7.2 标准明细表.....	56
7.3 重点标准介绍.....	63
8 大数据标准化工作建议.....	66
8.1 完善大数据标准化工作机制.....	66
8.2 夯实大数据标准体系建设.....	66
8.3 加强大数据标准宣传推广.....	66
8.4 构建标准符合性评估体系.....	67
8.5 推进大数据治理标准支撑.....	67
8.6 加大标准化人才培养力度.....	67
8.7 加快大数据标准的国际化步伐.....	68
9 附件：应用案例.....	69
9.1 案例：广州市公安局大数据平台.....	69
9.2 案例：咸阳市精准扶贫大数据平台.....	73
9.3 案例：国家生猪大数据中心.....	76
9.4 案例：中国航发南方工业公司 AEOS 运营管理平台.....	79
9.5 案例：广东移动大数据实时跨集群协同云计算及行业化快速支撑应用..	81
9.6 案例：生态环境标准化体系建设成效与案例.....	84
9.7 案例：政法跨部门大数据办案平台.....	85

9.8 案例：蚂蚁金服数巢一站式大数据共创服务平台.....	88
9.9 案例：济南健康医疗大数据平台建设及运营案例.....	91
9.10 案例：华西医院肿瘤专科临床科研智能大数据平台.....	95
9.11 案例：面向智慧城市数据融合共享的大数据治理平台.....	99
9.12 案例：招商局融资租赁（天津）有限公司数据治理项目.....	101
9.13 案例：中国进出口银行数据治理项目.....	104
9.14 案例：中国农业发展银行数据服务平台项目.....	108
9.15 案例：国家电网公司数据治理实践.....	110

全国信标委大数据标准工作组

# 1 前言

## 1.1 研究背景

大数据是新时代最重要的“数字金矿”，是全球数字经济发展的核心动能。数据资源如同农业时代的土地、劳动力，工业时代的技术、资本，已经成为信息时代重要的基础性战略资源和关键生产要素，是推动经济发展质量变革、效率变革、动力变革的新引擎，不断驱动人类社会加快在信息化时代中的前进步伐，逐步向智能化时代迈进。大数据作为数据资源价值挖掘的动力源，受到了世界各国政府和国际组织的高度重视，世界主要国家和地区竞相开展大数据战略布局，推动大数据技术创新研发与产业应用落地，旨在以大数据为抓手，抢占数字经济时代全球竞争制高点。

我国党中央、国务院高度重视并持续推进大数据发展。自 2015 年国务院印发《促进大数据发展行动纲要》以来，陆续发布《政务信息资源共享管理暂行办法》《大数据产业发展规划（2016-2020）》等政策文件，持续加强大数据、数字经济国家战略部署。2019 年 3 月 5 日国务院总理李克强在政府工作报告中强调“深化大数据、人工智能等研发应用，培育新一代信息技术、高端装备、生物医药、新能源汽车、新材料等新兴产业集群，壮大数字经济”，这是大数据连续第六年写入政府工作报告中。2019 年 10 月，习近平总书记在十九届四中全会中提出建立健全运用互联网、大数据、人工智能等技术手段进行行政管理的制度规则，将大数据的定位从构建数字经济延伸至国家治理，赋予大数据以新的历史使命。

我国拥有海量的数据资源和丰富的应用场景，具备大数据发展的先天优势。随着国家政策的日渐完善以及扶持力度的不断增强，我国大数据技术、产业得到长足发展。国内骨干企业已经具备自主开发建设和运维大规模大数据平台的能力，一批新产品、新服务、新模式陆续出现，大数据生态体系逐渐完善。然而，随着我国大数据应用的进一步深入，产业发展的痛点问题相继暴露：数据开放共享进程缓慢、数据质量不高、数据管理与治理缺失、垂直行业应用深度不足。面对当前大数据发展的瓶颈问题，不仅需要从技术、行业机制的角度寻求突破，也

亟需从标准化维度持续加强支撑。

前期，在工业和信息化部信息技术发展司、国家市场监督管理总局标准技术管理司的指导下，全国信标委大数据标准工作组组织国内相关产、学、研、用单位的专家、学者，在深入调研国内外大数据技术、产业及标准化需求的基础上，编制形成了《大数据标准化白皮书（2016版）》《大数据标准化白皮书（2018版）》，受到了业界关注和一致好评。2020版白皮书将在2018版的基础上更新国内外大数据领域最新政策法规，提出大数据技术发展趋势和重点领域应用实践，分析大数据发展热点方向，并在此基础上进一步完善大数据标准体系，更新大数据标准化相关成果，提炼最新标准化需求。

## 1.2 研究目标及意义

本白皮书梳理了国内外主要国家、地区大数据领域的发展战略，描述了大数据核心技术和产业应用现状，力图从政策、技术、产业等多个角度，勾画出当前大数据发展的整体脉络。同时结合大数据发展的最新态势及热点方向，迭代完善形成了新时期的大数据标准体系，并提出了我国大数据标准化工作建议。

本白皮书的发布，旨在与业界分享笔者在大数据领域的研究成果和实践经验，展示近年来大数据发展的技术、产业及标准化最新态势，呼吁社会各界共同关注大数据政策研究、技术投入、服务应用和标准化建设，加强对大数据标准化工作的重视程度，共同助力大数据引领国民经济高质量发展。



## 2 相关政策法规

### 2.1 国外政策

#### 2.1.1 美国大数据政策

早在 2011 年，美国国家科技委员会专门成立“大数据高级督导组（Big Data Senior Steering Group, BDSSG）”，负责确定联邦政府当前需要开展的大数据研发任务，做好部门间的工作协调，制定远景目标。随后，自 2012 年起，美国针对大数据技术革命实施了多轮政策行动。

2012 年 3 月，美国白宫科技政策办公室发布《大数据研究和发展计划》，成立“大数据高级指导小组”，旨在大力提升美国从海量复杂的数据集合中获取知识和洞见的能力。2013 年 11 月，美国信息技术与创新基金会发布《支持数据驱动型创新的技术与政策》，提出了“数据——知识——行动”计划，建议世界各国的政策制定者应采取措施，鼓励公共部门和私营部门开展数据驱动型创新。2014 年 5 月，美国总统行政办公室发布《大数据：把握机遇，保存价值》，对美国大数据应用与管理的现状、政策框架和改进建议进行了集中阐述，并就保护个人隐私的价值、数字时代负责任的教育创新、大数据与歧视、执法与安全保护、数据公共资源化提出建议。2016 年 5 月，美国总统科技顾问委员会发布了《联邦大数据研发战略计划》，主要围绕代表大数据研发关键领域的七个战略进行，包括促进人类对科学、医学和安全等方面的认识，确保美国在研发领域继续发挥领导作用，通过研发来提高美国和世界解决紧迫社会和环境问题的能力等。2019 年 12 月 23 日，美国白宫行政管理和预算办公室 (OMB) 发布《联邦数据战略与 2020 年行动计划》，以 2020 年为起始，描述了美国联邦政府未来十年的数据愿景，确立了政府范围内的数据使用框架原则，40 项具体数据管理实践以及 20 项 2020 年具体行动方案。

#### 2.1.2 欧盟大数据政策

2014 年欧盟推动形成《数据价值链战略计划》草案，发布了《数据驱动经

济战略》，旨在通过一个以数据为核心的连贯性欧盟生态体系，让数据价值链的不同阶段产生价值，以驱动欧洲经济繁荣。2018年，欧盟委员会发布政策文件《建立一个共同的欧盟数据空间》，聚焦公共部门数据开放共享、科研数据保存和获取、私营部门数据分享等事项。

2018年5月，欧盟《通用数据保护条例》（General Data Protection Regulation, 以下简称“GDPR”）正式实施。作为全球最严格的个人数据保护法，一方面GDPR强化了自然人的个人数据保护权，在已有自然人权属的基础上，增设了删除权（被遗忘权）、限制处理权、持续控制权（数据可携权）和拒绝权等一系列新的权利给数据主体，并对错误处理个人数据或侵犯数据主体权利的企业提出了高额的处罚条款；另一方面GDPR提出不能以保护处理个人数据中的相关自然人为由，对欧盟内部个人数据的自由流动进行限制或禁止，进一步推动欧盟境内数据交换共享。

2018年11月，欧洲议会和欧盟理事会共同颁布《非个人数据自由流动条例》（Regulation on the free flow of non-personal data），并于2019年5月28日正式实施。该《条例》旨在保障非个人数据在欧盟境内能够自由流动，并对数据本地化要求、主管当局的数据获取及跨境合作、专业用户的数据迁移等问题作了具体规定。该《条例》和GDPR一同奠定了欧盟的第五大自由，即在人、货物、服务和资本以外，数据也摆脱了成员国的边境、负担和障碍带来的限制，可以在欧盟范围内自由流动。

2020年2月，欧盟发布《欧洲数字战略》，旨在通过增强欧盟企业及公民数字能力建设、善用技术巨头市场力量及挖掘信息通讯技术可持续发展潜力，使欧盟成为世界上最具竞争力的数据敏捷型经济体。该《战略》提出建立统一治理框架，加强数据基础设施投资，提升个体数据权利和技能，打造公共欧洲数据空间等多项具体措施，推动欧盟各领域数据流通及深度挖掘，培育形成面向全球的数据单一市场。

### 2.1.3 英国大数据政策

2012年5月，世界上首个开放式数据研究所ODI(The Open Data Institute)

在英国政府的支持下建立，首批注资十万英镑。这是英国政府研究和利用开放式数据方面的一次里程碑式发展。未来，英国政府将通过这个组织来利用和挖掘公开数据的商业潜力，并为英国公共部门、学术机构等方面的创新发展提供“孵化环境”，同时为国家可持续发展政策提供进一步的帮助。

2013年1月，英国商业、创新和技能部宣布，将注资6亿英镑发展8类高新技术，大数据独揽其中的1.89亿英镑，将近三成。同年10月，英国商务、创新和技能部牵头发布《英国数据能力发展战略规划》，该战略旨在使英国成为大数据分析的世界领跑者，并使公民和消费者、企业界和学术界、公共部门和私营部门均从中获益。该战略在定义数据能力以及如何提高数据能力方面，进行了系统性地研究分析，并提出了举措建议。

#### 2.1.4 法国大数据政策

2011年7月，法国工业部长埃里克贝松宣布，启动“Open Data Proxima Mobile”项目，希望通过该项目实现公共数据在移动终端上的使用，从而最大限度的挖掘它们的应用价值。同年12月，法国政府推出的公开信息线上共享平台data.gouv.fr，该网站的数据由每个政府部门的专员统计和收集、持续更新，以便推动法国公共数据的开放共享。

2013年2月，法国政府发布《数字化路线图》，明确了大数据是未来要大力支持战略性高新技术，政府将以新兴企业、软件制造商、工程师、信息系统设计师等为目标，开展一系列的投资计划，旨在通过发展创新性解决方案，并将其用于实践，来促进法国在大数据领域的发展。同年7月，法国中小企业、创新和数字经济部发布了《法国政府大数据五项支持计划》，具体包括引进数据科学家教育项目；设立一个技术中心给予新兴企业各类数据库和网络文档存取权；通过为大数据设立原始扶持资金，促进创新；在交通、医疗卫生等纵向行业领域设立大数据旗舰项目；为大数据应用建立良好的生态环境，如在法国和欧盟层面建立用于交流的各类社会网络等。

### 2.1.5 日本大数据政策

2012年6月，日本IT战略本部发布电子政务开放数据战略草案，迈出了政府数据公开的关键性一步。政府将利用信息公开方式标准化技术实现统计信息、测量信息、灾害信息等公共信息，在紧急情况时可以较少的网络流量向手机用户提供信息，并尽快在网络上实现行政信息全部公开并可被重复使用，以进一步推进开放政府的建设进程。同年7月，日本推出了《面向2020年的ICT综合战略》，聚焦大数据应用所需的社会化媒体等智能技术开发，传统产业IT创新以及在新医疗技术开发、缓解交通拥堵等公共领域的应用。

2017年10月，日本公正交易委员会竞争政策研究中心发布了《数据与竞争政策研究报告书》。在这部报告书中，日本明确了运用竞争法对“数据垄断”行为进行规制的主要原则和判断标准。

### 2.1.6 俄罗斯大数据政策

2015年9月1日，俄罗斯的数据本地化立法（第242-FZ号联邦法中确定）正式生效，要求所有国内外公司在俄罗斯境内的服务器上存储和处理俄罗斯公民的个人信息。根据法律规定，任何存储俄罗斯国民信息的组织，无论是客户还是社交媒体用户，都必须将该数据移至俄罗斯服务器。

2016年12月1日，俄罗斯总统普京签发第642号总统令，批准《俄罗斯联邦科技发展战略》，这是当代俄罗斯第一份以总统令形式规划国家科技发展的纲领性文件。该《战略》提出，俄罗斯科技总体发展目标是通过整合国家科技资源，有机统一科技创新能力，使科技与国家经济社会体系充分融合，以保证国家独立和国际竞争力。该《战略》规划了未来10-15年俄罗斯科技创新优先发展方向，包括发展数字技术、人工智能，开发大数据、机器学习及人工智能系统等内容。2017年7月，俄政府发布《俄罗斯联邦数字经济规划》，推进利用现代数字技术，保证国家信息安全，主要包括大数据、神经网络技术和人工智能、分布式注册系统、量子技术、新型制造技术、工业互联网、机器人和传感器、无线技术、虚拟和增强现实技术等。

## 2.2 国内政策

### 2.2.1 国家和行业政策

近年来，我国党中央、国务院高度重视并不断完善大数据政策支撑。

2015年7月，国务院办公厅发布《关于运用大数据加强对市场主体服务和监管的若干意见》（国办发〔2015〕51号），肯定了大数据在市场监管服务中的重大作用，并在重点任务分工安排中提出要建立大数据标准体系，研究制定有关大数据的基础标准、技术标准、应用标准和管理标准，加快建立政府信息采集、存储、公开、共享、使用、质量保障和安全管理的技术标准，引导建立企业间信息共享交换的标准规范。同年8月，国务院发布《关于印发促进大数据发展行动纲要》（国发〔2015〕50号），全面系统部署我国大数据发展，并在政策机制部分中着重强调建立标准规范体系，推进数据关键共性标准制定和实施，开展标准验证和应用试点示范，积极参与相关国际标准制定工作。

2017年1月，工业和信息化部发布《大数据产业发展规划（2016—2020年）》（工信部规〔2016〕412号）。全面部署“十三五”时期大数据产业发展工作，加快建设数据强国，为实现制造强国和网络强国提供强大的产业支撑。作为未来五年大数据产业发展的行动纲领，《大数据产业发展规划（2016—2020年）》明确了“十三五”时期大数据产业发展的指导思想、发展目标、重点任务、重点工程及保障措施等内容，指引了从推进大数据技术产品创新发展、提升大数据行业应用能力、繁荣大数据产业生态、健全大数据产业支撑体系、夯实完善大数据保障体系五个方面开展工作。

2020年4月，中共中央、国务院印发《关于构建更加完善的要素市场化配置体制机制的意见》，首次将数据作为一种新型生产要素写入文件，并强调加快培育数据要素市场，推进政府数据开放共享，提升社会数据资源价值，加强数据资源整合和安全保护。数据作为生产要素与土地、劳动力、资本、技术并列，对我国大数据发展提出了更高的要求。

围绕国家政策，各部委和相关行业管理机构也陆续出台了一系列行业政策以促进大数据在各领域的深入应用，如表 2-1 所示。

表 2-1 部分领域大数据政策

序号	政策名称	发布日期	发文单位
1.	《智慧城市时空大数据平台建设技术大纲（2019 版）》	2019 年 02 月	自然资源部
2.	《银行业金融机构数据治理指引》	2018 年 05 月	中国银行保险监督管理委员会
3.	《关于深入开展“大数据 + 网上督察”工作的意见》	2017 年 09 月	公安部
4.	《智慧城市时空大数据与云平台建设技术大纲（2017 版）》	2017 年 09 月	国家测绘地理信息局办公室
5.	《大数据驱动的管理与决策研究重大研究计划 2017 年度项目指南》	2017 年 07 月	国家自然科学基金委员会
6.	《关于推进水利大数据发展的指导意见》	2017 年 05 月	水利部
7.	《大数据产业发展规划（2016—2020 年）》	2017 年 01 月	工信部
8.	《农业农村大数据试点方案》	2016 年 10 月	农业部
9.	《关于推进交通运输行业数据资源开放共享的实施意见》	2016 年 08 月	交通运输部
10.	《关于加快中国林业大数据发展的指导意见》	2016 年 07 月	国家林业局
11.	《关于印发促进国土资源大数据应用发展实施意见》	2016 年 07 月	国土资源部
12.	《生态环境大数据建设总体方案》	2016 年 03 月	环保部
13.	《关于组织实施促进大数据发展重大工程的通知》	2016 年 01 月	发改委

### 2.2.2 地方政策

围绕国家大数据战略，各省市也相继出台大数据政策规划，积极推动当地大数据产业发展。贵州省以建设社会信用体系与大数据融合发展试点省为契机，重点落实国家各项社会信用体系建设量化指标，助力“放管服”和供给侧结构性改革，优化营商环境，推动政府职能转变和社会治理能力提升，推进“诚信贵州”建设，大力推进社会信用体系与大数据融合发展试点省工作；浙江省以构建城市大脑为核心开展数据资源、应用服务、标准规范、安全保障、运营模式的探索与创新，带动大数据产业全面落地；山东省明确要加快数字园区建设，培育和支持软件产业园区发展，加快推动大数据产业创新发展，进一步健全大数据产业链条，形成创新协同、布局合理、产业配套、科学有序的产业生态体系。

部分省市出台的大数据政策规划见表 2-2。

表 2-2 部分省市大数据政策规划

省市	文件名称	发布时间
天津	《天津市促进大数据发展应用条例》	2019 年 01 月
重庆	《重庆市以大数据智能化引领的创新驱动发展战略行动计划（2018-2020 年）》	2018 年 03 月
	《关于开展大数据智能化领域 2018 年市级工程研究中心认定的通知》	2018 年 07 月
浙江	《浙江省“城市大脑”建设应用行动方案》	2019 年 06 月
广东	《广东省政务数据资源共享管理办法（试行）》	2018 年 11 月
山东	《数字山东发展规划（2018-2022 年）》	2019 年 02 月
	《数字山东 2019 行动方案》	2019 年 03 月
山西	《山西省促进大数据发展应用 2019 年行动计划》	2019 年 04 月
广西	《广西数字社会建设三年行动计划（2018—2020 年）》	2018 年 08 月
湖南	《湖南省大数据产业发展三年行动计划（2019-2021 年）》	2019 年 01 月
河南	《河南省促进大数据产业发展若干政策》	2018 年 09 月
贵州	《贵州省大数据战略行动 2019 年工作要点》	2019 年 03 月
	《贵州建设社会信用体系与大数据融合发展试点省 2019 年工作要点》	2019 年 04 月
四川	《四川省促进大数据发展工作方案的通知》	2018 年 01 月
成都	《成都市引进培育大数据人才实施办法》	2019 年 05 月
广州	《广州市人民政府办公厅关于推进健康医疗大数据应用的实施意见》	2018 年 12 月
福州	《福州市推进大数据发展三年行动计划（2018-2020 年）》	2018 年 09 月

基于政策规范，各地还纷纷加强机构设置，旨在凝聚资源、形成合力，推动大数据政策落实。2018 年省级机构改革后，山东、重庆、福建、广东、浙江、吉林、广西、贵州等省市明确设立大数据管理机构或部门，并在大数据管理部门之下设立了大数据中心作为技术支撑机构开展具体政务大数据建设工作。

部分省市大数据机构见表 2-3。

表 2-3 部分省市大数据机构

省份	机构	隶属机构
上海	上海市大数据中心	上海市政府
重庆	重庆市大数据应用发展管理局	重庆市政府
浙江	浙江省大数据发展管理局	浙江省政府
福建	福建省数字福建建设领导小组办公室(省大数据管理局)	福建省政府
贵州	贵州省大数据局	贵州省政府
山东	山东省大数据局	山东省政府
广东	广东省政务服务数据管理局	广东省政府

省份	机构	隶属机构
四川	四川省大数据中心	四川省政府
贵州	贵州省信息中心	贵州省大数据局
	贵州省量子信息和大数据应用技术研究院	
山东	山东省大数据中心	山东省大数据局
广东	广东省信息中心（改革中）	广东省政府
四川	四川省大数据中心	四川省政府
海南	海南省大数据管理局	海南省政府
成都	成都市大数据中心	成都市政务服务和网络理政办公室

## 2.3 安全隐私法律法规

随着大数据应用的不断深入，针对数据安全保护的需求越发强烈，数据安全相关法律法规建设已成为大数据发展的必然趋势。当前，全球针对数据安全的立法主要集中在美国、欧盟等发达国家和地区，是在个人信息保护立法基础上进行的完善。相较而言，我国在数据安全保护方面立法起步较晚，数据安全相关法律体系建设仍在探索阶段。

### （1）国外大数据立法

国外大数据安全隐私立法的切入点主要是建立在个人数据保护法律法规和政府信息公开体系基础之上。

美国没有发布综合性数据安全法律文件，而是采取分散式立法模式，与网络发展、社会发展相适应，逐步探索并不断完善数据保护体系。在互联网发展初期，美国的立法以尊重公民知情权、规范信息管理、促进信息公开和自由为重点，同时兼顾国家安全，先后颁布了《信息自由法》（1966）、《信息自由法修正案》（1975）等法案，使行政机关承担向民众提供包括电子信息在内的行政情报方面的义务。随着互联网进一步发展以及在使用信息的过程中侵犯个人隐私权事件的频繁发生，美国针对涉公领域与非涉公领域的个人隐私，分别采用立法和行业自律的方式进行数据保护：在涉公领域，美国采取分散式立法，逐步规范了美国政府对个人信息的使用方式，使其承担保护个人隐私的法律义务；在非涉公领域，采取灵活的行业自律模式，保持政府与商界及消费者团体的长期对话机制，制定行业准则。随着 9.11 事件发生，美国数据安全的立法重点逐步放在了监控国内



和全球信息流动，发布了《关键性基础设施信息法》（2002）、《爱国者法案》（2002）、《国家网络安全保护法案》（2014）等。2018年1月特朗普签署了《外国情报监视法案修正案》同意授权美国国家安全局监听外籍人士及其手机等与之相关的数据情报。同年3月特朗普签署《澄清境外数据合法使用法案》，这使得美国执法机构更容易跨境调取其公民海外信息和个人信息。

欧盟在数据安全立法方面重点关注个人数据隐私保护。欧盟早在2002年就通过了《电子通信领域个人数据处理和隐私保护的指令》（“ePD指令”，2002/58/EC），该指令已于2017年1月10日进行了最新的修订，规范了一系列专门针对电子通信领域个人信息处理和隐私权保护的特别规范。2015年欧盟通过《通用数据保护条例》（General Data Protection Regulation，简称“GDPR”），并于2018年5月9日正式生效，该条例提出史上最为严格的个人数据保护规定。在跨境数据流动方面，该条例增加了新的制度安排，并提出了高额的处罚标准，目前已有众多的互联网公司因违反条例而受到高额罚款。在上述两个构成欧盟数据保护法律框架的支柱指令外，2016年欧盟理事会通过了《网络与信息系统安全指令》（NIS指令），这是欧盟网络安全立法的第一部分，核心部分体现在第十四条即安全要求与事件通报，该指令要求2016年8月生效后，成员国必须于21个月之内将其转化为国内法，从而推动在欧盟范围内形成统一的网络法规标准。

除欧盟和美国之外，德国、英国、日本、俄罗斯等国都制定了本国的相关数据保护法案。德国2002年颁布了《联邦数据保护法》，并前后进行了多次修订，是全球范围内影响力较大的数据保护立法。日本既注重国家层面的立法，也注重发挥行业自律的机制，于2005年制定和通过了一系列个人数据信息保护的法律法规文件，包括《个人信息保护法》《信息公开与个人信息保护审查会设置法》等五部法律法规文件。英国的大数据法律范本，一部分是本国法律，另一部分则直接沿用了欧盟法律。1998年的《数据保护法案》是英国数据保护的基石，后被《2017年数据保护法案（草案）》取代，更新强化了数字经济和大数据时代的个人数据保护。2014年俄罗斯公布了《俄罗斯联邦〈关于信息、信息技术和信息保护法〉修正案及个别互联网信息交流规范的修正案》，该修正案中主要对

《俄罗斯联邦通讯法》等法律进行了修正，对互联网信息传播者的行为进行了规范。

## （2）国内大数据立法

我国在大数据安全隐私领域的立法仍处于探索阶段。**国家层面**，我国尚没有专门的数据安全保护或个人信息安全保护法律法规，数据安全相关法律制度零散分布于各部门法中。除了《中华人民共和国宪法》《中华人民共和国网络安全法》《中华人民共和国民法总则》等内容中涉及到的个人信息和隐私的保护外，主要通过一些指导意见和行业规范对数据安全保护提出宏观要求。例如我国国务院印发的《关于运用大数据加强对市场主体服务和监管的若干意见》中提出加强网络和信息安全保护，建立健全信息安全保障体系；中国人民银行制定并发布的《个人信用信息基础数据库管理暂行办法》中规定了个人信息基础数据的报送、整理、查询、异议处理、安全管理等；国家网信办制定的《个人信息和重要数据出境安全评估办法（征求意见稿）》中对在我国境内进行个人信息和重要数据收集的网络运营商提出了需进行安全评估等要求。**地方层面**，我国部分地方政府从地方性法规和规章层面出发，开展数据安全保护的积极探索。2017年，浙江省发布《浙江省公共数据和电子政务管理办法》，这是公共数据和电子政务领域的第一部省级政府规章；2018年，贵阳市发布的《贵阳市大数据安全管理条例》，对大数据安全定义、防风险安全保障措施、监测预警与应急处置、投诉举报等方面做出规定；2020年，深圳市印发《深圳经济特区数据条例（征求意见稿）》，从个人数据保护、公共数据管理和应用、数据要素市场培育、数据安全、法律责任等方面全面对自然人、法人和非法人组织数据相关权益进行规范，并首次提出“数据权”的概念。

从国内外数据立法情况来看，目前在全球范围内大数据立法主要集中在个人信息安全和隐私保护，以及数据跨境流通安全保护等方面。由于数据本身具有的多主体使用与复用性等特点，其数据权属难以界定，对大数据立法的推进造成了一定的阻碍，这也是后续大数据立法过程中亟需研究与突破的方向。

### 3 大数据技术发展现状

#### 3.1 大数据参考架构

大数据作为一种新兴和不断演进的技术，相关技术标准体系也在社会各界的广泛参与和关注中不断的完善。大数据标准工作组结合 NIST（美国国家标准技术研究所）和 ISO/IEC JTC 1/SC 32 数据管理和交换分技术委员会的研究成果，制定并发布了 GB/T 35589-2017《信息技术 大数据 技术参考模型》国家标准，给出了大数据参考架构，如图 3-1 所示。

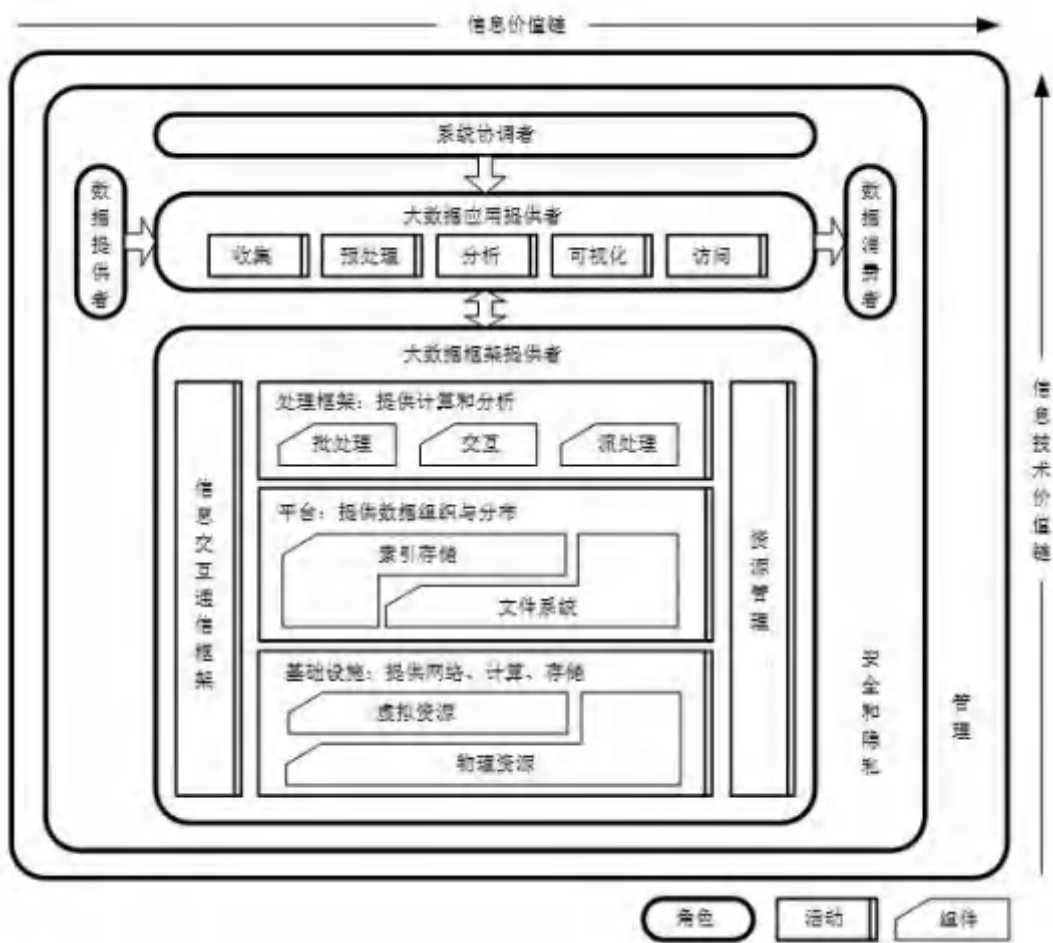


图 3-1 大数据参考架构图

大数据参考架构是一个通用的大数据系统概念模型，它表示了通用的、实现无关的大数据系统的逻辑功能构件及构件之间的互操作接口，可以作为开发各种具体类型大数据应用系统架构的通用技术参考框架。大数据参考架构采用构件层级结构来表达大数据系统的高层概念和通用的构件分类法。从构成上看，大数据

