

大数据标准化白皮书 (2018 版)

编写单位：中国电子技术标准化研究院
全国信息技术标准化技术委员会大数据标准工作组

二零一八年三月

《大数据标准化白皮书》（2018 版）

顾问指导组

梅 宏 谢少锋 戴 红 赵 波 林 宁

李冠宇 孙文龙 商 超 刘大山 傅永宝

侯建仁

全国信息技术标准化技术委员会大数据标准工作组

组长：梅 宏

副组长：孙文龙 杜小勇 吴建明 闵万里

秘书长：吴东亚

联络员：刘大山 傅永宝 侯建仁

编写单位（排名不分先后）

中国电子技术标准化研究院	江苏省经济和信息化委员会
勤智数码科技股份有限公司	成都市标准化研究院
神州数码系统集成服务有限公司	中兴通讯股份有限公司
北京大学	北京市金杜律师事务所
华为技术有限公司	成都四方伟业软件股份有限公司
浪潮软件集团有限公司	重庆大数据研究院有限公司
美林数据技术股份有限公司	江苏中堃数据技术有限公司
陕西省信息化工程研究院	深圳市华傲数据技术有限公司
北京东方国信科技股份有限公司	北京软件和信息服务交易所有限公司
湖北省标准化与质量研究院	星环信息科技（上海）有限公司
中电长城网际系统应用有限公司	九次方大数据信息集团有限公司
中国科学院信息工程研究所（信息安全国家重点实验室）	湖北省楚天云有限公司
	交通运输部科学研究院

编写组（排名不分先后）

代红	吴东亚	董建	张群	刘宇峰	李正	张慧敏
马珊珊	光亮	尹卓	黄先芝	于洋	马红霞	张勇
张永丽	谢秋琪	闵京华	郭亚茹	刘莎	马洪杰	李佳承
王东	卫凤林	叶润国	宁宣凤	周刚	刘斌	魏清
何运昌	于铁强	周洪明	胡荣	夏天	杨坤	赵茅
张巍	赵俊峰					

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	4
2.1.3	英国大数据政策	5
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	印度大数据政策	7
2.1.7	澳大利亚大数据政策	7
2.2	国内政策	8
2.2.1	国家和行业政策	8
2.2.2	国家大数据综合试验区	10
2.2.3	地方政策	13
2.3	安全隐私法律法规	15
2.3.1	国内外立法现状	16
2.3.2	大数据立法趋势和对产业的影响	19
3	大数据发展现状和趋势分析	21
3.1	大数据核心产业链	21
3.1.1	大数据产业生态商业角色构成	21
3.1.2	大数据产业生态商业模式分析	23
3.2	重点领域应用	24
3.2.1	社会管理与公共服务	24
3.2.2	农业	31
3.2.3	制造业	32
3.2.4	电力及水务领域	33
3.2.5	通信与电子商务	35
3.2.6	交通运输领域	37
3.2.7	邮政领域	38
3.2.8	金融领域	39
3.2.9	科学研究	40
3.2.10	教育领域	41
3.2.11	卫生领域	43

3.2.12 文化领域.....	44
3.3 大数据发展趋势分析.....	44
4 大数据参考架构.....	47
4.1 参考架构.....	47
4.2 参考架构中涉及的相关内容.....	49
4.2.1 系统协调者.....	49
4.2.2 数据提供者.....	49
4.2.3 大数据应用提供者.....	50
4.2.4 大数据框架提供者.....	52
4.2.5 数据消费者.....	54
4.2.6 安全和隐私.....	54
4.2.7 管理角色.....	57
5 大数据标准化现状.....	59
5.1 国外标准化现状.....	59
5.1.1 ISO/IEC JTC 1.....	59
5.1.2 ITU-T.....	61
5.1.3 IEEE BDGMM.....	62
5.1.4 NIST.....	63
5.2 国内标准化现状.....	64
5.2.1 国家标准现状.....	64
5.2.2 行业及地方标准现状.....	66
6 大数据标准体系.....	68
6.1 大数据标准体系框架.....	68
6.2 标准明细表.....	70
6.3 重点标准描述.....	75
7 我国大数据标准化工作建议.....	77
7.1 完善大数据标准体系建设.....	77
7.2 加强大数据标准的宣传推广工作.....	77
7.3 加强大数据重点标准应用示范工作.....	77
7.4 加强大数据标准化在数据治理领域的推进作用.....	78
7.5 加快大数据标准化人才培养.....	78
7.6 加快我国大数据标准的国际化步伐.....	78
8 附件：成功案例.....	79

1 前言

1.1 研究背景

全球已步入大数据时代，互联网上的数据量每两年会翻一番。截止到 2013 年，全球数据量为 4.3 泽字节，2020 年有望达到 40 泽字节。如果将数据视为一种生产资料，大数据将是下一个创新、竞争、生产力提高的前沿，是信息时代新的财富，价值堪比石油。大数据所能带来的巨大商业价值，被认为将引领一场足以与 20 世纪计算机革命匹敌的巨大变革。当前，世界各国政府和国际组织都认识到了大数据的重要作用，将开发利用大数据作为夺取新一轮竞争制高点的重要抓手，世界工业发达国家纷纷制定相关政策，积极推动大数据相关技术的研发与落实。

我国正处于数字经济发展的转型时期，信息的公开、共享与服务成为时代发展的主题。信息逐渐成为与物质和能源同等重要的资源，以开发和利用信息资源为目的的经济活动迅速扩大，逐渐占据或超越工业活动在国民经济活动中的地位。大数据的出现是跨学科技术与应用发展的结果。对于大数据，自然科学家强调在网络虚拟环境下对于密集型数据的研究方法，社会科学家则看重密集型数据后面隐藏的价值与推动社会发展的模式。目前大数据在支撑履行政府职能、保障公共安全、实施社会治理、支持重大决策和改进公共服务等方面发挥出越来越重要的作用。

党中央、国务院高度重视大数据发展，将大数据上升为我国国家战略之一。党的十九大明确提出“推动互联网、大数据、人工智能和实体经济深度融合”。国务院《促进大数据发展行动纲要》（国发〔2015〕50 号）明确指出要“建立标准规范体系”。工信部 2017 年初发布的《大数据产业发展规划（2016—2020 年）》（工信部规〔2016〕412）中部署了“推进大数据标准体系建设，加强大数据标准化顶层设计，逐步完善标准体系，发挥标准化对产业发展的重要支撑作用”的重点任务。2017 年 5 月 14 日，习近平在“一带一路”国际合作高峰论坛上发表讲话“要坚持创新驱动发展，加强在数字经济、人工智能、纳米技术、量子计算机等前沿领域合作，推动大数据、云计算、智慧城市建设，连接成 21 世纪的数字

丝绸之路。”2017年12月8日习近平在中共中央政治局第二次集体学习时强调“实施国家大数据战略加快建设数字中国”。

2016年，在工信部信软司和国标委工业二部的指导下，中国电子技术标准化研究院组织国内相关产、学、研单位的专家针对大数据应用、产业、技术与标准化需求进行了问卷调研。通过对调研数据的分析，初步形成了对于大数据应用、技术、产业发展以及标准化需求的分析成果。作为业界共同研究的基础，由30家单位共同编制形成并发布了《大数据标准化白皮书》2016版，得到了业界广泛关注。2018版白皮书在2016版的基础上更新了大数据政策，分析了大数据发展的最新趋势和重点领域的应用实践，完善了大数据标准体系，给出了最新的大数据标准化工作建议，大数据标准相关成果的推进在本版本中也有所体现。

1.2 研究目标及意义

本白皮书介绍了国内、国外主要国家在大数据领域的发展战略、发展现状和趋势，描述了大数据的核心产业链以及大数据重点领域的应用实践，力图从应用、技术、产业、标准等角度，勾画出大数据发展的整体轮廓；从数据生存周期的角度提出大数据参考架构；同时抛开其他影响因素，从数据自身的角度提出在不断创新应用与服务模式下的大数据标准体系及大数据标准化路线；最后提出了我国大数据标准化工作建议。

本白皮书立足于大数据产业发展的社会转型历史时期所具有的政策、经济与文化等特点，分析处于初期发展阶段的大数据产业对于经济、社会的作用和影响，旨在与业界分享我们在大数据领域的研究成果和实践经验，呼吁社会各界共同关注大数据的政策研究、技术投入、标准建设与服务应用，共同推动大数据的发展，提升社会整体决策与服务管理能力。

2 相关政策法规

2.1 国外政策

2.1.1 美国大数据政策

2009年5月，美国政府推出Data.gov，这是为了增加政府资料透明度而设立的一系列网站。宣布实施“开放政府计划”(OpenGovernmentInitiative)，这项计划提出利用整体、开放的网络平台，公开政府信息、工作程序和决策过程，以鼓励公众交流和评估，增进政府信息的可及性，强化政府责任，提高政府效率，增进与企业及各级政府间的合作，推动政府管理向开放、协同、合作迈进。联邦政府同时开通了旗舰级项目——“一站式”政府数据下载网站Data.gov，只要不涉及隐私和国家安全的相关数据，均需全部在该网站公开发布。Data.gov的上线意味着美国政府数据仓库的正式建立，标志着美国政府信息进一步公开与透明。

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。具体实现三个目标：(1) 开发能对大量数据进行收集、存储、维护、管理、分析和共享的最先进的核心技术；(2) 利用这些技术加快科学和工程学领域探索发现的步伐，加强国家安全，转变现有的教学方式；(3) 扩大从事大数据技术开发和应用的人员数量。

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》。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。指出“数据驱动型创新”作为崭新命题，所面临的包括新概念、新技术的挑战；并就政府如何支持数据型驱动的创新提出了建议：一是政府应大力培养所需的有技能的劳动力；二是政府要推动数据相关技术的研发。

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NITRD编写的《联邦大数据研究和开发战略计划》，该计划在已有基础上提出美国下一步的大数据七大发展战略，代表大数据研究和开发（R&D）的关键领域。包括在科学，医学和安全的各个方面促进人们的理解；确保国家在研发上的持续领导；提高国家应对社会压力的能力以及通过研究和开发面向国家和世界的环境问题。

2.1.2 欧盟大数据政策

2014年欧盟委员会发布了《数据驱动经济战略》，聚焦深入研究基于大数据价值链的创新机制，提出大力推动“数据价值链战略计划”，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。数据价值链的概念为数据的生命周期，从数据产生、验证以及进一步加工后，以新的创新产品和服务形式出现的利用与再利用。

2015年欧盟大数据价值联盟正式发布了《欧盟大数据价值战略研究和创新议程》（以下简称为《议程》），设定了欧盟国家和区域层面的发展目标，以实现未来欧洲在世界创造大数据价值中的领先地位。《议程》建议建立欧盟大数据契约的合同制公私伙伴（cPPP），以在欧盟2020地平线（Horizon 2020）、各国和地区计划中推行议程，增强泛欧的研究与创新工作，形成清晰的研究、技术发展和投资战略。议程从七个方面指出了在欧盟建立良好的大数据生态系统所需要解决的主要挑战。议程对大数据发展目标的预期影响进行了研究，设定了关键绩效指标，以评估预期影响。

2017年欧盟委员会发布《打造欧洲数字经济》报告，对数据驱动型经济的潜力、面临的障碍、解决方案等进行了分析总结。报告指出，大数据是经济增长、就业和社会进步的重要资源，2015年欧盟数据经济的价值是2720亿欧元，接近于欧盟地区生产总值的1.9%。如果有适当的政策和法律解决方案，数据经济的价值将会在2020年翻一番。

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。英国政府预计大数据将成为英国经济的主要驱动力,而四个大数据研究中心将确保英国在国际竞争中保持较强竞争力。到2017年,大数据分析将为英国创造5.8万个工作岗位,并带来2160亿英镑的经济收入。

2013年10月,由英国商务、创新和技能部牵头编制的《英国数据能力发展战略规划》发布。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。

2017年8月,由英国运输部与英国国家基础设施保护中心(CPNI)共同制定新的网络安全准则《车联网和自动驾驶汽车网络安全准则》出台。该准则隶属于英国政府道路安全与网络安全相关政策的一部分。

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“OpenDataProximaMobile”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。项目内容涉及交通、文化、旅游和环境等领域。

2011年12月,法国政府推出的公开信息线上共享平台 data.gouv.fr,上线当天发布的第一批资源中就包含352000组数据,且网站的数据由每个政府部门的专员统计和收集、持续更新。

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。

2013年4月，经济、财政和工业部宣布将投入1150万欧元用于支持7个未来投资项目，法国政府投资这些项目的目的在于“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”。

2013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》，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，并尽快在网络上实现行政信息全部公开并可被重复使用。

2012年7月，日本推出了《面向2020年的ICT综合战略》，提出“活跃在ICT领域的日本”的目标，重点关注大数据应用。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。

2013年6月，日本公布新战略：“创建最尖端IT国家宣言”。宣言阐述了2013-2020年期间以发展开放公共数据和大数据为核心的日本新IT国家战略，提出要把日本建设成为一个具有“世界最高水准的广泛运用信息产业技术的社会”。

2015年6月，日本政府经内阁会议决定了2014年度版《制造业白皮书》。白皮书中指出，日本制造业在积极发挥IT作用方面落后于欧美，建议转型为利用大数据的“下一代”制造业。

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 印度大数据政策

2012 年，印度批准了国家数据共享和开放政策，目的是在于促进政府拥有的数据和信息得到共享及使用。印度制定了一个一站式政府数据门户网站 data.gov.in，把政府收集的所有非涉密数据集中起来，包括全国的人口、经济和社会信息，截至目前，已包括 107 个部门的 4237 个数据目录，3535 个 API 和 135191 项数据资源。同时，印度政府还拟定一个非共享数据清单，保护国家安全、隐私、机密、商业秘密和知识产权等数据的安全。

2013 年 1 月，印度政府公布新的科技创新政策。新政策既着眼于形成新的创新视角，又提出了到 2020 年跻身全球五大科技强国的目标。新政策强调印度将加强科学、技术与创新之间的协同，使之全方位融入社会经济进程。印度政府还将 2010-2020 年作为“创新十年”，并组建了国家创新委员会，预计在 2017 年，该国研发投入占 GDP 的比例将提高到 2%。

2.1.7 澳大利亚大数据政策

2012 年 10 月，澳大利亚政府发布《澳大利亚公共服务信息与通信技术战略 2012-2015》，强调应增强政府机构的数据分析能力从而促进更好的服务传递和更科学的政策制定，并将制定一份大数据战略确定为战略执行计划之一。截至 2016 年底，其政府数据开放网站 data.gov.au 已包括 275 个组织的 23293 个数据集和 5625 个应用程序

2013 年 8 月，澳大利亚政府信息管理办公室 (AGIMO) 大数据工作组发布了《公共服务大数据战略》，以六条“大数据原则”为指导，旨在推动公共部门利用大数据分析进行服务改革，制定更好的公共政策，保护公民隐私，使澳大利亚在该领域跻身全球领先水平。

2016 年 5 月，澳大利亚信息专员办公室 (OAIC) 发布了《大数据指南和澳大利亚隐私原则》的草案，指南草案概述了关键的隐私要求，并鼓励实施隐私管

理框架，采用这种方法将在设计初始阶段就考虑将‘设计的隐私’嵌入在实体文化，系统和交互中。

2.2 国内政策

2.2.1 国家和行业政策

近年来，我国也相继出台了一系列相关政策推动大数据的技术、产业及其标准化的发展。（党中央、国务院相关政策见表 1）

表 1 国家大数据政策

排序	政策名称	发布日期	发文单位
1	关于运用大数据加强对市场主体服务和监管的若干意见	2015 年 7 月	国务院办公厅
2	关于印发促进大数据发展行动纲要	2015 年 8 月	国务院
3	政务信息系统整合共享实施方案	2017 年 5 月	国务院办公厅

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“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等；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。”

2015 年 8 月，国务院印发《促进大数据发展行动纲要》（国发〔2015〕50 号），系统部署了我国大数据发展工作，并在政策机制部分中着重强调“建立标准规范体系。推进大数据产业标准体系建设，加快建立政府部门、事业单位等公共机构的数据标准和统计标准体系，推进数据采集、政府数据开放、指标口径、分类目录、交换接口、访问接口、数据质量、数据交易、技术产品、安全保密等关键共性标准的制定和实施，加快建立大数据市场交易标准体系；开展标准验证和应用试点示范，建立标准符合性评估体系，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作用；积极参与相关国际标准制定工作。”

2017 年 5 月，国务院办公厅发布《政务信息系统整合共享实施方案》（国办发〔2017〕39 号），根据《国务院关于印发政务信息资源共享管理暂行办法的通

知》(国发〔2016〕51号)、《国务院关于印发“十三五”国家信息化规划的通知》(国发〔2016〕73号)等有关要求制定,明确了加快推进政务信息系统整合共享的“十件大事”。

党的十九大报告中重点提到了互联网、大数据和人工智能在现代化经济体系中的作用:“加快建设制造强国,加快发展先进制造业,推动互联网、大数据、人工智能和实体经济深度融合,在中高端消费、创新引领、绿色低碳、共享经济、现代供应链、人力资本服务等领域培育新增长点、形成新动能”。

围绕国家政策,我国各部委和相关行业也出台了一系列政策来促进推动大数据在各领域中的应用与方面相关发展。(相关政策见表2)