参考架构是由一系列在不同概念层级上的逻辑构件组成的。这些逻辑构件被划分为三个层级，从高到低依次为角色、活动和功能组件。五个主要的模型构件代表在每个大数据系统中存在的不同技术角色：系统协调者、数据提供者、大数据应用提供者、大数据框架提供者和数据消费者。另外两个非常重要的模型构件是安全隐私与管理，代表能为大数据系统其他五个主要模型构件提供服务和功能的构件。其中管理角色的功能尤其重要，被集成在任何大数据解决方案中。

大数据参考架构体系中的管理角色可以归类为系统管理、大数据管理和大数据治理这三个活动组。系统管理活动组包括调配、配置、软件包管理、软件管理、备份管理、能力管理、资源管理和大数据基础设施的性能管理等活动。大数据管理涵盖了大数据生存周期中所有的处理过程，其活动和功能是验证数据在生命周期的每个过程是否都能够被大数据系统正确地处理。大数据治理负责定义在数据全生存周期中如何访问和处理数据，从而实现更广泛的策略和指引，以确保数据管理的角色和责任的执行、维护数据的合规性、满足数据质量要求、标准化数据管理和利用、降低数据管理的低效率和成本、通过定义和验证数据访问要求来提高数据安全性、建立数据访问的过程以提高性能等目标的实现。

## 3.2 大数据通用技术概述

大数据技术应用于大数据系统端到端的各个环节，包括数据接入、数据预处理、数据存储、数据处理、数据可视化、数据治理，以及安全和隐私保护等。

(1) 数据接入。大数据系统需要从不同应用和数据源（如互联网、物联网等）进行离线或实时的数据采集、传输、分发。为了支持多种应用和数据类型，大数据系统的数据接入需要基于规范化的传输协议和数据格式，提供丰富的数据接口、读入各种类型的数据。

(2) 数据预处理。预处理是大数据重点技术之一。由于采集到的数据在来源、格式、数据质量等方面可能存在较大的差异，需要对数据进行整理、清洗、转换等过程，以便支撑后续数据处理、查询、分析等进一步应用。

(3) 数据存储。随着大数据系统数据规模的扩大、数据处理和分析维度的提升、以及大数据应用对数据处理性能要求的不断提高，数据存储技术得到持续

的发展与优化。一方面，基于大规模并行数据库（MPPDB）集群实现了海量结构化数据的存储与高质量管理，并能有效支持 SQL 和联机交易处理(OLTP) 查询。另一方面，基于 HDFS 分布式文件系统实现了对海量半结构化和非结构化数据的存储，进一步支撑内容检索、深度挖掘、综合分析等大数据分析应用。同时，数据规模的快速增长，也使得分布式存储成为主流的存储方式，通过充分利用分布式存储设备的资源，能够显著提升容量和读写性能，具备较高的扩展性。

（4）数据处理。不同大数据应用对数据处理需求各异，导致产生了如离线处理、实时处理、交互查询、实时检索等不同数据处理方法。离线处理通常是指对海量数据进行批量的处理和分析，对处理时间的实时性要求不高，但数据量巨大、占用计算及存储资源较多。实时处理指对实时数据源（比如流数据）进行快速分析，对分析处理的实时性要求高，单位时间处理的数据量大，对 CPU 和内存的要求很高。交互查询是指对数据进行交互式的分析和查询，对查询响应时间要求较高，对查询语言支持要求高。实时检索指对实时写入的数据进行动态的查询，对查询响应时间要求较高，并且通常需要支持高并发查询。近年来，为满足不同数据分析场景在性能、数据规模、并发性等方面的要求，流计算、内存计算、图计算等数据处理技术不断发展。同时，人工智能的快速发展使得机器学习算法更多的融入数据处理、分析过程，进一步提升了数据处理结果的精准度、智能化和分析效率。

（5）数据可视化。数据可视化是大数据技术在各行业应用中的关键环节。通过直观反映出数据各维度指标的变化趋势，用以支撑用户分析、监控和数据价值挖掘。数据可视化技术的发展使得用户借助图表、2D\3D 视图等多种方式，通过自定义配置可视化界面实现对各类数据源进行面向不同应用要求的分析。

（6）数据治理。数据治理涉及数据全生存周期端到端过程，不仅与技术紧密相关，还与政策、法规、标准、流程等密切相关（详见 5.1 节）。从技术角度，大数据治理涉及到元数据管理、数据标准管理、数据质量管理、数据安全治理等多方面技术。当前，数据资源分散、数据流通困难（模型不统一、接口难对接）、应用系统孤立等问题已经成为企业数字化转型最大挑战之一。大数据系统需要通过提供集成化的数据治理能力、实现统一数据资产管理及数据资源规划。

(7) 安全与隐私保护。大数据系统的安全与系统的各个组件、及系统工作的各个环节相关，需要从数据安全（例如备份容灾、数据加密）、应用安全（例如身份鉴别和认证）、设备安全（例如网络安全、主机安全）等方面全面保障系统的运行安全。同时随着数据应用的不断深入，数据隐私保护（包括个人隐私保护，企业商业秘密保护、国家机密保护）也已成为大数据技术重点研究方向之一。

### 3.3 大数据关键技术概述

#### (1) 分布式数据库技术

分布式数据库是指将物理上分散的多个数据库单元连接起来组成的逻辑上统一的数据库。随着各行业大数据应用对数据库需求不断提升，数据库技术面临数据的快速增长及系统规模的急剧扩大，不断对系统的可扩展性、可维护性提出更高要求。当前以结构化数据为主，结合空间、文本、时序、图等非结构化数据的融合数据分析成为用户的重要需求方向。同时随着大规模数据分析对算力要求的不断提升，需要充分发挥异构计算单元（如 CPU、GPU、AI 加速芯片）来满足应用对数据分析性能的要求。

分布式数据库主要分为 OLTP 数据库、OLAP 数据库、HTAP 系统。OLTP（联机事务处理）数据库，用于处理数据量较大、吞吐量要求较高、响应时间较短的交易数据分析。OLAP（联机分析处理）数据库，一般通过对数据进行时域分析、空间分析、多维分析，从而迅速、交互、多维度地对数据进行探索，常用于商业智能和系统的实时决策。HTAP（混合交易/分析处理）系统，混合 OLTP 和 OLAP 业务同时处理，用于对动态的交易数据进行实时的复杂分析，使得用户能够做出更快的商业决策，支持流、图、空间、文本、结构化等多种数据类型的混合负载，具备多模引擎的分析能力。

分布式数据库的发展呈现与人工智能融合的趋势。一方面基于人工智能进行自调优、自诊断、自愈、自运维，能够对不同场景提供智能化性能优化能力；另一方面通过主流的数据库语言对接人工智能，有效降低人工智能使用门槛。此外，基于异构计算算力，分布式数据库能基于对不同 CPU 架构（ARM、X86 等）的调度进行结构化数据的处理，并基于对 GPU、人工智能加速芯片的调度实现高维向

量数据分析，提升数据库的性能、效能。

## （2）分布式存储技术

随着数据（尤其是非结构化数据）规模的快速增长，以及用户对大数据系统在可靠性、可用性、性能、运营成本等方面需求的提升，分布式架构逐步成为大数据存储的主流架构。

基于产业需求和技术发展，分布式存储主要呈现 3 方面趋势。一是基于硬件处理的分布式存储技术。目前大多的存储仍是使用 HDD（传统硬盘），少数的存储使用 SSD（固态硬盘），或者 SSD+HDD 的模式，如何充分利用硬件来提升性能，推动着分布式存储技术进一步发展。二是基于融合存储的分布式存储技术。针对现有存储系统对块存储、文件存储、对象存储、大数据存储的基本需求，提供一套系统支持多种协议融合，降低存储成本，提升上线速度。三是人工智能技术融合，例如基于人工智能技术实现对性能进行自动调优、对资源使用进行预测、对硬盘故障进行预判等，提升系统可靠性和运维效率，降低运维成本。

## （3）流计算技术

流计算是指在数据流入的同时对数据进行处理和分析，常用于处理高速并发且时效性要求较高的大规模计算场景。流计算系统的关键是流计算引擎，目前流计算引擎主要具备以下特征：支持流计算模型，能够对流式数据进行实时的计算；支持增量计算，可以对局部数据进行增量处理；支持事件触发，能够实时上对变化进行及时响应；支持流量控制，避免因流量或高而导致崩溃或者性能降低等。

随着数据量的不断增加，流计算系统的使用日益广泛，同时传统的流计算平台和系统开始逐步出现一些不足。状态的一致性保障机制相对较弱，处理延迟相对较大，吞吐量受限等问题的出现，推动着流计算平台和系统向新的发展方向延伸。其发展趋势主要包括：更高的吞吐速率，以应对更加海量的流式数据；更低的延迟，逐步实现亚秒级的延迟；更加完备的流量控制机制，以应对更加复杂的流式数据情况；容错能力的提升，以较小的开销来应对各类问题和错误。

## （4）图数据库技术

图数据库是利用图结构进行语义查询的数据库。相比关系模型，图数据模型具有独特的优势。一是借助边的标签，能对具有复杂甚至任意结构的数据集进行

建模；而使用关系模型，需要人工的将数据集归化为一组表及它们之间的 JOIN 条件，才能保存原始结构的全部信息。二是图模型能够非常有效地执行涉及数据实体之间多跳关系的复杂查询或分析，由于图模型用边来保存这类关系，因此只需要简单的查找操作即可获得结果，具有显著的性能优势。三是相较于关系模型，图模型更加灵活，能够简便的创建及动态转换数据，降低模式迁移成本。四是图数据库擅于处理网状的复杂关系，在金融大数据、社交网络分析、推荐、安全防控、物流等领域有着更为广泛的应用。

全国信标委大数据标准工作组



## 4 大数据产业应用现状

### 4.1 大数据产业生态

近年来，我国大数据产业保持稳步增长，大数据技术逐步成熟、应用场景日益丰富，大数据产业生态初步形成。从相关利益方考虑，大数据产业生态中主要涵盖 8 种角色，如图 4-1 所示。

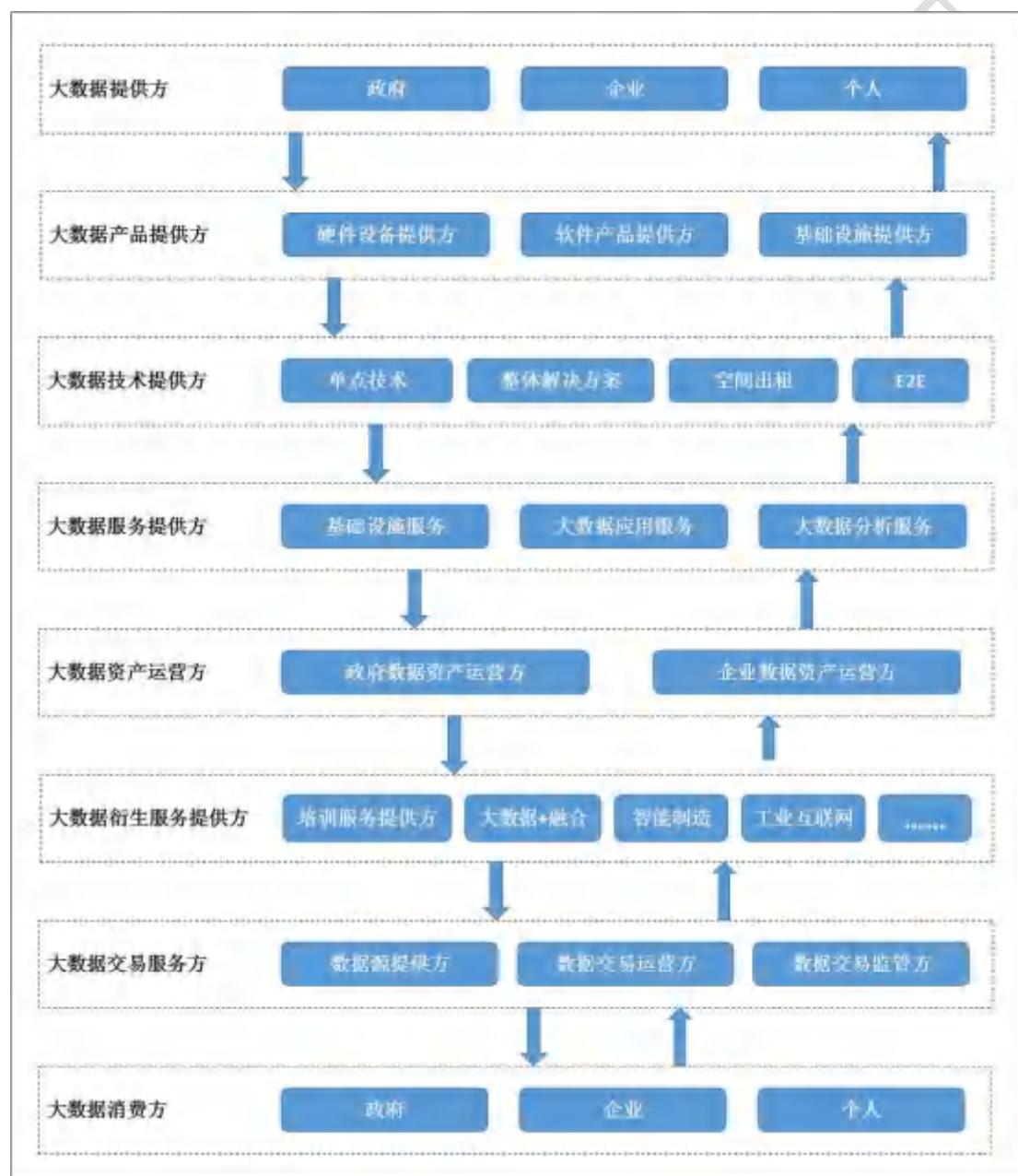


图 4-1 大数据生态参与角色

(1) 大数据提供方：大数据提供方是大数据产业链上的基础环节，其职责

是将数据和信息引入大数据系统，用以提供给其他参与方进行数据访问、处理和分析，主要以数据源提供商以及提供数据流通的平台等形态存在。

(2) 大数据产品提供方：提供大数据相关产品的单位，包括但不限于提供大数据应用软件、基础软件、相关硬件产品的企业等。

(3) 大数据技术提供方：提供大数据技术输出的单位，如各类单点技术输出、整体解决方案输出、E2E 在线大数据技术方案等。

(4) 大数据服务提供方：主要指基于大数据核心技术提供数据价值挖掘服务的企业。包括大数据存储服务提供方、大数据分析服务提供方、大数据基础设施服务提供方等。

(5) 大数据资产运营方：其主要思想是将数据作为一种全新的资产形态，并且以资产管理的标准和要求，来加强相关制度和应用。具体面向数据拥有方提供数据资源规划、数据资产运营管理、数据资产评估、数据资产增值等相关咨询和解决方案服务。

(6) 大数据衍生服务提供方：主要是大数据在各行业、各领域深度融合产生的新业态服务商。通过与各领域深度融合发掘数据价值，拓宽大数据应用场景。

(7) 大数据交易服务方：提供包括大数据资产交易、大数据相关的金融衍生数据设计及相关服务的单位。大数据交易服务方将为数据商开展数据期货、数据融资、数据抵押等业务，旨在建立交易双方数据信用评估体系，增加数据交易量，加快数据流转速度。

(8) 大数据消费方：数据消费方通过使用大数据交易商的分成平台，使得数据提供者获得相应的回报，这样可使大数据产业生态实现良性循环。

从商业模式角度，大数据产业拥有多元化的商业模式，并在此基础上扩展和衍生，具体包含数据买卖模式、信息服务模式、第三方数据服务模式、融合服务模式和软硬件销售模式。

(1) 数据买卖模式。数据买卖模式是指企业直接通过买卖数据取得收入。此类模式的主体是大数据经营商，对大数据的交易是其业务核心，对大数据的重复利用是其发展的原动力。这种公司同时具有很强大的大数据技术能力，多数时候大数据技术本身主要用于自身的运作，例如通过经营大数据交易平台和大数据

API 开发牟利的互联网企业。

(2) 信息服务模式。信息服务模式是指企业将通过隐含在信息服务中的大数据取得利润，这类企业往往具备多种技能，甚至同时具有大数据提供者+技术提供者+服务提供者的能力。这些企业既包括传统的信息技术服务和软件服务企业，也包括咨询、审计、财务、金融等非传统意义上的 IT 企业，信息服务模式是大数据核心产业和衍生产业的相互融合表现最突出的一种模式。

(3) 第三方数据服务模式。第三方数据服务模式是指企业既不是数据的提供者，也不是数据服务的应用者，而是专注通过提供第三方数据服务取得收入。其主体为数据中间商，本身不具有创造数据的能力，从各种地方搜集数据进行整合，通过搭建或提供数据交易平台，从数据中提取的有用信息进行利用或者交易，从而获取利润。

(4) 融合服务模式。融合服务模式是指企业将数据隐含在传统产品及服务中取得收入，这其中既包括提供信息服务的咨询、审计、财务等企业，也包括利用大数据在产业链上下游提供金融、物流等服务而获取利润的制造业企业。融合服务模式是大数据发展的重要方向。

(5) 软硬件销售模式。软硬件销售模式是指各类大数据产业链企业通过对服务和产品直接销售的方式盈利，对于大数据产品提供方来说，软硬件销售模式是他们主要的盈利方式。

## 4.2 重点领域应用

党的十九大报告指出，以更大力度推动大数据和实体经济深度融合。在国家政策的引领和支持下，大数据正在逐步融入国民经济各个领域，并已经形成了一定规模的应用效果。

### 4.2.1 社会管理与公共服务

#### 4.2.1.1 政府治理助力转型升级

##### (1) 助力政府服务转型升级

我国政府在放管服改革中运用数据治理方式，大大提升了政府治理能力，

服务型政府和高效政府建设成果已初步显现。发展过程中通常会面临传统体制、思维模式、基础设施、技术能力以及制度保障等可能的阻滞因素，因此必须针对这些薄弱环节进行改善，为实现数据治理创造有利的外部环境，从而使放管服改革能借助现代科技顺利推进，早日实现我国治理体系和治理能力的现代化。

#### (2) 打造多元化政府治理新体系

政府数据治理以“数据共享、互联互通、业务协同”为原则，运用大数据技术实现政府数据的开放与共享，达到改进政府管理、辅助政府决策、评估政府绩效及监督政府行为等目的，从而构建多元主体共治的政府治理新体系。“跨层级、跨地域、跨系统、跨部门、跨业务”的五跨数据治理正在成为趋势，通过政务服务数据“三张清单”（需求清单、供给清单、负面清单）的梳理、政务信息资源目录的编制以及数据共享交换平台的建设，有力推动了政府数据治理体系的形成，提升了政府对数据资产的管理能力，使政府部门清晰地掌握现有的数据资源情况，并且联动政府内各个部门，让每个人都参与到数据治理的过程中。政府数据治理促使政府决策模式从“经验决策”到“科学决策”、政府治理理念从“权威治理”向“数据治理”、政府治理方式从“模糊治理”向“精准治理”转变，帮助政府提升内部现代化管理水平和对外公众服务能力。

#### (3) 提升政府内部管理水平

在政府内部建立数据治理体系的过程中，通过确定数据治理的决策层、管理层和执行层，带动政府内部全员参与数据治理的过程，加强了组织机构内部的行动力；通过数据治理体系的建设，为数据标准落地、数据质量问题解决、元数据管理加强等提供途径；通过制定政府内部的数据标准、数据质量、元数据、数据安全等管理办法，并定期对数据治理各个阶段进行评估和考核，从整体上提升了政府内部管理水平。

### 4.2.1.2 社会管理与治理能效提升

#### (1) 社会公共安全管理智能化

在公安部“金盾工程”的统一规划下，我国公安信息化建设在近年来发展迅猛，公安信息资源服务平台建设极大的提升了数据资源的种类、质量和服务能力。

在社会服务方面，通过对社会治安总体状况、需求等大数据分析，预测警力的总量需求，分段用警、区域调剂，进而实现管理智能化；在指挥调度方面，建成“情报、指挥、勤务”一体化警务指挥联动平台，利用大数据分析研判警力部署、目标轨迹、案件热点、警情态势等，实现点对点调度、扁平化指挥，缩短响应时间；在警务预测方面，将各警种、各部门的案件登记信息汇集，进行以人、地、事、物、组织等五要素为关联的大数据分析处理，更好地洞察社会治安发展态势，合理布警，预防犯罪；在打击犯罪方面，通过犯罪时空轨迹分析等技术深度分析案件特点，以及对 DNA、指纹、音频等数据的处理、比对和分析，实现精确打击犯罪。未来如何更加充分挖掘警务信息数据的价值、更加进一步推动警务大数据的实战应用将是未来公安部门工作关注的重点。

### （2）社会治理高效化与精细化

随着公安机关投入的加大，已经逐步建立了完善的公民身份信息库、指纹库、档案库。通过人工智能技术可以对社会公共管理大数据进行科学分析，找到城市治理中的薄弱环节，如社会治理盲区、事故高发区域等。还可以科学划分城市不同区域的安全等级或风险等级，从而科学决策和指导城市级公共安全建设。人工智能技术提高了数据解析和挖掘的精细度以及准确性，使得大量有效数据为社会公共安全领域服务，将公共安全事务变得可视化、透明化，提升公共安全管理效率，抑制潜在的公安安全危害，为实现高效精细的城市级社会综合管理打下基础。

### （3）信用化监管降低信息不对称风险

信用大数据是通过大数据技术采集、处理、管理、分析多样且海量的信用信息，提取其中的信息价值，形成反映各类社会主体信用状况的大数据，是信用体系建设的基石。信用大数据可实现决策管理、商业智能、智能预警、风险防控等应用，从而推动社会信用体系建设的全面发展。通过大数据技术将分散的数据基于信用业务场景整合关联，促进信用信息基于信用价值链进行交换共享，缩小信用主体之间在空间与时间上的不对称，建多元化信用风险评估预测模型和信用知识图谱，将信用大数据广泛应用于经贸工商、交通出行、机构团体、文化休闲、医疗卫生、教育科技、社会民生、资源环境、公共安全、农业农村、社保就业、

财税金融、信用服务等社会管理与治理领域，助力社会管理与治理能力提升。

### 4.2.1.3 政府公共服务效能提升

#### (1) 提升社会治理精准度

在大数据时代，决策者可以通过对海量的实时数据的掌握和挖掘，把握数据中蕴含的规律性、倾向性问题。除了在提升行政效能方面，政府还可在智慧医疗、智慧交通、智慧城管、市场监管等方面做出创新及突破，并利用大数据精准解决城市痛点问题，通过数据之间的关联性分析，从社会发展的角度做好精准化管理，提高社会治理精准度。

#### (2) 助力公共服务高效化

我国“互联网+政务服务”深化发展，各级政府依托网上政务服务平台，推动线上线下集成融合，落地“一网通办”等一体化在线政务服务平台，让企业和群众办事“上一张网，办所有事，最多跑一次，一次能办成”，提升政务服务的质量和效率，切实落实优化城市营商环境的工作目标。

#### (3) 提升人力资源监管服务水平

我国是人力资源大国，随着市场经济的不断深化发展，人才流动成为常态。大数据技术可以帮助人力资源管理部门从凭借经验模式逐步向依靠事实数据模式转变，从而使人力资源的配置朝着依靠多维数据技术和联盟化的统一的大数据共享平台的方向发展，有效提高人力资源数据分析科学性、人力资源配置的精准度、人才培养及其智力成果转化与人才资源利用的效率，助力政府准确把握宏观就业形势、出台适宜的人才政策、优化人才引进和发展等问题。

#### (4) 助力政府审计工作智能化

随着科技和社会的发展，信息技术和计算机网络的兴起，审计数据呈现出数据体量大、数据类型多样性、数据价值密度低、数据处理速度快等大数据独有的特点，审计手段也由原来的查看账表进入了大数据审计。大数据审计旨在将“数据孤岛型、业务封闭型、模式单一型”向“数据一体化、业务一体化、模式一体化”的方向推进发展。通过多部门、多行业，多地区获取数据，互相对比、互为佐证，将审计范围横向和纵向大幅度延伸。通过对相关领域海量历史数据分析，

形成对相关政策实施的效果进行跟踪审计和评估，挖掘出某种群体行为的特点，警示某种社会现象的潜在规律，为政府制定政策提供关键依据，有利于提高政府预测以及应急响应能力，使审计更好服务于国家治理。

#### 4.2.2 农业

随着国家乡村振兴战略的提出，农业作为支撑国民经济发展的基础产业，正在被互联网、大数据、人工智能等新科技进行重构和升级。农业大数据已广泛应用于农业生产、经营、管理、服务等全产业链，涉及农业生态环境监测、农产品质量安全溯源、设施农业、精准农业以及农资投入、农机服务、农技服务、金融农投、农业保险等各涉农环节。

在农业生产中的应用。通过对气象、疫病及灾害数据进行分析，及时对疫病、灾害进行评估预警，提升对农业生产过程中病害、自然灾害的防控。通过监测气候、土壤和空气质量、农资使用情况、作物生长情况等，预测环境因素变化，推进精准农业，进行精准种植、养殖，最大限度优化农业生产效率。

在农业经营中的应用。通过加强农产品价格周期波动规律研究，一方面为农业生产主体提供较为精准的预测与预警信息，并指导制定生产计划，在市场价格波动中减损增效，降低生产风险；另一方面，对农产品整体供需关系、价格影响因素剖析、流通渠道和流向分析，形成产销过剩或紧缺预警，保障农产品供需平衡，优化产品结构及销售渠道，制定正确的经营战略。

在农业监管中的应用。通过农业全产业链数据资源统筹共享，提高监管部门工作效率和业务监管能力，实现对土地资源的实时监测，及时发现基本农田侵占、耕地红线低值预警、农村耕地水资源污染等情况，提高土地流转透明度，实现土地经营权管理和跟踪。

在农业服务中的应用。基于农业经营主体信用档案、行政许可与行政处罚、认证或登记、监督检查和奖励等信息，构建信用评分模型，实现对涉农生产主体信用评估，满足农业生产经营主体的信贷资金需求，拓展融资渠道。

### 4.2.3 制造业

在“中国制造 2025”国家战略的指导下，伴随着一批智能制造试点示范项目和工业互联网试点示范项目的建设，大数据已经在智能化产品与装备、智能化生产、智能化服务等领域形成了典型应用模式，并逐步深入到产品研发设计、生产制造、服务保障等核心业务环节，大数据已经成为制造业生产力、竞争力、创新能力提升的关键。

首先，基于统一标准化思路驱动的工业大数据产品研发设计，实现了研发过程的智能化，提升了创新能力、研发效率和设计质量。通过产品全生命周期数据的采集，工业大数据建模和数字仿真技术优化设计模型，及早发现设计缺陷，减少试制实验次数，降低研发成本、提升设计效率，缩短了产品研发周期。其次，综合制造过程中设备、效率、成本、耗能等数据展开建模分析，实现了运行过程的状态监测与优化工艺参数推荐。通过生产工艺过程参数，设备运行状态参数与产品质量性能、生产线排产负荷、耗能等数据进行关联性深度挖掘，形成数据闭环，可得出工艺参数的最优区间、车间排产计划的最优方案、厂房能效优化的最佳调控手段等。

工业大数据技术的发展和相关标准化工作的推进，也带来了制造业产业链上下游企业间各协同环节的信息共享和同步升级，企业可根据自身优劣势分析对业务进行重新取舍，整合资源实现平台化运营，优化价值链。另外，基于大数据构建的产品故障预测系统，能帮助用户实时掌握产品状态，在产品出现异常前展开预测性维修。基于数据标准化思路的企业全流程的数据集成贯通与工业大数据建模分析，支撑了大规模定制为代表的典型智能制造模式。基于研发知识库的大数据产品模块化分析，以及协同创新平台所整合的内外部产业链协同设计能力，可实现产品的个性化设计；基于工业生产大数据的互联工厂柔性化生产能力，保障了个性化设计订单低成本高效率的制造；结合物流大数据分析优化的物流配送系统，可充分保障个性化定制产品在最短时间内按承诺交付至用户。