表2 部分行业领域大数据政策

排序	政策名称	发布日期	发文单位
1	《关于组织实施促进大数据发展重大工程的通知》	2016年1月7日	发改委
2	《生态环境大数据建设总体方案》	2016年3月7日	环保部
3	《关于印发促进国土资源大数据应用发展实施意见》	2016年7月4日	国土资源部
4	《关于加快中国林业大数据发展的指导意见》	2016年7月13日	国家林业局
5	《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8月25日	交通运输部
6	《农业农村大数据试点方案》	2016年10月14日	农业部
7	大数据产业发展规划(2016—2020年)	2017年1月17日	工信部
8	中国大数据发展报告(2017)	2017年2月26日	国家信息中心
9	《关于推进水利大数据发展的指导意见》	2017年5月2日	水利部
10	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7月25日	国家自然科学基金委员会
11	智慧城市时空大数据与云平台建设技术大纲(2017版)	2017年9月6日	国家测绘地理信息局办公室
12	关于深入开展“大数据+网上督察”工作的意见	2017年9月8日	公安部

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)。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)》部署了7项重点任务,明确了8大重点工程,制定了5个方面保障措施,全面部署“十三五”时期大数据产业发展工作,为“十

三五”时期我国大数据产业崛起，实现从数据大国向数据强国转变指明了方向。

2.2.2 国家大数据综合试验区

国家大数据综合试验区的设立，旨在贯彻落实国务院《促进大数据发展行动纲要》，为大数据制度创新、公共数据开放共享、大数据创新应用、大数据产业聚集、大数据要素流通、数据中心整合利用、大数据国际交流合作等方面开展试验探索，推动我国大数据创新发展。

国家选择具有一定条件的地区开展试点工作，一方面可以以建设国家大数据综合试验区为抓手，探索大数据与传统产业、区域经济的融合发展，促进数据要素与其他生产要素的整合利用，提高产业组织效率，加速形成高质量、多层次的供给体系，重塑产业链供应链价值链，实现资源优化配置，全面释放数据红利，推动供给侧结构性改革。另一方面，可以把发展大数据的风险和试错成本控制在一定区域内，平稳有序的推进大数据发展进程。

综合试验区建设将发挥“三个作用”：一是示范带头作用；二是统筹布局作用；三是先行先试作用。在试验区内，开展面向应用的数据交易市场试点，鼓励产业链上下游间进行数据交换，探索数据资源的定价机制，规范数据资源交易行为，建立大数据投融资体系，激活数据资源潜在价值，促进形成新业态。

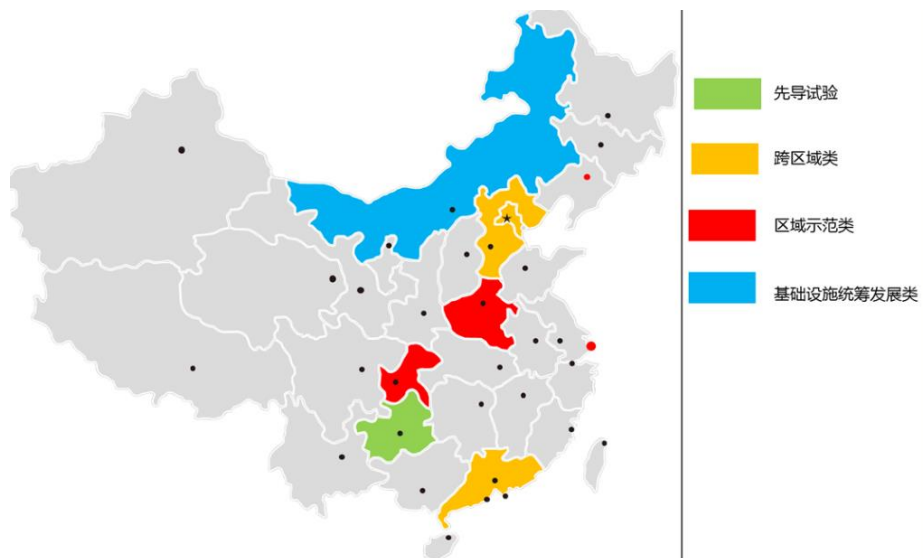


图 1 试验区分布图

目前我国共设有 8 个国家大数据综合实验区，其中先导试验型综合实验区 1

个，跨区域类综合试验区 2 个，区域示范类综合试验区 4 个，大数据基础设施统筹发展类综合试验区 1 个。

2.2.2.1 先导试验型综试区

国务院于 2015 年 8 月 31 日印发的《促进大数据发展行动纲要》(国发〔2015〕50 号)中，明确提出了“开展区域试点，推进贵州等大数据综合试验区建设”，贵州成为其中唯一明确提到的省份。

贵州国家大数据综合试验区：积极开展大数据综合性、示范性、引领性发展的先行先试，开展了一系列先行探索，积累了先试经验，围绕数据资源管理与共享开放、数据中心整合、数据资源应用、数据要素流通、大数据产业集聚、大数据国际合作、大数据制度创新等七大主要任务开展系统性试验，打破数据资源壁垒，通过不断总结可借鉴、可复制、可推广的实践经验，最终形成试验区的辐射带动和示范引领效应。

2.2.2.2 跨区域类综试区

跨区域类综合试验区定位是，围绕落实国家区域发展战略，更加注重数据要素流通，以数据流引领技术流、物质流、资金流、人才流，支撑跨区域公共服务、社会治理、和产业转移，促进区域一体化发展。目前我国已有的跨区域类综合试验区包括：

京津冀国家大数据综合试验区：2016 年 10 月获批，将充分发挥京津冀在大数据基础设施建设、数据开放共享、产业集聚发展等方面的示范带动作用，打破数据资源壁垒，发掘数据资源价值，在数据开放、数据交易、行业应用等方面开展创新探索，将其打造为“一心一地两区”的区域协同发展的典范。

珠三角国家大数据综合试验区：2016 年 10 月获批，将珠江三角洲地区打造成全国大数据综合应用引领区、大数据创业创新生态区、大数据产业发展集聚区，抢占数据产业发展高地，建成具有国际竞争力的国家大数据综合试验区，形成“一区两核三带”功能布局。

2.2.2.3 区域示范类综试区

区域示范类综合试验区定位是，积极引领东部、中部、西部、东北等“四大板块”发展，更加注重数据资源统筹，加强大数据产业集聚，引领区域发展，发挥辐射带动作用，促进区域协同发展，实现经济提质增效。目前我国已建设的区域示范类综合试验区包括：

上海国家大数据综合试验区：2016年10月获批，将围绕自贸区建设和科创中心建设两个战略，在四个方面推动大数据发展，包括推动公共治理大数据的应用、推动大数据的科技创新和基础性治理的工作和研究、推动大数据与公共服务基层社会治理相结合、在大数据方面进一步加强与长三角地区和长江经济带城市的合作。

河南国家大数据综合试验区：2016年10月获批，以深化大数据应用为主线，重点在管理机制创新、数据汇聚共享、重点领域应用、产业集聚发展等四个方面进行试点，提升大数据在促进转型发展中的引领支撑作用，形成一套适应大数据创新发展的管理机制和发展模式，基本建成“两区两基地”为支撑的综合试验区。

重庆国家大数据综合试验区：2016年10月获批，定位为引领东部、中部、西部、东北等“四大板块”发展，注重数据资源统筹，加强大数据产业集聚，发挥辐射带动作用，促进区域协同发展。

沈阳国家大数据综合试验区：2016年10月获批，以工业大数据应用引领两化深度融合，推动大数据在产品全生命周期、产业链全流程各环节的应用，促进传统产业转型升级，形成“一体两翼”的发展格局。

2.2.2.4 基础设施统筹发展类综试区

基础设施统筹发展类综合试验区定位是，在充分发挥区域能源、气候、地质等条件基础上，加大资源整合力度，强化绿色集约发展，加强与东、中部产业、人才、应用优势地区合作，实现跨越发展。

内蒙古国家大数据综合试验区：2016年10月获批，加大资源整合力度，强化绿色集约发展，向国内外提供数据存储服务，发挥数据中心的辐射作用，争取大数据农牧业、大数据政务、大数据精准扶贫等九大工程的顺利实施完成，力争

建成“中国北方大数据中心、丝绸之路数据港、数据政府先行区、产额融合发展引导区、世界级大数据产业基地”。

2.2.3 地方政策

在《促进大数据发展行动纲要》等国家政策的引领下，各地政府也高度重视大数据发展，多个省市出台专门的大数据相关政策文件（见表 3），部分地方专门设置了大数据管理机构或部门（见表 4）。

表 3 部分省市出台的大数据产业发展政策文件

编号	地区	政策标题	发布日期
1	成都	《成都市大数据产业发展规划(2017—2025年)》	2017年10月10日
2	贵州	《贵州省发展农业大数据助推脱贫攻坚三年行动方案(2017—2019年)》	2017年9月8日
3	成都	《成都市促进大数据产业发展专项政策》	2017年9月6日
4	辽宁	辽宁省政务信息系统整合共享实施方案	2017年8月16日
5	贵州	贵州省政府办公厅下发关于促进和规范健康医疗大数据应用发展的实施意见	2017年7月18日
6	江西	《江西省大数据发展行动计划》	2017年7月5日
7	内蒙古	内蒙古自治区人民政府办公厅关于印发《内蒙古自治区“十三五”科技创新规划》的通知	2017年7月4日
8	内蒙古	内蒙古自治区人民政府办公厅关于印发2017年自治区大数据发展工作要点的通知	2017年6月29日
9	广东	云浮市《促进云计算大数据产业发展优惠办法(试行)》	2017年6月26日
10	贵州	贵阳市人民政府办公厅关于印发关于支持区块链发展和应用的若干政策措施(试行)的通知	2017年6月7日
11	河南	河南省推进国家大数据综合试验区建设实施方案和若干意见	2017年5月25日
12	贵州	大数据+产业深度融合2017年行动计划	2017年5月24日
13	青岛	《关于促进大数据发展的实施意见》	2017年5月23日
14	河南	河南省云计算和大数据“十三五”发展规划	2017年5月4日
15	广东	《东莞市大数据发展规划(2016-2020年)》	2017年4月18日
16	东莞	《东莞市大数据发展规划(2016-2020年)》	2017年4月18日
17	贵州	贵阳市政府数据共享开放条例	2017年4月12日
18	河南	河南省人民政府关于印发河南省推进国家大数据综合试验区建设实施方案的通知	2017年4月8日
19	广东	广东省人民政府办公厅关于印发珠江三角洲国家大数据综合试验区建设实施方案的通知	2017年4月6日
20	内蒙古	内蒙古自治区人民政府关于加快推进“互联网+政务服务”工作的实施意见	2017年3月28日

编号	地区	政策标题	发布日期
21	广东	清远市大数据发展“十三五”规划	2017年3月6日
22	厦门	《厦门市促进大数据发展工作实施方案》	2017年3月2日
23	河北	关于加快发展“大智移云”的指导意见	2017年3月2日
24	河南	河南省推动云计算和大数据发展加快培育新业态新模式行动指南（2017—2020年）	2017年3月1日
25	湖北	《湖北省云计算大数据发展“十三五”规划》	2017年2月21日
26	贵州	贵阳市人民政府办公厅印发贵阳市大数据标准建设实施方案	2017年2月9日
27	广东	广东省人民政府办公厅关于促进和规范健康医疗大数据应用发展的实施意见	2017年2月8日
28	湖北	《湖北省人民政府办公厅关于促进和规范健康医疗大数据应用发展的实施意见》	2017年2月7日
29	合肥	《合肥市大数据发展行动纲要（2016—2020）》	2017年2月7日
30	安徽	《安徽省“十三五”软件和大数据产业发展规划》	2017年1月17日
31	广州	《关于促进大数据发展的实施意见》	2017年1月11日
32	广东	广州市人民政府办公厅关于促进大数据发展的实施意见	2017年1月7日
33	上海	市经济信息化委关于印发《上海市关于促进云计算创新发展培育信息产业新业态的实施意见》的通知	2017年1月4日
34	苏州	《苏州市大数据产业发展规划》	2016年12月27日
35	广西	《促进大数据发展的行动方案》	2016年11月11日
36	山东	《关于促进大数据发展的意见》	2016年10月27日
37	深圳	《深圳市促进大数据发展行动计划（2016—2018年）》	2016年10月25日
38	宁波	《关于推进大数据发展的实施意见》	2016年9月26日
39	湖北	《湖北省人民政府关于加快推进楚天云建设的意见》	2016年9月22日
40	上海	《上海市大数据发展实施意见》	2016年9月15日
41	江苏	《江苏省大数据发展行动计划》	2016年8月19日
42	北京	《北京市大数据和云计算发展行动计划（2016—2020年）》	2016年8月18日
43	淮南	《淮南市大数据产业发展三年行动计划（2016—2018年）》	2016年8月16日
44	哈尔滨	《哈尔滨市促进大数据发展若干政策（试行）》	2016年8月15日
45	重庆	《关于进一步促进大数据发展的实施意见（2016—2018年）》	2016年6月8日
46	陕西	《陕西省大数据与云计算产业发展五年行动计划》	2016年5月20日
47	广东	《广东省促进大数据发展行动计划（2016—2020年）》	2016年4月22日
48	浙江	《浙江省促进大数据发展实施计划》	2016年2月18日
49	沈阳	《沈阳市促进大数据发展三年行动计划（2016—2018年）》	2016年2月4日
50	武汉	《武汉市人民政府关于加快大数据推广应用促进大数	2015年4月15日

编号	地区	政策标题	发布日期
		据产业发展的意见》	
51	武汉	《武汉市大数据产业发展行动计划(2014—2018年)》	2014年7月18日
52	重庆	《重庆市大数据产业发展规划》	2013年6月28日

各地方的大数据产业发展政策的制定出台呈密集态势，对大数据产业的经济与社会意义进行了充分说明，对促进产业发展提出了具体举措，基于地方产业基础与经济特点进行了高适性匹配。具有认知深刻、创新灵活、匹配度高、管理到位、强调实效的特点。

表 4 部分地区的大数据管理机构

省份	机构	隶属机构
贵州	贵州省大数据局	贵州省政府
	贵阳市大数据发展管理委员会	贵阳市政府
	贵阳高新区大数据发展办公室	贵阳高新区管委会
广东	广东省大数据管理局	广东省经信委
	广州市大数据管理局	广州市工信委
辽宁	沈阳市大数据管理局	沈阳市经信委
四川	成都市大数据和电子政务管理办公室	成都市政府办公厅
甘肃	兰州市大数据社会服务管理局	兰州市政府
	兰州新区大数据管理局筹备办公室	兰州新区党工委、管委会
浙江	浙江省数据管理中心	浙江省政府
	杭州市数据资源局	杭州市政府
陕西	陕西省政务数据服务局	陕西省政府
	咸阳市大数据管理局	咸阳市政府
宁夏	银川市大数据管理服务局	银川市政府
湖北	黄石市大数据管理局	黄石市经信委
云南	昆明大数据管理局	昆明市工信委
	保山市大数据管理局	保山市工信委

在国家治理、经济发展等诸多领域，大数据都在发挥着至关重要的作用，地方大数据管理机构的成立有利于统筹产业规划，是行政体制上的一次灵活创新。

2.3 安全隐私法律法规

大数据的良性发展自然离不开大数据产业政策的大力支持，而大数据与公民个人数据关系异常密切，无论是数据收集的来源、数据使用和处理的范围和范围，都与个人数据保护等相关法律制度直接相关。因此，涉及公民个人数据保护的安全隐私政策，对于大数据产业的发展也起着至关重要的作用，具有举足轻重的价

值。

2.3.1 国内外立法现状

从针对大数据环境下个人数据保护的法律制度来看，目前欧盟模式和美国模式是全球最有影响的两种模式。

欧盟一直是数据保护领域的立法先驱，从启动时间到法律文件数量、领先概念和自我更新，欧盟都为其他司法辖区的数据保护立法工作提供了蓝本和榜样，影响不仅局限于其各成员国，还扩展到了其他国家和地区如日本、韩国和我国香港地区等。欧盟模式是统一的立法模式，通过综合立法确定个人数据保护的各项基本原则，并设立专门的机构来监督法律实施。

早在 1981 年，欧盟理事会就通过了《有关个人信息自动化处理保护公约》；1995 年欧盟通过了《关于个人数据处理保护与自由流动指令》（1995/46/EC），很快就成为世界各国个人信息隐私保护，以及数据保护领域法律文件和国际协议制定中的范例，后被 2016 年 5 月通过的《一般数据保护条例》（“GDPR”）替代；2002 年通过了《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），并于 2017 年 1 月 10 日进行了最新的修订。针对性的隐私保护指令——ePD 指令与一般性的综合数据指令——GDPR 共同构成了欧盟数据保护法律框架的两大支柱，为欧盟公民的个人数据权利和隐私权保护提供坚实的保障基础，赋予数据主体包括访问权、纠错权、被遗忘权、限制处理权、反对权、拒绝权和自决权等权利，对数据控制者和处理者构建了相应的义务体系，并通过相关的监督机构设置、域外效力条款和高昂的罚则充分保障了数据保护法律制度的实施，具有极强的法律震慑力和适用性。

美国则是行业自律模式的倡导者，成文立法散见于联邦、各州的各行业规定之中，辅之以行业内部的行为规则、规范、标准和行业协会的监督，充分保证个人数据自由流动的基础上保护个人数据，实现行业内个人数据保护自律和行业利益保护的平衡。美国早期数据保护方面的立法主要是 1974 年通过的《隐私法》和 1986 年的《储存信息保护法》，《公平信用报告法》中也有信贷和消费者信用行业的特殊规定。2015 年 10 月，美国通过了《网络安全信息共享法》，明确规

定了个人隐私、自由等私权利的保护。

此外，针对金融、医疗、电信、教育、娱乐、消费者保护和儿童隐私保护等高危行业，美国立法也遵循“公平信息实践法则”，采取“告知与同意”框架，按照行业领域进行细分。而针对大数据安全方面的复杂性，2015年美国国家标准与技术研究院（NIST）大数据工作组下属安全与隐私小组针对大数据安全与隐私发布了第一版框架性草案，从安全与隐私的维度对大数据的几个关键特征——多样性、规模性、真实性、高速性、有效性进行了阐述，提供了大数据领域安全与隐私保护的参照性蓝本。

相对于国外的安全隐私政策立法进程而言，我国对于个人数据保护的立法起步较晚，目前还没有专门的《个人信息保护法》。

表 5 国内关于个人信息保护相关的主要规定

年份	法律法规	主要内容
2017 年	2017 年 5 月 8 日最高人民法院、最高人民检察院发布了《关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》（以下简称“《两高解释》”）	进一步明确侵犯公民个人信息罪行的适用条件
2017 年	2017 年 3 月 15 日全国人民代表大会第五次会议通过《中华人民共和国民法总则》（以下简称“《民法总则》”）	其中第一百一十一条规定，自然人的个人信息受法律保护。任何组织和个人需要获取他人个人信息的，应当依法取得并确保信息安全，不得非法收集、使用、加工、传输他人个人信息，不得非法买卖、提供或者公开他人个人信息。
2016 年	2016 年 11 月 7 日全国人大常委会颁布《中华人民共和国网络安全法》（以下简称“《网络安全法》”）（自 2017 年 6 月 1 日起施行）	首次从立法层面定义“个人信息”，对“个人信息”进行了不完全列举
2015 年	2015 年 8 月 29 日全国人大常委会颁布《中华人民共和国刑法修正案（九）》	将“违反规定，向他人出售或者提供公民个人信息”的行为定性为犯罪行为
2013 年	2013 年 7 月 16 日工业和信息化部发布《电信和互联网用户个人信息保护规定》	对电信业务经营者、互联网信息服务提供者收集和使用个人信息作出规定
2013 年	2013 年 1 月 21 日中华人民共和国国务院发布《征信业管理条例》	对征信业务相关的个人信息的收集、使用、存储、加工作出规定
2012 年	2012 年 12 月 28 日全国人民代表大会常务委员会发布《关于加强网络信息保护的決定》	将“能够识别公民个人身份和涉及公民个人隐私的电子信息”纳入保护范围

2007 年	2007 年 6 月 22 日公安部、国家保密局、国家密码管理局和国务院信息化工作办公室联合颁布《信息安全等级保护管理办法》	根据信息系统的重要程度和破坏后果实施等级保护制度
--------	--	--------------------------

除了以上已正式发布的相关规定，2017 年 5 月 2 日，国家互联网信息办公室还正式发布了《网络产品和服务安全审查办法（试行）》。该办法明确了在审查网络产品和服务的安全性和可控性时，应当充分考虑“产品和服务提供者利用提供产品和服务的便利条件非法收集、存储、处理、利用用户相关信息的风险”这一重要因素。2017 年 4 月 11 日，国家互联网信息办公室发布关于《个人信息和重要数据出境安全评估办法（征求意见稿）》，明确了网络运营者因业务需要，在向境外提供在中国境内运营中收集和产生的个人信息和重要数据前，需进行安全评估的详细内容和流程，为我国跨境数据保护制度打下了基础。

此外，国家质量监督检验检疫总局、国家标准化管理委员会也已发布或正在制定一系列与个人信息保护相关的国家推荐性标准，包括但不限于已发布的《信息安全技术 个人信息安全规范》和尚未正式发布的《信息安全技术 数据出境安全评估指南》。尽管国家推荐性标准不具有强制性效力，但在《网络安全法》配套措施尚未完善的情况下，其在一定程度上反映监管态度，为个人信息保护提供了具有操作性的指引。

具体而言，2018 年 1 月 2 日正式发布的国家推荐性标准《信息安全技术 个人信息安全规范》包含了个人信息及其相关术语基本定义，个人信息安全基本原则，个人信息收集、保存、使用以及处理等流转环节以及个人信息安全事件处置和组织管理要求等。2017 年 8 月 25 日发布的《信息安全技术 数据出境安全评估指南（征求意见稿）》对数据出境安全评估流程、评估要点、评估方法、重要数据识别指南等内容进行了具体规定。

在法律责任方面，侵犯个人信息可能需要承担民事、行政甚至刑事责任，从而为个人信息的保护提供了有力的法律制度保障。

1) 民事责任:

《网安法》第七十四条第一款规定，“违反本法规定，给他人造成损害的，依法承担民事责任。”因此，未经同意收集、使用他人个人信息，数据主体可依据《民法总则》、《侵权责任法》和《消费者权益保护法》等相关法律规定主张赔

偿，相关侵权者将承担民事侵权责任。

2) 行政责任:

未经同意收集、使用他人个人信息，尚不足《两高解释》“情节严重”要求时，根据《网安法》第六十四条的规定，面临的处罚将包括警告、没收违法所得、罚款、暂停相关业务、停业整顿、关闭网站、吊销相关业务许可证和吊销营业执照等；对直接责任人员的罚款数额可达十万元，对网络运营者的罚款数额更可高达一百万元。

3) 刑事责任:

未经同意收集、使用他人个人信息，达到《两高解释》“情节严重”要求时，则构成刑法“侵犯公民个人信息罪”，应承担刑事责任，包括对直接负责的主管人员和其他直接责任人员定罪处罚，并对单位判处罚金。

2.3.2 大数据立法趋势和对产业的影响

我国大数据立法虽然起步较晚，但随着《民法总则》、《网络安全法》的实施以及后续的配套法律法规建设完善，我国个人信息保护相关制度越发健全，取得了卓著成果，有效地改善了过去“守法成本高、违法成本低”的怪象。万众期待下出台的《网络安全法》不仅赋予了数据主体收集同意权、被遗忘权、纠错权等一系列权利，还首次构筑了网络运营者在收集、使用个人信息数据过程中相应的保障义务；在罚则方面侵犯个人信息更是有可能触发民事、行政甚至刑事责任的风险，从而为个人信息的保护提供了有力的法律制度执行保障。

不过，在制度越发健全和完善的同时，必须指出的是，我国个人信息保护的立法仍然寄居于网络安全和国家安全保护立法的大框架之下，并没有针对公民个人和隐私进行专门的立法保护，相较于欧盟的 GDPR、ePD 和美国的分行业自律来看，在数据保护领域我国立法仍有相当的可完善空间。

同时，值得一提的是，大数据产业的发展和个人信息及隐私保护存在某种程度上的分歧。不加限制的催熟大数据产业，可能导致企业在利益吸引下铤而走险，选择以侵害公民隐私等数据权益的方式或手段发展大数据业务。反过来说，一味强调公民的个人信息保护，则可能从数据源头、收集和利用等环节对企业开展大

数据业务提出了过高的挑战，阻碍大数据产业的发展成熟。大数据产业的兴起和发展证明了大数据前所未有的巨大社会、经济价值，在未来的立法进程中，如何平衡大数据产业的发展和个人信息的合理保护将会成为立法宗旨上的一个重要考量角度。

此外，大数据时代的一个重要基石——数据权属尚未确定时，无论是从合理立法角度考虑还是从产业长期发展角度出发，大数据方面的安全隐私立法进程都应当谨慎稳健。唯有逐渐在目前的数据安全隐私保护框架上添砖加瓦，稳步推进，才能以清晰、可预测的法律制度指引企业行为，以明确、统一的处罚措施治理违法违规行为，引导企业自觉遵守相关的法律制度；进而以良好的法律制度促进产业的良性发展，增强守法合规企业的竞争力，创建“良币战胜劣币”的有序竞争环境。

总而言之，大数据系统安全体系的建立是一个系统性的工程，需要国家从法律法规、行业规范以及技术手段等多方面对个人隐私数据进行监管和保护，也需要社会公众、企业等各方主体切实遵守相关规定，承担相应责任，在自身权利义务范围内进行行为决策，方能使大数据行业安全、合法地行走阳光下。

3 大数据发展现状和趋势分析

大数据对经济社会发展的推动作用已经成为社会各界的共识，大数据促进用户对网络基础设施建设的创新和信息产品升级换代的需求，具有外溢效应；大数据连接生产者和消费者，帮助生产者拉动新的市场需求，具有市场效应；大数据通过连接不同领域的信息，增强了信息的“活力”，可以创造出新的社会经济价值，具有连接效应；大数据有效利用了社会资源，打破了行业垄断，使得消费者可以用同样的价钱获得更多的社会服务，具有福利效应。本章对大数据在我国产业和应用领域发展现状和未来趋势进行分析。

3.1 大数据核心产业链

在社会认知、政策环境、市场规模和产业支撑能力等多方面，我国的大数据产业已经具备了一定的基础，并取得了积极的进展，在大数据资源建设、大数据技术、大数据应用领域涌现出一批新型企业。在龙头企业引领下，上下游企业互动、核心产业融合发展的产业格局初布，与日趋成熟的产业生态相对应的商业模式也日渐明晰。

3.1.1 大数据产业生态商业角色构成

3.1.1.1 大数据提供者

拥有数据的公司、个人、社会团体以及政府机构等，此类角色属于大数据产业链上的基础环节，包括数据源提供者、数据流通平台提供者和数据 API 提供者。目前我国大数据提供者包括政府管理部门、企业数据源提供商、互联网数据源提供商、物联网数据源提供商、移动通讯数据源提供商、提供数据流通平台服务和数据 API 服务的第三方数据服务企业、社会团体或者个人等。

3.1.1.2 大数据产品提供者

提供直接应用于大数据产品的企业，包括提供大数据应用软件、大数据基础软件、大数据相关硬件产品的企业。

大数据应用软件产品提供者，包括提供整体解决方案的综合技术服务商，也包括大数据计算基础设施上（与云结合），从简单的文件存储的空间租售模式，逐步扩展到提供数据聚合平台，进而扩展到为客户提供分析业务的服务上。

大数据基础软件提供商，此类企业搭建大数据平台、提供相关大数据技术支持、云存储、数据安全等，此类公司在某些垂直行业或者区域掌握大数据入口与出口，并能对一些数据进行采集、整合和汇集。这样的企业包括传统的 IT 企业、设备商以及新兴的云服务相关企业。

大数据相关硬件产品提供商，此类企业提供大数据采集、接入、存储、传输、安全等硬件产品和设备。

3.1.1.3 大数据服务提供者

以大数据为核心资源，以大数据应用为主业开展商业经营的企业。包括大数据应用服务提供者、大数据分析服务提供者、大数据基础设施服务提供者。这类企业挖掘数据价值，处于大数据产业链的下游，它们通过发掘隐藏在大数据中的价值，不断推动大数据产业链中各个环节的发展和成熟。从某种角度上说正是此类公司创造了大数据的真正价值，具体包括：

应用服务提供者，基于大数据技术，对外提供大数据服务。

分析服务提供者，提供技术服务支持、技术（方法、商业等）咨询，或者为企业提供类似数据科学家的咨询服务。

大数据基础设施服务提供者，提供面向大数据技术和服务提供者的培训、咨询、推广等的基础类通用类的服务提供者。

从商业角度看，大数据产业链可以用如下图示意。

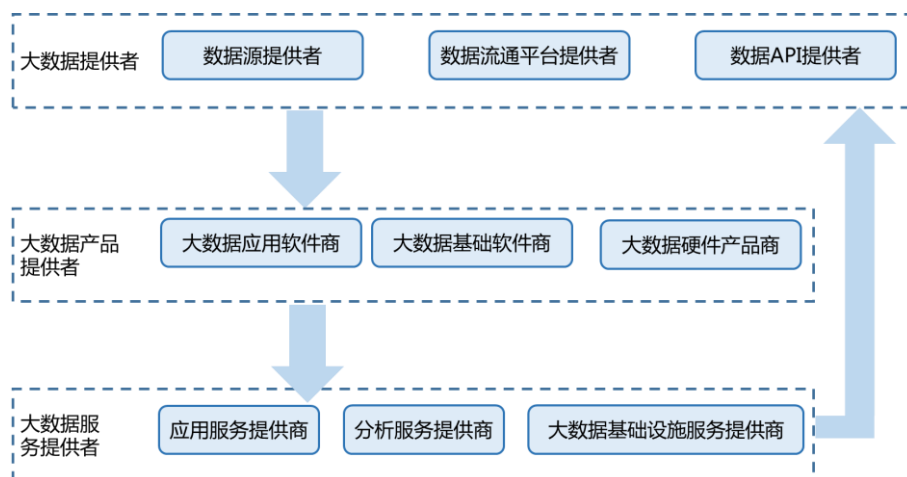


图 2 大数据生态商业角色图

3.1.2 大数据产业生态商业模式分析

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。

3.1.2.1 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者，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。

3.1.2.2 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。

3.1.2.3 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。

3.1.2.4 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。

3.1.2.5 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，对于大数据硬件提供方和大数据基础设施服务提供方来说，软硬件销售模式是他们主要的盈利方式。

3.2 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，大数据的驱动力和引领作用正在给国民经济的各个领域带来革命性变化，在国家政策引导和支持下，各级政府、企事业单位、民间组织与个人顺应互联网、大数据、人工智能和实体经济深度融合发展的技术和应用趋势，积极探索各领域大数据应用实践。在推进标准先行的基础上，大数据标准化活动在各领域发展中应用范围逐步扩大，应用程度逐渐加深，应用效果不断显现。

3.2.1 社会管理与公共服务

大数据应用能够揭示传统技术方式难以展现的关联关系，近年来“用数据说话、用数据决策、用数据管理、用数据创新”的共识逐步达成。在推进“互联网

+政务服务”工作过程中，各级政府正加快推进政务大数据的应用，提高社会治理的精准性和有效性，提升政府宏观调控、风险防范能力、政府数据资产的管理能力和城市科学化管理水平，在建设与社会主义市场经济体制和中国特色社会主义事业发展相适应的法治政府、创新政府、廉洁政府和服务型政府的过程中发挥重要作用。

3.2.1.1 政府数据资产治理能力的提升

3.2.1.1.1 促进政府数据资产的全面管理

随着信息技术的高速发展，逐步实现数据与产业、经济、社会、文化的深度融合，大数据已经作为一种新兴的战略性资源登上了历史舞台。一个城市拥有数据的规模、开发运用能力将成为城市综合竞争力的重要组成部分。因此，对于数据资产的全面掌握与监管成为了一项非常重要的工程。

要实现数据资产的全面掌握与监管就必须先对辖区的信息资源进行梳理，“摸清家底”，做到“心中有数”。信息资源用元数据描述其主要特征，多以信息资源目录的方式呈现，通过信息资源目录的编辑，让数据拥有者直观清晰地掌握所拥有的信息资源；数据使用者也可以通过目录发现自己所需要的数据现状，并发出需求申请。大数据建设，信息资源目录是资产管理的关键数据，是管理核心，是数据使用者和拥有者之间沟通的主要桥梁。

信息资源梳理是对信息资源元数据的采集过程，通过信息资源分类信息、信息资源名称、信息资源提供方、信息资源提供方代码、信息资源摘要、信息资源格式、信息项信息、共享属性、开发属性、更新周期、发布日期、关联资源等基础特征进行采集，构建元数据库。再利用工具对元数据库进行分类管理，依据国家标准形成信息资源目录。

通过信息资源的梳理与编目便可以全面的掌握数据资产现状，为数据的挖掘分析和开发运用提供了准确、全面的数据支撑。数据拥有者通过对目录及数据访问权限的设定，便可以很好地管控数据资源。再结合大数据采集、处理、存储、传输、分析挖掘、安全等各领域的技术实现，便可非常有效地推动数据的共享与开放，实施科学、精准的数据监管。

3.2.1.1.2 推进政务数据共享开放体系的建设

政府数据资产的全面管理可以有效推进政务数据共享开放体系的建设速度。国家发展改革委印发的《“十三五”国家政务信息化工程建设规划》(以下简称《规划》)中明确提出了构建形成大平台共享、大数据慧治、大系统共治的顶层设计,建成全国一体化的国家大数据中心,有利促进网络强国建设的一体化设想,这是我国政务信息化发展的必然要求。《规划》突出了基于政务信息资源目录、共享开放和服务平台,有效推进政务数据共享开放和服务的客观需求。

按照“谁主管,谁提供,谁负责”、“谁经手,谁使用,谁管理,谁负责”的原则,明确政务数据共享开放和服务的主管机制、提供机制、使用机制和管理机制等;按照制定政务数据共享开放和服务的目录清单、推动政务数据共享开放和开展政务数据大普查的要求,明确政务数据共享开放和服务的牵头部门、提供部门和使用部门的职责,形成国家统筹、部际协调、部门统一的工作局面;按照政务数据共享开放的财政、审计、纪检监察等部门的考核机制,对于政务数据共享开放和服务好的部门予以奖励,对于工作拖沓、作风疲软甚至拒不执行共享开放和服务的部门予以问责。

按照政务数据“共享是原则、不共享是例外”、“开放是常态、不开放是例外”的原则,积极组织建设政务数据的共享开放和服务目录;按照建立疑义、错误信息快速校核机制,明确使用部门和社会公众对获取的共享开放服务信息有疑义或发现有明显错误的及时反馈、提供部门的数据校核和数据提供;按照建立跨部门、跨地区协同治理大系统的要求,明确各系统、各项目的建设必须满足数据共享、开放和服务的需要,实现政务数据共享开放与社会大数据融合应用的需要。

按照“覆盖全国、统筹利用、统一接入”的要求,明确政务国家数据共享交换工程、国家公共数据开放网站和服务平台主管部门的工作职责,落实政务数据共享、开放和服务的部门职责;按照数据存储、数据共享、数据开放和数据服务的要求,明确各平台主管部门与信息提供部门的数据提供程度,保障以数据为支撑的治理能力,提升宏观调控、市场监管、社会管理和公共服务的精准性和有效性。