## 4.2.4 能源领域

目前，大数据在电力、石油天然气、煤炭和新能源等能源行业的应用，主要集中在通过综合采集、处理、分析与应用能源领域生产经营数据以及人口、地理、气象等诸多相关领域数据，企业可以优化库存、合理调配产品供给，并对数据实时分析，加速推进能源产业发展及商业模式创新，大数据在能源领域的成功应用充分体现了能源生产和消费方式与大数据理念的深度融合。

### 4.2.4.1 电力

电力随着智能电网的加速建设，电力行业通过传感器、智能设备、视频监控设备、音频通信设备、移动终端等各种数据采集渠道，广泛收集发电、输电、变电、配电、用电和调度等生产利用环节以及天气、气象、交通、经济等外部支持系统的海量业务数据。通过集中式和分布式混合架构的企业级电力大数据平台，重点推进大数据在输变电智能化、智能配用电、源网荷协调优化、智能调度控制、企业经营管理和信息通信等六大领域的支撑作用。在电力生产方面：通过关联内外部数据对设备状态、电能负载等数据进行分析挖掘与预测，开展电力生产设备精准调度、故障判断和预测性维护，提高发电设备利用效率和安全稳定运行水平。在电力设备运行维护方面：增强大数据技术与变电、输电和配电等运检全业务流程深度融合，消除数据壁垒，实现数据资源的集中和精益化管理，推动运检业务管理模式变革，实现运电流程高效协同。在电力信息通信方面：通过基于机器学习的数据质量提升、自动巡检和远程诊断的设备智能检修以及网络运行状态智能诊断和辅助优化等应用过程中逐渐积累的海量数据，提升对通信传输设备运行状态的深度分析、预警和预控能力，实现电力通信网的智能运维和辅助决策。在电力营销方面：帮助电力公司调配电力供给，通过错峰限电，引导用户在电力成本低的时间段用电，为客户提供最佳用电方案；通过可开放容量等数据的挖掘分析，实现合理规划配电网，优化业扩报装流程，满足用户新增负荷需求，避免因信息掌握不准确造成接入新负荷后出现设备过载现象。

#### 4.2.4.2 石油天然气

石油天然气领域的大数据应用主要应用于优化业务链、设备运营状态诊断、市场营销分析、能源管理分析等领域。通过对原油和天然气等两条业务链除开采环节外，实现上游产能建设布局优化分析、下游炼化物流优化与排产决策等，提高生产运行效率、降低生产成本并实现节能降耗。基于广泛实时采集自动控制和物联网数据分析各类重要设备的运行参数，根据预置及通过学习获取的各类分析模型对设备故障进行预警，通过预防性维护有效降低设备运维成本，避免不必要的停工损失。针对行业内的成品油、天然气、化工等产品进行市场及客户分析，帮助制定有针对性的营销方案，更好满足不同类型的客户需求。通过自动采集并汇总分析各类型用能、用水数据分析，查找存在的“跑冒滴漏”等异常问题，结合用能设备类型开展能耗分析评价，实现节能降耗。

#### 4.2.4.3 煤炭

中国煤炭工业协会会同中国煤炭运销协会于 2017 年底推出“中国煤炭大数据平台”并上线运行，平台全面整合全国四十余万家煤炭生产、消费及相关企业最新基本信息，实时把控煤炭流向，精确锁定竞争对手，从煤炭供给、煤炭消费、煤炭库存及煤炭进出口等角度进行深度整合与解析，全面解读各地区煤炭供应、消费和库存结构，动态反映煤炭产业链下游用户心理行为的下游接货意愿指数等功能。煤炭行业已实现煤炭生产、运输、销售、安全、资源等数据集成覆盖以及与相关市场主体的数据共享。

#### 4.2.4.4 新能源

通过构建数据汇集、存储、服务、运营一体的新能源大数据创新平台，接入风力发电场、清洁能源电站数据，开展数值天气预报和功率预测，通过对比分析美国、欧洲以及中国气象部门多元气象数据，以高精度数值预报，可提供新能源发电量精准预估，使系统在全国范围内具备提供数据服务能力，实现不同风速、风轨迹以及辐照度等多种气象要素预测，确保电网安全稳定运行，促进新能源消纳。平台通过市场化手段，吸引各类能源企业、工业企业和不同服务提供商等入

驻平台，实现新能源电站集控监控、集中功率预测、生产管理、节能降耗、设备健康诊断、业务流程优化等业务，初步构建了覆盖全产业链的能源互联网生态圈。

#### 4.2.5 通信领域

通信领域具有典型的超大规模海量数据特征，利用大数据分析技术，采取虚拟化的存储途径，使得数据各种结构类别都能在同一平台中进行整合存储，节约资源减少成本。通过数据挖掘技术对用户的行为进行数据分析和预测，不仅可以分析用户的喜好与购买能力，还可以预测用户日后的购买方向，从而实现经济效益最大化。在移动站点布局、网络结构优化方面，为使网络运行顺畅，达到最好的状态效果，基于大数据技术支撑，通过遗传算法、神经网络、模拟退火算法等算法模型，找出最合适的基站建设点，实现干扰分析、话务预测、掉话处理等预测分析。同时，为应对通信业务对大数据平台在数据规模、性能、扩展性等方面的需求，通信大数据系统也呈现出敏捷、智慧、可信的发展趋势。

#### 4.2.6 交通运输领域

随着交通运输服务与管理信息化、智能化的发展，大数据在交通运输领域中的应用成为热点。大数据的应用使人们重新认识了交通需求以及交通运行的内在规律，同时也改变了交通运输的规律。要推进交通运输领域大数据应用的发展，各级交通运输主管部门间、政府与企业间需密切合作，建立数据定义、处理、交换等标准的完整体系，才能取得共赢。为落实国务院《关于促进大数据发展的行动纲要》和《国务院关于印发政务信息资源共享管理暂行办法的通知》的有关要求，攻克交通运输领域在数据资源共享利用中的难题，交通运输部印发了《交通运输部办公厅关于推进交通运输行业数据资源开放共享的实施意见》（交办科技〔2016〕113号）。随后，交通运输部印发了《交通运输部政务信息资源共享管理办法（试行）》。在国家层面，通过打造国家综合交通运输信息平台，落实交通运输行业实施国家大数据战略的工作任务，实现了行业相关业务系统、平台的统一整合。同时，交通运输部印发的《国家综合交通运输信息平台总体技术方案》中，对国家综合交通运输信息平台建设应遵循的标准规范给出了统一要求。为进

一步深入推进交通运输信息资源开放共享和融合应用，强化信息资源整合共享工作中的应用导向起到了积极推动作用。

#### 4.2.7 生态环境领域

党中央、国务院高度重视大数据在推进生态文明建设中的地位和作用。习近平总书记明确指出，要推进全国生态环境监测数据联网共享，开展生态环境大数据分析。李克强总理强调，要在生态环境等重点领域引入大数据监管，主动查究违法违规行为。运用现代信息技术加强政府公共服务和市场监管，推动简政放权和政府职能转变，构建“互联网+”绿色生态，实现生态环境数据互联互通和开放共享。通过推进督察信息化建设，加大卫星遥感、无人机、大数据等技术应用，不断提高生态环境督察质量和效率。在党中央、国务院和部党组重视下，生态环境信息化建设已取得一定的进展，2018年独立设置信息中心，2019年组建网络安全与信息化领导小组及其办公室，2019年3月14日召开了生态环境部网络安全与信息化领导小组2019年第1次全体会议，通过强化生态环境信息化标准体系建设工作，将“统一标准”列为生态环境信息化工作“四统一、五集中”总体要求之一。网络安全与信息化领导小组办公室设置综合管理组，专门负责统筹生态环境信息化标准工作，切实推进和落实统一规划、统一标准、统一建设、统一运维、数据集中、人员集中、技术集中、资金集中、管理集中等具体工作要求。

#### 4.2.8 司法领域

新一代信息技术的广泛应用，对中国司法体制、机制、工作方式等产生了深刻影响。通过推动大数据、人工智能等科技创新成果同司法工作深度融合，实现了检察大数据应用、法院大数据应用等具体工作实践。

在检察大数据应用领域，提升服务司法办案效率，建设了智能检索、类案推送、量刑建议等应用，可有效促进检察监督职能作用的发挥。在服务管理决策方面，建设了绩效管理、法律主体（案件当事人、检察官）画像、大数据专题分析等应用，可提供客观、精确的决策支持；在服务司法为民方面建设了全方位的检务公开应用，有效提升了检务公开水平。

在法院大数据应用领域，提升服务人民群众的满意度，基于对当事人诉求、案情的认知，自动生成诉状，并给予当事人诉情解析、诉讼风险评估、推送已决相似案件。在服务审判执行方面，根据案件诉讼材料、当事人诉请等，为立案法官提供案由自动识别服务，智能关联整合历史数据，自动填写案件结构化信息；对案件进行智能案情分析，辅助法官快速了解案情全貌；对控辩双方提出的事实和证据进行审查，为非法证据排除提供知识辅助服务；为书记员提供庭审笔录自动生成服务；为法官办案提供法律条款、相似案例、判决参考等数据支撑；自动生成裁判文书，提高文书编写的效率和质量。在服务司法管理方面，提供庭审自动化巡查；辅助相关人员对裁判量刑进行把控，避免出现量刑偏差；通过对卷宗随案生成情况自动检查，对文书质量智能校验，为违规案件的督办以及案件的规范化管理提供帮助。

在智慧监狱大数据应用方面，可对罪犯进行精准画像和危险性评估，对罪犯劳动改造、教育改造的成效进行评估，对监狱安全风险进行评估；在智慧戒毒大数据应用方面，可对戒毒人员进行画像以及危险性、违纪率、逃脱率等指标分析，并形成毒品犯罪等专题分析报告；在智慧调解大数据应用方面，可对案件、调解员和当事人进行画像，可对区域、调委会和调解员的调解效能进行指数分析。

在司法协同大数据应用方面，建设政法跨部门大数据办案平台，横向打通公、检、法、司等政法部门业务协同流程，实现刑事案件跨部门网上协同办理以及政法各单位之间电子文书、电子卷宗、结构化数据的共享交换；对刑事案件、治安情况的变化趋势、涉案人员地域分布、涉案财物流转信息进行智能分析，为执法办案和社会治理提供参考，为重点人员管理提供线索，精准预测风险。

#### 4.2.9 市场监管

在政府机构改革浪潮中，新组建的市场监管部门整合了原工商、食药、质检等多部门职能，其工作开展面临着监管任务重、要求高、市场形势复杂等一系列挑战。按照国务院《“十三五”市场监管规划》和《关于运用大数据加强对市场主体服务和监管的若干意见》等政策文件要求，市场监管部门需要积极运用大数据、云计算等先进手段，提升市场监管与综合服务能力，促进市场经济健康和谐

发展。通过对市场监管内外部数据的汇总关联融合，强化数据动态分析比对，预警系统性、区域性的市场异常，有效提升了市场监管效能。

药品安全监管作为市场监管的重要组成部分，是关乎民生的重大工程，直接关系到百姓生命健康和社会和谐稳定，2019年5月，国家药品监督管理局印发《国家药品监督管理局关于加快推进药品智慧监管的行动计划》，文件指出要坚持以人为本、创新发展、统筹整合、共享开放的基本原则，通过监管工作与云计算、大数据、“互联网+”等信息技术的融合发展，构建监管“大系统、大平台、大数据”，推进信息技术与监管工作深度融合，形成“严管”加“巧管”的监管新局面。在药品监管信息化建设中，采用国家局主导统筹、省局协同参与的方式，加强信息化标准的宣贯与应用，建设药品监管应用平台和数据共享平台，促进药品监管信息共享、业务协同和大数据应用。

#### 4.2.10 金融领域

如今，“数字蝶变”席卷金融行业各个领域。随时、随地、随人、随需的普惠金融服务也成为数字金融时代的标志之一，金融机构转型已经是行业共识。大数据、云计算、人工智能、区块链等技术正在逐步与金融业务进行深度融合，推动行业进行转型升级。金融大数据在供应链金融、智能营销、信用评估、风险定价、反欺诈、风控、股市行情预测、黑产防范等具体业务中都有着广泛的应用。大数据技术逐渐向金融领域各个业务场景进行深入渗透，带来了整体效率的提升以及服务模式的转变，促进了金融领域的创新发展。银行业、保险业、证券业等领域都对于金融大数据的应用进行了广泛的探索。

大数据在银行业的典型应用，主要体现在智能营销、风险防控、供应链金融管理等领域。在精准的用户画像基础上，银行通过金融大数据能够开展有效的智能精准营销，对用户流失进行预警，深层次理解用户特征和风险偏好，智能预测用户需求，进行交叉和个性化推荐等。金融大数据在风控中的应用主要体现身份验证、授权、贷中监控等环节，协助整合企业内外部数据，对数据进行实时动态挖掘，更准确估算用户价值、信用额度和预测违约概率等。供应链金融授信主体是整个链条，银行将金融大数据应用在供应链金融中，可以根据企业间关联以及

企业法人和股东的关联关系，挖掘企业关系图谱，利于供应链金融的风险控制和企业关系分析。

大数据在证券业的典型应用，主要体现在股价预测、用户关系管理等领域。随着大数据技术在证券行业的应用，对非结构化的数据以及结构化的数据都能进行收集和分析，使得对于市场情绪分析变为可能，实现股价预测的辅助决策。通过对用户的账户状态、交易习惯、风险偏好等进行对用户聚类 and 细分，为用户提供个性化服务；通过历史数据进行建模对用户流失进行预警。

大数据在保险业的应用，主要体现在骗保识别、精细化运营等领域。通过金融大数据结合企业数据，保险企业能够对于保险欺诈进行建模，从而能够较为准确分析和预测欺诈等非法行为。借助金融大数据可为用户提供个性化解决方案，进行用户关联销售，流失用户预警、潜在用户挖掘、用户生命周期管理等。

网络泄密等数据安全问题出现，推动大数据在金融领域的发展重心逐步向数据安全防护转变，国家有关部门也出台相应政策加强金融行业数据安全，数据安全必然是金融大数据下一步重要发展方向。同时，随着金融大数据与实际业务的紧密结合，产生了更为明晰的产业链分工以及更为专业化的服务内容，整体呈现服务专业化、多样化的趋势。

#### 4.2.11 科学研究

国家层面相继发布了《科学数据管理办法》、《国家科技资源共享服务平台管理办法》，旨在规范管理国家科技资源共享服务平台，推进科技资源向社会开放共享，提高资源利用效率，加强和规范科学数据管理工作。多个省市、机构先后发布了相关实施细则。2019年6月5日，科技部、财政部联合发布了国家科技资源共享服务平台优化调整名单的通知，共形成20个国家科学数据中心和30个国家生物种质与实验材料资源库。国家科学数据中心承载着国家科学数据顶层设计实施重大使命，以及资源汇聚整合、资源开发应用与分析挖掘等工作的方向性布局，将对全国的科学数据工作起到示范性引导作用。可以说，国家科学数据中心既是当前国家创新体系的基础要素，又是国家创新体系的重要引擎之一，也是变革未来创新模式的重要推手。

科学大数据工程作为中国科学院十三五信息化专项建设五大工程之一，是中国科学院 30 多年科学数据工作的继承，也是大数据新时代的发展。十三五期间，科学大数据工程继续完善科学数据资源体系，建成空间科学领域、生物信息领域、第三级环境、资源学科领域等 4 个大数据驱动的创新示范平台；喀斯特地形地貌区重点数据库、动物学重点数据库、化学学科领域重点数据库、农业多尺度病虫害图像重点数据库、南海海洋重点数据库、环境微生物多样性重点数据库、光学天文重点数据库等 7 个重点数据库；湖泊学科领域特色专题数据库、中国土壤特色数据库、多民族语言资源特色数据库等 20 个特色数据库。截止 2018 年底，累计积累数据资源总量超过 2.8PB。

行业科研数据异军突起、蓬勃发展。2018 年 2 月，国家烟草专卖局印发通知，启动实施烟草科研大数据重大专项。探索烟草科研大数据中心可持续发展模式，为行业科研大数据工程研究中心和国家级大数据工程实验室分中心建设奠定基础。2018 年 10 月，由自然资源部中国地质调查局主持研发的国家地质大数据共享服务平台——“地质云 2.0”宣布上线服务，旨在通过数据资源整合和信息系统集成，全面提升地质调查数据采集、汇聚、处理、分析、共享与服务能力，为新时期地质调查工作转型升级提供核心动力，以信息化带动地质调查现代化。

目前，以科技部国家科技基础条件平台、中国科学院科学数据等为代表的基础科学数据资源建设和应用服务已经取得了一定成效，未来将尽快加强权威科学数据资源建设，构建基于知识网络的一体化数据基础设施；加快数据核心技术研发布局，重视面向应用的自主管理软件开发，部署关键技术攻关，最终打破国外科学数据资源垄断，实现我国从数据大国向数据强国的转变，助推我国产生若干具有世界领先水平的重大科学发现与原创技术突破，推动科技创新能力和社会经济发展。

#### 4.2.12 卫生健康

医疗健康大数据的应用主要围绕三方面开展：一是围绕建设和完善以居民电子健康档案、电子病历、电子处方等为核心的基础数据库；二是建设统一权威、云联云通的人口健康信息平台，推动大数据资源共享开放；三是建设统一的大数



据分析和应用系统，主要包括管理决策大数据应用、健康医疗临床和科研大数据应用、公共卫生大数据应用、健康管理大数据应用等。通过推进大数据在医疗健康领域的应用，一方面，可以让决策者多角度全局性的掌握医疗机构运营的总体情况，实现医院精细化管理；另一方面，能够对体制改革进行合理的监测与评估，使优势资源“下得去”，助力实现分级医疗效果的科学评估，合理进行资源配置，更好地推动分级诊疗落地。

在诊疗大数据应用场景方面，医疗临床及科研大数据应用是通过人工智能技术对影像、基因和诊疗文书等多源异构数据进行统一集成、清洗、加工等多道数据治理工序，结合医学术语库、知识图谱实现临床数据向科研数据高效转化，让科研项目可以低成本使用全维度数据，为疾病科研提供基础数据支撑。其次，通过智能筛选、智能预分析，挖掘相关性和有效性证据，实现一键式验证科研思路，提升科研效率。最后，基于海量医疗数据的深度有效利用，可开展人工智能辅助诊断、临床辅助决策、疾病预测等研究，优化诊疗路径，提升诊治水平。

在公共卫生大数据应用场景方面，通过大数据能够监测和分析疾病模式、追踪疾病爆发及传播方式和途径，有效进行疾病的监测评估、预防与快速干预，提高对公共卫生事件的监测和反应速度，进而更快更准确地制定公共卫生政策并协调开展相应的服务。

在健康管理大数据应用场景方面，主要包括建设覆盖全生命周期的预防、治疗、康复和健康管理的一体化大数据健康服务。一是基于数据引导居民进行疾病的有效预防；二是通过大数据实现个性化健康管理、慢病管理，让居民在医院、社区及线上的服务连续性得到保障；三是不断增强“自主健康”服务体验，让健康数据“多跑路”，让人们群众“少跑腿”，给百姓带来更加便捷的应用与服务。

## 4.2.13 文化、教育和旅游领域

### 4.2.13.1 文化和出版

大数据已成为驱动文化出版领域知识服务转型的重要技术力量，知识聚合、链接数据、知识图谱、知识管理、知识服务等大数据及人工智能领域的新兴技术

热点正被快速应用，推动文化出版从“数字出版”向“大数据出版”转变，服务知识驱动下的创新发展，实现知识相关业态深入转型融合。文化出版业覆盖数据采集、输送、加工、编辑制作、分析、存储、使用、传播、服务等全过程。

在大众出版相关的应用场景中，大数据技术推动图书出版数字化转型从“内容数字化”走向“出版流程数字化”，对社会热门事件、热点主题传播的频度和热度等进行大数据分析，降低前期人力投入，提高选题策划的含金量。在印刷发行环节，实现智能印刷，按需出版，精准分发，达到消除库存、减轻资本压力的目标。在营销传播阶段，挖掘读者深层次需求，提供个性化推荐清单，将知识内容及时、精准、定向推送。通过数字阅读平台与在线支付平台的无缝衔接，推动知识付费，帮助出版机构转型创收。

在教育出版相关的应用场景中，数字化教育是教育类出版社谋求优化升级的布局重点。在数字教育中引入大数据技术，动态地对学习行为、过程和结果进行深度分析，全面诊断学生的知识需求、知识水平、认知能力和学科素养，有针对性的提供学习报告、学习建议、学习方案和学习内容，开展个性化学习和自主学习。同时，为教师推送教学建议、教学方案及学情报告，支撑实现因人施教、因材施教。大数据技术使得数字教育能够有针对性地提供个性化、差异化的学习和教学服务，帮助教师和学生获取优质教学资源和学习资源，消除“信息鸿沟”，促进教育公平、均衡发展。

在专业与学术出版中的应用场景中，大数据技术能够加速学术传播，推动出版流程各个环节之间的数据传递，专业与学术出版领域已出现创新网络首发、增强出版、协创出版、数据出版、碎片化出版、云出版等多种出版方式，在出版流程中采用集大数据智能校验、学术不端检测、大数据科学评价等工具为一体的全过程探究式学习与科研模式，将海量知识大数据与科研、学习等业务过程融合，在实现精准知识服务的同时，推动“产学研用一体化”的知识创新发展服务。

在新闻传媒领域的应用场景中，新闻传媒领域以大数据技术为支撑，将内容数据、流程数据与用户数据进行全面聚合、关联和深度挖掘，重塑业务流程与机制，打通、整合“报、网、端、微”的采访、编辑和技术力量，全力打造集“融合策划、融合采集、融合加工、融合传播”为一体的融媒体中央厨房，实现“一

次采集、多种生成、多元传播”。

#### 4.2.13.2 文化旅游

“十三五”时期是中国文化和旅游产业进入大数据时代的历史机遇期，新业态不断涌现，文化复兴与旅游繁荣已成为新时代的重要特征。文化是内容，旅游是载体，科技是手段，“文化+旅游+科技”的新模式正在重构文旅产业链，不断丰富人们的文化和旅游新体验。新时代人民群众日益增长的美好生活需要，大众旅游时代广大游客的个性化、自主化、多元化需求，都在快速推动文化和旅游产业的数字化转型。大数据与文旅产业的紧密融合，是促进我国文旅产业经济发展与弘扬传统文化的重要途径。

为满足新时代文化和旅游业快速发展过程中对数据的需求，我国各地已开始陆续建立国家-省（直辖市）-市（州）-县（区）-文旅企事业单位（文化场馆、景区等）五级联动的各级文旅数据中心。为各级文旅管理部门、文化场馆、文化企业、涉旅企业提供了行业监管、数据分析与预测、精准营销与服务、招商引资与产业发展等信息服务与数据支撑，对于创新产品供给、改善经营管理、提高服务质量、促进精准营销等产生了重大意义。

文旅大数据在文旅管理部门中的应用。通过整合全国或区域范围内的文旅数据，实现文旅管理部门的数据采集、数据共享交换和数据分析决策。建立面向应用的多层级文旅大数据中心，建设文旅大数据基础平台、文旅大数据共享交换平台、文旅大数据应用服务平台，实现产业全景监控，对文旅产业的运行状况进行动态监测，为客流实时监测与预警、应急指挥调度、政府管理与服务决策、舆情监测、市场分析与品牌推广等应用服务提供数据支撑。

文旅大数据在文旅企事业单位中的应用。通过大数据应用，实现对文旅企事业单位（博物馆、景区、文旅特色小镇、乡村旅游点等）的基础资源、实时游客流量、环境与自然灾害、营销数据等运行状况进行实时监测。通过将用户各种属性数据、消费过程中各类数据采集整合，基于大数据的企业画像分析、游客画像分析、营销分析，为文旅企事业单位实现游客个性化服务、产品精准营销提供数据支撑。通过基于电子地图的文旅公共信息服务平台，为游客提供个性化、定制

化、智能化的信息推送，将信息查询、导游导览导购导航、产品推荐、在线预订、营销宣传、投诉评价等功能集为一体，实现“一部手机游”的智慧服务体验。

全国信标委大数据标准工作组

## 5 大数据热点领域概述

### 5.1 大数据治理

随着大数据技术在各领域应用的不断深入，数据价值变现能力越来越高，数据确权、数据质量、数据安全、数据流通等问题受到业内关注，并引发各界深度思考，如何做好大数据治理工作，成为大数据产业生态系统中一个新的热点。

#### (1) 大数据治理体系

数据治理是对数据资产行使权利和活动控制的集合，是数据管理体系的核心，并用于评估、指导和监督其他相关数据管理职能的执行。大数据治理相比于数据治理，在数据治理对象、数据处理架构、治理组织职能、数据管理措施、数据应用范围等方面呈现多层次、多形式、大范围等特点。围绕数据资产、共享开放、安全与隐私保护等的大数据技术应用的新需求，大数据治理不再仅限于单一组织数据治理范畴，而是要从国家层次、行业层次、组织层次构建形成一个自上而下、多元共治的数据治理体系。大数据治理体系框架如图5-1所示。

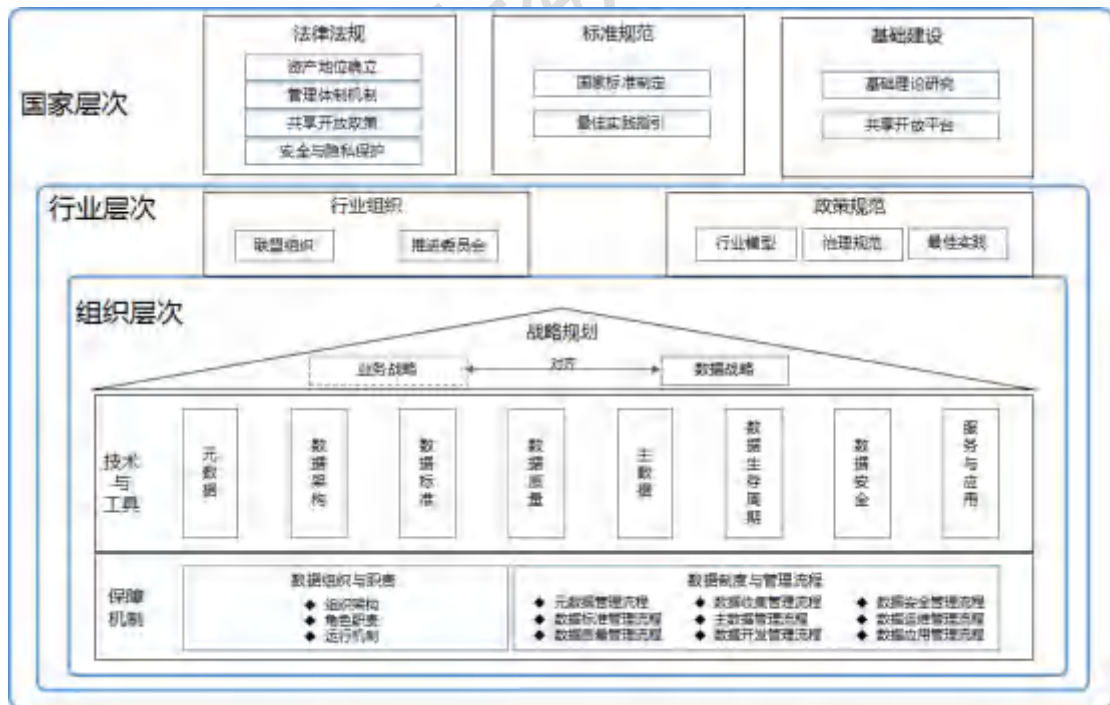


图 5-1 大数据治理体系框架

国家层面，需要通过政策法规支撑大数据治理建设。一是需要从国家法律法

规层面明确数据资产地位，确定数据权属规则，完善数据隐私保护，为大数据治理提供安全可靠的政策、法律环境。二是需要通过国家标准规范数据管理机制，构建业内协调统一的数据治理标准体系，保障数据产业的健康有序发展。三是需要通过建设政府主导的数据开放共享平台推动业内数据流通，深化数据资源应用，实现数据价值挖掘。

行业层面，在国家相关法律法规和标准体系建设的基础之上，需要面向金融、制造、能源等各领域具体需求，建立完善行业大数据治理指引，引导完善行业内部数据共享与开放规则，推动行业标准、治理模型建设，开展行业内部最佳实践积累，逐步形成面向行业业务需求的数据治理体系。

组织层面，需要明确企业数据资产核心地位，构建数据治理、数据管理体系。一是确立企业的业务战略和数据战略；二是建立数据组织、明确管理职责，制订数据管理制度和管理流程，形成大数据治理体系保障机制；三是依据企业数据现状和业务现状规范元数据、数据架构、数据标准、数据质量、数据安全、数据应用等具体管理活动并明确相关管理职能。

## **(2) 大数据治理研究现状**

当前，全球各国数据治理的主要工作大多集中在企业数据治理和行业数据治理领域。2008年12月，Gartner发布了企业信息管理成熟度模型(the EIM Maturity Model)，把数据管理成熟度划分为六个阶段来进行定义，帮助IT管理者识别、确认各自企业信息管理的水平。2010年9月，IBM发布了《IBM数据治理统一流程》(The IBM Data Governance Unified Process)，描述了企业进行数据治理工作的必要和非必要步骤，指导企业推动数据治理实施。2015年2月，EDM Council基于众多实际案例的经验总结编写并发布了数据管理能力评估模型(Data Management Capability Assessment Model, DCAM)，定义了数据能力成熟度评估涉及的能力范围和评估的准则，从战略、组织、技术和操作的最佳实践等方面描述了如何成功地进行数据管理。此外ISO/IEC JTC 1/SC 40 (IT服务管理与IT治理分技术委员会)推出了ISO/IEC 38505-1《信息技术 信息技术治理 数据治理 第1部分：ISO/IEC 38500在数据治理领域的应用》、ISO/IEC TR 38505-2《信息技术 信息技术治理 数据治理 第2部分：ISO/IEC 38500在数据管理方面的执行》

等多项数据治理国际标准，提出了数据治理的意义、原则、模型和特征，明确了数据治理的任务、实施导则和应用等内容。

我国数据治理起步晚、发展快。在借鉴国际理论经验的基础上，结合自身实际探索数据治理模型和方法，2018年4月，全国信标委大数据标准工作组编制的GB/T 36073-2018《数据管理能力成熟度评估模型》国家标准正式发布，为我国各行业领域的有关单位开展数据管理、数据治理建设提供了重要依据。基于该标准研制基础，金融、电力、军工等多个领域开展数据管理行业规范制定以及数据治理实施，构建完善的数据治理产业氛围。部分行业数据治理发布文件见表5-1。

表5-1 我国部分行业大数据治理相关文件

行业	发布文件	发布单位	发布时间
银行	银行业金融机构数据治理指引	中国银监会	2018年5月
军工	军工行业数据管理能力成熟度评估指南	国防科技工业局	2018年12月
证券	证券公司数据治理操作指引	中国证券业协会	编制中
电力	电网企业数据管理能力成熟度评估模型	中国电子工程学会	2019年2月

## 5.2 大数据与人工智能

大数据与人工智能作为现代计算机技术的重点发展方向，是众多垂直领域应用解决方案的重要支撑技术。大数据技术演化的总体目标是高效收集、存储、处理与分析大规模、多源数据，并满足业务需求。近年来，大数据技术路线从批处理架构，逐渐演化为内存计算架构、流处理架构、批流融合处理架构、图数据处理架构等。人工智能关注的技术重点是人工智能算法，即如何通过大数据构建机器学习模型，如何高效训练、评估、测试人工智能模型，并解决人工智能的应用问题；具体而言，包括算法的技术突破、算法的性能和效率提升等。

### (1) 大数据为人工智能提供了大规模多源异构的数据资源

在大数据时代，人工智能使用的不再是样本数据，而更多的是全量数据。高价值数据体量越大，预测结果越准确，对人类思维模拟程度越高。正是基于大数据的数据规模体量，人工智能才得以在算法、算力提升的基础上实现重大突破。

### (2) 统一的数据分析与人工智能平台成为发展趋势

传统大数据平台主要提供基于CPU与内存的分布式数据处理架构，但近年来随着人工智能技术与应用的发展，新型大数据平台支持GPU、GPU/CPU混合计算等新的计算架构。此外，新型大数据平台逐步开始支持TensorFlow、PyTorch等人工智能编程框架，统一数据分析与人工智能平台成为趋势。如Intel推出了面向Apache Spark的统一数据分析与人工智能平台Analytics Zoo，Databricks联合Microsoft推出了MLFLOW，方便用户快速开发、验证、部署人工智能应用。

### (3) 大数据与人工智能技术关联融合

大数据分析的核心技术是SQL、统计分析、图分析与机器学习，而人工智能的核心技术则包括以深度学习为代表的机器学习、知识图谱、逻辑规划和专家系统等，两者在技术上存在明显重合，如大数据与人工智能都需要应用机器学习技术，人工智能领域对知识图谱数据进行分析将与图分析进行结合，如图5-2所示。

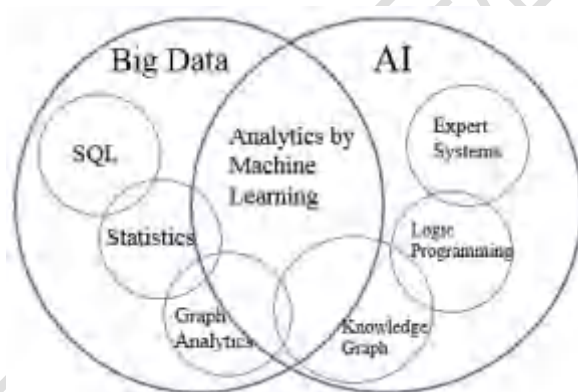


图 5-2 大数据与人工智能技术关联

### (4) 人工智能拓展了大数据应用场景

传统大数据分析主要针对结构化、半结构化数据，缺乏对非结构化数据，如图像、视频、语音的处理能力。数据驱动的人工智能技术则提供了高维非结构化数据的分析能力。在大数据框架下通过不断补充完善与人工智能相关的视频、图像、语音等非结构化数据类型，实现多源异构数据的统一分类、处理与解析，并基于多源异构数据形成统一索引，在各种媒体资源的语义与计算结果之间建立关联，向人工智能大数据智慧应用提供数据服务。

随着大数据与人工智能的深度融合，以及在各行业应用的不断加深，未来大数据和人工智能必将迎来新的增长浪潮并不断产生新模式、新业态。在此情形下，将对大数据、人工智能标准化的协调发展提出新的发展需求。



## 6 大数据标准化现状

### 6.1 国际标准化现状

#### 6.1.1 ISO/IEC JTC 1

##### (1) ISO/IEC JTC 1/SC 32 数据管理和交换分技术委员会

ISO/IEC JTC1/SC 32 数据管理和交换分技术委员会（以下简称 SC 32）是与大数据关系最为密切的国际标准化组织之一。SC 32 持续致力于研制信息系统环境内及之间的数据管理和交换标准，为跨行业领域协调数据管理能力提供技术性支持。其标准化技术内容主要包括：协调现有和新生数据标准化领域的参考模型和框架；负责数据域定义、数据类型和数据结构以及相关的语义等标准；负责用于持久存储、并发访问、并发更新和交换数据的语言、服务和协议等标准；负责用于构造、组织和注册元数据及共享和互操作相关的其他信息资源（电子商务等）的方法、语言服务和协议等标准。SC 32 下设 WG 1 电子业务工作组、WG 2 元数据工作组、WG 3 数据库语言工作组、WG 4 SQL 多媒体和应用包工作组。

2014 年 6 月 SC 32 在北京召开的全会上，批准了 4 项为大数据提供标准化支持的新工作项，分别为《SQL 对多维数组的支持》《数据集注册元模型》以及《数据源注册元模型》3 项国际标准，以及《SQL 对 JSON 的支持》1 项技术报告。其中《SQL 对多维数组的支持》由中国专家担任编辑。

2015 年 5 月 SC 32 在英国召开的全会上，中国提案“SQL 对 MapReduce 及与之相关的流数据处理的支持”得到 SC 32 专家的高度肯定，并在次年 6 月 SC 32 美国全会上通过全会决议，确认以 WG 3 “数据库语言工作组”和中国国家成员体的名义联合申报“数据库语言新技术设计说明第 1 部分：SQL 对流数据的支持”新工作项目（英文名：Design Notes for New Database Language Technologies, Part 1: SQL Support for Streaming Data）。2016 年 10 月在 SC 32 秘书处发布的 2016-2017 年业务计划中明确了该新工作项正式立项，这也是我国第 1 项获得立项的大数据相关领域国际标准。

##### (2) ISO/IEC JTC 1/SC 42/WG 2 人工智能分委员会/数据工作组(原 ISO/IEC

## JTC 1/WG 9)

ISO/IEC JTC 1 于 2013 年 11 月全会上成立了 ISO/IEC JTC 1/SG 2 大数据研究组,主要负责开展大数据领域关键技术、参考模型以及用例等基础标准研究,确定大数据领域应用术语和定义,评估当前大数据标准具体需求等。2014 年 11 月 JTC 1 全会上,ISO/IEC JTC 1/SG 2 向 JTC 1 提交了研究报告,并建议成立独立的 ISO/IEC JTC 1 大数据工作组。该建议得到 JTC 1 批准并于此次全会上成立了 ISO/IEC JTC 1/WG 9 大数据工作组(以下简称 WG 9)。WG 9 工作重点包括:开发大数据基础性标准,包括参考架构和术语;识别大数据标准化需求;同大数据相关的 JTC1 其他工作组保持联络关系;同 JTC 1 外其他大数据相关标准组织保持联络关系。

2017 年 10 月,在 ISO/IEC JTC 1 第 32 次全会上决议成立 ISO/IEC JTC 1/SC 42 人工智能分委会,并决议在 JTC 1/SC 42 正式成立后,解散 JTC 1/WG 9 大数据工作组并将其研究项目转移至 JTC 1/SC 42 中。

2018 年 4 月,ISO/IEC JTC 1/SC 42 人工智能分委会第 1 次全会在北京召开。JTC 1/SC 42 正式向 JTC 1 申请将 JTC 1/WG 9 大数据研究项目转移至 JTC 1/SC 42 中。2018 年 10 月,ISO/IEC JTC 1/SC 42 人工智能分委会第二次全会在美国加利福尼亚召开。会议上正式成立 WG 2 大数据工作组,确定了工作组召集人,并将在新的工作组内继续推进 ISO/IEC DIS 20546《信息技术 大数据 概述和词汇》、ISO/IEC TR 20547-1《信息技术 大数据参考架构 第 1 部分:框架和应用进程》、ISO/IEC DIS 20547-3《信息技术 大数据参考架构 第 3 部分:参考架构》等相关标准研制工作。

2020 年 4 月,在 ISO/IEC JTC 1/SC 42/WG 2 大数据工作组会议上,由我国提交的国际提案《信息技术 人工智能 用于分析和机器学习的数据质量 数据质量过程框架》(英文名“Information technology — Artificial intelligence — Data quality for analytics and ML — Data quality process framework”)得到 WG 2 专家的一致认可,并以 WG 2 的名义向分委会申请发起新工作项目投票,目前该项目正在投票过程中。

截止目前,ISO/IEC JTC 1/SC 42/WG 2 已开展 6 项大数据领域国际标准研

制工作，其中 4 项标准已发布，相关标准状态见表 6-1。

表 6-1 中国参与大数据领域国际标准情况

序号	国际标准号	国际标准名称（英文）	国际标准名称（中文）	阶段
1	ISO/IEC 20546:2019	Information technology -- Big data -- Overview and vocabulary	信息技术 大数据 概述和术语	Published
2	ISO/IEC CD TR 20547-1	Information technology -- Big data reference architecture -- Part 1: Framework and application process	信息技术 大数据参考架构 第1部分：框架与应用	FDIS
3	ISO/IEC TR 20547-2:2018	Information technology -- Big data reference architecture -- Part 2: Use cases and derived requirements	信息技术 大数据参考架构 第2部分：用例和衍生需求	Published
4	ISO/IEC 20547-3	Information technology -- Big data reference architecture -- Part 3: Reference architecture	信息技术 大数据参考架构 第3部分：参考架构	Published
5	ISO/IEC TR 20547-5:2018	Information technology -- Big data reference architecture -- Part 5: Standards roadmap	信息技术 大数据参考架构 第5部分：标准路线图	Published
6	ISO/IEC AWI 24668	Information technology -- Artificial intelligence -- Process management framework for Big data analytics	信息技术 人工智能 大数据分析过程管理 框架	CD

### (3) ISO/IEC JTC 1/SC 27/WG 4 信息安全、网络空间安全和隐私保护分委员会/安全控制与服务工作组

ISO/IEC JTC 1/SC 27/WG 4 在大数据安全方面正在编制 ISO/IEC 20547-4 《信息技术 大数据参考架构 第 4 部分：安全与隐私保护》（状态 FDIS）、ISO/IEC 27045《信息技术 大数据安全与隐私保护 过程》(状态 WD)和 ISO/IEC 27046 《信息技术 大数据安全与隐私保护 实现指南》（状态 WD），均为我国专家担任项目编辑主导的国际标准。

#### (4) ISO/IEC JTC 1/AG 9 数据使用咨询组

2018年11月在瑞典斯德哥尔摩举行的JTC 1全体会议上通过成立数据使用咨询组(Data Usage Advisory Group, 以下简称AG 9)。AG 9主要负责开展数据使用相关研究,包括对数据使用的内涵、概念达成共识;识别与数据使用相关的JTC 1下设机构及现有工作;识别数据使用的要点,以及现有标准与实现现存要点之间的距离。2020年3月,AG 9向JTC 1提交了数据使用报告。报告中给出了数据共享和数据使用框架;对数据使用相关标准基础和实际标准化需求之间的差距进行了识别;并提出下一步数据使用相关研究工作建议,包括建议成立新的数据使用工作组(Data Usage Working Group)并申请立项“数据使用指南”、“数据使用术语和用例”等新工作项。目前AG 9提出的《信息技术 数据使用 术语和用例》(Information technology - Data Usage - Terminology and use cases)、《信息技术 数据使用 数据使用指南》(Information technology - Data Usage - Guidance for data usage)两项新工作项已在JTC 1层面开展立项投票。

#### 6.1.2 ITU-T

ITU在2013年11月发布了题目为《大数据:今天巨大,明天平常》的技术观察报告,该技术观察报告分析了大数据相关的应用实例,指出大数据的基本特征、促进大数据发展的技术,在报告的最后部分分析了大数据面临的挑战和ITU-T可能开展的标准化工作。

从ITU-T的角度来看,大数据发展面临的挑战包括:数据保护、隐私和网络安全,法律和法规的完善。根据ITU-T现有的工作基础,开展的标准化工作包括:高吞吐量、低延迟、安全、灵活和规模化的网络基础设施,汇聚数据机和匿名,网络数据分析,垂直行业平台的互操作,多媒体分析,开放数据标准等。目前,ITU-T大数据标准化工作主要集中在SG13(第13研究组:未来网络包括云计算、移动和下一代网络)、SG16(第16研究组:多媒体编码、系统和应用)、SG17(第17研究组:安全)以及SG20(第20研究组:物联网(IoT)和智能城市及社区(SC&C))等开展。

SG13 下设的 Q7/13（第 7 课题组）正在开展 Y. bDDN-fr “基于深度报文检测的大数据驱动网络框架”标准、Y. Sup-bDDN-usecase “大数据驱动网络的用户案例和应用场景”研究报告、Y. bDDN-req “大数据驱动网络的需求”标准、Y. bDDN-MNTMP “大数据驱动的移动网络流量管理及规划”标准、Y. bDPI-Mec “应用于网络大数据中的深度报文检测机制”标准等方面的工作。Q17/13（第 17 课题组）于 2015 年 11 月正式发布 ITU-T Y. 3600 (Y. BigData-reqts) “基于云计算的大数据需求和能力”标准；2017 年 11 月完成 ITU-T Y. 3601(Y. BigDataEX-reqts) “大数据交换框架和需求”标准；此外，正在开展 Y. bdp-reqts “大数据溯源需求”，Y. bdi-reqts “大数据集成概览和功能需求”，Y. bdm-sch “大数据元数据框架和概念模型”等标准的研制工作。Q18/13（第 18 课题组）正在研制 Y. BDaaS-arch “大数据即服务的参考架构”标准；目前也在讨论如何开展 Y. BD-arch “大数据的功能架构”标准项目，该标准预计 2017 年 11 月完成立项；同时，Q18/13 也将 Y. BigDataEX-arch “大数据交换功能架构”作为 2018 年将要开展的标准工作内容。Q19/13（第 19 课题组）在开展 Y. bddp-reqts “大数据数据保存概览和需求”标准研制工作。

SG17 下设的 Q8/17（第 8 课题组：云计算和大数据基础设施安全）正在开展 X. 1643 (X. sgtBD) “电信大数据的数据生存周期管理安全指南”、X. 1644 (X. GSBaaS) “云计算大数据平台服务安全指南”和 X. sgBDIP “大数据基础设施和平台安全指南”等标准研制工作；Q7/17（第 7 课题组：安全应用服务）于 2018 年 11 月正式发布 X. 1147 “移动互联网服务中大数据分析的安全需求和框架”；Q13/17（第 13 课题组：智能交通系统安全方面）正在开展 X. mdcv “基于大数据分析的联网车辆安全相关不良行为检测机制”标准研制工作。

此外，SG16 下设的 Q21/16（第 21 课题组）正在开展 F. 743. 20 “大数据基础设施评测框架”和 ITU-T F. 743. 21 “数据资产管理框架”两项标准研制，即将正式发布；Q24/16（第 24 课题组）正在开展 F. VSBD “视频监控系统中大数据应用的需求”标准研制；SG20 下设的 Q2/20（第 2 课题组）于 2017 年 7 月正式发布 ITU-T Y. 4114 (Y. IoT-BigData-reqts) “针对大数据的物联网具体需求和能力要求”标准。

### 6.1.3 IEEE BDGMM

在 IEEE 新倡议委员会 (NIC) 的 IEEE 大数据倡议 (BDI) 下, IEEE 大数据治理和元数据管理 (BDGMM) 于 2017 年 6 月成立, 主导大数据标准化工作。BDGMM 的工作是指导如何开展大数据治理和大数据交换工作, 使得大数据消费者能更好地了解 and 访问可用数据, 帮助大数据生产者正确设定期望值并确保按照期望值维护和共享数据集, 帮助拥有大数据的组织做出如何存储、策划、提供和治理大数据的决策, 以便更好地服务于大数据消费者和生产者。BDGMM 每两周召开一次远程会议。

BDGMM 的目标是能够整合来自不同领域的异构数据集, 通过可器读和可操作的规范的基础设施, 使数据可发现、可访问和可利用。

BDGMM 期望的可交付成果包括: (1) 通过 IEEE 发起的研讨会和 Hackathons 或者其他会议收集、分析和识别相关用例、要求和解决方案, 并形成文档; (2) 基于上述文档, 更详细地框定问题、找出课题, 形成白皮书; (3) 来自大数据元数据管理相关最佳实践的参考架构概念和解决方案, 用以规划数据互操作基础设施, 使不同领域数据库之间的数据集成为可能; (4) 识别和启动大数据元数据管理相关的 IEEE 标准活动 (包括建议的实践、指南)。

### 6.1.4 NIST

NIST (美国国家标准技术研究所) 是最早进行大数据标准化研究的机构之一。专门成立了大数据公共工作组 (NBD-PWD) 对大数据的发展和应用, 及标准化进行研究。工作组最重要的输出是被广泛参考的大数据互操作性框架 (NBDIF) 报告。大数据互操作性框架的核心是面向各个角色 (系统协调者、数据提供者、大数据应用提供者、大数据框架提供者、数据消费者等) 定义一个由标准接口互联的、不绑定技术和厂商实现的、模块可替换的大数据参考架构 (NBDRA)。这一报告目前有两个版本。第一个版本 (已经发布) 包括七卷: 定义、分类、用例和要求、安全和隐私、架构调研白皮书、参考架构、标准路线。第二个版本在征求意见阶段, 除了修改、完善第一个版本中的内容, 又增加了两卷: 大数据参考

架构接口，采用和（传统系统的）现代化。NIST 的这一系列报告，包括对大数据术语的定义、参考架构、应用案例、标准路线等的分析是大数据标准化工作的重要参考。

## 6.2 国内标准化现状

### 6.2.1 国家标准化现状

大数据领域的标准化工作是支撑大数据产业发展和应用的重要基础，为了推动和规范我国大数据产业快速发展，建立大数据产业链，与国际标准接轨，在工业和信息化部，国家标准化管理委员会的领导下，社会各界朋友关心支持之下，2014 年 12 月 2 日全国信标委大数据标准工作组正式成立。2016 年 4 月，全国信安标委大数据安全标准特别工作组正式成立。

#### (1) 全国信标委大数据标准工作组（以下简称“工作组”）

工作组主要负责制定和完善我国大数据领域标准体系，组织开展大数据相关技术和标准的研究，申报国家、行业标准，承担国家、行业标准制修订计划任务，宣传、推广标准实施，组织推动国际标准化活动，对口 ISO/IEC JTC 1/SC 42/WG 2（原 ISO/IEC JTC 1/WG 9）大数据工作组工作。

工作组组长由中国人民解放军军事科学院副院长梅宏院士担任，副组长为中国电子技术标准化研究院副院长孙文龙、中国人民大学教授杜小勇、华为 IT 技术开发部部长吴建明、原阿里云首席科学家闵万里，秘书处设在中国电子技术标准化研究院，联络员由工业和信息化部电子信息司副巡视员侯建仁、工业和信息化部信息技术发展司调研员傅永宝和国家市场监督管理总局标准技术管理处处长刘大山担任。

基于实际需求，为了更好的保障标准化对技术、产业的支撑能力，不断加强大数据标准化工作，2019 年 7 月工作组完成了第三届专题组换届工作并重新调整了组织架构，下设总体专题组、国际专题组、技术专题组、产品和平台专题组、工业大数据专题组、政务大数据专题组、服务大数据专题组、生态环境大数据专题组、电力大数据专题组、大数据治理专题组等 10 个专题组，负责大数据领域

不同方向的标准化工作。目前，工作组已开展 33 项大数据国家标准的研制工作，其中已发布国家标准 24 项，在研 9 项。详见表 6-2。

表 6-2 全国信标委大数据标准工作组标准研制情况

序号	标准号	标准名称	状态	所属专题组
1.	GB/T 35295-2017	信息技术 大数据 术语	发布	总体专题组
2.	GB/T 35589-2017	信息技术 大数据 技术参考模型	发布	总体专题组
3.	GB/T 34952-2017	多媒体数据语义描述要求	发布	技术专题组
4.	GB/T 35294-2017	信息技术 科学数据引用	发布	技术专题组
5.	GB/T 34945-2017	信息技术 数据溯源描述模型	发布	技术专题组
6.	GB/T 36073-2018	数据管理能力成熟度评估模型	发布	总体专题组
7.	GB/T 36343-2018	信息技术 数据交易服务平台 交易数据描述	发布	总体专题组
8.	GB/T 37728-2019	信息技术 数据交易服务平台 通用功能要求	发布	总体专题组
9.	GB/T 36344-2018	信息技术 数据质量评价指标	发布	技术专题组
10.	GB/T 36345-2018	信息技术 通用数据导入接口规范	发布	产品平台专题组
11.	GB/T 37721-2019	信息技术 大数据分析系统功能要求	发布	产品平台专题组
12.	GB/T 37722-2019	信息技术 大数据存储与处理系统功能要求	发布	产品平台专题组
13.	GB/T 38672-2020	信息技术 大数据 接口基本要求	发布	总体专题组
14.	GB/T 38667-2020	信息技术 大数据 数据分类指南	发布	技术专题组
15.	GB/T 38673-2020	信息技术 大数据 大数据系统基本要求	发布	产品平台专题组
16.	GB/T 38676-2020	信息技术 大数据 存储与处理系统功能测试要求	发布	产品平台专题组
17.	GB/T 38643-2020	信息技术 大数据 分析系统功能测试要求	发布	产品平台专题组
18.	GB/T 38675-2020	信息技术 大数据 计算系统通用要求	发布	产品平台专题组
19.	GB/T 38633-2020	信息技术 大数据 系统运维和管理功能要求	发布	产品平台专题组
20.	GB/T 38664.1-2020	信息技术 大数据 政务数据开放共享 第 1 部分：总则	发布	政务大数据专题组
21.	GB/T 38664.2-2020	信息技术 大数据 政务数据开放共享 第 2 部分：基本要求	发布	政务大数据专题组
22.	GB/T 38664.3-2020	信息技术 大数据 政务数据开放共享 第	发布	政务大数据



序号	标准号	标准名称	状态	所属专题组
		3 部分：开放程度评价		专题组
23.	GB/T 38666-2020	信息技术 大数据 工业应用参考架构	发布	工业大数据专题组
24.	GB/T 38555-2020	信息技术 大数据 工业产品核心元数据	发布	工业大数据专题组
25.	20180988-T-469	信息技术 工业大数据 术语	草案	工业大数据专题组
26.	20182054-T-339	智能制造 工业数据空间模型	草案	工业大数据专题组
27.	20182040-T-339	智能制造 多模态数据融合系统技术要求	草案	工业大数据专题组
28.	20182053-T-339	智能制造 工业大数据平台通用要求	草案	工业大数据专题组
29.	20182052-T-339	智能制造 工业大数据时间序列数据采集和存储框架	草案	工业大数据专题组
30.	20190841-T-469	信息技术 大数据 面向分析的数据存储与检索技术要求	草案	技术专题组
31.	20190842-T-469	信息技术 大数据 政务数据开放共享 第 4 部分：共享评价	草案	政务大数据专题组
32.	20190840-T-469	数据管理能力成熟度评估方法	草案	总体专题组
33.	20194186-T-469	信息技术 大数据 数据资源规划	草案	大数据治理专题组

此外，工作组积极研究和参与大数据领域国际标准化工作，全面参与 ISO/IEC JTC 1/SC 42/WG 2、ISO/IEC JTC 1/SC 32、ISO/IEC JTC 1/AG 9 相关研究工作。此外，工作组还重点关注 NIST NBD-PWG 大数据公共工作组，并对 ITU 的动态进行研究和跟踪。

## (2) 全国信安标委大数据安全标准特别工作组（以下简称“数据安全标准特别工作组”）

数据安全标准特别工作组组长由清华大学软件学院院长王建民教授担任，副组长为四川大学网络空间安全研究院常务副院长陈兴蜀教授，秘书为清华大学软件学院金涛博士。特别工作组目前已开展 8 项大数据安全领域国家标准研制工作，其中 6 项国家标准已发布，2 项国家标准在研，如表 6-3 所示。

表 6-3 全国信安标委大数据安全标准特别工作组大数据安全相关标准研制情况

序号	标准号	标准名称	状态
1.	GB/T 35273-2017	信息安全技术 个人信息安全规范	发布
2.	GB/T 35274-2017	信息安全技术 大数据服务安全能力要求	发布
3.	GB/T 37973-2019	信息安全技术 大数据安全管理指南	发布
4.	20180840-T-469	信息安全技术 个人信息安全影响评估指南	征求意见
5.	GB/T 37964-2019	信息安全技术 个人信息去标识化指南	发布
6.	GB/T 37988-2019	信息安全技术 数据安全能力成熟度模型	发布
7.	GB/T 37932-2019	信息安全技术 数据交易服务安全要求	发布
8.	20173853-T-469	信息安全技术 数据出境安全评估指南	征求意见

此外,数据安全标准特别工作组陆续启动了《大数据基础软件安全技术要求》《大数据业务安全风险控制实施指南》《数据安全分类分级实施指南》《区块链安全技术标准研究》等研究工作,并在 2018 年编写发布了《大数据安全标准化白皮书(2018 版)》。

## 6.2.2 地方标准化现状

基于大数据产业发展区域特点,全国各地地区纷纷成立了地方性的大数据标准化技术委员会,逐步开展大数据地方标准研制工作,旨在形成安全可靠、统一规范、便捷高效的地方大数据标准体系建设,服务当地大数据产业发展。本白皮书梳理了部分省级标准化技术委员会,如表 6-4 所示。

表 6-4 部分省级大数据标准化技术委员会情况

序号	标准化技术委员会名称	成立时间
1.	贵州省大数据标准化技术委员会	2017 年
2.	广东省大数据标准化技术委员会	2017 年
3.	内蒙古自治区云计算与大数据标准化技术委员会	2017 年
4.	山东省大数据标准化技术委员会	2018 年
5.	山西省网络安全和大数据信息技术标准化技术委员会	2019 年
6.	上海市公共数据标准化技术委员会	2020 年

依托省级标准化技术委员会,贵州省面向政务数据研制形成了 DB52/T 1123-2016《政府数据 数据分类分级指南》、DB52/T1406-2019《政府数据 数据

开放工作指南》、DB52/T1407-2019《政府数据 开放数据核心元数据》等 10 余项政务数据地方标准，对人口、法人、空间地理、非物质文化遗产、宏观经济等领域数据的核心元数据进行统一规范，推动政府数据共享与开放。山东省加快推进山东省农业供给侧结构性改革，研制形成 DB37/T 3431-2018《农业大数据标准体系》、DB37/T 3432-2018《农业大数据 数据处理基本要求》、DB37/T 3433.1-2018《农业大数据 基础数据元 第 1 部分：公共》等 10 项农业大数据地方标准，聚焦大数据与农业的融合应用。内蒙古自治区基于“云上北疆”云平台建设，研制形成了 DB15/T 1590-2019《大数据标准体系编制规范》、DB15/T 1873-2020《大数据平台 数据接入质量规范》、DB15/T 1874-2020《公共大数据安全管理指南》等地方标准，推动政府数据共享交换和公共数据高质量开放。此外，上海市推动研制《公共数据共享交换工作规范 平台管理规范》，湖北省发布了《政务数据服务度量计价规范》，陕西省面向平台、应用、管理、隐私等方面开展大数据标准体系的建设，并重点在气象、铁路、车联网、城市运行管理等行业应用领域组织开展大数据标准研究工作。