按照“坚守底线、确保安全”的要求,明确政务数据共享开放的信息安全等

级保护；按照守住政务信息化工程全过程信息安全和监督机制的要求，明确安全责任边界，落实网络安全工作责任，形成跨部门、跨地区的条块融合的安全保障工作联动机制；按照加强关键软硬件产品自主研制与安全审查的要求，明确云计算服务、数据中心托管、大数据分析存储等方面的企业准入等自主保障能力。

3.2.1.2 政府社会管理能级的提高

3.2.1.2.1 强化社会公共安全保障

作为社会公共安全保障的核心职能单位，依托公安部门“金盾工程”的建设成果，形成了面向部门、警种的各种类型业务系统，以及面向多部门协作和底层信息支撑的综合业务系统，建成了可以全国范围内共享的八大业务信息资源库。进而在公安业务系统内部积累了庞大的数据资源。在大数据时代，走科技强警、信息化强警的大数据之路是解决警力不足等实际问题的重要途径。

为积极应对大数据、云计算时代对下一代公安信息化建设的挑战，公安部多措并举，积极推进公安行业大数据应用实践，并在建立了基于大数据架构的公安信息化应用公安部重点实验室和大数据警务云计算中心建设工程，推动了智能化全时空大数据预警、新一代超级智能化搜索引擎等大数据应用的建设，促进了大数据和云计算技术融合的新一代公安行业信息化平台建设。伴随着公安领域的大数据应用实践在提高反恐能力、预测犯罪趋势、推进案件侦破、破解交通难题等方面取得了比较丰硕的成果，各地积极探索践行大数据警务战略，自主研发了各类警情热点分布图等辅助指挥技术；积极探索利用大数据系统破解大城市的交通难题，确保重要会议或者活动的交通保障任务；各级公安部门也应抓住国家在全国设立国家级大数据综合试验区的契机，通过深入挖掘数据价值，加快警务机制内生性变革，积极推动“大数据+打击、防范、管理、服务、监督”五大主题应用，形成“大数据+”打防管控综合应用模式。总体来看，各级公安机关的大数据应用实践有效地提升了各警种的实战能力，大数据技术正在成为驱动和引领警务改革的关键要素。

3.2.1.2.2 创新市场监管能力

2017年7月，国务院办公厅印发《关于运用大数据加强对市场主体服务和监管的若干意见》。《意见》要求，以社会信用体系建设和政府信息公开、数据开放为抓手，充分运用大数据、云计算等现代信息技术，提高政府服务水平，加强事中事后监管，维护市场正常秩序，促进市场公平竞争，释放市场主体活力，进一步优化发展环境。

市场监管领域牵涉部门众多、系统异构、数据繁多，通过归集整合工商、质监、税务、法院、人民银行、水电气等多个部门和公共企事业单位掌握的身份、业绩、提示、警示和企业年报、自主申报信用记录，形成市场主体（企业、个体工商户、中介机构和重点人群）大数据。在信用监管领域，利用市场监管主体登记信息、资质信息、信用行为信息等数据开展信用主体动态监控，并探索在政府采购、招标投标、日常监管等工作中使用信用相关信息，提升市场监管工作，实现城市精准管理。在联合惩戒方面，利用监管主体登记信息、司法信息、行政执法信息等数据开展跨部门协作互动，实现对失信市场主体的联合惩戒的精准互动。在社会共治方面，利用监管主体信息、资质信息、信用行为信息等数据，依托“部门+社区”、“部门+社会机构”等服务模式，引入社会力量参与市场监管，实现社会参与多元共治。

以市场监管数据资源为核心，利用“标准化+大数据技术”对提高市场主体的监管效率、规范市场秩序，构建新型监管模型和系统性风险防控机制具有巨大影响力。

3.2.1.2.3 提升城市智能运营管理能力

越来越多的城市面临着环境污染、交通拥堵、房价虚高、管理粗放、应急迟缓等问题，这些“城市病”给市民工作和生活带来了许多不便，降低了人们的幸福感。但随着物联网、大数据和云计算技术的发展和成熟，为治疗“城市病”带来了机遇，从社交媒体到交通流量，从产业大数据到工业经济运行质态监测与分析，从人的移动性到外来人口管理，从工业排放到环境治理等，依托大数据分析，

实现城市的人、地、产业、资源等的优化配置。科学开展城市规划、建设、运行、管理、治理、管控和服务等已经成为了构建城市智能运营管理体系的一种必然趋势。

大数据的发展促进了城市智能化管理体系的建设速度，通过构建一个城市智能运营中心（IOC），依托一套城市综合的数据融合体系、一套城市数据标准规范和统一的城市大数据支撑平台的建立，实现跨部门、异构数据的采集、管理、共享、分析、建模和展示，提供数据跨域共享、数据挖掘和城市运行体征的综合服务，提高城市“综合性”问题的解决能力。通过一个开放的体系架构、一套天地一体的城市服务网格和一个高效的城市运行中心的建立，实现“网格大巡查、数据大智慧、政府大服务、政府大治理、公众大参与”的新型城市服务应用生态，解决跨部门城市事件和城市业务协同“联动”的问题。

城市智能运营中心通过城市大数据的分析与建模，实现城市运行体征监测、城市空间布局与资源承载优化、工业经济运行质态监控与宏观分析、城市资产全生命周期管理、人口分析与民生服务、环境监测与应急处置等“综合性”功能，为城市管理者提供辅助决策和科学治理的依据。通过政务数据开放和数据服务，实现政务办理、社会事务、市场监管、社会综治等服务，打造“互联网+”政务、信息惠民等城市服务新业态。

3.2.1.3 政府公共服务效能的提升

3.2.1.3.1 提高政务服务和管理能级

随着国家大数据战略的推进和大数据应用的深化，政府为民服务的大数据时代已经来临。对政府公共服务而言，大数据之“大”，不仅仅在于其容量之大、类型之多，更为重要的意义在于用数据创造更大的公共价值，提升政府网上服务能力，形成政民融合、互动的“互联网+政务服务”新格局。2016年4月发布的《国务院办公厅关于转发国家发展改革委等部门推进“互联网+政务服务”开展信息惠民试点实施方案的通知》（国办发〔2016〕23号）针对困扰基层群众的“办证多、办事难”等问题，提出了以实现“一号一窗一网”为目标的“互联网+政务服务”新模式。而“一号一窗一网”的关键就在于利用大数据技术和方法创新

政府网络服务模式，有效整合、开放共享和深化利用政务服务大数据，为公众提供个性化、精准化便捷服务。

优化政务服务大数据应用，以大数据创新网络服务模式，从“供给端”和“需求端”两个方面做好政务服务大数据应用工作。积极运用大数据手段推动政府监管模式转型，强化事中事后监管，进一步优化政务服务供给机制；利用大数据精准感知用户需求和体验，对公众行为数据、电子证照库、数据共享交换平台数据库等数据资源进行整合，充分利用大数据技术对跨领域、跨渠道的多元数据展开分析挖掘，全面掌握用户访问行为特征和规律，深入了解用户的服务需求，从而根据用户需求和体验来优化服务资源配置，丰富服务内容，改进服务方式，为用户提供个性化、精准化的服务，变“被动服务”为“主动服务”，不断提升政务服务水平和群众满意度。

3.2.1.3.2 提高政府资金使用效率

由于我国的社会保障制度建立时间不长，制度本身又是采取分人群设计、分部门管理、分地区实施，同时各部门大多实行封闭监管模式，彼此之间管理信息不能有效共享，数据、凭证传递不及时，各方账目、数据常有出入，造成社会保障信息管理协调难度加大，信息共享不足导致资金使用效率不高。面对这些问题，大数据的深入应用在信息采集、信息资源利用、跨地域社会保障统一管理等方面有效地提高了资金使用效率：

1) 大数据提高信息采集效率。通过大数据共享平台的建设，原本存在于不同部门的信息比对将成为可能，对数据的加密脱敏技术，将解决出于信息保密原因造成的信息共享困难，国家和各省市各级政务大数据中心将有序归集数据资源，解决由于信息系统建设权属和数据存储方式造成的共享难题，从根本上解决信息的有效采集。

2) 大数据解决资源利用难题。利用多源异构数据处理技术，不同部门采集的数据不一致，数据源存储格式不一的政务信息资源可以得到共享。同时，来自各部门各专业系统的人群信息可以得到良好的比对和识别，以退休职工、在职职工、农村低保、优抚对象、特困救助对象等不同保障对象的数据信息可以分别储

存、统合使用。数据的比对和分析，可以彻底解决因为信息分散而导致的冒领、错领、套取、虚报等问题，确实提高资金的使用效率。

3) 大数据解决跨地域问题。目前我国社会保障资金政策业务区域性强, 跨地域流动障碍大, 容易导致保障资金重复享受, 浪费大量财政补贴等问题。大数据的应用将有助于利用信息化手段消除农村劳动力进城和跨地区就业的限制, 完善参保人员社会保障关系转移、衔接等政策措施, 同时大数据将有助于建立唯一身份标识性公民身份画像, 建立各个领域各地域社会保障性资金资助过程和结果分析体系, 彻底解决跨地域重复参保或领取待遇的问题。

3.2.2 农业

物联网、大数据等技术已经涉及到耕地、育种、播种、施肥、植保、收获、储运、农产品加工、销售、畜牧业生产等各环节, 可以实现对作物种植、培育、成熟和销售等环节的管理, 使得传统的粗放式农业生产模式迈向集约化、精准化、智能化、数据化。

农业大数据是指现代农业生产、经营、管理等各种活动中形成的, 满足统一的数据技术、管理标准, 具有潜在价值的、海量的、活的数据。目前农业大数据的应用主要集中在精准农业、农资流通、农产品价格与农产品流通、土地流转、农产品质量追溯、农业经营者征信等方面。精准农业是大数据在农业中应用最普遍的领域之一, 通过对气候、土壤和空气质量、作物成熟度, 甚至是设备和劳动力的成本及可用性方面的实时数据收集与预测分析, 有助于做到精准种植、养殖, 减少资源浪费和成本投入。农资流通大数据可判断某个品类农产品的生产规模, 作为调整规模的依据。农产品价格与农产品流通大数据可帮助调节农产品生产规模、调整生产品类, 同时通过 B2B、B2C 电子商务平台促使农产品供求信息对接, 能拓展销售市场, 提升农产品销量。土地流转大数据的应用价值在于盘活土地, 提高土地流转透明度, 实现土地经营权管理和追踪, 有效降低土地监管和纠纷处理成本。农产品质量追溯大数据可监管农业生产全过程, 通过对监管数据进行分析 and 综合利用, 为农产品质量安全监管服务; 同时可以消除消费者对农产品质量的疑虑, 提高农产品的购买率。农业经营者征信大数据可作为发放贷款、设置农

业保险的信用依据，以此推动金融和农业的融合。

3.2.3 制造业

伴随着“智能制造 2025”国家战略的实施，大数据应用已成为制造业生产力、竞争力、创新能力提升的关键，是驱动制造过程、产品、模式、管理及服务标准化、智能化的重要基础，体现在产品全生命周期中的各个阶段，工业大数据正在加速制造业的转型升级。

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。

3.2.4 电力及水务领域

随着大数据技术的广泛应用，大数据已成为推动社会产业经济快速发展和创新的有力工具，将大数据技术应用于电力及水务等能源领域，是技术发展的必然趋势。将能源消费数据、智能设备数据、客户信息等数据相结合，充分挖掘客户行为特征，发现用户消费规律，从而提升企业运营效率，已成为能源大数据研究的重心。

3.2.4.1 电力生产和供应

电力大数据主要来源于电力生产和电能使用的发电、输电、变电、配电、用电和调度各个环节，未来电网的发展是以信息通信系统与电力基础设施深度融合为主要特征，融合的过程必将产生大量的数据，如何管理和利用这些数据是我们面临的重大挑战，另一方面，这些数据的融合、挖掘为我们认识和驾驭未来电网提供了有效的手段。

电力大数据应用涉及电力企业的各个专业，典型的如人、财、物、营销、规划、调度运行等。大数据技术在电力公司的核心价值主要体现在两个方面：一是将数据视作人、财、物一样的企业核心资产，建设统一的标准化数据管理体系，通过复杂的关联分析，让数据创造新的价值，提升精细化管理水平，促进管理方式和商业模式创新；二是将大数据技术应用于电力生产和电能使用的各个环节，通过技术变革，优化电网生产方式，提升生产效率，推动智能电网创新发展。具体包括：

1) 搭建技术平台，建设满足大数据应用、符合大数据特点的数据中心。利用云计算、云存储、HADOOP 体系架构等技术，建设公司级数据集市和数据仓库，实现公司数据的快速获取与应用，为大数据应用提供技术支撑。

2) 夯实数据基础，构建满足大数据应用的数据资源管理体系。加快数据匹配和业务融合，实现数据资源的互联互通。加快整合数据资源、促进数据共享、消除信息孤岛，形成统一规范的数据定义。建立数据资产登记制度和数据资产查询规则，建立数据资产元数据标准，规范数据资产说明，建立和发布数据资产目录，建立数据资源管理与服务机制。

3) 开展数据应用，助推公司生产经营管理水平提升。在生产、营销等公司主营业务中开展大数据应用场景的探索、实践和印证，为公司大数据应用积累成功经验。在成功探索基础上，积极探索大数据分析在电网规划、运维检修、电力交易中的应用，强化数据说话，持续提升数据管理能力，提高公司运营管理水平。加强公司外部数据的获取和应用，拓展数据渠道，丰富数据关联，提升公司对外部环境的感知能力，为公司战略决策提供有效参考。

3.2.4.2 水务管理

随着社会经济和市场化的快速发展，水资源作为国家能源之一，水源、水质、供水服务越来越受到社会民众的关注。随着水务企业生产监测系统、营业收费管理系统的实施建设，极大的提高了供水、水质的安全可靠性和水务企业的管理水平。以水务监测、运行、管理等数据为基础，充分利用大数据分析及人工智能研究成果，科学有效地开展水资源的合理开发、高效利用、优化配置、全面节约、有效保护和综合治理，对推动水务智能调度决策，推动水力企业业务创新有着重要的意义。

水务监测、运行、管理等基于统一的标准化的数据管理，充分利用大数据分析及人工智能研究成果，实现供水及需水分析预测、峰谷经济用能分析、泵组运行效率分析，提升水务数据的深层次价值。

1) 供水及需水需求预测分析

采用回归分析法和时间序列分析法，通过对历年供水量、历年气温、历年降雨量、历年系统检修情况、竞争对手历年供水情况、地方人口、经济历史数据的分析，分析系统上日供水量、降雨预测、客户自有水源、竞争对手供水量、水库运行水位、水质、设备设施检修、地方人口经济情况影响因子对供水及需水的影响度，寻求影响供水及需水的关键因素，运用影响因子实现对供水方案的优化。最终形成对客户月供水量、区间次日供水量、次日系统抽水量的预测。

2) 峰谷经济用能分析

采用回归分析法、时间序列分析法，针对影响泵组运行的流量扬程、机组流量、功率、机组大修、改造等因素进行分析和研究，寻求影响泵组运行效率的关键影响因子，通过对关键影响因子的调节和修正，提升泵组运行效率，达到降低

机组能耗的目的。

3.2.5 通信与电子商务

3.2.5.1 通信领域

通信行业拥有巨大的数据资源，发展大数据有得天独厚的优势。首先运营商拥有的数据涉及范围广，不仅涉及财务收入、业务信息等结构化数据，也会涉及图片、文本、音频、视频等非结构化数据。此外，运营商拥有的数据涵盖全业务、全用户和全渠道，信息完整。同时，运营商拥有的数据记录周期长，数据延续性好，覆盖用户从入网到离网前的连续时间。

要利用巨大的数据资源充分发展广泛的通信业务，大数据平台要解决很大的挑战。运营商系统内多个子系统（业务支撑系统 BSS，运营支撑系统 OSS，管理支撑系统 MSS）内的数据表现形式、数据结构和定义完全不同。如何融合这三个域的数据资源，支持运营商业务，成为通信大数据首要解决的问题。此外，数据量大（已从 TB 发展至 PB，甚至更大量级），响应时延的要求高，数据多份存储高冗余等，也是通信大数据平台需求解决的问题。

为了解决这些行业需求，特别是对分割的业务数据统一处理及分析，需要构建汇聚网格、业务、终端、客户行为等多维度的数据分析平台，遵循行业内统一的数据技术、管理等相关标准，提供融合数据存储、统一数据访问等跨平台的数据能力，包括：

1) 多样化的数据采集：

支持跨领域业务、多种格式（表、文件、消息等）数据的分布式批量采集，及实时增量数据采集。性能上要比基于传统 ETL 的采集性能有数量级的提升。支持与各种数据库技术的对接，实现跨平台开发和数据管理。

2) 跨域、跨业务数据融合，跨平台的数据分析能力

能够将不同子域（BSS、OSS、MSS）、不同业务的数据融合统一地存储、组织及管理，使能统一的数据访问、整合和分析。支持主流计算框架和计算引擎（如 MapReduce、Spark、Flink），能够高效处理海量非结构化、半结构化数据，同时满足批处理、流处理等不同计算场景需求。

3) 集成多种挖掘算法，提高数据挖掘能力

基于通信行业丰富的数据资源，利用大数据平台强大的计算能力，通过多种建模算法（朴素贝叶斯、稀疏线性矩阵、决策树、随机森林、逻辑回归、Kmeans、社交网络推荐、影响力传播、协调过滤、线性回归等），提升数据价值。并支持系统自动建模。

4) 图形化开发，降低大数据应用门槛

集成易用的开发和维护工具，支持图形化界面进行二次开发。降低大数据应用的开发难度，提升效率。

3.2.5.2 电子商务领域

大数据开启了电子商务的时代转型。电商经营中获取的海量信息，如商家、用户信息和产品使用体验等，都蕴藏着具有巨大价值的用户需求和竞争情报。这些信息随着交易不断积累，渗透到电商交易的各个环节。