# 7 大数据标准体系

## 7.1 大数据标准体系框架

通过调研我国大数据技术、产业发展现状，分析大数据与实体经济融合带来的新标准化需求，本白皮书在《大数据标准化白皮书（2018版）》提出的大数据标准体系框架基础上进行修订，形成新的大数据标准体系框架，如图7-1所示。



图 7-1 大数据标准体系框架图

大数据标准体系由7个类别标准组成，分别为基础标准、数据标准、技术标准、平台/工具标准、治理与管理标准、安全和隐私标准、行业应用标准。

### (1) 基础标准

基础标准为大数据其他部分的标准制定提供基础遵循，支撑行业间对大数据达成统一理解，主要包括术语、参考架构类标准。

### (2) 数据标准

数据标准主要针对底层数据相关要素进行规范，包括数据资源和交换共享两类。其中数据资源标准面向数据本身进行规范，包括数据元素、元数据、参考数据、主数据、数据模型等标准；交换共享标准面向数据流通相关技术、架构及应用进行规范，包括数据交易和开放共享标准。

### (3) 技术标准

技术标准主要针对大数据通用技术进行规范。包括大数据集描述、大数据生存周期处理技术、大数据开放与互操作技术、面向领域的大数据技术四类。其中，大数据集描述标准主要针对多样化、差异化、异构异质的不同类型的数据建立标准的度量方法，以衡量数据质量；大数据生存周期处理技术标准主要针对大数据产生到其使用终止这一过程的关键技术进行标准制定，包括数据采集、数据预处理、数据存储、数据分析、数据可视化、数据访问等标准；大数据开放与互操作标准主要针对不同功能层次功能系统之间的互联与互操作机制、不同技术架构系统之间的互操作机制、同质系统之间的互操作机制等相关标准以及通用数据开放共享技术框架等标准进行研制；面向领域的大数据技术标准主要针对电力行业、医疗行业、电子政务等领域或行业的共性且专用的大数据技术标准进行研制。

#### （4）平台/工具标准

平台/工具标准主要针对大数据相关平台及工具产品进行规范，包括大数据系统产品和数据库产品。其中大数据系统产品标准主要针对业内主流的用于实现数据全生存周期处理的大数据产品的功能和性能进行规范；数据库产品标准则主要面向不同类型的数据库的功能和性能进行要求。此外，该类标准还包括相关产品功能及性能的测试方法和要求。

#### （5）治理与管理标准

治理与管理标准贯穿于数据生存周期的各个阶段，是大数据实现高效采集、分析、应用、服务的重要支撑。该类标准主要包括治理标准、管理标准和评估标准三部分。其中，治理标准主要对数据治理的规划和具体实施方法进行标准研制；管理标准则主要面向数据管理模型、元数据管理、主数据管理、数据质量管理、数据目录管理以及数据资产管理等理论方法和管理工具进行规范；评估标准则在治理标准和管理标准的基础上之上，总结形成针对数据管理能力、数据服务能力、数据治理成效、数据资产价值的评估方法。

#### （6）安全和隐私标准

数据安全和隐私标准同样贯穿于整个数据生存周期的各个阶段，主要包括应用安全、数据安全、服务安全、平台和技术安全四部分。其中，应用安全主要对大数据与其他领域融合应用中存在的安全问题进行规范；数据安全主要围绕个人

信息安全、重要数据安全以及跨境数据安全标准进行研制，保障数据主体所拥有数据不被侵害；服务安全主要包括数据安全治理、服务安全能力和交换共享安全，面向数据产品和解决方案的安全性进行要求；平台和技术安全则针对大数据平台，以及以大数据平台为底座的应用平台的系统安全、接口安全、技术安全进行标准研制。

### (7) 行业应用标准

行业应用标准主要面向通用领域应用以及工业、政务、电力、生态环境等垂直行业领域应用开展标准研制。通用领域应用标准主要从大数据在通用领域中所提供的共性服务出发，开展应用方法、能力评估等标准研制；垂直行业应用标准主要从大数据为各个垂直行业所能提供的服务角度出发，是各领域根据其领域特性产生的专用数据标准，包括工业大数据、政务大数据、电力大数据、生态环境大数据等领域应用标准。

## 7.2 标准明细表

根据大数据标准体系框架，整理出发布、报批、在研以及拟研制的大数据领域相关国家标准 94 项，如下表所示：

表 7-1 大数据标准明细表

序号	一级分类	二级分类	国家标准编号	标准名称	采用标准号及采用程度	状态
1.	基础	术语	GB/T 35295-2017	信息技术 大数据 术语		发布
2.		参考	GB/T 35589-2017	信息技术 大数据 技术参考模型		发布
3.		架构	GB/T 38672-2020	信息技术 大数据 接口基本要求		发布
4.	数据	数据资源	GB/T 18142-2017	信息技术 数据元素值表示格式记法	修订GB/T 18142-2000:ISO/IEC FDIS 14957:2009	发布
5.			GB/T 18391.1-2009	信息技术 元数据注册系统 (MDR) 第1部分：框架	ISO/IEC11179-1: 2004, IDT	发布
6.			GB/T 18391.2-2009	信息技术 元数据注册系统 (MDR) 第2部分：分类	ISO/IEC11179-2: 2005, IDT	发布

序号	一级分类	二级分类	国家标准编号	标准名称	采用标准号及采用程度	状态
7.			GB/T 18391.3-2009	信息技术 元数据注册系统 (MDR) 第3部分：注册系统元模型与基本属性	ISO/IEC11179-3: 2003, IDT	发布
8.			GB/T 18391.4-2009	信息技术 元数据注册系统 (MDR) 第4部分：数据定义的形成	ISO/IEC11179-4: 2004, IDT	发布
9.			GB/T 18391.5-2009	信息技术 元数据注册系统 (MDR) 第5部分：命名和标识原则	ISO/IEC11179-5: 2005, IDT	发布
10.			GB/T 18391.6-2009	信息技术 元数据注册系统 (MDR) 第6部分：注册	ISO/IEC11179-6: 2005, IDT	发布
11.			GB/Z 21025-2007	XML使用指南		发布
12.			GB/T 23824.1-2009	信息技术 实现元数据注册系统内容一致性的规程 第1部分：数据元	ISO/IEC TR20943-1: 2003, IDT	发布
13.			GB/T 23824.3-2009	信息技术 实现元数据注册系统内容一致性的规程 第3部分：值域	ISO/IEC TR20943-3: 2004, IDT	发布
14.			GB/T 32392.1-2015	信息技术 互操作性元模型框架 (MFI) 第1部分：参考模型		发布
15.			GB/T 32392.2-2015	信息技术 互操作性元模型框架 (MFI) 第2部分：核心模型		发布
16.			GB/T 32392.3-2015	信息技术 互操作性元模型框架 (MFI) 第3部分：本体注册元模型		发布
17.			GB/T 32392.4-2015	信息技术 互操作性元模型框架 (MFI) 第4部分：模型映射元模型		发布
18.			GB/T 32392.5-2018	信息技术 互操作性元模型框架 (MFI) 第5部分：过程模型注册元模型		发布
19.			GB/T 32392.7-2018	信息技术 互操作性元模型框架 第7部分：服务模型注		发布

序号	一级分类	二级分类	国家标准编号	标准名称	采用标准号及采用程度	状态
				册元模型		
20.			GB/T 32392.8-2018	信息技术 互操作性元模型 框架 第8部分：角色与目标 模型注册元模型		发布
21.			GB/T 32392.9-2018	信息技术 互操作性元模型 框架 第9部分：按需模型选 择		发布
22.			GB/T 30881-2014	信息技术 元数据注册系统 (MDR) 模块	ISO/IEC 19773:2011	发布
23.		交换	GB/T 36343-2018	信息技术 数据交易服务平 台 交易数据描述		发布
24.		共享	GB/T 37728-2019	信息技术 数据交易服务平 台 通用功能要求		发布
25.	技术		GB/T 32909-2016	非结构化数据表示规范		发布
26.			GB/T 34952-2017	多媒体数据语义描述要求		发布
27.		大数据集 描述	GB/T 38667-2020	信息技术 大数据 数据分 类指南		发布
28.			GB/T 36344-2018	信息技术 数据质量评价指 标		发布
29.			GB/T 35294-2017	信息技术 科学数据引用		发布
30.			GB/T 34945-2017	信息技术 数据溯源描述模 型		发布
31.		大数据生 存周期处 理技术	GB/T 32908-2016	非结构化数据访问接口规 范		发布
32.			GB/T 36345-2018	信息技术 通用数据导入接 口规范		发布
33.			20190841-T-4 69	信息技术 大数据 面向分 析的数据检索与存储技术 要求		在研
34.			_____	信息技术 大数据 分析总 体技术要求		拟研 制



序号	一级分类	二级分类	国家标准编号	标准名称	采用标准号及采用程度	状态
35.		大数据开放与互操作技术	_____	信息技术 大数据 互操作技术指南		拟研制
36.	平台/ 工具	大数据系统产品	GB/T 37722-2019	信息技术 大数据存储与处理系统功能要求		发布
37.			GB/T 38676-2020	信息技术 大数据 存储与处理系统功能测试要求		发布
38.			GB/T 37721-2019	信息技术 大数据分析系统功能要求		发布
39.			GB/T 38643-2020	信息技术 大数据 分析系统功能测试要求		发布
40.			GB/T 38673-2020	信息技术 大数据 大数据系统基本要求		发布
41.			GB/T 38675-2020	信息技术 大数据 计算系统通用要求		发布
42.			GB/T 38633-2020	信息技术 大数据 系统运维和管理功能要求		发布
43.			_____	信息技术 大数据 批流融合计算系统技术要求		拟研制
44.		GB/T 28821-1012	关系数据管理系统技术要求		发布	
45.		GB/T 30994-2014	关系数据库管理系统检测规范		发布	
46.		GB/T 34949-2017	实时数据库C语言接口规范		发布	
47.		GB/T 32633-2016	分布式关系数据库服务接口规范		发布	
48.		GB/T 32630-2016	非结构化数据管理系统技术要求		发布	
49.		_____	信息技术 大数据 图数据库系统技术要求		拟研	

序号	一级分类	二级分类	国家标准编号	标准名称	采用标准号及采用程度	状态
						制
50.			GB/T 12991-2008	信息技术 数据库语言SQL 第1部分：框架	ISO/IEC9075-1 : 2003, IDT	发布
51.	治理 与 管 理	治理	20194186-T-4 69	信息技术 大数据 数据资 源规划		在研
52.			_____	信息技术 大数据 数据治 理体系建设指南		拟研 制
53.			_____	城市数据治理能力成熟度 模型		拟研 制
54.			_____	信息技术 大数据 数据治 理工具基本要求		拟研 制
55.		管理	GB/T 36073-2018	数据管理能力成熟度评估 模型		发布
56.			_____	信息技术 大数据 主数据 管理规范		拟研 制
57.		评估	20190840-T-4 69	数据管理能力成熟度评估 方法		在研
58.			_____	信息技术 大数据 数据服 务能力评估 第1部分：评估 模型		拟研 制
59.			_____	信息技术 大数据 数据服 务能力评估 第2部分：评估 过程		拟研 制
60.			_____	信息技术 大数据 数据治 理绩效评价		拟研 制
61.	_____		信息技术 大数据 数据资 产评估		拟研 制	
62.	安全 和隐 私	应用 安全	_____	大数据业务安全风险控制 实施指南		拟研 制

序号	一级分类	二级分类	国家标准编号	标准名称	采用标准号及采用程度	状态
63.		服务安全	GB/T 35274-2017	信息安全技术 大数据服务安全能力要求		发布
64.			GB/T 37932-2019	信息安全技术 数据交易服务安全要求		发布
65.			GB/T 37988-2019	信息安全技术 数据安全能力成熟度模型		发布
66.			—————	大数据服务安全可控评价指标		拟研制
67.		数据安全	GB/T 35273-2020	信息安全技术 个人信息安全规范		发布
68.			GB/T 37964-2019	信息安全技术 个人信息去标识化指南		发布
69.			GB/T 34978-2017	信息安全技术 移动智能终端个人信息保护技术要求		发布
70.			GB/T 37988-2019	信息安全技术 数据安全能力成熟度模型		发布
71.			20180840-T-4 69	信息安全技术 个人信息安全影响评估指南		在研
72.		平台和技术安全	GB/T 20009-2005	信息安全技术 数据库管理系统安全评估准则		发布
73.			GB/T 20273-2006	信息安全技术 数据库管理系统安全技术要求		发布
74.			GB/T 22080-2008	信息技术 安全技术 信息安全管理体系要求	ISO/IEC 27001:2005, IDT	发布
75.			GB/T 22081-2008	信息技术 安全技术 信息安全管理体系实用规则	ISO/IEC 27002:2005, IDT	发布
76.			GB/T 31496-2015, IDT	信息技术 安全技术 信息安全管理体系实施指南	ISO/IEC 27003:2010, IDT	发布
77.			GB/Z 28828-2012	信息安全技术 公共及商用服务信息系统个人信息保护指南		发布
78.			GB/T 37973-2019	信息安全技术 大数据安全管理指南		发布
79.			—————	信息安全技术 大数据中的		拟

序号	一级分类	二级分类	国家标准编号	标准名称	采用标准号及采用程度	状态
				隐私保护框架		研制
80.	行业应用	通用领域应用	-----	基于大数据的数字化转型能力要求		拟研制
81.		工业大数据	GB/T 38666-2020	信息技术 大数据 工业应用参考架构		发布
82.			GB/T 38555-2020	信息技术 大数据 工业产品核心元数据		发布
83.			20180988-T-469	信息技术 工业大数据 术语		在研
84.			20182054-T-339	智能制造 工业数据空间模型		在研
85.			20182040-T-339	智能制造 多模态数据融合系统技术要求		在研
86.			20182053-T-339	智能制造 工业大数据平台通用要求		在研
87.			20182052-T-339	智能制造 工业大数据时间序列数据采集和存储框架		在研
88.			-----	信息技术 工业大数据 数据分类		拟研制
89.		政务大数据	GB/T 38664.1-2020	信息技术 大数据 政务数据开放共享 第1部分：总则		发布
90.			GB/T 38664.2-2020	信息技术 大数据 政务数据开放共享 第2部分：基本要求		发布
91.			GB/T 38664.3-2020	信息技术 大数据 政务数据开放共享 第3部分：开放程度评价		发布
92.			20190842-T-469	信息技术 大数据 政务数据开放共享 第4部分：共享评价		在研
93.		生态环境大数据	-----	信息技术 生态环境大数据 数据分类指南		拟研制

序号	一级分类	二级分类	国家标准编号	标准名称	采用标准号及采用程度	状态
94.		据	_____	信息技术 生态环境大数据数据体系		拟研制

### 7.3 重点标准介绍

前期,依托全国信标委大数据标准工作组已自主研制形成一批大数据领域国家标准,并开展了试验验证、试点应用工作。

#### (1) GB/T 35589-2017《信息技术 大数据 技术参考模型》

GB/T 35589-2017《信息技术 大数据 技术参考模型》国家标准规范了大数据的基础通用模型,包括大数据角色、活动、主要组件及其之间的关系。适用于理解大数据领域的复杂操作,是讨论需求、结构和操作的有效工具,并为大数据系列标准的制定提供了架构依据。大数据技术参考模型提供了一个构件层级分类体系,用于描述技术参考模型中的逻辑构件以及定义逻辑构件的分类。大数据技术参考模型中的逻辑构件被划分为三个层级,从高到低依次为角色、活动和组件。最顶层级的逻辑构件是代表大数据系统中存在的五个角色,包括系统协调者、数据提供者、大数据应用提供者、大数据框架提供者、数据消费者五个角色。另外两个非常重要的逻辑构件是安全和隐私以及管理,它们为大数据系统的五个角色提供服务和功能。第二层级的逻辑构件是每个角色执行的活动。第三层级的逻辑构件是执行每个活动需要的功能组件。该模型可以用于表示由多个大数据系统组成的堆叠式或链式系统,其中一个系统的数据消费者可以作为后面一个系统的数据提供者。该模型支持各种商业环境,包括紧密集成的企业系统和松散耦合的垂直行业,有助于理解大数据系统如何补充并有别于已有的分析、商业智能、数据库等传统的数据应用系统。

#### (2) GB/T 36073-2018《数据管理能力成熟度评估模型》

GB/T 36073-2018《数据管理能力成熟度评估模型》国家标准给出了数据管理能力成熟度评估模型以及相应的成熟度等级,适用于组织和机构对内部数据管理能力成熟度进行评估。该标准通过对组织、制度、流程和技术的有效整合将组

织或机构内部数据管理能力划分为数据战略、数据治理、数据架构、数据应用、数据安全、数据质量、数据标准和数据生存周期等 8 项一级能力域，基于数据全生存周期，以数据战略为导向，建全数据治理组织管理体系，对数据架构、数据标准、数据质量、数据应用、数据安全进行全方位管控。同时，在此基础上对每个一级能力域进行了二级能力项的划分（共计 28 个能力项），描述了每个组成部分的定义、功能、目标和标准（共计 441 项指标），并将每个能力项以及总体数据管理能力划分为初始级、受管理级、稳健级、量化管理级和优化级等 5 个成熟度等级，是针对一个组织或机构数据管理、应用能力的自上而下、分工科学、协作紧密、流程清晰的数据管理能力评估框架。该标准在研制过程中在浙江移动、天津天臣、广东佛山、鹏博士集团等单位进行了充分验证，并于标准发布后在贵州、上海、济南等地区以及电力、通信、金融等行业进行了广泛的推广宣贯及试点应用，对我国产业数据管理能力的整体摸底及提升提供了重要基础。

### （3）GB/T 38673-2020《信息技术 大数据 大数据系统基本要求》

GB/T 38673-2020《信息技术 大数据 大数据系统基本要求》国家标准主要对大数据系统的功能要求及非功能要求进行了规范，适用于各类大数据系统，可作为大数据系统设计、选型、验收、检测的依据。从功能要求出发，该标准将大数据系统划分数据收集、数据预处理、数据存储、数据处理、数据分析、数据访问、数据可视化、资源管理、系统管理等 9 个功能模块，并形成大数据系统框架；从非功能要求考虑，该标准对大数据系统整体的可靠性要求、兼容性要求、安全性要求、可扩展性要求、维护性要求、易用性要求进行了规范。该标准在研制过程中基于华为、阿里云、百分点、海康威视、新华三、中兴、南大通用等企业的大数据系统产品开展了的多次试验验证工作，助力企业完善其大数据产品的功能。基于该标准形成的标准符合性测试能力，国家认证认可监督管理委员会批复成立了“国家大数据系统产品质量监督检验中心”。

### （4）《信息技术 大数据 开放共享》系列标准

GB/T 38664.1-2020《信息技术 大数据 政务数据开放共享 第 1 部分：总则》、GB/T 38664.2-2020《信息技术 大数据 政务数据开放共享 第 2 部分：基本要

求》、GB/T 38664.3-2020《信息技术 大数据 政务数据开放共享 第3部分：开放程度评价》系列标准对我国政务数据开放共享提出了整体要求及具体指导。其中，《第1部分：总则》提出了政务数据开放共享系统参考架构，包括网络设施、数据资源、平台设施、安全保障和管理评价五部分组成的，适用于国家、省、市各级数据共享交换系统的建设，同时满足各级系统逐级对接要求。《第2部分：基本要求》围绕第1部分中提出的参考架构，对政务数据开放共享的网络设施、数据资源、平台设施和安全保障的基本要求进行了规范，适用于政务部门实施政务数据开放共享工程，为企事业单位、其他组织参与公共数据开放共享工程建设、验收和运营提供参考，规范引导政务数据开放共享和开发利用。《第3部分：开放程度评价》提出了政务数据开放程度评价指标体系，包括网络设施、数据资源、平台设施、安全保障、管理评价和应用成效6项一级指标，以及19项二级指标。同时基于评价指标，该标准给出了指标权重表，可以用于对评价对象的具体指标进行打分，从而量化评估各地区政务数据开放水平。

开放共享系列标准为参与或实施数据开放共享的机构开展开放共享工程规划、建设、验收和运营的活动提供依据。后续将在已有标准基础上推进《信息技术 大数据 政务数据开放共享 第4部分：共享评价》（20190840-T-469）国家标准研制工作，支撑我国政务数据开放共享实施。

## 8 大数据标准化工作建议

### 8.1 完善大数据标准化工作机制

随着大数据在各行业应用的不断深入，大数据标准化工作逐步由互联网企业主导演变至各领域协同参与的新态势。为此，需进一步加强全国信标委大数据标准工作组组织建设，统筹考虑各领域大数据标准化需求，加大产学研用资源凝聚力度，广泛吸纳相关单位参与国家大数据标准化工作，构建通用互联网、工业、政务、电力、生态环境等多领域沟通协调的标准化工作机制，加强大数据标准化工作对大数据技术在垂直领域深入应用的支撑力度。

### 8.2 夯实大数据标准体系建设

大数据与实体经济融合已成为当前大数据产业发展的重要趋势，在驱动大数据技术不断创新，产业持续变革的同时，也必将对大数据标准化工作提出新的要求。面对新业态、新模式下的新需求，需要紧密结合国内外大数据技术、产业发展动态，以“基础统领、应用牵引”为原则，以加快我国数字经济建设为目标，持续提炼梳理各应用领域大数据标准需求，并在现有标准体系的基础上迭代更新，保障大数据标准对大数据技术、产业创新发展的持续推动作用。

### 8.3 加强大数据标准宣传推广

标准的研制需要全社会的广泛参与。近年来我国大数据标准化工作稳步推进，多项国家标准陆续发布，需要各级政府及相关部门充分发挥引导作用，加强对大数据标准化工作的宣传推广工作。一方面通过宣传帮助产业各方深刻理解大数据标准化的发展现状、未来规划和对产业发展的重要推动作用，提升整体标准化意识水平；另一方面通过加强对大数据标准的宣传推广，有助于吸引更多的企业和组织机构参与到需求调研、研究编制、意见征集等工作中，持续壮大我国大数据标准化国家队。



## 8.4 构建标准符合性评估体系

“研以致用”是开展大数据标准化工作的重要目标。需要在现有大数据国家标准研制的基础上，面向重点行业、重点地区开展大数据标准试验验证和试点示范工作。围绕数据开放共享、大数据系统测试、数据管理、工业大数据、数据安全等领域的核心标准，优先选择大数据综合试验区和大数据产业集聚区等优势地区，开展标准验证应用工作，汇聚优秀应用案例，积累行业先进经验。同时以此为基础建立并不断完善标准符合性评估体系，培育大数据标准符合性评估、测试、咨询能力，充分发挥标准在培育服务市场、提升服务能力、支撑行业管理等方面的重要作用。

## 8.5 推进大数据治理标准支撑

数据治理是提升组织内部数据质量、推动数据广泛共享、强化数据安全保障、盘活数据资源价值的重要手段。目前我国政府部门、企事业单位的数据治理能力整体不足，尚未掌握科学的数据治理方法、不具备良好的数据管理能力、缺乏专业的数据管理人才，严重限制着我国数字产业化、产业数字化进程，阻碍数字经济建设。下一步需要深入开展国内外数据治理理论研究，并结合我国产业发展特点，建设符合我国社会主义特色的数据治理体系框架，并通过标准化手段固化研究成果，指导各个组织机构完善数据治理机制，提升产业整体数据治理能力。

## 8.6 加大标准化人才培养力度

大数据人才是大数据领域发展的核心资源，标准作为技术、产业发展的顶层支撑，亟需加快相关人才队伍培养。建议在全国范围内组织开展大数据标准宣贯培训，加强对业内技术人员的大数据标准理论和实施方法指导，提升标准化意识；加强标准化机构与相关企业、高校合作，强化兼顾大数据技术、行业经验和标准化能力的复合型人才培养机制；编制数据管理能力标准宣贯培训教材，指导第三方机构，依据标准制定数据管理从业人员能力培养和评价方法，形成市场化的从业人员能力培养和评价机制。

## 8.7 加快大数据标准的国际化步伐

标准是国际大数据竞争的重要战场之一。近年来我国积极对接ISO/IEC JTC 1等国际标准组织，持续加强大数据国际标准参与力度。下一步建议依托全国信标委大数据标准工作组，组织国内核心标准化机构和重点企业，深度参与国际标准化制修订工作，积极贡献国际标准提案。同时支持我国专家争取工作组召集人等国际标准化组织职务，承办大数据相关国际标准化活动，加强我国大数据标准化组织与相关国际组织的交流与合作，深化大数据领域的多边国际互利共赢。

## 9 附件：应用案例

附表 1 案例与应用领域的对应表

序号	应用领域	案例名称
1.	社会管理与公共服务	广州市公安局大数据平台
2.	社会管理与公共服务	咸阳市精准扶贫大数据平台
3.	农业领域	国家生猪大数据中心
4.	制造领域	中国航发南方工业公司AEOS运营管理平台
5.	通信领域	广东移动大数据实时跨集群协同云计算及行业化快速支撑应用
6.	生态环保领域	生态环境标准化体系建设成效与案例
7.	司法领域	政法跨部门大数据办案平台
8.	金融领域	蚂蚁金服数巢一站式大数据共创服务平台
9.	健康医疗	济南健康医疗大数据平台建设运营案例
10.	健康医疗	华西医院肿瘤专科临床科研智能大数据平台
11.	数据治理领域	面向智慧城市数据融合共享的大数据治理平台
12.	数据治理领域	招商局融资租赁（天津）有限公司数据治理项目
13.	数据治理领域	中国进出口银行数据治理项目
14.	数据治理领域	中国农业发展银行数据服务平台项目
15.	数据治理领域	国家电网公司数据治理实践

### 9.1 案例：广州市公安局大数据平台

#### 应用领域：社会管理与公共服务

广州市公安局大数据平台汇集来自公安、政务、社会不同层面的海量数据资源，实现了公安网、政务网、互联网、物联网等各类信息资源的汇集和整合，形成了跨业务领域的人、地、事、物、组织等多种维度公共数据集合的统一视图。在服务层面上，实现了跨警种、跨区域、跨部门的信息共享机制，为情报、刑侦、治安等各个业务警种提供实时、全方位的信息资源服务支撑。在应用层面上，实现了综合查询、全文检索、各类电子档案、人员轨迹分析、可视化情报分析等基础大数据应用功能，并建立灵活的大数据自助式分析机制，使数据资源得到有效的利用。该项目在数据资源标准化归集与应用、大数据标准体系与服务体系建设、大数据资产管理等方面进行了探索，具体体现在五个方面：

#### 1、以数据标准指导大数据体系建设

广州市公安局大数据平台建设之初即将标准体系建设的重要性提升至首要

地位，提出“标准先行”的要求。项目建设初期开展了深入的数据资源及标准体系的调研工作，梳理了公安信息数据种类与平台标准规范，参考国标及公安行业标准，对广州市公安局各警种、各系统所使用的数据字典进行归纳整理，建立了广州公安数据资源目录和平台标准规范体系。在建设过程中实现了地市数据标准化平台和省厅数据标准化平台的对接及双向的标准同步机制的建立，实现部、省、市三级标准的联动体系，并建立了物联网、互联网数据资源整合及应用相关的标准规范，对数据元进行扩充。全面汇集情报、户政、刑侦、消防、交警等各业务部门的现有数据字典标准并进行梳理，按照国标、部标、行标、专业标准以及本地标准进行分类归并，面向全业务警种提供服务支撑。同时，建立了数据标准管理系统，对数据标准进行管理及维护。

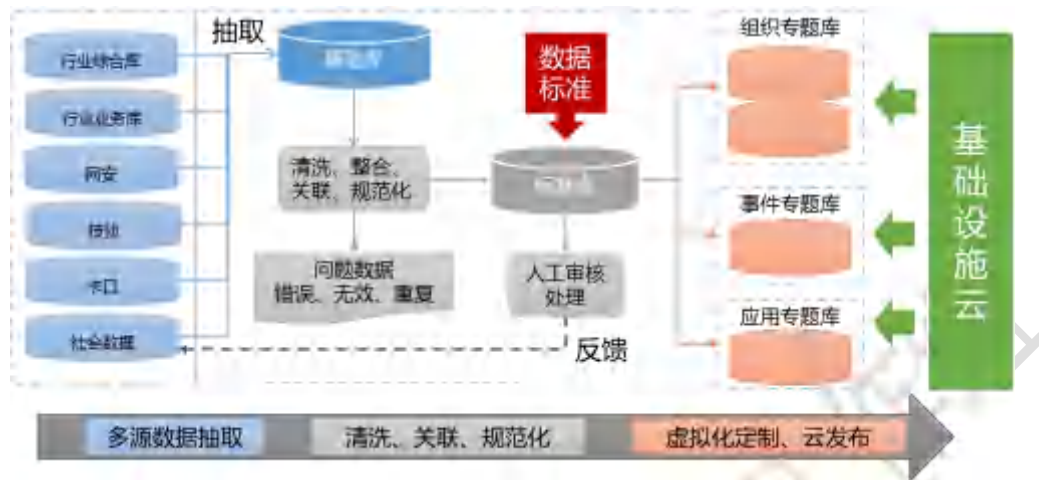
截至 2019 年 6 月底，该项目总共收集了 1400 多项公安数据元标准，4200 多个字典定义，共 163 万个字典项，覆盖了公安、社会、政务、交委、互联网、物联网等各个领域。数据对标工作为 11873 个非标准的属性项进行了标准化，最终清洗了上千亿条的核心数据资源。