从需求挖掘上看，以阿里巴巴、京东为代表的全品类、综合性平台，凭着一站式满足全部消费需求提高了平台的利用效率，也通过整合需求获得了更高的规模经济优势和更多的大数据来源；从预测市场上看，大数据预测技术通过对数据的甄别与分析，勾勒用户消费习惯、能力的“用户画像”，获取产品在各区域、时间段和消费群的销售情况与市场趋势等，实现电商企业在开发、生产到销售的全产业链中更精准和迅速的反应；从营销环节上看，急剧增加的消费数据使得电商企业更加理解客户，通过大数据应用划分消费群体，进行个性化、智能化的推广，有效提升了营销行为转化成购买行为的比例，带来了更高的经营效率；从仓储物流环节上看，基于行业内统一的数据标准化管理体系，电商（及合作的物流企业）依靠对客户数据的分析，选择更合理的派送方式、路径，更科学、智能地调配仓储，提供分时配送等服务，大大降低了仓储物流环节的存货和时间压力，提升了物流服务质量及交易的即时性、便捷性；从定制服务上看，电商企业通过“用户画像”等数据技术，为用户提供差异化、定制化产品和服务，如定制咨询应答策略、针对性商品推荐和个性化关怀等，以个性化服务大大提升了用户体验，同时也有利于改善电商行业竞争越发同质化的现状，避免过度竞争的问题。

综上，依靠着电商业务和数据之间天然的紧密联系，借助大数据挖掘与分析

技术，电商企业可在经营的各个环节中充分利用这些信息提高行为效率，获得更好的经营效益，也能为消费者提供更丰富、更优质的产品和服务，促进电商领域长期的良好发展。

3.2.6 交通运输领域

交通运输是国民经济中基础性、先导性、战略性产业，是重要的服务性行业，与经济发展和百姓生活紧密相关。当前，交通运输行业正处于交通运输基础设施发展、服务水平提高和转型发展的黄金时期。在这一历史进程中，交通运输行业积累了体量巨大、类型繁多、来源多样的数据资源，大数据应用需求活跃，大数据产业蓬勃发展。

在公路领域，基于统一的数据技术、管理标准，利用公路动态监测数据、收费数据等，开展公路基础设施使用性能评价、公路网运行监测预警、公路养护决策支持等；在水路领域，利用船舶位置数据和港口码头运行数据，开展水路运输监测预警、港口布局优化、水上搜救应急指挥等；在城市交通领域，利用 IC 卡数据、手机信令数据、车辆 GPS 数据和移动互联网众包数据等，开展交通流仿真及预测、城市群交通出行特征分析、公交能耗排放动态监测、公交线网评价及优化等；在铁路领域，利用铁路客票数据和动态检测数据，开展铁路旅客用户出行特征分析、铁路设施设备动态监测等；在民航领域，利用海量的旅客数据，开展民航旅客精准“画像”和服务体验提升分析等；在邮政领域，利用单证数据、车辆位置数据等，开展邮政车辆动态调度、运输网络仿真及优化、物流成本与绩效评价等。在出行服务领域，依托政企合作模式，聚焦出行导航、订票、约租车、物流、汽车后服务、航运和船舶信息服务等领域，推动商业化的交通大数据产品呈现爆发式增长，在创新出行信息服务模式，改善用户体验的同时，又汇集形成新的交通运输大数据资源。

依托统一的准化管理思路，充分利用交通运输行业各领域汇集的海量数据资源，深度挖掘数据资源价值，不仅可以为交通各领域的发展提供更多机遇，也可为社会公众提供更加便捷、安全、高效、绿色的出行环境，从而推动整个交通运输行业的发展。

3.2.7 邮政领域

邮政体制改革十年来，邮政管理部门已逐步形成了邮政普遍服务“大数据”的基础。2016年7月1日，国务院办公厅《关于运用大数据加强对市场主体服务和监管的若干意见》正式出炉，提出要建立大数据标准体系。邮政业是一个数据高度密集性的企业，掌握着可能是全球最庞大的数据，在大数据时代应是遇到了极好的机遇。

按照国家对政府简政放权、加强事中事后监管、加强行业安全监管等相关要求，为有效解决三级邮政管理部门监管人少、事多、覆盖面大、监管基础能力薄弱、信息化基础较差等问题，依靠统一的标准化思路、信息化手段创新监管方式、提升监管能力，促进企业落实安全主体责任，提高政府履职能力，国家邮政局向国家发展改革委申报了“邮政业安全生产监管信息化工程”的建设，并被列入了国家安全生产监管信息化工程（一期）的建设内容。安全生产监管信息化工程（一期）国家邮政局建设部分项目于2016年6月获得批复并正式启动建设。大数据资源建设是工程的重要建设内容之一，主要实现邮政行业相关数据资源整合利用。

国家邮政总局运用大数据技术构建了邮政大数据平台，整合国内海量的快递物流数据，通过数据治理平台实现对邮政行业数据标准的管理、元数据管理、数据质量管理等功能，提升了对大数据平台海量数据管控能力，保证了物流数据的有效性、完整性和准确性，充分运用大数据挖掘分析技术，建立科学合理的仿真模型，主动发现违法违规现象，提高政府科学决策和风险预判能力。同时，通过邮政大数据平台作为数据支撑基础，综合评估各级邮政管理部门的监管与服务绩效，并根据评估结果及时调整和优化，提高各级邮政管理部门在邮政服务监管的有效性。另外，以邮政大数据平台为基础，构建分析模型，为财政资金补贴使用提供科学分析，通过大数据判断网点有否重复建设、补贴买车辆数量是否合理、新网点位置是否达标等等，确保资金用到实处，提高补贴资金使用效果。

以大数据体系建设为基础的信息化改革浪潮，将不断完善邮政服务监管数据，进一步提升邮政管理部门监管能力，打造智慧邮政，提升邮政普遍服务的可持续发展能力。

3.2.8 金融领域

金融是通过对现有资源的重新整合，实现价值和利润的等效流通，是人们在不确定环境中进行资源跨期的最优配置决策行为。纵观国内外，以互联网金融为代表的金融科技飞速发展，银行、券商等金融机构通过大数据平台等基础设施建设，快速的将发展重心转向基于大数据的业务价值探索和应用实践。

金融大数据以客户为中心，全面分析“学、选、买、用”的金融交易全流程，以高价值金融交易数据为核心，逐步按需进行跨行业、跨领域综合数据应用。金融大数据典型应用领域包括：历史查询、客户管理、营销管理、运营管理、绩效管理、风险管理等。

金融大数据解决方案则以数据技术标准、管理标准、大数据平台为基础，聚焦金融业务场景，围绕客户认知、价值提升、全生命周期的持续运营，切实提升业务营销人员、业务管理人员、经营决策人员的实际工作效能，提升辅助决策水平。通过金融大数据应用类项目建设，将聚焦以下方面持续发展：

1) 优化金融服务和产品创新、提升服务层次

围绕客户需求，提升综合金融服务解决方案能力，通过更高水平的智能投顾，提供差异化服务价值，提升客户粘性。

2) 促进金融生产、管理模式转型，提升运营水平

打造数据化业务运营体系，运用大数据思维和互联网技术，促进金融行业生产模式、管理模式的转型和创新发展的。

3) 实现金融和类金融数据资源集成与共享

建立健全法律法规、行业制度、个人保护、技术和管理标准化体系等约束，完善金融机构内部跨业务系统的数据整合，持续探索金融行业间数据共享、以及类金融的电商平台、工商、税务、电信运营商等外部数据融合。

另外，互联网金融的首个纲领性指导意见是 2015 年 7 月由人民银行、银监会、财政部等十部委联合出台的《关于促进互联网金融健康发展的指导意见》。两年间，《互联网金融专项整治实施方案》、《股权众筹风险专项整治工作实施方案》、《网络借贷信息中介机构业务活动管理暂行办法》、《网络借贷资金存管业务指引》、《网络借贷信息中介机构业务活动信息披露指引》等行业整顿、监管文件

陆续出台，指导并规范金融行业向智能化、信息化发展。

3.2.9 科学研究

大数据时代的科学研究是一个大科学、大需求、大数据、大计算、大发现的过程，数据密集型科学发现已经成为继实验科学、理论推演、计算机模拟之后的第四种科学范式。基于对科学大数据的处理和分析将成为从科学数据到科学发现的重要桥梁，能够为诸多学科领域的科学新发现提供坚实的技术基础，加速具有国际影响力的科技成果产出。

在天文研究方面，基于世界最大口径单体射电望远镜 FAST 每日产生的海量天文数据，开展海量天文数据整合分析、天体分析和挖掘、天文大数据可视化等研究工作，推动天文科学研究和探索由假设驱动向数据驱动转变；在气候环境方面，基于“地球数值模拟装置”系统对自然观测数据的积累，开展自然观测数据的分析和挖掘等研究工作，实现对未来气候环境变化趋势的预测和预报；在雾霾治理方面，基于从各个环境监测点汇集的海量环境监测数据，利用大数据技术不断模拟雾霾的形成、演绎雾霾的走向，获得精准的空气质量计算模型，实现精准的雾霾预报与预警；在精准医学领域，建立多层次精准医疗知识库体系和国家生物医学大数据共享平台，开展新一代基因测序技术、组学研究和大数据融合分析技术等精准医疗核心关键技术的研究，推动医学诊疗模式的变革。

科学大数据在推动科学研究、促进各行业领域科学发现和技术创新方面有着非常大的潜力。当前，中国科学院已启动“十三五”信息化专项科学大数据工程项目“大数据驱动学科创新示范平台”，基于“十三五”建设的“中国科技云”和全院科学大数据公共基础环境平台，在生命科学、空间科学、天文科学、高能物理等学科领域，建设大数据驱动学科创新示范平台，实现学科内数据资源的深度集成整合，构建科学大数据分析应用环境，创新大数据驱动科学发现的研究模式。未来，科学大数据必将是人类科研革命和社会进步的重要支撑。

科技竞争正在成为国际综合国力竞争的焦点，世界各国纷纷把推动科技创新和发展新兴产业作为国家战略，科技创新的速度和转化为现实生产力的步伐进一步加快，创新资源配置全球化的特点更为突出，区域竞争重心已转向创新创业生

态的竞争，以开放、移动和可持续为特征的新一轮产业技术革命推动全球创业活动大幅提升。开放平台、大数据、云计算使创业成本达到历史新低，大数据与其他产业的融合创造出许多新应用，推动全球科技创业活动进入新一轮热潮。利用大数据技术打造新型双创孵化平台可以为创业科技人员和处于创业初期的科技型中小企业提供必要的资源和服务，降低创业成本，提高创业成功率，促进科技成果转化，培育科技型企业和企业家，对推动高新技术产业发展、完善国家和区域创新体系、繁荣经济发挥着重要的作用，是科技型中小企业生存与成长所需要的制度安排，具有重大的社会经济意义。

基于大数据的科技双创孵化活动紧密围绕科技创业活动，以促进科技成果转化、培养高新技术企业和企业家为宗旨，是培育战略性新兴产业源头企业和建设创新型国家的战略工具，是培养产业领军人才的有效载体，是支撑区域科学发展、优化发展的科技服务创新业态。大数据科技双创孵化为创新链条提供全过程综合服务，提供创新活动所需的资金流、信息流、知识流、人才流等，提高创新效率。

3.2.10 教育领域

随着“互联网+”时代的到来，物联网、移动互联网、大数据、人工智能等相继进入校园工作和生活中，教育已经进入一个新的时代。教育大数据作为教育信息化发展的更高阶段，成为促进教育公平、提高教育质量、推动教育改革的有力抓手和有效手段。信息技术与教育的深度融合，使得全方位、深度追踪、量化学习过程，采集和汇聚教育场景中各类数据、乃至其它各种跨界数据成为可能。由于教育业务复杂，无标准化的操作流程和模式，创新人才的培养又需要更多元化、创新性的教学模式与方法。缺少标准化的业务流程以及学习方式的多样性和学习地点的不确定性，导致教育大数据的采集变得异常复杂。学校作为教育信息化落地的主体被迫要求扮演更加多元的角色，也为学校的信息化建设提出了更高的要求。

学校需要遵循行业统一的数据技术、管理标准，在原来以流程为核心的系统建设基础上，加强以数据为核心的系统体系建设。教育大数据的建设不是一套软件工具和几个场景应用的搭建就能简单的实现的，而是在进行顶层设计的规划后，

分步骤持续建设的过程，根据学校目前应用系统和数据的现状，并结合实际的具体需求，建设一个能够支撑开展多源异构数据汇聚、系统化监测与评估、教与学过程的综合建模和行为分析、数据可视化等技术的研发和工程化的教育大数据应用技术创新平台，支撑未来学校的发展与变革。

面对具体教育业务，教育大数据技术对教育的管理、教学、学习、科研和评价等都产生了很大的影响。目前学校里可利用的数据一共分为三大块，业务数据、机器数据以及互联网数据。其中前面两块属于校内数据，互联网数据属于校外数据。各个业务系统的数据需要对接，各种半结构化数据(设备产生的日志)和非结构化数据(监控的视频、照片以及各种文本数据)也在大量产生，这些是传统的业务系统所处理和承载不了的数据，但是对于学校来说又是非常有价值的。教育大数据具有具备转换层次的四个过程特征：数据、信息、知识、智慧。教育大数据技术采集的海量教育元数据，经过抽取、转换、加载，联机分析处理和统计分析等过程，转换为教育信息，又经过数据可视化将教育数据呈现出来，最后形成教育决策来指导教育者和学习的教学，这个过程数据完成了从数据到信息、知识和智慧的演变。

当前最重要的是整合第一块的业务数据，通过建立传统的数据仓库来将这些数据进行统一和集中的管理，可以直接横向打通各个业务系统的壁垒。但是对于后面两块的数据，学校重视程度不够，机器数据实际上也是学校要实现智慧校园所不可或缺的数据基础，广泛分布于学校的无线网、门禁、视频等，都属于这一类数据，机器数据不仅仅数量庞大，同时记录的信息也足够全，合理整合及利用机器数据，是智慧校园建设的重中之重。而互联网媒体数据则是存在于校园网之外，但是又跟学校紧密程度比较高的一类数据，在利用好业务数据和机器数据的同时，将互联网数据也纳入到整个学校的数据体系中，也是智慧校园建设的一个关键指标。通过大数据平台可以为学校把所有业务系统数据打通，开发利用现有的机器数据，同时再纳入互联网数据，全面激活学校数据价值，释放学校数据力量。

“互联网+教育”的真正意义在于促进教育的转型升级。教育大数据将重构教育生态系统，宏观层面为教育决策提供科学依据，中观层面推进教学管理和评价的创新实践，微观层面为个性化教学提供精准支持。

3.2.11 卫生领域

随着我国城镇化进程加快以及人口老龄化的加速，民众健康意识不断增强，人民群众对医疗健康的需求也越来越高。日益增长的医疗健康需求和我国原有分级诊疗、以药养医的僵化体制的矛盾日益激烈，以大数据应用为突破口的医疗卫生信息化升级，给解决上述问题带来了期望。

大数据在医疗个性化服务中的应用。利用好“大数据”可以提升医疗价值，形成个性化医疗，即基于基因科学的医疗模式，同时通过对居民健康影响因素进行分析，对患者健康信息进行整合，为疾病的诊断和治疗提供更好的数据证据，进行居民健康知识库的积累，从而改进居民健康。在个性化药物的开发领域，研发药物时，大数据技术可通过考察遗传变异、对特定疾病的易感性和对特殊药物的反应三者之间的关系，然后在药物研发和用药过程中考虑个人的遗传变异因素，针对不同的患者采取不同的诊疗方案，或者根据患者的实际情况调整药物剂量，可以减少副作用。

大数据在个人健康管理方面的应用。依托行业统一的数据技术、管理标准，利用先进的大数据技术，对个人健康进行全生命周期管理，实现在任何时间、任何地点都可以访问相关信息，从而保证了健康信息的一致性、连续性，利于健康分析人员能够有效地对个人健康状况进行分析并及时进行干预。利用大数据技术对所有产生数据进行分析，汇总成一个健康风险指数，用户可以看到自己的健康风险指数和同龄、同性别人群的平均风险指数，并且能明确自己的健康风险在同龄人群中的排位，可使用户对自己健康状况有个初步的评估并对行为方式或生活习惯等进行调整。

大数据时代电子病历的应用。医院电子病历管理系统所产生和保存的病案信息是医疗卫生“大数据”的重要信息来源，以大数据信息管理理念构建医疗卫生信息共享中心，能利用大数据客观、完整、连续地记录患者的诊疗经过、病情变化、治疗效果等，在医院医疗、教学、科研领域，具有丰富最佳治疗途径、提高诊疗水平、防控流行病疫情等方面都具有重大意义。

未来，随着对海量医疗卫生大数据的结构化处理的深化，大数据会在精准医疗、医药供给侧改革、商业健康保险等领域进一步深入发展，会产生巨大的商业

价值，并且使得整个医疗产业链受益。

3.2.12 文化领域

在文化版权领域，当前的网络版权侵权现象十分严重，网络版权侵权难于发现，更难于取证，传统工作模式下很难对网络版权进行有效的监管。但是，随着大数据时代的到来，网络版权保护有了新的契机，国家新闻出版广电总局和贵州省人民政府共同推动了“中国文化（出版广电）大数据产业”项目，依托大数据技术，建立数字音像互联网实时监测系统，探索出版广电行业在新形势下跨区域、跨行业、跨网络、跨终端、跨所有制的发展新模式，形成“云、管、端”一体化的技术、市场新体系，助推传统媒体与新兴媒体融合发展。