## 2、跨行业、跨领域、跨地域的海量数据一体化处理流程

广州市公安局标准化大数据平台将公安、政务、社会不同层面的数据进行整合化、标准化、关联化、专题化，构成大数据资源库，形成了跨业务领域的人、地、事、物、组织等多种维度公共数据集合的统一视图。平台将以上各类的数据通过数据库抽取、文件交换等多种形式进行汇集，最终汇集的数据资源再通过数据资源汇集库汇集，并在此基础上构建可信关联库和业务专题库，为不同的应用场景提供数据服务。

在标准化的数据基础之上，广州市公安局对汇聚的海量数据资源进行了全面的数据可信化清洗与关联整合。截至 2019 年 6 月底，广州市公安局大数据平台已汇集上千类、几千亿条来自公安、政务、社会等不同来源的数据，包括来自市属相关部门社保、税务、工商等在内的几百类、上千亿的社会数据资源。并经过主数据分析与整合，形成人员、机动车、组织机构、地址资源、电话号码资源等主数据库，对原始数据资源存在的冗余、冲突、无效、错误数据进行了全面的分析，极大地提升了全局的数据质量。数据交换的频率大多为每天更新，非政务信

息方面一次性数据也有定期获取，如燃气信息等。数据体量十分庞大，时效性强，数据质量高。



附图 1 案例一：数据资源汇集流程图

### 3、以标准化、即开即用、安全可控的服务体系支撑业务应用

广州市公安局大数据平台通过对外统一的服务总线对所有服务进行统一管理，通过资源服务中心实现所有服务的对外共享和发布。并参考公安数据元标准，对服务接口进行了规范化设计，使所有的数据服务资源具备统一的规范。

该项目的对外开放服务接口包含数据统计、查询、比对、分析、可视化服务、应用集成等服务种类，覆盖市局、分局、派出所各级公安机关，以及网监、治安、情报、刑侦、出入境等多个业务警种和政府部门，支撑着出入境办证、无犯罪记录核查等重要业务应用。同时，建立了数据平台向其它业务警种以及政法网的各类线索推送体系，共享线索情报资源，并基于数据交换总线建立业务协同机制。完善对服务资源中心的服务资源管理流程，各个单位用户通过服务资源中心申请、审批获得数据、应用服务资源。

### 4、推动低价值的“数据”向高可信的“信息”的转化与提升

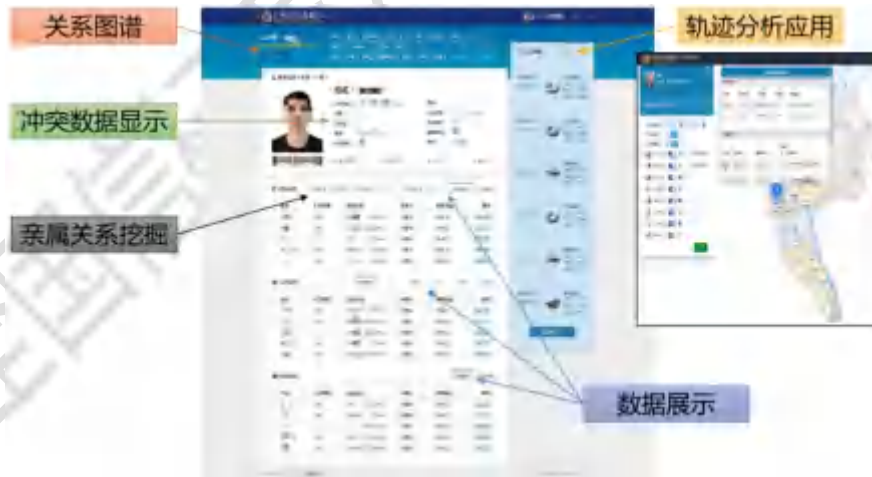
该项目建立了自助式的大数据分析机制和标签体系，将数据资源转化为知识，使资源得到了有效的利用。民警可登陆平台开展自助式大数据分析；业务部门可通过拖拽的方式可视化的定制大数据分析模型，根据需要在自助式大数据分析平台直接发布服务接口，在授权机制的支撑之上，开放给相应的业务系统进行调用。

标签体系以公安“五要素”信息模型体系理论为基础，结合警务工作的重点，从业务出发建立涉事案人员、案事件、涉事案组织等多个标签主题，并针对主题对象进行标签体系建设，进行多维度分析、快速分析用户特征、查询用户画像等。



附图 2 案例一：数据分析流程图

在建立了主数据分析、数据标签化、自助式分析等机制的基础之上，该项目实现了综合查询、全文检索、人员电子档案、机动车电子档案、组织机构电子档案、人员轨迹分析、可视化情报分析等基础大数据应用功能，并面向基层建设了新市派出所的大数据应用。并结合广州公安大数据平台门户，按照 AppStore 的模式建设公安行业的应用商店，实现了大数据应用的共享发布和管理机制。



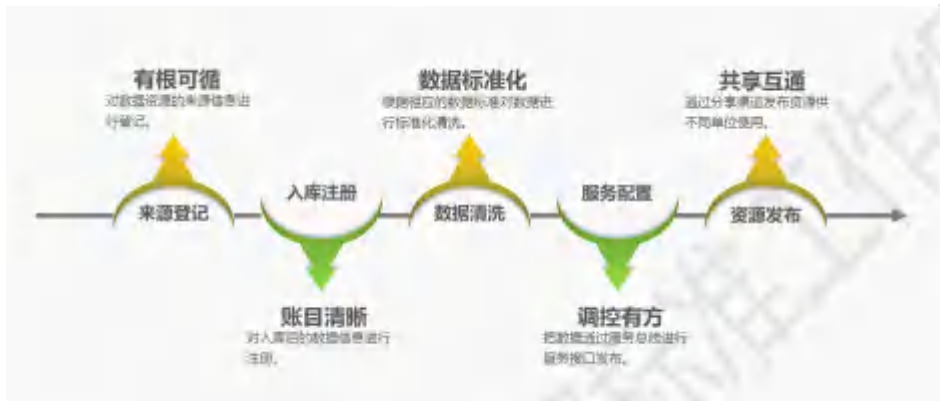
附图 3 案例一：系统应用功能展示

## 5、建立完善的大数据治理流程

广州市公安局大数据平台通过建立全局数据资源治理体系，实现对数据资产的全生命周期管理。以数据资产管理为核心，在数据管理和使用层面上进行规划、监督和控制，包括数据资产、数据标准、数据质量、数据安全、元数据、数据生

存周期等管理。通过规范化的数据治理保证数据资源的透明、可管、可控，完善数据标准的落地、形成完整的数据资源目录、规范数据处理流程、提升数据质量、保障数据的安全使用、促进数据流通与价值提炼。

通过建立数据资源加工链，每类新数据都需要经过来源登记、入库注册、数据清洗、服务配置、资源发布等加工环节，形成了流水线一样的工厂模式，大大提高数据处理效率。



附图 4 案例一：数据资源加工链将数据处理从作坊模式变为工厂模式

通过建立完善的数据运维管理体系，对数据治理全过程进行实时监控。实现了平台运行的全面监控报警，提高了数据的实时更新频率，确保提供最新最实时的数据资源，成功支撑了十九大期间广州市的治安保障工作。

## 9.2 案例：咸阳市精准扶贫大数据平台

应用领域：社会管理与公共服务——精准扶贫

### 1、项目背景

党的十八大以来，习近平总书记站在全面建成小康社会、实现中华民族伟大复兴中国梦的战略高度，把脱贫攻坚摆到治国理政突出位置，并提出了“六个精准”，即扶持对象精准、项目安排精准、资金使用精准、措施到户精准、因村派人精准、脱贫成效精准，确保各项政策好处落到扶贫对象身上。

咸阳市积极响应国家号召，将脱贫攻坚作为首要政治任务和第一民生工程，大力推进精准扶贫，打造了全市统一的“精准扶贫大数据平台”，解决了扶贫基础数据掌握不全面、扶贫对象识别不准确、帮扶施策不科学、缺乏动态监管和智能分析等问题，脱贫工作取得显著成效。



附图 5 案例二：咸阳市精准扶贫大数据平台业务场景

## 2、平台的大数据标准体系研究与建设

为实现“六个精准”，咸阳市精准扶贫大数据平台一方面需要采集全市各类扶贫基础数据，另一方面需要汇聚公安、财政、教育、人社、卫生、民政、住建、国土等相关部门业务数据，再将两类数据进行比对，围绕咸阳市建档立卡贫困户进行逐人核查，从而发现存在矛盾数据的贫困户，并将这些问题分发到镇到村，安排人员跟踪核查问题，确保实际情况与“国网系统”记录、纸质档案、贫困群众口述、帮扶队员掌握情况“五个一致”。





附图6 案例二：咸阳市精准扶贫大数据平台界面

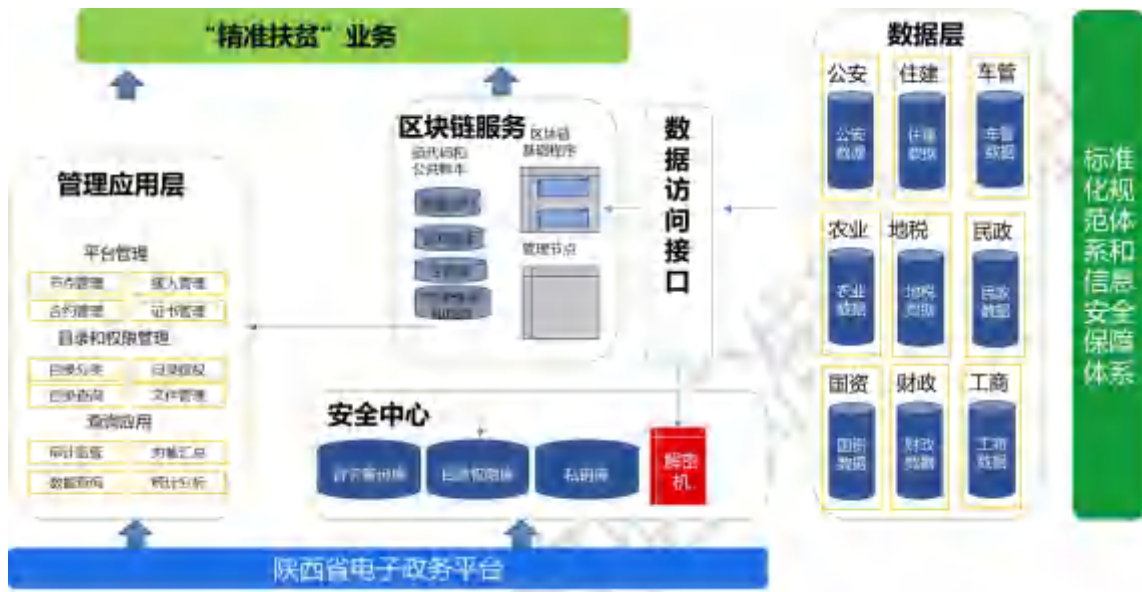
在贫困户筛查的过程当中，数据发挥了核心作用。但平台中的数据来源广泛，存在多源异构的问题，数据重复、数据错误、数据不一致等冲突现象明显，需要一套统一的标准体系来予以规范，实现数据有效处理、交换、共享与应用等。因此，咸阳市在参考《信息技术 大数据 参考架构》《大数据存储与处理平台技术要求》《信息技术 数据质量评价指标》等国家标准的基础之上，建设了信息资源、接口和安全三大类标准。其中，信息资源类标准包括《数据分类与编码标准》、《核心元数据标准》《数据处理规范》《数据质量规范》等；接口类标准包括《数据交换共享接口规范》《数据分析与应用接口规范》等；安全类标准包括《数据安全规范》等。这些标准为平台的数据采集、处理、存储与分析应用全过程提供了规范依据，为平台建设实施奠定了基础。

### 3、基于标准的创新应用

围绕“精准识别、精准投放”的工作思路，平台将陕西省政务云平台区块链数据服务引入精准扶贫，各部门数据无需上传至中心数据库，使用沙盒模式比对数据，结果获取后沙盒数据自动销毁，最大程度保障数据安全使用，保证数据权

属不变。

同时，对“精准扶贫”从认定、帮扶，到施策、脱贫的全流程都记录在区块链上不可篡改，方便业务部门对扶贫工作全面监管，防止弄虚作假、徇私舞弊。针对贫困原因，对扶贫资源进行针对性按需投放，从根本缓解致贫因素；扶贫资金的使用可追溯，使扶贫更加精准、高效、透明、公正。



附图7 案例二：咸阳市精准扶贫大数据平台逻辑架构

#### 4、平台实施效果

通过精准扶贫大数据平台的应用，咸阳市将社保、工商、税务、银行、公安等 15 个部门的 67 项数据上链，实现扶贫信息全程可溯、可管、可控，精准定位贫困户 42155 户，贫困人口 123379 人；实现扶贫资金低成本、高效率覆盖；通过人工智能筛选比对，发现问题数据 55577 条，与扶贫办等部门对接反馈相关结果，全面核查整改。全年实现精准脱贫 11758 户，44783 人。政务云+区块链+人工智能的数据共享应用模式，服务了咸阳近 500 万人口，市县镇三级 720 多个部门，应用成效这一平台是运用大数据解决百姓民生难点、痛点问题，化解社会关注的焦点、热点工作的有效手段，能够起到积极推动和促进为民服务水平、提升社会治理能力的双重作用。

### 9.3 案例：国家生猪大数据中心

应用领域：农业

## 1、项目背景

猪肉是我国居民的重要食品，我国既是养猪大国，又是猪肉消费大国。但我国生猪产业发展面临着生猪供需错位、政府对供应链投资的引导方向不明确、污染处理技术与粪便资源供给对接渠道不通畅等问题。一方面，政府方需要释放政府数据资源价值，提升对生猪产业监测监管的掌握能力。另一方面，生产主体需要数据指导调整生产计划，提高生产效率和流通率。此外，信贷机构也需要数据支持为生产主体提供信贷服务。基于以上现状，迫切需要建立国家生猪大数据中心，汇聚生猪产业链数据，释放数据价值，为各方决策提供数据资源与服务。

## 2、项目目标、技术路线及应用效果

国家生猪大数据中心的建设，一是通过数据采集和共享交换等方式，汇聚生猪养殖、交易、加工、流通、贸易及饲料、兽药、养殖机械等相关行业运行的数据信息，实现政府部门、相关机构的生猪数据共享交换；二是构建生猪产业监测预警的系列模型，为国家精准调控生猪产业发展和国民经济宏观调控提供精准的决策依据；三是在保障数据安全、不泄露国家经济秘密和企业商业秘密的前提下，充分开发生猪数据的商业价值，服务生猪产业主体；四是建设生猪大数据服务门户，打造国内生猪产业的权威信息发布平台。

本项目的技术路线如下：

(1) 以生猪产业链为依据，梳理涉猪数据资源目录，对生猪产业链各环节涉及到的数据进行采集，在数据可采集范围内，数据采集主要方式包括在线填报、手动导入（excel、word 等格式文件）、数据共享、网络抓取等方式；

(2) 数据加工处理，因采集到的数据来源广，数据格式、数据结构、数据类别、数据标识等不一，对采集到的涉猪数据进行加工处理，以便数据的后续利用；

(3) 实现数据交换共享，在数据资源汇聚并加工处理完成后，建设统一的数据共享交换标准和接口规范，实现生猪大数据中心和农业部、政府内部以及与外部单位之间的数据交换共享；

(4) 数据分析与应用，利用数据建模、商业智能、OLAP、GIS 等工具挖掘掩藏在数据资源表面下的业务特点、规律，实现对生猪产业上的工作进行决策支

持，提供数据监测、分析、预警、预测等服务。

通过对涉猪数据资源目录的梳理，以及对涉猪数据资源的采集和汇聚，深化生猪单品种大数据的应用发展；通过建设统一的涉猪数据共享交换标准和接口规范，促进生猪数据资源流通，充分挖掘生猪大数据价值，提升生猪管理水平；通过对涉猪数据的分析与数据应用平台的搭建辅助农业有关部门更准确地对生猪产业发展相关情况进行监控预警，指导生猪产业健康绿色发展。

本项目在开展过程中，通过对涉猪大数据资源目录的梳理，开展了数据采集、数据清洗等相关标准的研制工作。同时，依据生猪大数据中心的特点，对大数据中心总体技术架构进行规划，形成涵盖术语、网络、安全、应用等体系的整体框架和模型，以及业务承载模式，并依托大量的数据建模，为构建涉猪大数据中心标准体系建设提供支持。

### 3、案例实施情况及意义

国家生猪大数据中心一期（生猪大数据管理与应用平台）建设工程已顺利启动，一期建设内容包括：“1个体系、3个平台”，1个体系是指生猪数字资源体系；3个平台是指生猪大数据管理平台、生猪数据资源展示平台、生猪大数据应用平台。目前数据采集及基础平台建设稳步推进，数据中心平台建设也已进入设计开发阶段。

开展国家生猪大数据中心建设。从政府监管层面，通过对生猪全产业链数据的整合和把控，帮助政府有关部门释放数据价值，提升对生猪产业监测监管的掌握能力；同时利用大数据云平台的搭建以及大数据的深入分析，科学辅助检验检疫部门决策，提升检疫系统的数据处理能力，缓解检力不足。从产业链服务方面，帮助生猪养殖户及时了解生猪市场价格波动规律和价格走势，调整生产计划，降低养殖风险；帮助饲料厂、兽药厂等厂商提前获取相关信息，对资源进行合理配置调整和规划，并利用此类数据服务产生收益，助力农业供给侧结构性改革需要。从资金信贷方面，为信贷机构提供更多且更可靠的数据，解决其与农业信贷主体信息不对称的问题，助力改善金融机构对新型农业经营主体“惜贷”、“慎贷”的现状。

## 9.4 案例：中国航发南方工业公司 AEOS 运营管理平台

应用领域：制造——离散军工制造、航空工业

### 1、应用背景

中国航发南方工业有限公司（以下简称“南方公司”）经过近十几年的信息化建设，已经形成了以各业务应用系统为核心的业务平台，但现阶段各系统业务数据彼此孤立，呈孤岛形式。同时，各业务系统数据不一致，集成融合困难，数据没有转化成业务管理、决策所需要的价值信息，已经成为影响公司信息化水平与管理能力提升的重要瓶颈，难以为公司的经营管理提供准确支撑。

### 2、建设思路

流程是主线，数据是核心，集成是关键，应用是目的。美林数据助力南方公司以 AEOS 运营管理体系为指导，梳理公司关键业务流程，运用数据仓库、大数据分析等手段提高数据的关联重用与价值挖掘。基于数据驱动业务的原则，建成了一套面向公司决策层、车间管理层、生产业务层于一体的运营管理平台，将管理要求真正输送到末梢。同时，开展了航空发动机动态关键路径分析、零件期量标准计算与优化、安全库存动态预测等大数据分析应用，针对性调整和优化了生产业务过程，达到提质增效的目的。



附图 8 案例四：AEOS 运营管理平台架构图

### 3、项目总体实施内容和效果

(1) 构建业务主题数据库，逐步实现企业数据资产管理与建设。

以工业互联网和业务运营数字化为基础，参考《信息技术 大数据参考架构》、《信息技术 数据质量评价指标》等国家标准，通过采集、清洗、整合产品的结构和非结构化数据，建立基于大数据存储计算架构的业务主题数据仓库、数据集市，建设南方公司企业级的数据中心，全面将数据作为资产展开管理，以经营管理、产品设计、生产制造、供应商管理、服务保障业务为分析目标，发挥数据的价值，将数据变为信息、变为知识，提高决策准确性和预见性。

(2) 梳理公司共性业务，打造公司级一体化业务平台

南方公司不同的车间、加工中心业务也不尽相同，各中心为满足自身业务所需，建设了各自的业务系统，造成了现阶段烟囱式的信息化架构。美林数据通过对南方公司各加工中心的共性与个性的业务进行统筹与拆分，将共性业务汇聚，构建了覆盖多个加工中心的一体化业务应用平台，集中将共性的、关键的业务数据和分析过程统一调用展示，更好地沉淀和整合了业务能力，增加了分析服务的重用。

(3) 利用大数据技术开展工业大数据应用试点

通过零件生产周期相关数据的积累、统计和挖掘，优化公司生产周期期量标准，指导生产计划精确排产和异常订单追溯。一方面，将发动机 BOM 信息与订单交付周期进行关联梳理统计，利用大数据可视化技术，建立了产品关键路径动态分析展现方法，实现某类型发动机及单台发动机动态关键路径识别与剩余加工周期分析，帮助生产管理人员准确识别管控关键点，及时调整生产管控策略。另一方面，根据需求量、供应商交付周期、期望服务水平，分析当期库存对生产需求的满足情况，构建了基于业务机理指导的安全库存预测模型，有效减少了生产缺货情况，实现了库存的精细化管理。结合项目实施过程中的现场问题和解决过程为工业大数据领域拟编制的《信息技术 大数据 分析系统基本功能要求》，《信息技术 大数据 分析系统功能测试规范》等标准规范提供了信息反馈和经验积累。

随着南方公司信息化应用的不断深入，数据已成为其重要资产，是未来面向

智能工厂转变的重要基础，基于数据建立的运营管控平台将不断向横向与纵向扩展，直至覆盖公司全部经营管理、产品设计、生产制造、供应商管理、服务保障等各个业务域。项目的落地实施为行业内 AEOS 运营管理体系实施做出了示范，支持向国防装备精密制造型企业进行应用推广。

## 9.5 案例：广东移动大数据实时跨集群协同云计算及行业化快速支撑应用

**应用领域：通信**

### 1、项目背景

伴随着 5G 和物联网的高速发展，通信领域业务对海量数据的规模化、实时化处理、业务应用需求提出更高的要求。广东移动运营作为全世界最大的移动通信子网，总用户数达到 1.25 亿，每天通话次数达 73 亿次，每天用户上网流量高达 2600TB，需要存储的数据容量 400TB。由于数据的快速增长，一个数据中心无法存储所有数据，因此规划在三水数据中心的大数据集群存储 B 域、O 域和 M 域的数据，汕头数据中心的大数据集群存储家庭宽带的用户数据。为了解决异地跨中心实时协同处理的业界难题和行业快速应用支撑的瓶颈，广东移动、华为技术有限公司联合开展了《大数据实时跨集群协同云计算及行业化快速支撑应用》项目研究。

该项目基于大数据集群跨域部署现状和大数据平台体系，开展关键技术创新研究及应用研发，涵盖统一 SQL 查询、分布式计算下推、跨 DC 高速传输优化、内存管理技术、大数据应用平台构建、大数据应用服务提供等多个方面。同时基于机器学习技术，实现挖掘流程自动化、智能化建模机制。

项目成果实现了大数据技术的多项关键创新和能力提升，解决通信业务在大数据运算、大数据存储、大数据应用等诸多难点和痛点。同时以能力开放的形式支撑了智慧政务、精准扶贫、智能公安、智慧交通、文化旅游等多个行业领域的应用，产生了显著的经济效益和社会效益。

### 2、系统方案介绍及标准建设

广东移动实时跨集群协同处理的大数据平台下图所示：



附图 9 案例五：跨集群协同处理的大数据基础平台架构

**数据层：**深度研发大数据跨域集群协同高性能云计算组件，引入基于流式处理的实时计算平台，全面实现海量实时大数据的提供。

**服务层：**基于 PaaS 大数据平台的“三大开放能力”，即“资源开放”、“数据开放”、“应用开放”，闭环式全流程地提升业务支撑能力和用户体验。融合机器学习专区及微服务组件群，提供专业化智慧化的数据服务能力和业务运营能力。

**应用层：**开放对内对外应用快速支撑窗口，提供多样化一体化应用服务。面向企业内部运营、外部行业应用，分别开放相应的对内对外应用快速支撑窗口。

本项目是广东移动在大数据平台跨集群协同技术上的创新和实践。项目实现了在大数据运算、大数据存储、大数据应用等多方面的技术创新和能力提升，构建和完善大数据产业生态，推动高新技术和产业升级转型。系统的大数据平台建设遵循 GB/T 35295-2017《信息技术 大数据 术语》、GB/T 35589-2017《信息技术 大数据 技术参考模型》等基础性国家标准。并在架构设计、模型算法上做出自主创新，将通信业务和相关行业应用中提炼的需求和技术创新，向《信息技术 大数据 存储与处理系统功能要求》《信息技术 大数据 大数据系统基本要求》等国家标准提供有效输入，支撑标准研制工作。



### 3、主要创新点

#### (1) 提升异地跨集群协同效率

通过数据虚拟化的统一 SQL 查询、分布式计算下推、跨 DC 高速传输优化等技术实现跨集群云计算协同处理。在数据访问上，对多个异地集群设计了新的数据访问接口、统一视图，应用可以透明访问异地集群中的数据；在数据同步上，可以把子任务运算结果数据同步，同步前还对数据进行压缩，而现有技术方案需先将数据源全部同步；在协同计算上，项目支持把复杂 SQL 查询下推到不同的数据源执行计算并完成碰撞分析，最大程度地调动闲置资源，同时动态选择最优合并结果集数据，合理均衡利用了每个集群的计算能力；在运算结果存储上，通过全新开发的上层应用中统一协同计算组件，突破了以往对数据存放集群的限制。

#### (2) 高效的大数据实时流计算能力：

对分布式缓存 Redis(一种键值对存储数据库)进行优化，提升后端系统的性能和可靠性，配合流处理应用场景。对 Flink(一种开源流处理框架)内存管理技术进行优化，减小存储空间。对在流式计算中进行模式匹配的语法进行改造，使业务人员能够用熟悉的 SQL 编写复杂的 Flink 流应用(Flink CEP on SQL)，降低了开发门槛。

#### (3) 高并发、快速组装服务的大数据应用平台：

基于大数据实时跨集群协同云计算基础平台，搭建了高并发、可快速组装服务的大数据应用平台，实现了超过 2000 个接口的实时接入，6000 个指标的共享开放，开发了“蜂巢政信”“岭南优品”“蜂巢精销”“蜂巢热图”等 10 多个标准化产品，从而形成大数据行业应用生态链，为超过 1.2 亿手机用户提供一体化快速应用服务支撑。

### 4、应用成果及对地方的指导意义

项目成果通过能力开放的形式，迅速推广至广东移动 21 个市公司；通过应用开放的形式与数十个政府部门和重点行业携手合作，包含智慧政务、精准扶贫、智能公安、智慧交通、文化旅游等多个行业领域，提供专业化、智能化、一体化快速支撑。此外，还帮助合作伙伴实现业务的快速开发和部署，提升效率，降低

成本。这个项目的创新和成功实践有利于促进大数据产业发展，有利于推动国家数字经济发展，有效提升国家治理现代化水平和社会治理创新，全面提升公共服务水平。

**在民生方面**，提供公益信息服务，包括台风应急短信告知、精准扶贫、大数据防诈骗监控项目等。例如 2018 年台风“山竹”登陆前，广东移动充分发挥大数据能力优势，紧急开通“蜂巢热力图”，协助政府、警方进行人、船定位，利用政信通及时向失联市民发送告警引导短信，以引导市民安全撤离，确保群众生命财产安全。大数据平台在应急救援中发挥的作用受到广泛关注和肯定。

**在政务方面**，支撑政府工作，包括春运安全保障、广东省制造业大数据指数 MBI、珠海市政务大数据平台项目等。例如自 2016 年起每年向省政府提供春运大数据报告，自 2017 年起向春运指挥部提供交通枢纽实时客流大数据。春运期间，蜂巢政信持续为广大春节返乡人员发送交通指引、安全防诈提醒等短信，为广大人民群众提供安全保障服务。广州市交通委员会向中央媒体介绍春运大数据，展示了广州交委春运保障的能力和职责。

**在智慧交通、智慧旅游、智慧林业方面**，与广东省交通厅、广东省文旅厅等多个单位联合创建了交通、公安、旅游、城市规划等政务大数据实验室，推进广泛应用于人口分析、城市规划、交通指挥、综合治理、应急保障等行业及领域，并与各级政府单位开展深度合作，用大数据展示人口管理、安全维稳、交通运输、民生服务、应急指挥等方面的实时动态，为市民提供各种便利，提高城市管理和服务的运作水平，提升相关企业的经济效益。

## 9.6 案例：生态环境标准化体系建设成效与案例

### 应用领域：生态环境

生态环境部网络安全与信息化领导小组发布 8 项生态环境领域信息化标准，以需求为导向，制定发布《排污单位编码规则》《环境专题空间数据加工处理技术规范》《中国地表水环境水体代码编码规则》《生态环境信息基本数据集编制规范》《暴露参数调查基本数据集》5 项数据资源类标准规范，为生态环境部数据资源采集、传输、存储提供支撑。同时，结合信息化发展现状与趋势，发布《环