大数据技术在网络版权保护中具体体现在：1、文化版权使用信息、文化版权授权信息、文化版权盗版信息等相关文化版权的数据，在互联网的数据海洋中浩如烟海，通过搜索引擎和网络爬虫技术，可以实现全网文化版权数据的自动化、智能化获取，解放人力资源，实现以往需要耗费大量人力也难以触及的目标；2、捕获到文化版权信息后，构建相关的数据技术、管理标准，通过智能分析对比，可以完成对版权信息的初步匹配和深度匹配，自动化发现疑似侵权目标，将疑似侵权目标交由人工进行确认，大大减少了人力投入；3、互联网版权侵权内容涵盖文字、图片、音频、视频等几类数据，大多数为非结构化数据，而对于海量的非结构化数据的存储与检索必须采用大数据技术才能得以实现，通过高性能的分布式存储计算能力，满足海量数据的存储与实时计算性能要求的同时，还需要实现信息的快速检索；4、建立数据共享服务机制，依托行业内统一的大数据标准体系，打通政府各部门之间的“数据高速公路”，使侵权存证信息快速推送到对应执法部门，为执法监管提供有利条件，最终形成各部门之间的联合惩戒。5、借助大数据可视化技术，使文化版权监管做到全过程溯源，全流程监督，真正做到实时监督，全面掌握网络文化版权态势。

3.3 大数据发展趋势分析

我国作为数据大国，在互联网、工业制造、金融、医疗等各个领域均有着庞

大的数据基础，整体数据量大、数据品种丰富，这为我国大数据领域的发展提供了重要的基础支撑。当前我国大数据领域已步入快速推进期，核心技术逐步突破、涉及行业不断拓展，产业应用逐渐深入，呈现出资源集聚、创新驱动、融合应用、产业转型的新趋势。

(1) 大数据与实体经济深度融合

大数据作为新一代信息技术的重要标志，对生产制造、流通、分配、消费活动以及经济运行机制、社会生活方式和国家治理能力均产生重要影响。面对我国目前新旧动能转化的关键时期，通过大数据与实体经济深度融合，利用大数据采集、大数据存储、大数据处理、大数据分析、数据管理等技术提升产业能效，加速传统行业经营管理方式变革、服务模式和商业模式创新及产业价值链体系重构，最终完成产业转型，培育新动能。

(2) 大数据发展开启数字中国建设

大数据重塑传统经济形态，数字经济将成为未来发展的新模式；大数据推动了国家治理模式的创新，提高了政府宏观调控、社会管理和市场监管能力，促进了政府决策科学化、社会治理精准化、市场监管高效化；大数据促进了民生服务精细化、均等化和普惠化，提升了民生水平；大数据为实现网络安全态势的全面感知、深度洞察和超前研判提供了海量信息支撑，保障了网络安全。大数据正逐渐与经济、政务、民生、安全等各个领域全面融合，推动了信息化发展模式的变革创新，开启了数字中国建设的新时代。

(3) 大数据将向智能化、智慧化发展

在以大数据驱动的人工智能时代，海量数据经过预处理之后抽取出可用信息，可用信息经过加工后成为知识。大数据是人工智能的基石，目前的深度学习主要是建立在大数据的基础上，即对大数据进行训练，并从中归纳出可以被计算机运用在类似数据上的知识或规律。人工智能将成为大数据生态中的重要组成部分，相关方面的应用将呈现爆发态势，并将在医疗健康、金融、零售、广告、交通、教育等细分领域取得突破，最终从“大数据”演变为“大智慧”。

(4) 数据治理将成为重点发展领域

伴随大数据的热潮，大量的组织对大数据平台建设、分析应用等方面盲目投入，缺乏对大数据资源的整体规划和综合治理。随着一些项目实施的终止和失败，

以及数据量的继续激增，数据治理的重要性逐步得到业界的共识。治理是基础，技术是承载，分析是手段，应用是目的，随着国家政策支撑以及产业实际需求的增长，如何通过数据治理提升组织数据管理能力，消除数据孤岛，挖掘数据潜在价值将成为重点发展领域。

4 大数据参考架构

4.1 参考架构

大数据作为一种新兴技术，目前尚未形成完善、达成共识的技术标准体系。本章结合 NIST 和 JTC1/SC32 的研究成果，结合我们对大数据的理解和分析，提出了大数据参考架构（见图 3）。

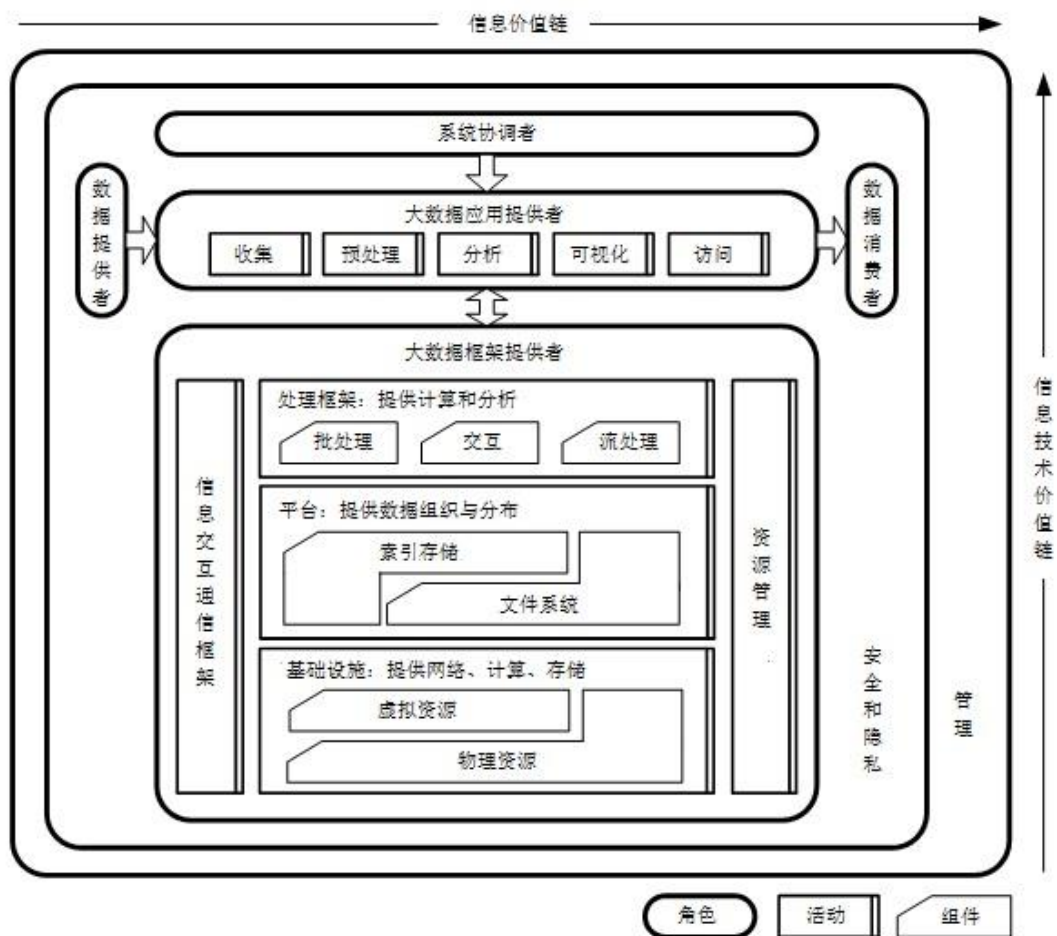


图 3 大数据参考架构图

大数据参考架构总体上可以概括为“一个概念体系，二个价值链维度”。“一个概念体系”是指它为大数据参考架构中使用的概念提供了一个构件层级分类体系，即“角色—活动—功能组件”，用于描述参考架构中的逻辑构件及其关系；“二个价值链维度”分别为“IT 价值链”和“信息价值链”，其中“IT 价值链”反映的是大数据作为一种新兴的数据应用范式对 IT 技术产生的新需求所带来的价值，“信息价值链”反映的是大数据作为一种数据科学方法论对数据到知识的处理过

程中所实现的信息流价值。这些内涵在大数据参考架构图中得到了体现。

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、技术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。其目标是建立一个开放的大数据技术参考架构，使系统工程师、数据科学家、软件开发人员、数据架构师和高级决策者，能够在可以互操作的大数据生态系统中制定一个解决方案，解决由各种大数据特征融合而带来的需要使用多种方法的问题。它提供了一个通用的大数据应用系统框架，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。

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。最顶层级的逻辑构件是角色，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、安全和隐私、管理。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。

大数据参考架构图的整体布局按照代表大数据价值链的两个维度来组织，即信息价值链（水平轴）和 IT 价值链（垂直轴）。在信息价值链维度上，大数据的价值通过数据的收集、预处理、分析、可视化和访问等活动来实现。在 IT 价值链维度上，大数据价值通过为大数据应用提供存放和运行大数据的网络、基础设施、平台、应用工具以及其他 IT 服务来实现。大数据应用提供者处在两个维度的交叉点上，表明大数据分析及其实施为两个价值链上的大数据利益相关者提供了价值。

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。这两个关键模型构件的功能极其重要，因此也被集成在任何大数据解决方案中。

参考架构可以用于多个大数据系统组成的复杂系统(如堆叠式或链式系统),这样其中一个系统的大数据使用者可以作为另外一个系统的大数据提供者。

4.2 参考架构中涉及的相关内容

4.2.1 系统协调者

系统协调者的职责在于规范和集成各类所需的数据应用活动,以构建一个可运行的垂直系统。系统协调者角色提供系统必须满足的整体要求,包括政策、治理、架构、资源和业务需求,以及为确保系统符合这些需求而进行的监控和审计活动。系统协调者角色的扮演者包括业务领导、咨询师、数据科学家、信息架构师、软件架构师、安全和隐私架构师、网络架构师等。系统协调者定义和整合所需的数据应用活动到运行的垂直系统中。系统协调者通常会涉及到更多具体角色,由一个或多个角色扮演者管理和协调大数据系统的运行。这些角色扮演者可以是人,软件或二者的结合。系统协调者的功能是配置和管理大数据架构的其他组件,来执行一个或多个工作负载。这些由系统协调者管理的工作负载,在较低层可以是把框架组件分配或调配到个别物理或虚拟节点上,在较高层可以是提供一个图形用户界面来支持连接多个应用程序和组件的工作流规范。系统协调者也可以通过管理角色监控工作负载和系统,以确认每个工作负载都达到了特定的服务质量要求,还可能弹性地分配和提供额外的物理或虚拟资源,以满足由变化/激增的数据或用户/交易数量而带来的工作负载需求。

4.2.2 数据提供者

数据提供者的职责是将数据和信息引入到大数据系统中,供大数据系统发现、访问和转换。数据提供者角色为大数据系统提供可用的数据。数据提供者角色的扮演者包括企业、公共代理机构、研究人员和科学家、搜索引擎、Web/FTP 和其他应用、网络运营商、终端用户等。在一个大数据系统中,数据提供者的活动通常包括采集数据、持久化数据、对敏感信息进行转换和清洗、创建数据源的元数据及访问策略、访问控制、通过软件的可编程接口实现推或拉式的数据访问、发

布数据可用及访问方法的信息等。

数据提供者通常需要为各种数据源(原始数据或由其它系统预先转换的数据)创建一个抽象的数据源,通过不同的接口提供发现和访问数据功能。这些接口通常包括一个注册表,使得大数据应用程序能够找到数据提供者、确定包含感兴趣的数据、理解允许访问的类型、了解所支持的分析类型、定位数据源、确定数据访问方法、识别数据安全要求、识别数据保密要求以及其他相关信息。因此,该接口将提供注册数据源、查询注册表、识别注册表中包含标准数据集等功能。

基于对大数据的 4V 特性和系统设计方面的考虑,暴露和访问数据的接口需要根据变化的复杂性采用推(push)和拉(pull)两种软件机制。这两种软件机制包括订阅事件、监听数据馈送、查询特定数据属性或内容,以及提交一段代码来执行数据处理功能。由于需要考虑大数据量跨网络移动的经济性,接口还可以允许提交分析请求(例如,执行一段实现特定算法的软件代码),只把结果返回给请求者。数据访问可能并不总是自动进行,可以让人类角色登录到系统提供新数据应传送的方式(例如,基于数据馈送建立订阅电子邮件)。

4.2.3 大数据应用提供者

大数据应用提供者的职责是通过在数据生命周期中执行的一组特定操作,来满足由系统协调者规定的要求,以及安全性、隐私性要求。大数据应用提供者通过把大数据框架中的一般性资源和服务能力相结合,把业务逻辑和功能封装成架构组件,构造出特定的大数据应用系统。大数据应用提供者角色的扮演者包括应用程序专家、平台专家、咨询师等。大数据应用提供者角色执行的活动包括数据的收集、预处理、分析、可视化和访问。

大数据应用程序提供者可以是单个实例,也可以是一组更细粒度大数据应用提供者实例的集合,集合中的每个实例执行数据生命周期中的不同活动。每个大数据应用提供者的活动可能是由系统协调者、数据提供者或数据消费者调用的一般服务,如 Web 服务器、文件服务器、一个或多个应用程序的集合或组合。每个活动可以由多个不同实例执行,或者单个程序也可能执行多个活动。每个活动都能够与大数据框架提供者、数据提供者以及数据消费者交互。这些活动可以并行

执行，也可以按照任意的数字顺序执行，活动之间经常需要通过大数据框架提供者的消息和通信框架进行通信。大数据应用提供者执行的活动和功能，特别是数据收集和数据访问活动，需要与安全和隐私角色进行交互，执行认证/授权并记录或维护数据的出处。

收集活动负责处理与数据提供者的接口和数据引入。它可以是一般服务，如由系统协调者配置的用于接收或执行数据收集任务的文件服务器或 Web 服务器；也可以是特定于应用的服务，如用来从数据提供者拉数据或接收数据提供者推送数据的服务。收集活动执行的任务类似于 ETL 的抽取 (extraction) 环节。收集活动接收到的数据通常需要大数据框架提供者的处理框架来执行内存队列缓存或其他数据持久化服务。

预处理活动执行的任务类似于 ETL 的转换 (transformation) 环节，包括数据验证、清洗、标准化、格式化和存储。预处理活动也是大数据框架提供者归档存储的数据来源，这些数据的出处信息一般也要被验证并附加到数据存储中。预处理活动也可能聚集来自不同的数据提供者的数据，利用元数据键来创建一个扩展的和增强的数据集。

分析活动基于数据科学家的需求或垂直应用的需求，确定处理数据的算法来产生新的分析，解决技术目标，从而实现从数据中提取知识。这需要有特定的数据处理算法对数据进行处理，以便从数据中得出能够解决技术目标的新洞察。分析活动包括对大数据系统低级别的业务逻辑进行编码 (更高级别的业务流程逻辑由系统协调者进行编码)，它利用大数据框架提供者的处理框架来实现这些关联的逻辑，通常会涉及到在批处理或流处理组件上实现分析逻辑的软件。分析活动还可以使用大数据框架提供者的消息和通信框架在应用逻辑中传递数据和控制功能。

可视化活动提供给最终数据消费者处理中的数据元素和呈现分析功能的输出。可视化活动的任务是将分析活动结果以最利于沟通和理解知识的方式展现给数据消费者。可视化的功能包括生成基于文本的报告或者以图形方式渲染分析结果。可视化的结果可以是静态的，存储在大数据框架提供者中供以后访问。更多的情况下，可视化活动经常要与数据消费者、大数据分析活动以及大数据提供者的处理框架和平台进行交互，这就需要基于数据消费者设置的数据访问参数来提

供交互式可视化手段。可视化活动可以完全由应用程序实现，也可以使用大数据框架提供者提供的专门的可视化处理框架实现。

访问活动与可视化和分析功能交互，响应应用程序请求，通过使用处理和平台框架来检索数据，并响应数据消费者请求。访问活动主要集中在与数据消费者的通信和交互。与数据收集活动类似，访问活动可以是由系统协调者配置的一般服务，如 Web 服务器或应用服务器，用于接受数据消费者请求。访问活动还可以作为可视化活动、分析活动的界面来响应数据消费者的请求，并使用大数据框架提供者的处理框架和平台来检索数据，向数据消费者请求做出响应。此外，访问活动还要确保为数据消费者提供描述性和管理性元数据，并把这些元数据作为数据传送给数据消费者。访问活动与数据消费者的接口可以是同步或异步的，也可以使用拉（pull）或推（push）软件机制进行数据传输。

4.2.4 大数据框架提供者

大数据框架提供者的职责是为大数据应用提供者在创建具体应用时提供使用的资源和服务。大数据框架提供者的角色扮演者包括数据中心、云提供商、自建服务器集群等。大数据框架提供者的活动包括基础设施、平台、处理框架、信息交互/通信和资源管理。

基础设施为其他角色执行活动提供存放和运行大数据系统所需要的资源。通常情况下，这些资源是物理资源的某种组合，用来支持相似的虚拟资源。资源一般可以分为网络、计算、存储和环境。网络资源负责数据在基础设施组件之间的传送；计算资源包括物理处理器和内存，负责执行和保持大数据系统其他组件的软件；存储资源为大数据系统提供数据持久化能力；环境资源是在考虑建立大数据系统时需要的实体工厂资源，如供电、制冷等。

平台包含逻辑数据的组织和分布，支持文件系统方式存储和索引存储方法。文件系统方式实施某种级别的 POSIX 标准以获取权限，进行相关的文件操作。索引存储方法无需扫描整个数据集，便可以迅速定位数据的具体要素。平台通过相关的应用编程接口（API）或其他方式，提供数据的逻辑组织和分发服务。它也可能提供数据注册、元数据以及语义数据描述等服务。逻辑数据组织的范围涵盖

从简单的分隔符平面文件到完全分布式的关系存储或列存储。数据访问方式可以是文件存取 API 或查询语言（如 SQL）。通常情况下，实现的大数据系统既能支持任何基本的文件系统存储，也支持内存存储、索引文件存储等方式。

处理框架提供必要的基础软件以支持实现的应用能够处理具有 4V 特征的大数据。处理框架定义了数据的计算和处理是如何组织的。大数据应用依赖于各种平台和技术，以应对可扩展的数据处理和分析的挑战。处理框架一般可以分为批处理（batch）、流处理（streaming）和交互式（interactive）三种类型。

信息交互/通信为可水平伸缩的集群的结点之间提供可靠队列、传输、数据接收等功能。它通常有 2 种实现模式，即点对点（point-to-point）模式和存储-转发（store-and-forward）模式。点对点模式不考虑消息的恢复问题，数据直接从发送者传送给接收者。存储-转发模式提供消息持久化和恢复机制，发送者把数据发送给中介代理，中介代理先存储消息然后再转发给接收者。

资源管理对计算、存储及实现两者互联互通的网络连接进行管理。主要目标是实现分布式的、弹性的资源调配，具体包括对存储资源的管理和对计算资源的管理。资源管理活动负责解决由于大数据的数据量和速度特征而带来的对 CPU、内存、I/O 等资源管理问题。有两种不同的资源管理方式，分别是框架内（intra-framework）资源管理和框架间（inter-framework）资源管理。框架内资源管理负责框架自身内部各组件之间的资源分配，由框架负载驱动，通常会为了最小化框架整体需求或降低运行成本而关闭不需要的资源。框架间资源管理负责大数据系统多个存储框架和处理框架之间的资源调度和优化管理，通常包括管理框架的资源请求、监控框架资源使用，以及在某些情况下对申请使用资源的应用队列进行管理。特别的，针对大数据系统负载多变、用户多样、规模较大的特点，应采用更加经济有效的资源构架和管理方案。目前的大数据软件框架，其亮点在于高可扩展性，而本质诉求仍然是如何实现并行化，即对数据进行分片、并为每一个分片分配相应的本地计算资源。因此，对于基础架构而言，为了支持大数据软件框架，最直接的实现方式就是将一份计算资源和一份存储资源进行绑定，构成一个资源单位（如，服务器），以获得尽可能高的本地数据访问性能。但是，这种基础架构由于计算同存储之间紧耦合且比例固定，逐渐暴露出资源利用率低、重构时灵活性差等问题。因此，未来应通过硬件及软件各方面的技术创

新，在保证本地数据访问性能的同时，实现计算与存储资源之间的松耦合，即：可以按需调配整个大数据系统中的资源比例，及时适应当前业务对计算和存储的真实需要；同时，可以对系统的计算部分进行快速切换，真正满足数据技术（DT）时代对“以数据为中心、按需投入计算”的业务要求。

4.2.5 数据消费者

数据消费者通过调用大数据应用提供者提供的接口按需访问信息，与其产生可视的，事后可查的交互。与数据提供者类似，数据消费者可以是终端用户或者其它应用系统。数据消费者执行的活动通常包括搜索/检索、下载、本地分析、生成报告、可视化等。数据消费者利用大数据应用提供者提供的界面或服务访问他感兴趣的信息，这些界面包括数据报表、数据检索、数据渲染等。

数据消费者角色也会通过数据访问活动与大数据应用提供者交互，执行其提供的数据分析和可视化功能。交互可以是基于需要（demand-based）的，包括交互式可视化、创建报告，或者利用大数据提供者提供的商务智能（BI）工具对数据进行钻取（drill-down）操作等。交互功能也可以是基于流处理（streaming-based）或推（push-based）机制的，这种情况下消费者只需要订阅大数据应用系统的输出即可。

4.2.6 安全和隐私

4.2.6.1 针对安全和隐私的考虑

大数据参考架构图中的五个主要角色，即系统协调者、数据提供者、大数据框架提供者、大数据应用提供者、数据消费者，其活动都要受到安全和隐私角色的影响，因此要求各个角色在各自的安全和隐私管理领域，通过不同的技术手段和安全措施，构筑大数据系统全方位、立体的安全防护体系，同时应提供一个合理的灾备框架，提升灾备恢复能力，实现数据的实时异地容灾功能，最终满足不同的安全和隐私要求。

因此，如何平衡大数据与隐私之间的关系，已成为各国立法、司法、执法部

门，以及企业必须面临与思考的共同难题。

4.2.6.2 主要角色安全与隐私的基本要求

1) 系统协调者

在大数据体系中，系统协调者的主要职责在于垂直系统构建，因此从安全和隐私角度出发，系统协调者实质上是安全隐私保障的总体规划者，应从上至下地制定系统性、完善的安全隐私保障制度。虽然系统协调者与大数据之间具体、直接的接触相对较少，但其宏观调配功能相对突出，涉及到大数据系统中所有数据的指派和调配；由于后续的处理分析涉及到相应操作主体和人员对于所原始数据的接触，会产生数据泄露和灭失等风险，这要求系统协调者应当设置完备的权限管理，通过明确的权限授权保证相应操作的安全，达到业务治理调配和风险控制相协调的目的。

2) 数据提供者

考虑到数据提供者的主要活动是将数据和信息引入到大数据系统中，供大数据系统发现、访问和转换，而这些活动均是数据使用过程中尤为关键的环节，无疑对数据提供者提出了相当高的安全与隐私要求。鉴于数据引入的过程是大数据产业链的源头，数据的真实性和有效性将对最终的数据价值产生根本影响，因而数据提供者在此过程中应注意对数据来源进行甄别和验证，保证数据的合法性、真实性和有效性。

如上所述，数据提供者通常还会通过创建数据源的元数据及访问策略、访问控制、通过软件的可编程接口实现推或拉式的数据访问、发布数据可用及访问方法等，由于数据将从数据提供者的储存介质输入到特定的大数据系统之中，这种传输过程的保密性、可靠性需求也要求数据提供者采取必要的保障制度和措施来确保数据提供过程的安全性。

3) 大数据应用提供者

大数据应用提供者的主要活动是根据系统协调者规定的要求执行特定操作，因此大数据应用提供者某种程度上职责更接近于系统协调者的具体执行人。因而从安全与隐私的要求上讲，大数据应用提供者应当切实地根据系统协调者制定的统一要求执行操作，按照既定的安全与隐私规程进行数据的收集、预处理、分析、

可视化和访问等活动，而相关的规程应当依据法律法规、国家与行业标准、行业规范等具有较高参考意义的文本进行制定。

此外，由于大数据应用提供者还会构造出特定结构的大数据应用系统，因此在数据的收集、传输、处理等过程当中进行符合要求的加密等技术保障，以确保数据在整个使用过程中的安全。

4) 大数据框架提供者

数据提供者将数据输入大数据系统后，大数据框架提供者将在一定程度上承担着对数据的存储与管理的义务，这要求其必须保证的数据安全性、可用性、完整性。并使用已有的技术手段来确保所储存数据的私密和安全。在落实其基础设施、平台、处理框架、信息交互/通信和资源管理的活动中，大数据框架提供者更应当着重维护数据的物理安全，即以自身的安全技术措施确保在相关活动中，数据不会遭受泄露、毁损、灭失等安全事件。

此外，在预防外部因素对数据安全构成危险外，大数据框架提供者还应在产品与设备供应链上防范类似的风险，这要求大数据框架提供者更注重用以为大数据服务的产品与设备的安全性能，以便从产品与设备的源头隔绝可能存在不安全因素。

5) 数据消费者

数据消费者执行的活动主要是对大数据应用分析和处理后的数据进行使用。这种数据使用过程，也要求数据消费者应当满足相应的安全与隐私要求，例如数据消费者应当对计算结果进行评估，评估其有效性、可用性，如必要则进行相应的优化调整，以防分析结果对现实中的个人主体或企业主体的基本权益产生显著影响。

同时，任何去标识化和匿名化都可能因技术不成熟存在一定漏洞与缺陷，因此应用阶段也有必要采取一定的安全技术措施与策略，如访问控制、认证授权等对分析结果进行保护，一方面保护具有价值的大数据分析成果，另一方面也防止分析结果泄露，或遭反向追溯引发数据与隐私泄露的风险。

4.2.7 管理角色

管理角色包括二个活动组：系统管理和大数据生命周期管理。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。大数据生命周期管理涵盖了大数据生命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。

由于大数据基础设施的分布式和复杂性，系统管理依赖于两点：使用标准的协议把资源状态和出错信息传送给管理组件；通过可部署的代理或管理连接子（connector）允许管理角色监视甚至控制大数据处理框架元素。系统管理的功能是监视各种计算资源的运行状况，应对出现的性能或故障事件，从而能够满足大数据应用提供者的服务质量（QoS）需求。在云服务提供商提供能力管理接口时，通过管理连接子对云基础设施提供的自助服务、自我调整、自我修复等能力进行利用和管理。大型基础设施通常包括数以千计的计算和存储节点，因此应用程序和工具的调配应尽可能自动化。软件安装、应用配置以及补丁维护也应该以自动的方式推送到各结点并实现自动地跨结点复制。还可以利用虚拟化技术的虚拟映像，加快恢复进程和提供有效的系统修补，以最大限度地减少定期维护时的停机时间。系统管理模块应能够提供统一的运维管理，能够对包括数据中心、基础硬件、平台软件（存储、计算）和应用软件进行集中运维、统一管理，实现安装部署、参数配置、系统监控等功能。应提供自动化运维的能力，通过对多个数据中心的资源进行统一管理，合理的分配和调度业务所需要的资源，做到自动化按需分配。同时提供对多个数据中心的 IT 基础设施进行集中运维的能力，自动化监控数据中心内各种 IT 设备的事件、告警、性能，实现从业务维度来进行运维的能力。

大数据生命周期管理活动负责验证数据在生命周期中的每个过程是否都能够被大数据系统正确地处理，它覆盖了数据从数据提供者那里被摄取到系统，一直到数据被处理或从系统中删除的整个生命周期。由于大数据生命周期管理的任务可以分布在大数据计算环境中的不同组织和个体，从遵循政策、法规和安全要求的视角，大数据生命周期管理包括以下活动或功能：政策管理（数据迁移及处

置策略)、元数据管理(管理数据标识、质量、访问权限等元数据信息)、可访问管理(依据时间改变数据的可访问性)、数据恢复(灾难或系统出错时对数据进行恢复)、保护管理(维护数据完整性)。从大数据系统要应对大数据的 4V 特征来看,大数据生命周期管理活动和功能还包括与系统协调者、数据提供者、大数据框架提供者、大数据应用提供者、数据消费者以及安全和隐私角色之间的交互。

大数据管理需要提供大规模集群统一的运维管理系统,能够对包括数据中心、基础硬件、平台软件和应用软件进行集中运维、统一管理,实现安装部署、参数配置、监控、告警、用户管理、权限管理、审计、服务管理、健康检查、问题定位、升级和补丁等功能。

大数据管理需要具有自动化运维的能力,通过对多个数据中心的资源进行统一管理,合理的分配和调度业务所需要的资源,做到自动化按需分配。同时提供对多个数据中心的信息技术基础设施进行集中运维的能力,自动化监控数据中心内各种信息技术设备的事件、告警、性能,实现从业务维度来进行运维的能力。此外,对主管理系统节点及所有业务组件中心管理节点实现高可靠性的双机机制,采用主备或负荷分担配置,避免单点故障场景对系统可靠性的影响。

5 大数据标准化现状

5.1 国外标准化现状

5.1.1 ISO/IEC JTC 1

1) ISO/IEC JTC1/SC32 数据管理和交换分技术委员会

ISO/IEC JTC1/SC32 数据管理和交换分技术委员会（以下简称 SC32）是与大数据关系最为密切的标准化组织。SC32 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，其标准化技术内容涵盖：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。SC32 下设 WG1 电子业务工作组、WG2 元数据工作组、WG3 数据库语言工作组、WG4SQL 多媒体和应用包工作组。

2014 年 6 月 SC32 在北京全会上，批准了 4 项为大数据提供标准化支持的新工作项：国际标准《SQL 对多维数组的支持》、技术报告《SQL 对 JSON 的支持》、国际标准《数据集注册元模型》和国际标准《数据源注册元模型》。其中《SQL 对 JSON 的支持》由中国专家担任编辑。

SC32 在 2014 年北京全会期间举办了主题为“大数据标准化”的开放论坛，为国内外大数据领域的专家学者和产业管理部门人员、IT 界的骨干企业提供了一个开放交流的平台。来自于国内外大数据研究、应用及服务提供领域的专家学者做了相关主题报告，展现了当前大数据技术与标准的发展和应用前景。

2015 年 5 月 SC32 在英国全会上，批准了 2 项为大数据提供标准化支持的新工作项：技术报告《SQL 对多态表功能的支持》和技术报告《SQL 对多维数组的支持》。此外，本次会议上我国提案“SQL 对 MapReduce 及与之相关的流数据处理的支持”得到 SC32 专家的高度肯定，WG3 会议召集人将组织国际专家参与我国工作，共同完善该提案。

2016年6月13日-17日，ISO/IEC JTC1/SC32“数据管理与交换”分技术委员会2016年全会在美国孟菲斯召开，本次会议上，中国代表团汇报了“SQL对MapReduce及与之相关的流数据处理的支持”提案内容，并胜利通过了SC32全会决议，确认以WG3“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3日SC32秘书处发布的2016-2017年的业务计划中明确了新工作项ISO/IEC TR 29075-1: Design Notes for New Database Language Technologies, Part 1: SQL Support for Streaming Data。

SC32现有的标准制定和研究工作为大数据的发展提供了良好基础。

2) ISO/IEC JTC1/WG9 大数据工作组工作情况

ISO/IEC JTC1于2013年11月全会上成立负责大数据国际标准化的大数据研究组（ISO/IEC JTC1/SG2，以下简称SG2）。工作重点包括：调研国际标准化组织（ISO）、国际电工委员会（IEC）、第1联合技术委员会（ISO/IEC JTC1）等在大数据领域的关键技术、参考模型以及用例等标准基础；确定大数据领域应用需要的术语与定义；评估分析当前大数据标准的具体需求，提出ISO/IEC JTC1大数据标准优先顺序；向2014年ISO/IEC JTC1全会提交大数据建议的技术报告和其他研究成果。

2014年11月，SG2向ISO/IEC JTC1全会提交了研究报告，其中包括建议成立独立的ISO/IEC JTC1大数据工作组，需要标准化的大数据技术点。ISO/IEC JTC1于此次全会上成立了ISO/IEC JTC1/WG9大数据工作组（以下简称WG9）。

为了推动大数据标准的研制工作，WG9于2014年成立以来召开了八次会议。

WG9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JTC1外其他大数据相关标准组织保持联络关系。目前，WG9正在研制6项国际标准，相关标准状态见表6。

表 6 制定参与国际标准情况

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	Information technology - Big Data - Overview and Vocabulary	信息技术 大数据 概述和术语	CD
2	ISO/IEC TR20547-1	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	CD
3	ISO/IEC TR20547-2	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：用例和需求	CD
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	CD
5	ISO/IEC 20547-4	Information technology - Big Data Reference Architecture - Part 4: Security and Privacy Fabric	信息技术 大数据参考架构 第 4 部分：安全和隐私	WD
6	ISO/IEC TR20547-5	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	CD

5.1.2 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和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，大数据发展面临的巨大挑战包括：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）、SG16（第 16 研究组）、SG17（第 17 研究组）以及 SG20（第 20 研究组）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。

此外，SG16 下设的 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；SG17 下设的 Q7/17（第 7 课题组）正在开展 X.srfb “移动互联网中大数据分析的安全需求和框架”以及 Q8/17(第 8 课题组)正在开展 X.GSBDaaS “大数据即服务安全指南”等标准研制；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。

5.1.3 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作出如何存储、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）。

5.1.4 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）对大数据的发展和應用，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。

5.2 国内标准化现状

5.2.1 国家标准现状

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（以下简称“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

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/WG9 大数据工作组。

工作组组长由北京理工大学副校长梅宏院士担任，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、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。秘书长为中国电子技术标准化研究院信息技术研究中心副主任吴东亚。联络员为国家标准化管理委员会工业二部刘大山处长、工业和信息化部信软司傅永宝调研员和工业和信息化部电子信息司侯建仁处长。

根据大数据产业发展现状和标准化需求，为更好的开展相关标准化工作，2017年7月工作组在第二届组长会议上决议下设7个专题组，包括：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，负责大数据领域不同方向的标准化工作。目前，工作组已发布6项国家标准，3项国家标准正在报批阶段，15项国家标准正在研制，详见表7。

表7 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 34945-2017	信息技术 数据溯源描述模型	发布	技术专题组
5	GB/T 35294-2017	信息技术 科学数据引用	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	20141200-T-469	信息技术 数据交易服务平台 交易数据描述	报批	总体专题组
8	20141201-T-469	信息技术 数据交易服务平台 通用功能要求	通过评审	总体专题组
9	20141203-T-469	信息技术 数据质量评价指标	报批	技术专题组
10	20141204-T-469	信息技术 通用数据导入接口规范	报批	产品平台 专题组
11	20160597-T-469	信息技术 大数据 分析系统基本功能要求	征求意见	产品平台 专题组
12	20160598-T-469	信息技术 大数据 存储与处理平台技术要求	草案	产品平台 专题组
13	20171083-T-469	信息技术 大数据 基于参考架构下的接口框架	草案框架	总体专题组
14	20171082-T-469	信息技术 大数据 分类指南	草案框架	技术专题组
15	20171082-T-469	信息技术 大数据 系统通用规范	草案	总体专题组
16	20171081-T-469	信息技术 大数据 存储与处理系统功能测试规范	草案框架	产品平台 专题组
17	20171065-T-469	信息技术 大数据 分析系统功能测试规范	草案框架	产品平台 专题组
18	20171066-T-469	信息技术 大数据 面向应用的基础计算平台基本性能要求	草案框架	产品平台 专题组
19	20171067-T-469	信息技术 大数据 开放共享 第1部分：总则	草案	总体专题组
20	20171068-T-469	信息技术 大数据 开放共享 第2部分：政府数据开放共享基本要求	草案	总体专题组