保物联网 总体框架》《环保物联网 术语》《环保物联网 标准化工作指南》3项信息化基础设施类标准规范，指导生态环境物联网建设。

发布的8项标准在生态环境部信息化建设中发挥了积极作用，取得初步成效。其中：

《生态环境信息基本数据集编制规范》（HJ966-2018）填补了我国生态环境信息领域基本数据集编制无标准可循的空白。通过指导各级生态环境管理部门编制各自业务相关的基本数据集，包括《暴露参数调查基本数据集》《固定污染源基本信息基本数据集》等标准的编制，规范各数据集编制的程序、方法、表达格式统一，实现环境业务数据的“标准化”，对生态环境信息化应用、环境业务信息系统集成、环境信息数据交换与共享等方面起到积极促进作用。

《排污单位编码规则》（HJ 608-2017 代替 HJ 608-2011）在生态环境部全国固定污染源统一数据库建设中推广应用。该编码作为全国固定污染源的唯一标识，建立唯一的固定污染源信息名录库，对污染源进行统一编码管理，实现污染源排放信息整合共享，有效推进协同治理，开启“一证式”污染源管理新模式。成为解决固定污染源数出多门的有效手段，在落实国务院政务信息系统整合共享工作发挥了重要作用。

《环境专题空间数据加工处理技术规范》（HJ 927-2017）在我部生态环境信息“一张图”建设得以贯彻实施。通过对各类数据进行标准化处理，将《环境专题空间数据加工处理技术规范》要求的环境专题空间数据图示方法及环境专题空间数据符号、配色方案，应用于地表水环境监测、大气环境监测专题图中，用户可更形象、直观的了解到站点分布以及各地环境质量状况。

## 9.7 案例：政法跨部门大数据办案平台

**应用领域：司法**

### 1、项目背景

随着政法行业信息化建设的深入推进和发展，针对跨部门办案过程中线下办理耗时长、信息多次重复录入、案多人少压力大、案件办理过程不规范、无法有效监督等问题，华宇结合在公、检、法、司等政法行业深厚的业务积累和丰富的

项目建设经验，协助某单位建设政法跨部门大数据办案平台，通过整合各政法部门已有的信息化建设成果，实现政法业务工作的全流程信息化管理，提高政法工作智能化水平，构成横向连接公检法司等政法部门的全流程协作，纵向深入到所有政法机关基层干警的全贯通的政法工作新格局。



附图 10 案例七 政法信息综合管理平台

## 2、政法跨部门大数据办案平台中大数据标准体系建设及应用

### (1) 标准体系规划促进平台规范化建设

政法跨部门大数据办案平台建设初期进行了顶层设计和统筹规划，建立政法跨部门大数据办案平台业务协同信息化标准体系，规划并组织编制了平台建设、网上办理业务流程、数据、电子卷宗流转等业务协同信息化标准，促进政法跨部门大数据办案平台标准化、规范化建设。

### (2) 标准提升了数据资源流转效率和深度利用水平

统一的业务流程标准打通并规范了公、检、法、司等政法部门之间的信息通道，实现了案件跨部门网上在线流转、办理及监督；从数据利用的角度出发，对公检法司的业务进行统筹考虑和“归一”处理，将业务术语、业务指标、业务规则、统计口径等相关内容整合到数据标准中，从而实现数据资源的深度利用。

### (3) 大数据标准化应用推动社会治理水平提升

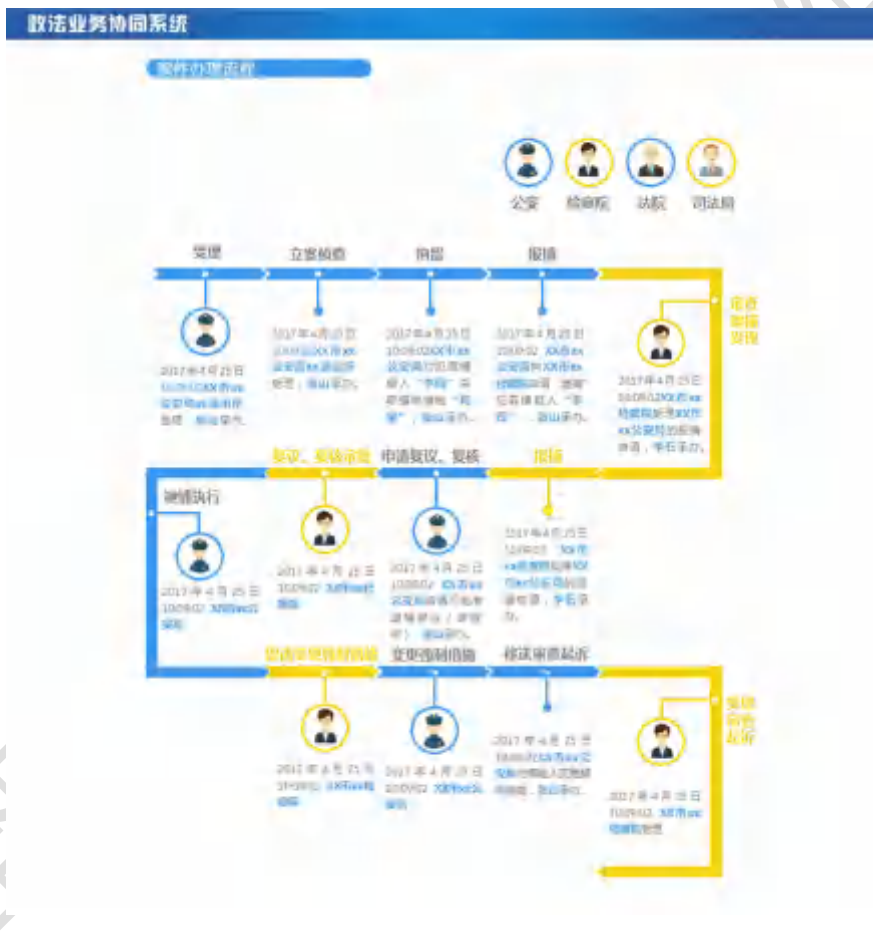
建立生产数据与数据交换标准映射模型，通过数据源的配置，依据模型实现数据交换，并且形成符合数据标准的信息资源库，再基于信息资源库，按业务域

建立业务主题库，从而可针对特定业务域提供更有针对性、更灵活快速的数据服务，有效推动社会治理水平的提升。

### 3、政法跨部门大数据办案平台运行质效

#### (1) 打通信息通道，提高跨部门协同办案质效

平台打通公、检、法、司等政法单位间的信息通路，取代线下人工办理，根据政法业务需求，案件信息一次录入，多次重复利用，解放了政法干警的机械重复性工作，切实解决案多人少的矛盾，打破政法部门界限，提高案件办理质效，降低案件协调成本，极大提升业务能力与工作效率。



附图 11 案例七 平台支撑案件办理

#### (2) 运用智慧科技，人工智能点对点支持办案

平台引入智能化办案组件，在刑事案件协同办理过程中统一标准规范，综合运用跨部门全流程案件信息，形成刑事案件知识图谱，智能辅助办案，进一步提升政法跨部门统一协作能力和智能化水平。平台在梳理全国千万量级的刑事案件案情基础上形成了刑事案件的要害库，基于案件特征词、案情要素进行智能分析

检索，向案件承办人推送全国范围内的相似案例。

### (3) 采用数据分析，挖掘信息富矿辅助科学决策

平台依托大数据技术，汇聚政法各单位业务数据，通过数据清洗、整合，形成政法数据资源池，对案件数据进行深度发掘，对刑事案件的变化趋势、涉案人员地域分布进行智能分析，为办案执法和社会治理提供参考，为重点人员管理提供线索，精准预测风险。

## 9.8 案例：蚂蚁金服数巢一站式大数据共创服务平台

**应用领域：金融**

### 1、项目背景

互联网时代，一切基于数据。数据的质量和数量成为影响金融大数据模型以及数据分析效果的重要影响因素，因此通过数据共享扩充数据量，从而提升模型效果的需求也变得越来越强烈。但数据共享问题重重，如何在满足用户隐私保护、数据安全和政府法规的要求下进行数据联合使用和建模，成为行业面临的难题。顺应时代发展、在国家推动“数据共享”的背景下，金融行业数据的整合、共享和开放正在成为趋势，给金融行业带来了新的发展机遇和巨大的发展动力。

### 2、数巢大数据智能服务平台简介

蚂蚁金服数巢大数据智能服务平台（以下简称：数巢智能服务平台）可以在安全、可信、公允的数据环境中完成数据共享，解决了数据共享与隐私数据保护的问题；并能够提供数据交换、数据连接、数据加工、数据挖掘等一站式数据服务能力，覆盖了大数据探索和研究的全链路需求。

数巢智能服务平台的技术架构图如下图所示：



附图 12 案例八：蚂蚁金服数巢大数据智能服务平台的架构图

数巢智能服务平台能够帮助金融机构实现数据的融合与共享，其架构图如下图所示：



附图 13 案例八：基于数巢智能服务平台实现跨机构数据共享的架构图

该平台上线以来，帮助金融机构进一步释放数据价值，产生较好的经济社会价值。

### 3、数巢智能服务平台的技术创新点

数巢智能服务平台帮助不同机构在满足用户隐私保护、数据安全和政府法规的要求下进行数据联合使用和建模，破解了数据共享和隐私保护难以平衡的难题，打破数据孤岛，发挥数据核变效应，具体的功能创新点有：

(1) 防止数据滥用：基于安全沙箱、多方安全协议技术、TEE 技术、远程认证技术，确保数据只能按约定好的行为进行使用，避免数据滥用；

(2) 隐私保护：基于 SGX 硬件加密等技术、可信计算集群化技术、主键保密技术等，确保数据无法被平台或者其他参与方窥探，避免隐私泄露；

(3) 易用性：多数据源接入、实现可视化建模、模型一键发布部署、完备的 API 可支持发起训练任务等。

#### 4、数巢智能服务平台的应用实践情况

数巢智能服务平台通过多方数据安全高效融合共创的方式，全面连接数据、渠道、客户及合作伙伴等创造商业价值。

在银行业应用中，数巢智能服务平台搭建了一套基于多方安全计算技术（Secure Multi-party Computation, MPC）下的数据融合、联合建模以及模型发布一体化平台方案，为蚂蚁微贷与银行的合作提供更完备的大数据风控能力支持。该联营合作模式，在具备用户端授权，隐私数据受保护的前提下，实现双方丰富变量的融合建模，帮助银行提升了风控效果以及数据处理的能力，符合政策监管要求，助力行方实现科技自主。

在保险行业应用中，数巢智能服务平台为蚂蚁保险与传统保险公司的联合运营提供精准权益策略，提高风险识别率的安全合规共享环境，孵化的车险分应用可以显著提升车险的差异化权益能力，在通过购险前的用户授权条件下，帮助保险制定更好的销售策略。通过车险定价模型实际评测的效果看，通过双方和其他合作方数据，车险分“从人”（从车主信息）因素能够细分不同风险的用户，对车主进行精准画像和风险分析，实现‘千人千面’的精准权益策略。

#### 5、数巢智能服务平台与大数据标准的协同及创新

蚂蚁金服数巢智能服务平台贯彻大数据标准体系框架，在遵循现行大数据标准规范之下，同时结合金融行业特性，将蚂蚁金服金融大数据实践积累的规范融入该平台。数巢共创平台以标准化的方式，从数据采集、数据融合、数据挖掘和数据呈现等全生命周期，根据不同类别数据的安全级别，设计、执行、复查、改进蚂蚁金服在各项云计算环境下的安全管理和技术控制规范措施，形成了更为严格的标准规范。数巢智能服务平台遵循“数据价值不侧漏、数据可用不可见、客户隐私不侵犯”的安全策略，覆盖从数据存储、数据访问、数据传输到数据销毁等多个环节的数据安全控制要求，并实现了与《数据分析与应用接口规范》《数据共享与服务接口规范》《信息安全管理体系实施指南》等大数据相关标准的对齐。此外，数巢智能服务平台作为金融级产品对于安全有更高的要求。平台实现对于国家以及国际安全标准的对齐，通过了 ISO 27001 全球安全标准的认证、可信云服务（TRUCS）认证以及中国国家信息技术安全研究中心的信息技术安全性



检测认证。

在合规以及数据安全保护前提下，国内跨机构的数据共享融合仍然处在初期的探索阶段。数巢智能服务平台实现了数据的多方协同和授权共享，得到更准确高效的模型和决策。基于实践积累，数巢智能服务平台形成了跨机构数据共享的标准规范，为金融机构进行数据融合提供了标准借鉴。

## 9.9 案例：济南健康医疗大数据平台建设与运营案例

**应用领域：健康医疗**

### 1、项目背景

在国家健康医疗大数据应用发展总体规划的“1+5+X”（一个国家数据中心，五个区域中心，若干个应用发展中心）中，山东省承担了国家健康医疗大数据北方中心建设任务。2018年4月，国家健康医疗大数据北方中心在济南签约成立，成为首家通过国家卫健委评估和授权的健康医疗大数据区域中心，济南市成为全国首个启动国家健康医疗大数据中心建设的试点城市。2018年6月，济南市卫健委与浪潮集团签署全面战略合作协议，依法依规授予浪潮集团对健康医疗大数据的汇聚、治理、应用及运营等相关权责，制定相关数据标准规范和安全体系搭建，启动了济南健康医疗大数据平台的建设和运营工作。

### 2、建设健康医疗大数据平台

参照国家大数据标准 GB/T 35589-2017《信息技术 大数据 技术参考模型》建设了国内首个健康医疗大数据平台 HDSP，采用全国首创的数据采集技术 CMSP，汇集医疗相关数据、政府数据、社会数据、互联网数据、环境学等医疗相关全量数据，形成健康医疗数据湖，并对汇聚的健康数据进行专项治理。打造数据计算平台，结合一码通主索引连接所有治理后的数据，根据应用需求形成分类应用，支持临床辅助、支持科研、支持超级档案检索。



附图 14 案例九：健康医疗大数据平台架构图

### 3、制定健康医疗大数据标准，汇聚全量数据

基于国家卫生行业标准和省级参考规范，结合济南实际创新制定《济南市健康医疗大数据目录》，共计 11 大类、60 个亚目、320 个细目，16717 个数据项，接入济南 23 个政府部门健康医疗相关数据，包含公安、民政、人社、农业、环保、市场监管、检验检疫、保险监管、安全监管、教育、体育、科技、统计、气象、残联等部门数据。汇聚完成了济南区域内 64 家公立医院、20 家社会办医院、20 家驻济体检机构、64 家社区卫生服务中心、47 个乡镇卫生院、191 家社区卫生服务站、2625 家村卫生室的健康医疗核心业务数 191.53 亿条、影像数据 3.81 亿张、电子病历 2794.63 万份，汇集整合数据量达到 170TB。形成了国内最大的健康医疗大数据资源池，在专业的数据治理技术支持下，构建了四大类 108 个数据集，数据汇聚治理总条数达到 200 亿条。

序号	数据集名称	数据项数 (条)	来源	更新时间	数据量 (TB)	更新频率	更新方式	更新周期	更新内容	更新时间	数据量 (TB)	更新频率	更新方式	更新周期	更新内容	更新时间	数据量 (TB)
1	济南市卫生健康委员会	1795	卫生健康	2018-12-31	130	实时	增量	每日	济南市卫生健康委员会	2018-12-31	130	实时	增量	每日	济南市卫生健康委员会	2018-12-31	130
2	济南市疾病预防控制中心	13725	卫生健康	2018-12-31	542	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	542	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	542
3	济南市疾病预防控制中心	79	卫生健康	2018-12-31	570	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	570	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	570
4	济南市疾病预防控制中心	12251	卫生健康	2018-12-31	54	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	54	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	54
5	济南市疾病预防控制中心	1799	卫生健康	2018-12-31	5	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	5	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	5
6	济南市疾病预防控制中心	6172	卫生健康	2018-12-31	10	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	10	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	10
7	济南市疾病预防控制中心	64282	卫生健康	2018-12-31	36	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	36	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	36
8	济南市疾病预防控制中心	5200	卫生健康	2018-12-31	34	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	34	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	34
9	济南市疾病预防控制中心	661	卫生健康	2018-12-31	141	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	141	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	141
10	济南市疾病预防控制中心	201	卫生健康	2018-12-31	700	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	700	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	700
11	济南市疾病预防控制中心	4	卫生健康	2018-12-31	20	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	20	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	20
12	济南市疾病预防控制中心	278	卫生健康	2018-12-31	1068	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	1068	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	1068
13	济南市疾病预防控制中心	1778	卫生健康	2018-12-31	2762	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	2762	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	2762
14	济南市疾病预防控制中心	1927	卫生健康	2018-12-31	10	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	10	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	10
15	济南市疾病预防控制中心	362	卫生健康	2018-12-31	38	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	38	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	38
16	济南市疾病预防控制中心	508	卫生健康	2018-12-31	3	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	3	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	3
17	济南市疾病预防控制中心	16	卫生健康	2018-12-31	114	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	114	实时	增量	每日	济南市疾病预防控制中心	2018-12-31	114
数据集总计 (条) : 131,625																	

附图 15 案例九：个人健康相关数据汇聚细目

#### 4、基于健康医疗大数据平台开展数据运营与应用服务

依托浪潮平台+生态模式，打造平台生态型业务架构模式，与合作伙伴一起构建新型健康服务应用体系，以数据运营带动生态发展，以平台支撑应用创新，面向政府、医疗机构、基层社区、居民提供医养健康创新应用，共同推进医养健康产业链快速发展。



附图 16 案例九：健康医疗大数据技术及应用服务模式架构

##### (1) 助力政府

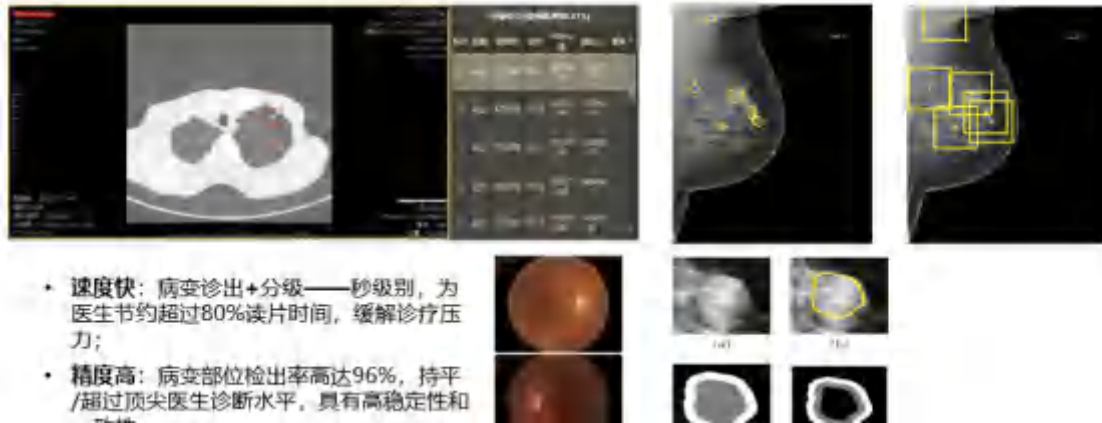
建成济南市基于健康医疗大数据的全民健康信息平台，为政府提供基于健康医疗大数据的决策监管服务，能够精准的为政府提供公卫数据分析、慢病综合管理、居民健康一码通等服务。



附图 17 案例九：济南市全民健康医疗大数据概览图

## (2) 赋能医疗

面向医疗机构，对已经治理的数据进行分类，目前在专病方面已经开放了21种专病队列，常见病200余种，全部病种两万六千余种，这将为医院在临床路径研究、医药研究、科学研究等方面提供坚实基础。并与合作伙伴开展基于大数据的人工智能影像辅助判读分析服务，提高医生识别病灶的准确率，降低误判风险，目前在乳腺癌和肿瘤以及糖网筛查方面有成功应用。



附图 18 案例九：人工智能阅片图示

## (3) 便民惠民

打造爱健康互联网服务平台，为基层居民和患者提供互联网一站式服务，并通过数据汇聚打通市各医疗机构数据，实现居民就医一码通服务，所有数据打通的医院间即可实现患者数据共享，居民可在爱健康 APP 查阅个人健康档案，也可授权医生读取个人的健康档案，真正实现跨医疗机构间的数据互联互通和共享，避免了患者重复检查造成的成本浪费，实现便民惠民。



附图 19 案例九：居民健康一卡通服务

## 9.10 案例：华西医院肿瘤专科临床科研智能大数据平台

应用领域：医疗

### 1、项目背景

现代医学已进入循证时代，基于严谨的科学研究过程产生的医学证据是优化和改良当前医疗决策的最优解之一。多年来医疗机构通过医疗信息化系统沉淀了海量的诊疗数据，这些数据是开展权威的医学科研，辅助诊疗决策优化，协助医院高效管理的重要支撑，具有非常高的应用价值。但受限于技术的瓶颈和高效解决方案的缺乏，这一海量的数据资产未被充分的挖掘和应用。

华西医院通过建设智能大数据平台，完成全周期全维度全模态病患数据采集，实现临床业务数据向标准化科研数据的智能转化、统一存储、处理、分析。该平台支持现有以及未来产生的各类医疗数据，包括文本数据，影像数据，基因组学数据等，并通过建立相关疾病专科数据标准，基于相应标准建立疾病专科患者诊疗模型，疾病专科疾病模型，形成相对应疾病的单病种库。

### 2、平台标准化研究方案及应用体系建设

肿瘤专科临床科研智能大数据平台建设的核心目的是对数据的分析利用，大量低质量的数据很难支撑深层次的临床科研应用。因此，系统建设之初规划、建设统一的规范术语标准体系。系统建设参考并遵循了《医院信息系统基本功能规范》、《电子病历系统功能应用水平分级评价方法及标准》、《电子病历共享文档规范》、《关于促进和规范健康医疗大数据应用发展的指导意见》、《医院信息安全等级评测规范》、《计算机软件工程规范》以及国际、国内相关标准 ICD、MeSH、LOINC、DICOM3.0、HL7、IHE、电子病历基本数据集等。项目建立遵循以上规范的医学术语标准体系，借鉴国际医疗信息化建设经验，优先采纳国际、国内、行业及医院标准，通过规范的术语管理标准体系，确保采集临床数据的质量，为科研数据的分析、挖掘提供支持。

肿瘤专科临床科研智能大数据平台围绕特定疾病继续建设科研专病数据库的基础平台。针对肺癌疾病特征，构建具有华西特色的疾病科研模型和自动化解析模型，有效贴合华西医院资深的临床数据特征，完成临床数据的多学科集成与

内容解析，将多模态的、异源异构的临床数据转化为有效的、结构化的、高质量的科研数据。

在平台上提供多种面向科研的具体应用，支持临床研究课题中肺癌相关患者的队列建设、支持大数据回顾性临床研究、支持大数据横断面研究、支持基于临床数据的人工智能辅助诊断、临床辅助决策、疾病预测等创新课题的研究。完成从数据集成、数据治理到科研支撑、临床应用的闭环支撑体系。



附图 20 案例十：科研平台架构图

### 3、数据治理标准研究的创新点

#### (1) 多模态数据处理引擎

通过依图 YT-Magic Box 将数据对接集成后接入平台的自由文本、影像数据及语音类数据,通过自然语言处理、影像识别、语音识别技术进行信息提取与处理。通过前沿的人工智能深度学习技术,结合医学术语库、知识图谱完成对临床自由文本数据的后结构化处理和标准化处理,同时对影像图谱进行自动化的病灶信息提取与病灶信息结构化处理,整合文本、影像、语音等多模态的医疗数据,为多种应用模型进行全面的信息解析。

#### (2) 智能数据治理

按照互联互通标准化成熟度、电子病历系统应用水平分级评价等相关规范,制定符合医院数据利用的标准规范。将医院 HIS、LIS、EMR、PACS 等系统原始数

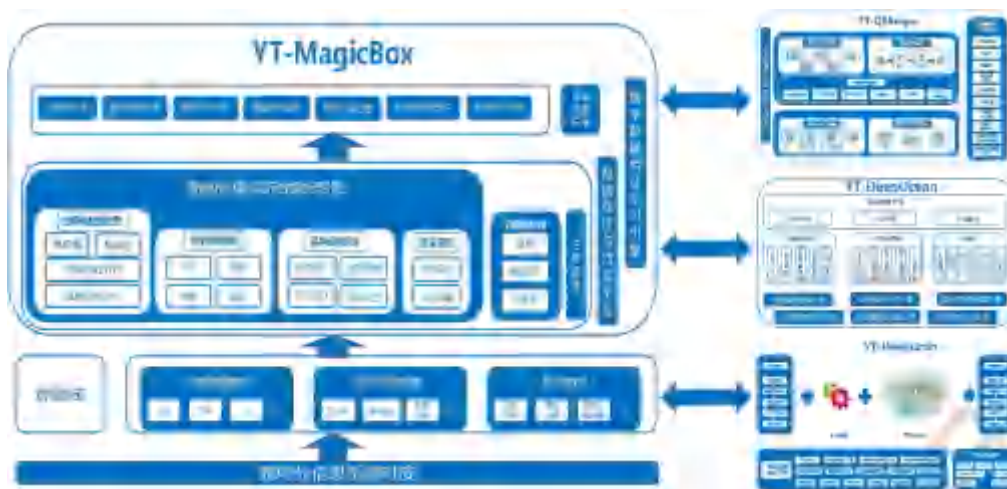
据通过 ETL 技术实现多学科多系统的临床数据采集和汇总。

制定数据清洗规则，包括非空检验、主键重复、非法代码清洗、非法值清洗、数据格式检验、记录数检验。对于空缺值、离群值和不一致等脏数据，采用人工检测、聚类、分类、关联规则等方法实现数据清洗。将相同的患者信息合并，对患者进行主索引处理。最终，基于数据清洗等技术完成患者的目标数据识别、数据的跨系统映射、关联、患者主索引识别等数据治理。

由于临床数据的不规范性、随意性等特点，需要对数据进行标准化处理。YT-Health Brain 系统参照国内外相关医学数据行业标准，无论医生书写的是标准词汇或非标词汇，均可通过平台转换，充分利用数据。对于较复杂的文本，通过自然语言处理技术和知识图谱，分析文本的隐含语义与上下文结构关系，实体识别、关系构建、医学知识库，进行模型适配，数据入组、模型比对、适配调试、结果比对和模型优化。借助人工+机器标注，发现更多医疗语义规则。最终，形成知识图谱构建推荐引擎、自主学习引擎、医学术语标注化配置引擎，基于自然语言处理、医学知识图谱、机器学习等 AI 技术完成各类临床文本数据的结构化、标准化和归一处理。

自主研发基于深度神经网络的影像结构化提取模型，实现医学影像数据降维与结构化，覆盖病灶类型、病灶大小、解剖学位置、影像学征象等影像诊断的核心内容，自动实现历史影像对比，为临床研究提供全维度数据。

最终完成院级科研平台的搭建，并完成全院诊治过的肺癌患者数据的治理工作，转化为可以支撑癌症相关临床研究的单病种科研平台；在平台上提供数据筛选、数据导出、数据可视化统计、数据建模分析、数据质控等一系列组件。



附图 21 案例十：智能数据治理图

### (3) 智能质控

依图智能质控 YT QA-Keeper 系统参照了医疗行业数据治理的基础标准，旨在结合通用模型标准，定义质量标准，面向通用数据产生结果，保证从原始数据到通用数据生产过程及结果的质量控制。智能质控主要包括制定数据清洗流程、流程控制、清洗质量控制、清洗过程管理等。通过规范流程和规则库，基于流程引擎构建统一可配置的数据转换、清洗、比对、关联、融合等加工处理过程。通过对异构异源海量离散的数据资源加工生产，形成全链路多级质控，保证数据完整性、准确性、唯一性、一致性、及时性。

### 4、标准应用效果及意义

智能单病种数据库建设完成后，实现万量级数据 AI 自动化提取入库少于 1 小时，将数据提取精度提升至 99.3%。以肺癌为例，成为国内首个含临床、影像、病理等多维度指标、数据全结构化的顶级智能肺癌科研病种库，为华西医院肺癌诊治提供大数据决策支撑；促进科研产出与成果转化应用，稳步提升华西医院的肺癌诊治水平。通过对该库的多中心运营，学科影响力进一步扩大，多中心学科建设与运营成本也将显著下降。同时，通过本项目研制肺癌人工智能辅助诊断标准，来引导行业发展、促进临床应用，提升医疗综合诊疗能力。



## 9.11 案例：面向智慧城市数据融合共享的大数据治理平台

**应用领域：数据治理**

### 1、项目背景

某市发展改革委作为社会治理、宏观调控的重要责任部门，近年来在委领导的高度重视下，已经基本实现从办公自动化为主的“电子发改”，向业务协同为主的“信息发改”转变。多年的电子政务建设，虽打下了有良好的信息化基础，但因数据规模和数据源种类丰富，导致数据标准混乱，数据质量层次不齐、各条块数据孤岛化较为严重，阻碍了数据应用。如同一项目在批复文件、项目系统上名称不同，无法关联，难以对项目全流程的有效管理；宏观经济数据来源复杂，数据标准不统一，难以快速形成宏观经济数据汇聚，支撑经济形势运行分析等工作；政务信息归集共享水平不高，信息孤岛现象依然存在；信息系统整合力度不够，业务协同应用亟待突破。为进一步落实《国家发展改革委“智慧发改”建设规划》，创新发展改革工作，打破处室边界，真正实现科学决策、精准调控、智慧监管和便捷服务，迫切需要加大系统整合力度，提升委内数据质量，加强数据管控，加快数据资源共享，加强业务统筹协调，实现项目全流程管理、宏观经济运行监测等。

### 2、平台介绍

大数据治理平台立足于国内智慧城市建设需求，采用开放、先进的互联网、大数据技术，遵循国家智慧城市顶层设计指南数据架构标准和信息技术服务治理数据治理规范设计，构建从数据采集、治理、融合到应用再生产的闭环式全生命周期管理，并利用大数据资源服务总线技术构建跨部门和跨业务的资源服务体系，充分发挥城市大数据资源的服务能力，为城市决策和各委办局构建具有实战应用价值的大数据应用，为办公协调应用、决策保障应用、政务服务应用、社会管理应用、民生服务应用、企业服务应用提供数据服务技术支撑。

### 3、实施内容：

(1) 数据采集，采用统一的信息资源标准和技术规范标准，建设多元数据采集系统，支持将政务数据、行业数据、互联网数据和其他各类不同来源的数据

资源采集并存储到数据全图，为数据挖掘、数据共享提供基础资源；

(2) 数据全图，包括各类数据汇聚形成的归集库、标准化和“一数一源”梳理后的中心库、以及数据加工处理后生成的各类基础库、主题库数据资源；提供城市级数据全图，扩大数据采集范围，实现全市的跨部门、跨平台、跨应用系统的数据采集；全面整合结构化数据、非结构数据及半结构化数据，整合数据，实现分布式、异构、异源数据的集成整合，为数据的互联互通及数据驱动应用智能化提供支持；

(3) 数据治理管理平台，实现对汇聚到大数据平台的数据进行集中治理和管控，针对发改数据特点，内置 7 种建库标准，包括字段标准、元数据模型、数据质量标志等，并基于标准管理子系统，加强了数据标准化、统一化；通过数据稽核，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；通过元数据管理的血缘/影响分析功能，快速定位数据问题，加强了对数据处理流畅的管控；内置 5 种核检维度、30 多个质量校验规则，通过数据质量子系统为数据管控能力、数据质量的提高提供长期稳定的服务和支撑，促进数据价值发挥；

(4) 数据资源目录，是针对所有委办局、资源中心已有和生成的数据资源统一编目形成的数据资源目录，用于数据检索、共享、开发使用与管理；实现自动化数据资源发现及主动数据资源注册功能，并提供智能判断资源的唯一性和资源质量，确保数据资源的质量；支持自动编目和手工编目，实现对目录服务中心的目录内容和目录服务运行的管理；通过自动化接口封装，对共享库中数据进行快速封装并提供应用，形成数据共享支撑能力；通过数据共享门户，对数据字典、数据名目、数据来源情况、数据调用情况等，进行统一可视化展示。

#### 4、主要亮点

结合业务场景形成统一的基础标准和各行各业可直接使用的主题标准，提升数据标准化程度，促进政务数据资源在跨层级、跨地域、跨部门、跨应用系统、跨网络的集成与共享；通过对数据质量进行识别、监控，提供数据质量问题报告，从而改善数据质量，为提升数据资产业务价值提供有力支撑，确保数据资产的准确性、权威性；通过建立宏观经济库、主题库等，支撑智慧发改产业分析、预测

经济形势、宏观经济研判、事中事后监管等业务分析需求，减少决策风险；将主要经济社会发挥指标数据与地理空间相结合，以“地图+图表+分析”的方式提供直观精确信息。

## 5、应用成果

为了有效解决企业扶持资金重复申报，申报后效果难以追踪的问题，构建委内统一的高质量大数据中心势在必行。通过汇集融合全委 38 套业务系统数据，并融合市场委、社保局等多个委办局的数据，结合互联网数据构建统一的法人库和项目库，通过大数据治理平台构建统一的数据标准，统一的数据融合机制和数据质量评价机制，并通过平台化进行落地保证，支撑项目申报唯一性自动识别，企业项目资助后持续跟进效果评估，确保政府扶持资源使用效率和效果。

大数据治理平台助力工信委实现了“一体高效运行”，即将原 38 个信息系统整合为一体的智慧服务平台。从数据采集、归集分析应用，对数据全生命周期进行高效治理，为“信息孤岛”的打通，实现“一数全面汇聚”奠定了坚实基础。

## 9.12 案例：招商局融资租赁（天津）有限公司数据治理项目

### 应用领域：数据治理

#### 1、项目背景

2018 年 5 月 21 日，银保监会印发《银行业金融机构数据治理指引》，要求银行业金融机构将数据治理纳入公司治理范畴，并将数据治理情况与公司治理评价和监管评级挂钩。在内受公司运营管控需求驱动的同时，外受监管机构管控驱动，招商局租赁在此情境下，立足眼下，着眼未来，规划大数据战略方案，“实行数据治理、建设数据仓库、逐步按需推进数据应用落地”，并依托北京华宇软件股份有限公司支撑开展“数据治理及数据仓库项目”建设，以统筹规划、分阶段实施的策略开展，逐步实现公司“用数据说话、用数据决策、用数据管理、用数据创新”的经营目标。

#### 2、项目目标

(1) 满足内部分析决策的需要。有效地整合利用已经积累的业务数据信息。通过科学的手段将数据转化为能支持客户关系管理、风险管理、财务分析、运营

绩效的决策支持数据。

(2) 满足提升风险管理的需要。租赁公司是经营风险的企业，通过本项目，密切关注资产分类，动态掌握客户的信用风险，准确的客户数据能切实提高企业资产质量，减少非预期损失。

(3) 满足业务创新的需要。高质量的数据是业务创新基础。公司推出创新业务产品，拓新业务模式，都需要对客户信息和业务数据进行挖掘，比较分析，基于数据的精准、详实。

(4) 满足系统建设的要求。数据标准化、数据治理等管控策略和体系的建设是整个信息系统建设的不可缺少的环节。通过数据治理，各个应用系统之间，统一数据接口，统一数据字典，统一数据报表，统一数据生存周期。消除信息系统的孤岛，进而提供数据共享服务。

(5) 满足监管的要求。2018年5月21日，银保监会印发《银行业金融机构数据治理指引》，要求银行业金融机构将数据治理纳入公司治理范畴，并将数据治理情况与公司治理评价和监管评级挂钩。通过开展“数据治理及数据仓库项目”建设，落实《银行业金融机构数据治理指引》，满足监管机构要求。

### **3、项目价值**

数据治理是全公司各部门各业务条线的责任和义务。做好数据治理工作，可为全员提供可靠支撑。对业务部门而言，数据治理成果是数据字典、业务规范及口径标准，是租赁业务说明书，是绩效的可靠保证；对信息化部门而言，数据治理成果是系统开发唯一的数据指导依据，是各系统数据参数关联映射的地图；对公司领导而言，数据治理的是公司的重要价值资产，可为经营决策，提供权威、可靠的数据支撑。

### **4、建设内容**

本项目采用统筹规划，分步实施的建设方法。



附图 22 案例十二：招商局融资租赁（天津）有限公司数据治理建设框架

### (1) 建立数据治理架构

成立数据治理领导小组及工作小组，明确岗位职责，协助制定数据治理管理制度，协助制定数据治理考核机制及办法。



附图 23 案例十二：招商局融资租赁（天津）有限公司数据治理战略架构

### (2) 建立数据标准

完成数据标准框架定义，开展了 580 多项基础数据项标准及 140 多项指标数据项标准的梳理，实现数据标准在数据仓库进行落标。搭建数据标准管理平台，

实现标准线上管理及应用。

### （3）数据质量管理

根据数据标准，协助对招商局租赁核心业务系统、财务系统等核心系统开展数据检核工作；通过数据检核，保证数据的准确性、时效性和完整性；完成数据访问权限等的建立；进行数据质量考核。

### （4）数据仓库建设

通过规范化、标准化手段，对内部数据及外部数据进行统一存储与分层建设。通过成熟 ETL 及调度工具“数据工厂 DataFactory”，构建数据模型，为管理决策提供权威、直观的数据支撑依据，促进公司业务长远发展。

### （5）数据应用建设

运用报表工具 BI@Report，建设工作报表、管理驾驶舱等系统，通过可视化手段，为用户提供各类数据服务，辅助领导决策，全面提升工作效能。

## 5、成果评估

现阶段“数据治理及数据仓库项目”建设已取得卓越的成绩。从管理层面看，招商局租赁已形成全司数据治理体系及各知识领域管理办法，并通过宣贯将治理体系和管理办法传导到个人。从操作层面看，配备专人，就前期完成的数据标准及质量检核规则进行数据检核，形成《数据治理错误数据确认单》，提交业务部门及工作小组确认，作为后续考核依据。从考核层面看，由战略发展部负责牵头，建立和完善数据质量考核体系。考核制度遵循源头负责制、及规范差错确认制，能够做到：谁创建、谁负责；谁审批，谁督导；谁使用，谁监督。每一笔错误数据由责任人确认。从信息系统建设层面看，通过数据治理平台搭建，辅助数据治理工作开展；同时，依据《数据错误确认单》问题，对数据在业务系统进行修复，保证数据源头的准确性，实现企业数据质量良性提升。

## 9.13 案例：中国进出口银行数据治理项目

### 1、项目背景

中国进出口银行（以下简称“进出口银行”）数据治理体系于 2016 年开始启动实施的基础类管理体系，是行内科技规划建设的最基本、也是最重要内容之

一。为了应对数据管理的挑战、满足监管部门的要求、有效提升行内的数据服务质量，并结合行内 IT 蓝图架构规划，建立数据治理管理体系，理顺内部协作关系，制定企业级数据标准，提高数据的准确性和一致性，提升数据共享水平，建立元数据管理、数据安全、数据传输与分发等管理流程，涵盖数据生存周期各个阶段，提高数据管控能力，进而提升银行在业务决策、客户营销、风险管理、内控合规和金融创新等方面的精细化、科学化管理水平。

## 2、工作历程

为落实行内第二次战略委员会中关于“进一步明确加强行内基础数据治理工作，确保数据的及时性、准确性和完整性，做好对存量数据的调整工作”的指示精神和行领导关于限时完成数据质量整改的要求，信息科技部于 2016 年开始逐步搭建全行层面的数据治理机制，并于 2018 年正式启动数据治理项目，搭建“一个机制，五个领域，五个平台”的数据治理体系，通过数据管控体系建设，加强对全行的数据管理，从全企业的视角协调、统领各个层面的数据管理工作，建立人、数据、系统之间的和谐关系，保证银行内部各岗位人员在正确的时间、正确的环境得到正确的数据支持及服务。

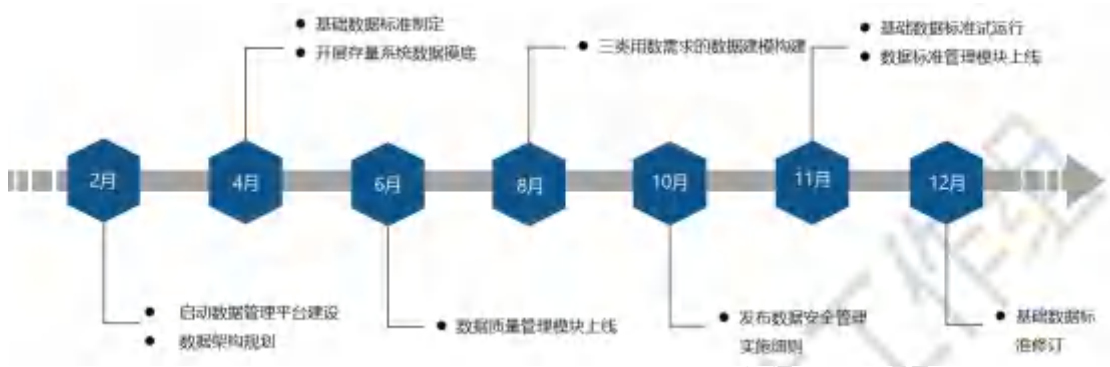
全行数据治理发展历程如下：

**2016 年数据质量摸索。**行内信息科技部于 2016 年初在报表系统中完成并上线数据质量检核功能，第一批次共开发上线了 19 个检核指标。随后以客户信息为主题，上线了 5 个批次，完成了行内主要客户信息相关业务系统如信贷管理系统、核心系统以及报表系统数据的一系列检核，有效提高了行内系统存量数据中客户信息相关数据的准确性和完整性以及行内系统存量数据中客户主体的唯一性。另外一方面，通过报送数据与前台业务数据的一致性检核，将数据质量检核进一步前置下探到前台系统内部的冗余数据一致性问题，从数据的源头彻底纠正问题数据，提高了前台业务系统数据的质量，进而优化了报表系统取数逻辑，同时大大减轻了报送补录期业务人员的补录工作量。

**2017 年数据治理起步。**随着数据质量工作的顺利开展，行内信息科技部加大了数据质量工作力度。一方面，考核频次上将原按季度考核的模式改为按月度考核，考核对象上纳入了优惠贷款部、转贷部，资金营运部、巴黎分行以及新建

分行等，数据检核覆盖系统扩大至资金系统、海外信贷系统的数据，不再仅仅局限于信贷、核心、报表等系统。较往年的数据质量增加了考核覆盖面和整改力度，完善了考核工作机制。

2018年数据治理工作全面铺开。下图为2018年数据治理重点工作进展情况。



附图 24 案例十三：中国进出口银行数据治理进展

### 3、工作框架

数据治理工程一期项目，从组织架构、制度建设、数据标准、数据质量、元数据管理、平台建设等方面开展工作。具体建设思路如下：

**一个机制：**搭建了一个高层决策、数据治理管理团队统筹、各部门和分行参与的数据治理长效机制，切实推进各项数据治理工作。

**五个领域：**数据模型、数据标准、数据需求、数据质量和元数据。一方面通过实现数据标准的落地，推动业务系统数据质量的提升；另一方面，推动行内元数据管理，减少系统变更风险，降低信息的复杂性，提高系统的开发、设计效率。

**五个平台：**依托基础系统建设，搭建数据标准管理平台、数据质量管理平台、元数据管理平台、数据需求管理平台、数据模型管理平台，通过自动化和系统规范减少手工操作，提高数据质量，管理银行数据资产。

### 4、工作目标

通过分析行内的战略规划，梳理银行现有的数据情况，初步建立全行数据治理的长效机制，包括完成相关管理架构的搭建，确定部门分工及职责，制定数据管控相关制度和流程；确定各部门数据标准、完成稽核问题的数据质量整改问题：

(1) 组织架构与制度流程。建立以信息科技管理委员会为决策层、数据治理工作小组为管理层、各业务部门和科技部门为执行层的三层数据治理管理架



构，理顺跨部门、跨领域的数据协作关系和协作流程，制订数据标准管理、数据质量管理、元数据管理相关的制度和流程。

(2) 数据需求管理。为解决后台系统报表开发过程中，遇到行内数据无法满足开发需求，组织供数需求沟通，制订解决方案，跟踪问题落实，并以此建立完善行内数据需求的工作流程。

(3) 数据模型管理。数据模型是作为数据与业务的一个连接，通过数据模型有效的了解金融机构的数据结构和业务规则，是 IT 和业务人员沟通的桥梁。

(4) 数据标准管理。根据数据治理委员会的相关精神，信息科技部将逐步承接由资产负债管理部负责的数据标准的制定发布工作，并于 2018 年底完成全行的数据标准的制订和颁布。

(5) 数据质量管理。以解决“数据质量清理情况通报”中的问题为基本目标，梳理分析问题成因、定位问题根源、制订解决方案，跟踪问题落实，评价整改成效，并以此建立完善行内数据质量问题整改的工作流程和评价考核机制。搭建数据质量检核平台，建立数据质量检核体系，继续深挖存量数据的质量问题，以稽核问题整改的工作流程和机制为基础，持续改进提升全行数据质量。

(6) 元数据管理。全面梳理全行系统的数据字典，建立各系统数据字典的基线版本，并纳入元数据管理平台。建立以元数据管理平台为中心的元数据管理流程，严格控制数据库结构变更操作，实现 T+1 日内获取各系统的数据库结构变更，建立完善相关配套的制度，优化上下游系统数据库结构变更的流程和机制。

## 5、项目价值

(1) 更好地服务于银行决策。在良好的数据治理环境下，未来管理层需要借助数据分析，实现敏锐的业务洞察能力，更好地适应经营过程中的不确定性因素、提升银行竞争力。

(2) 提供银行风险管理能力。银行在经营过程中可能会面对各种风险，比如对手可能违约的信用风险、误操作或流程管控缺失带来的操作风险、市场波动带来的市场风险以及政策合规性等方面的风险。通过对这些风险因素进行分类、识别、计量与风险，可以有效地管理和降低风险，为银行持续稳健经营提供保障。

(3) 提升银行运营管理水平。分析技术可以应用在行内的费用报销、人力

资源以及营销资源投放等领域，帮助行内识别冗余、低效的流程环节并加以改进，最终达到提高运营效率、降低成本的目的。

## 9.14 案例：中国农业发展银行数据服务平台项目

### 1、项目背景

伴随着市场开放、技术变革、产业社会形态变化的趋势，面临着盈利模式单一、经营模式同质、监管模式粗放等运营现状的管理压力，迫切需要借助数据资产的整合分析和深度挖掘，促进业务创新和价值发现，用数据洞察真相，驱动决策。同时，随着此类系统的持续投资、建设和完善，的确满足了相关领域所需要的一些报表也提供了部分数据分析能力，然而也出现了报表重复实现、数据标准不一、用户对于分析需求来自何处缺乏清晰了解、报表陈旧等各类问题，导致业务人员无法通过此类系统获得及时准确的数据认知，也难以进行数据深度利用。

基于以上对行内现状的分析，数据服务平台一期已于 2017 年正式上线运行，平台以数据服务为宗旨，全方位提升数据分析与服务水平，助力农发行业务发展向信息化、多元化、精细化转型。

### 2、数据服务能力建设的关键点

数据服务能力的构建首先需要回答：数据服务包含哪些内容，这些内容是用什么样的方式服务哪些最终的使用者。构建数据服务体系应包含三大要素：服务目录（内容）、服务模式（模式）与服务对象（用户）。服务目录是指对农发行数据服务内容的划分方式，为便于规范整个数据服务体系的编制、查询、使用及管理而建立的分类体系。服务模式是指数据服务呈现给最终用户的方式，呈现的方式往往通过技术手段实现。服务对象则是数据服务最终需要满足的用户，一般而言可按照不同的组织层级和业务条线或者对数据分析的功能要求等进行划分。

### 3、行内数据服务能力建设

本期项目主要从引入可视化数据分析能力、优化需求管理模式、建立统一的数据服务门户等几个方面构建数据服务能力：

#### （1）可视化分析

可视化分析打破了这种传统的报表模式，直接面向业务用户开放数据资源，

由用户采用简单的拖拽操作即可动态的制作统计报表，很好的满足了日常工作中数据灵活查询和分析的业务场景。数据资源的组织结合服务目录分类及业务部门的需求，明确不同领域下所关注的统计类指标和分析维度，基于指标和维度的自由组合可进行灵活的组合分析，提升业务人员数据分析效率。

## （2）需求管理

按照需求复杂度、业务紧急程度，对需求管理模式划分为传统固定报表类需求、灵活查询类需求、即时服务需求三类。对于全行广泛使用的固定报表，沿用行内需求管理规范，从报表需求的提出、数据口径定义、实施控制、报表验证等全生命周期的流程采用严格的管控，避免重复的报表并严控需求变更；对于用户灵活的查询和分析需求，可采用提供基础数据由使用者自行加工甚至自行管理的方式即可可视化数据分析功能进行管控；对于业务急需的数据应用需求提供即时服务，由技术人员通过 sql 语句或者平台中的报表模块为该需求定制个性化查询。

## （3）数据服务门户

门户首页是数据服务平台所有数据服务的统一入口，通过集成服务目录，为各类用户访问相应的数据服务提供统一指引。提供用户个性化设置，做到操作方便、快捷、人性化，满足用户对数据资产各种查询场景的要求。

### 1) 数据应用整合

在数据服务平台中集成农发行现有综合报表系统、元数据管理系统、数据质量管理体系、可视化平台、高管驾驶舱、反洗钱系统、统计集中管理系统、通用台账系统等应用系统，并建设用户的统一身份认证，实现业务应用的统一访问。

### 2) 构建数据服务目录

对农发行常用业务场景、业务条线、应用系统等分类进行总结，构建不同的专题领域，包括客户分析、运营分析、财务分析、监管报送，客户一部报表、基础设施贷款检测，标准化存贷款、信贷风险等，业务用户可以通过数据服务目录的分类索引，便捷的获取所需要的报表分析应用或服务功能。

### 3) 数据资产检索

通过资源整合实现数据资产化梳理，面向用户提供开放式搜索框，支持按类型筛选，按关键字检索等方式便捷定位。针对用户没有访问权限的资源也可检索

到但是不能查看详情，同时提供权限申请提示，促进数据资产的有效共享。

#### 4、项目价值

本期数据服务平台的建设作为农发行数据服务能力建设路径的第一步，在规划上充分考虑战略性和前瞻性，在设计上时刻贯彻以业务用户为中心的服务理念，在实现上遵循“统一登陆、统一整合、统一指标、统一展现、统一管理”的五大统一原则，为农发行未来开展数据资产深化应用打下了坚实基础。

## 9.15 案例：国家电网公司数据治理实践

### 1、项目背景

“十一五”以来，国网公司全面实施 SG-ERP 信息化工程，大力推进信息化建设，建成全球规模最大的一体化集团级信息系统，全面覆盖公司各级单位、各项业务和各类人员，实现了“从无到有、从线下到线上、从分散到集中、从孤岛到集成”的巨大转变。随着信息化应用深入推进，国家电网公司积累了大量的数据资源，从终端用户的日常电力消费数据到电网建设、运行、检修等数据，进而衍生到充电桩、储能设备等新能源领域数据，通过这些数据不但可以预测国家宏观经济的走势，提升电网公司自身的精细化管理水平，也可以有效支撑各类基于数据创新应用的新兴业务，数据已经成为国家电网公司的战略资产。

国家电网公司聚焦建设世界一流能源互联网企业建设，守正创新、担当作为，提出了“三型两网、世界一流”的战略目标，将能源流、业务流和数据流“三流合一”作为两网融合的重要内容，在这个过程中，数据资产的管理和应用就成为公司战略推进的关键之一。为支撑公司战略目标的落实推进，国家电网组建了互联网部，在总部成立互联网部（内设数据管理处），并在各部门设立数据处，组建国网大数据中心，体现了公司党组对数据管理的高度重视。为发挥数据资产的价值，支撑好公司“三型两网、世界一流”建设要求，国网公司组织开展了数据治理深化建设的工作。

### 2、建设思路

国家电网公司一直比较重视数据治理的工作，从早期 SG-ERP 时代的企业级数据模型建设，到近几年数据质量、主数据等工作的开展，为公司的管理和发

夯实了数据基础。但是过程中也存在很多的问题，特别是数据定义不全面、重点数据不突出等。在对国内外数据相关理论和实践全面研究的基础上，结合国家电网公司数据管理的现状，创新性的提出了“盘、规、治、用”的数据治理总体思路，坚持“统一管理、夯实基础、服务应用、持续完善”的工作原则，加快构建公司数据管理体系，使分散、孤立的数据成为汇集、共享的数据，构建公司统一的数据资产目录，推动公司数据有重点的管理，促进数据在各部门之间的共享和应用。



附图 25 案例十五：数据治理总体思路

1) 盘：对公司各系统数据进行盘点，明确数据治理的重点目标，梳理数据的定义和管理要求，构建公司数据资产目录，建立数据认责体系，推动数据的共享、整合，奠定数据管理和应用的基础。

数据盘点是数据整合与应用的基础，也是很多单位数据管理工作中的痛点，国家电网公司基于以往经验的总结，提出了总部及省公司两级协同的工作机制，由总部提出总体的数据盘点工作机制，形成体系化的工作方法和配套的工作模板，然后制定总体的工作计划和工作分工，省公司负责按照统一的工作方法要求推动本单位盘点工作，通过两级数据盘点的协同开展，国家电网公司全面盘点了数据家底，理清了各类数据的定义，构建了两级数据目录，实现了数据资产的可视化，为数据管理和共享应用奠定了基础。

2) 规：结合泛在电力物联网建设，聚焦业务需求和业务痛点，强化数据标准建设，建立统一数据标准，完善公司数据管理办法，为公司数据规范管理提供基本遵循。

数据标准、数据规范是数据治理的基础，由于省市之间业务差异及基础数据量大等原因，国网部分重要业务系统均为省公司二级部署，统一的数据标准是支撑各省公司在统一框架下开展信息系统建设与数据归集应用的基础。在此基础

上，公司还建立了常态的标准运行维护机制，结合应用情况持续对相关标准进行修订完善。同时，公司坚持“统一管理、分级负责，集中共享、服务应用、夯实基础、持续推进”的原则，组织制定了统一的数据管理规章制度，要求各单位、各部门共同落实，推进数据管理体系的落地。

3) 治：构建公司统一的、覆盖数据全生命周期的数据管理体系，推动数据责任体系的落实，狠抓重点数据的质量治理，建立并落实数据认责考核体系，提升公司基础数据质量。

数据质量治理是开展数据共享应用的基础，国家电网公司首先从组织层面进行调整，在公司总部、各省公司的互联网部设立了数据管理处，在各级业务部门也明确了对应的数据处，组建了专业支撑机构大数据中心，实现了数据管理组织的横向协同、纵向贯通。同时，以公司级重点数据应用建设为抓手，开展相关重点数据质量核查，建立健全数据认责评价机制，以应用成效检验数据管理成绩，推动数据治理体系的扎实落地。

4) 用：建立公司级的数据需求响应制度，以负面清单为基础推进公司内部数据的共享应用，培养公司的数据运营团队，挖掘数据应用的场景，提升公司数据管理和服务的水平，推动数据敏捷、高效应用。

数据的共享应用是数据价值的重要体现，也是数据管理工作的主要服务目标，国家电网公司组织系统研究，提出了以负面清单为基础的内部数据共享应用机制，改变原来“一事一议”的数据授权模式，基于最小化原则梳理数据共享负面清单，推动负面清单外数据在公司内部的高效共享。同时，组织数据应用团队主动挖掘公司级的数据应用场景，探索通过数据来推动公司管理、业务运营和数据对外服务的能力，实现从数据到资产的转化。

### 3、工作成绩

在“盘、规、治、用”思想的指导下，经过一年多各单位的协同努力，国家电网公司的数据治理工作取得了显著的成效，主要体现在以下四点：

(1) 理清数据资产的家底，建立了可视化的数据资产目录。数据盘点是国网公司 2019 年的重点工作，公司各部门在数据处的领导下全力协作，对公司的一级系统、二级系统进行了全面的梳理，形成了两级的数据资产目录，彻底打开

了各个系统的黑盒子，实现了数据资产的可视化、显性化。

(2) 推动了数据的高效共享。坚持“以共享为原则、不共享为例外”，完善公司内部数据共享机制，组织总部各部门及省公司以数据目录为基础，梳理数据共享负面清单，原则上除列入负面清单的数据外，其余数据均可在公司内部共享使用，同时改变当前由业务部门分别授权的方式，有效支撑基层单位和各部门的数据共享应用。

(3) 建立了新型的数据治理组织。国家电网公司充分参考国际数据管理协会 DMBOK 中数据治理组织的架构模型，同时参考各行业数据治理组织的建设经验，明确了统筹公司数据管理工作的部门，作为公司数据资产的第一责任方代表公司行使数据管理和应用的权限，同时在业务部门中也设立了数据处，强化了各业务部门在数据治理工作中的参与程度。

(4) 逐步推进数据文化的建设。国家电网公司从数据治理工作建设之初就认识到数据文化的重要性，公司数据资产的管理和应用依赖于公司各个部门、各个层级人员对于数据的理解、认识和应用。为此，公司从多渠道开展数据文化的建设工作，通过与各行业数据专家的交流以及对各单位先进实践的学习，以及赴外部开展数据管理专项调研，提升大家对数据管理工作的认识；其次，加强对公司的数据管理制度、标准和管理要求的宣贯，提升大家对于数据工作的重视；最后，层层落实数据责任，明确各业务部门、各基层单位在数据工作中的职责，加强各方人员的参与，进而实现全员参与的数据环境，逐步建立公司的数据文化。

# 创泽智能机器人集团主要产品



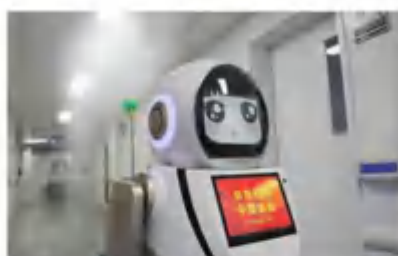
智能服务机器人



智能陪护机器人



安防巡检机器人



消毒机器人



智能党建机器人



智能教育机器人



智能导诊机器人



银行智能机器人



室外智能消毒机器人



多功能消毒机器人



全自动智能消毒杀菌机器人



智能医用消毒机器人



了解更多登录官网

[www.chuangze.cn](http://www.chuangze.cn)