21	20171069-T-469	信息技术 大数据 开放共享 第3部分： 开放程度评价	草案	总体专题组
22	20173818-T-469	信息技术 大数据 系统运维和管理功 能要求	草案框架	产品平台 专题组
23	20173819-T-469	信息技术 大数据 工业应用参考架构	草案框架	工业大数据 专题组
24	20173820-T-469	信息技术 大数据 产品要素基本要求	草案框架	工业大数据 专题组

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 WG9 和 SC32 相关工作。此外，工作组还重点关注 NIST NBD-PWG 大数据公共工作组，同时，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

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。目前，工作组正在研制的国家标准有 13 项，其中 2016 年在研标准《大数据服务安全能力要求》和《个人信息安全规范》进入报批稿阶段，《大数据安全管理指南》进入送审稿阶段；2017 年在研标准《数据安全能力成熟度模型》、《数据交易服务安全要求》和《个人信息去标识化指南》进入送审稿阶段，《数据出境安全评估指南》处于征求意见稿阶段，《个人信息安全影响评估指南》处于草案阶段；2017 年启动了《大数据基础软件安全技术要求》、《数据安全分类分级实施指南》、《大数据业务安全风险控制实施指南》、《区块链安全技术标准研究》等标准研究项目。

5.2.2 行业及地方标准现状

中国通信标准化协会（英文译名为：China Communications Standards Association，缩写为：CCSA）是国内开展通信技术领域标准化活动的非营利性法人社会团体。目前该协会有 TC1 WG6 工作组专门从事大数据方面的标准化工作，重点研究大数据技术产品标准化，数据资产管理制度、工具，数据开放与流通交易相关等方面的标准规范。除此之外还有 TC8 下的多个工作组也在开展大数据安

全方面的标准规范研究。目前已经有多项电信互联网大数据管理、大数据处理、大数据平台测试等方面的标准规范正在编制过程中。

各地发展大数据产业各有特色，上海市、广东省、湖北省、山东省、贵州省、四川省、陕西省、江苏省、内蒙古自治区等地方形成 30 余项地方标准，主要集中于资源开放共享、政务大数据领域、重点行业等。如贵州省出台的《政府数据数据分类分级指南》、《政府数据数据脱敏工作指南》、《贵州省政府数据 第 1 部分：元数据》等；成都就数据采集、数据共享、数据开放和安全方面制定了项目标准，目前正在修订四川省（区域性）地方标准《成都市政务信息资源交换标准体系》；湖北省发布了《政务数据服务度量计价规范》；陕西省也在平台、应用、管理、隐私等方面开展大数据标准体系的建设工作，并重点在气象、铁路、车联网、城市运行管理等行业应用方面组织大数据标准研究。

6 大数据标准体系

6.1 大数据标准体系框架

结合国内外大数据标准化情况、国内大数据技术发展现状、大数据参考架构及标准化需求，根据数据全周期管理，数据自身标准化特点，当前各领域推动大数据应用的初步实践，以及未来大数据发展的趋势，提出了大数据标准体系框架，如下图所示。

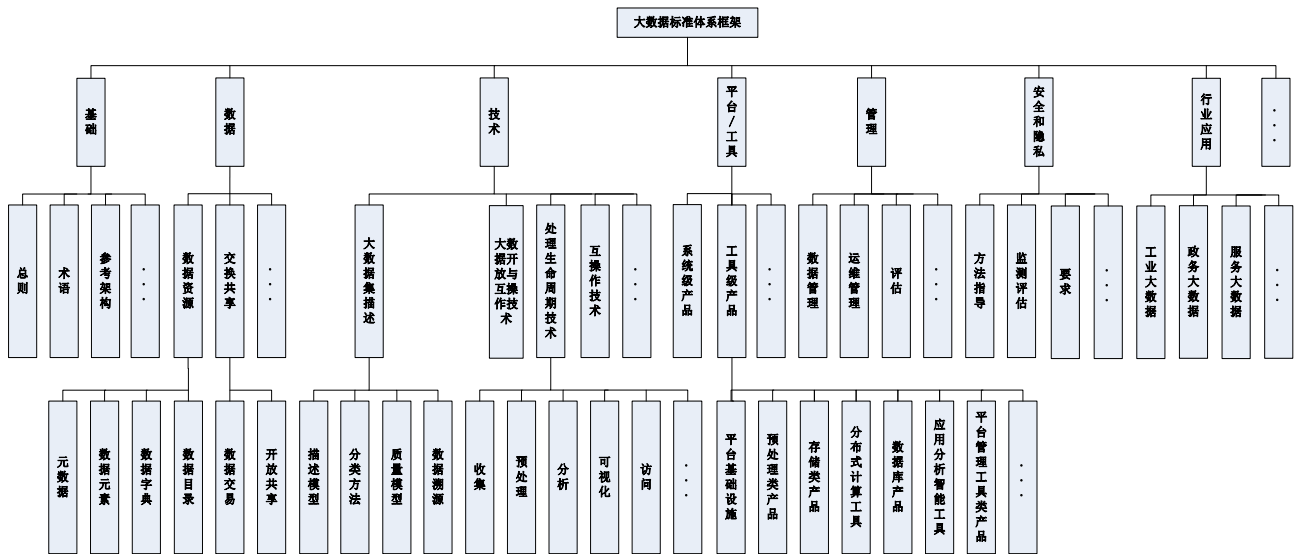


图 4 大数据标准体系框架示意图

大数据标准体系由七个类别的标准组成，分别为：基础标准、数据标准、技术标准、平台和工具标准、管理标准、安全和隐私标准、行业应用标准。

1) 基础标准

为整个标准体系提供包括总则、术语、参考模型等基础性标准。

2) 数据标准

该类标准主要针对底层数据相关要素进行规范。包括数据资源和数据交换共享两部分，其中数据资源包括元数据、数据元素、数据字典和数据目录等，数据交换共享包括数据交易和数据开放共享相关标准。

3) 技术标准

该类标准主要针对大数据相关技术进行规范。包括大数据集描述及评估、大数据处理生命周期技术、大数据开放与互操作、面向领域的大数据技术四类标准。

其中，大数据集描述及评估标准主要针对多样化、差异化、异构异质的不同类型数据建立标准的度量方法，以衡量数据质量，同时研究标准化的方法对多模态的数据进行归一处理，并根据我国国情，制定相应的开放数据标准，以促进政府数据资源的建设。大数据处理生命周期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，包括数据产生、数据获取、数据存储、数据分析、数据展现、数据安全与隐私管理等阶段的标准制定。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的标准化进行研制。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) 平台和工具标准

该类标准主要针对大数据相关平台和工具进行规范，包括系统级产品和工具级产品两类，其中系统级产品包括(实时计算产品(流处理)，数据仓库产品(OLTP)，数据集市产品(OLAP)，数据挖掘产品，全文检索产品，非结构化数据存储检索产品，图计算和图检索产品等)；工具级产品包括平台基础设施、预处理类产品、存储类产品、分布式计算工具、数据库产品、应用分析智能工具、平台管理工具类产品的技术、功能、接口等进行规范。相应的测试规范针对相关产品和平台给出测试方法和要求。

5) 管理标准

管理标准作为数据标准的支撑体系，贯穿于数据生命周期的各个阶段。该部分主要是数据管理、运维管理和评估三个层次进行规范。其中数据管理标准主要包括数据管理能力模型、数据资产管理以及大数据生命周期中处理过程的管理规范；运维管理主要包含大数据系统管理及相关产品等方面的运维及服务等方面的标准；评估标准包括设计大数据解决方案评估、数据管理能力成熟度评估等。

6) 安全和隐私标准

数据安全和隐私保护作为数据标准体系的重要部分，贯穿于整个数据生命周期的各个阶段。大数据应用场景下，大数据的4V特性导致大数据安全标准除了关注传统的数据安全和系统安全外，还应在基础软件安全、交易服务安全、数据分类分级、安全风险控制、电子货币安全、个人信息安全、安全能力成熟度等方

向进行规范。

7) 行业应用标准

行业应用类标准主要是针对大数据为各个行业所能提供的服务角度出发制定的规范。该类标准指的是各领域根据其领域特性产生的专用数据标准，包括工业、政务、服务等领域。

6.2 标准明细表

根据大数据标准体系框架，整理出已发布、已报批、已立项、已申报、在研以及拟研制的大数据相关国家标准 104 项，如下表所列：

表 8 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	总则		信息技术 大数据标准化指南		拟研制	
2.		术语	GB/T 35295-2017	信息技术 大数据 术语		已发布	
3.		参考架构		GB/T 35589-2017	信息技术 大数据 技术参考模型		已发布
4.					信息技术 大数据参考架构 第 1 部分：框架和应用指南		拟研制
5.					信息技术 大数据参考架构 第 2 部分：用例和需求		拟研制
6.					信息技术 大数据参考架构 第 5 部分：标准路线图		拟研制
7.				20171083-T-469	信息技术 大数据 基于参考架构下的接口框架		在研
8.	数据	数据资源	GB/T 18142-2000	信息技术 数据元素值格式记法	ISO/IEC 14957: 1996, IDT	已发布	
9.			20101507-T-469	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	在研	
10.			GB/T18391.1-2009	信息技术 元数据注册系统(MDR) 第 1 部分：框架	ISO/IEC11179-1: 2004, IDT	已发布	
11.			GB/T18391.2-2009	信息技术 元数据注册系统(MDR) 第 2 部分：分类	ISO/IEC11179-2: 2005, IDT	已发布	
12.			GB/T18391.3-2009	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	已发布	
13.			GB/T18391.4-2009	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	已发布	
14.			GB/T18391.5-2009	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	已发布	

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
15.			GB/T18391.6-2009	信息技术 元数据注册系统(MDR) 第6部分:注册	ISO/IEC11179-6: 2005, IDT	已发布
16.			GB/Z 21025-2007	XML 使用指南		已发布
17.			GB/T23824.1-2009	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/IECTR2094 3-1: 2003, IDT	已发布
18.			GB/T23824.3-2009	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	已发布
19.			GB/T32392.1-2015	信息技术 互操作性元模型框架(MFI) 第1部分: 参考模型		已发布
20.			GB/T32392.2-2015	信息技术 互操作性元模型框架(MFI) 第2部分: 核心模型		已发布
21.			GB/T32392.3-2015	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		已发布
22.			GB/T32392.4-2015	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		已发布
23.			20132340-T-469	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		在研
24.			20132341-T-469	信息技术 互操作性元模型框架 第7部分: 服务模型注册元模型		在研
25.			20132342-T-469	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		在研
26.			20132343-T-469	信息技术 互操作性元模型框架 第9部分: 按需模型选择		在研
27.			GB/T 30881-2014	信息技术 元数据注册系统(MDR) 模块	ISO/IEC 19773:2011	已发布
28.			GB/T 30880-2014	信息技术 通用逻辑(CL): 基于逻辑的语言族框架	ISO/IEC 24707:2007	已发布
29.			2010-3325T-SJ	信息技术 元数据属性		在研
30.		交换共享		信息技术 大数据 开放数据集基本要求		拟研制
31.				信息技术 大数据 开放数据集标识管理		拟研制
32.			20171067-T-469	信息技术 大数据 开放共享 第1部分: 总则		在研
33.			20171068-T-469	信息技术 大数据 开放共享 第2部分: 政府数据开放共享基本技术要求		在研
34.			20171069-T-469	信息技术 大数据 开放共享 第3部分: 开放程度评价		在研
35.				信息技术 大数据 开放共享 第4部分: 政府资源目录体系		拟研制
36.			20141201-T-469	信息技术 数据交易服务平台 通用功能要求		在研
37.			20141200-T-469	信息技术 数据交易服务平台 交易数据描述		在研
38.				信息技术 数据交易 通用概念描述		拟研制
39.				信息技术 数据交易 交易流程描述		拟研制
40.				信息技术 数据交易 数据管理规范		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	
41.				信息技术 数据交易 技术规范		拟研制	
42.				信息技术 数据交易 风险评估		拟研制	
43.				信息技术 数据交易 交易质量评估		拟研制	
44.				信息技术 数据交易 数据价值评估指引		拟研制	
45.		大数据集描述	GB/T 34952-2017	多媒体数据语义描述要求		已发布	
46.			20171082-T-469	信息技术 大数据 分类指南		在研	
47.			20141203-T-469	信息技术 数据质量评价指标		在研	
48.				信息技术 数据质量检测		拟研制	
49.			GB/T 35294-2017	信息技术 科学数据引用		已发布	
50.			GB/T 34945-2017	信息技术 数据溯源描述模型		已发布	
51.		处理生命周期技术	20141204-T-469	信息技术 通用数据导入接口规范		在研	
52.				信息技术 通用数据导入接口测试规范		拟研制	
53.				信息技术 大数据 分析总体技术要求		拟研制	
54.				信息技术 大数据 可视化工具通用要求		拟研制	
55.			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	已发布	
56.		互操作技术		信息技术 大数据 互操作技术指南		拟研制	
57.		平台和工具	系统级产品	20160598-T-469	信息技术 大数据 存储与处理平台技术要求		在研
58.				20171081-T-469	信息技术 大数据 存储与处理系统功能测试规范		在研
59.				20160597-T-469	信息技术 大数据 分析系统基本功能要求		在研
60.				20171065-T-469	信息技术 大数据 分析系统功能测试规范		在研
61.	20171082-T-469			信息技术 大数据 系统通用规范		在研	
62.	工具级产品		20171066-T-469	信息技术 大数据 面向应用的基础计算平台基本性能要求		在研	
63.			GB/T 28821-1012	关系数据管理系统技术要求		已发布	
64.			GB/T 30994-2014	关系数据库管理系统检测规范		已发布	
65.			GB/T 32633-2016	分布式关系数据库服务接口规范		已发布	
66.			20121409-T-469	非结构化数据表示规范		在研	
67.			20121410-T-469	非结构化数据访问接口规范		在研	
68.			GB/T	非结构化数据管理系统技术要求		已发布	

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	
			32630-2016				
69.			20141183-T-469	实时数据库通用接口规范		在研	
70.				非结构化数据查询语言		拟研制	
71.				智能硬件通用大数据接口规范		拟研制	
72.	管理	数据管理		信息技术 大数据 资产管理指南		拟研制	
73.		运维管理	20173818-T-469	信息技术 大数据 系统运维和管理功能要求		已立项	
74.		评估		信息技术 大数据 解决方案基本评估规范		拟研制	
75.			GB/T 36073-2018	数据管理能力成熟度评估模型		已发布	
76.	大数据安全和隐私	要求	GB/T 20009-2005	信息安全技术 数据库管理系统安全评估准则		已发布	
77.			GB/T 20273-2006	信息安全技术 数据库管理系统安全技术要求		已发布	
78.			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	已发布	
79.			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	已发布	
80.			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	已发布	
81.					信息安全技术 大数据参考架构 第4部分: 安全和隐私		拟研制
82.					信息安全技术 大数据安全分级指南		拟研制
83.					信息安全技术 大数据安全参考架构		拟研制
84.					信息安全技术 数据脱敏指南		拟研制
85.					信息安全技术 大数据平台安全技术要求		拟研制
86.					信息安全技术 大数据跨集群安全技术框架		拟研制
87.				20130323-T-469	信息安全技术 个人信息保护管理要求		在研
88.				20130338-T-469	信息安全技术 移动智能终端个人信息保护技术要求		在研
89.			检测评估		信息安全技术 隐私保护评估方法		拟研制
90.			方法指导		信息安全技术 大数据中的隐私保护框架		拟研制
91.					信息安全技术 个人信息保护指南		在研
92.		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		已发布	
93.		行业应用	工业大数据		信息技术 大数据 工业应用术语		拟研制
94.				20173819-T-469	信息技术 大数据 工业应用参考架构		已立项
95.				20173820-T-469	信息技术 大数据 产品要素基本要求		已立项
96.				信息技术 工业大数据 工业订单元数据规范		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
97.			20170057-T-469	智能制造 对象标识要求		已立项
98.			20173805-T-339	智能制造 制造对象标识解析体系应用指南		已立项
99.		政务大数据		信息技术 电子商务大数据 采集规范		拟研制
100.				信息技术 电子商务大数据 仓库模型规范		拟研制
101.				信息技术 电子商务大数据 应用指标体系		拟研制
102.		服务大数据		信息技术 服务大数据 运维服务元数据		拟研制
103.				信息技术 服务大数据 教育行业督导平台技术规范		拟研制
104.				信息技术 服务大数据 电力行业运行数据运维技术规范		拟研制

通过对现有大数据国家标准进行分析可以看出：

1) 在数据资源方面，我国已具备一定的标准基础，相关国家标准在大数据领域下同样适用。下一步需要根据大数据技术、产业现状适时修订已有国家标准，保证标准紧跟技术、产业发展；同时推进相关数据资源标准的推广与应用，保证标准的落地实施。

2) 在交换共享方面，依托全国信标委大数据标准工作组，已经开展了相关标准的研制工作：发布数据交易国家标准 2 项，在研开放共享国家标准 3 项。下一步需要推进相关数据开放共享国家标准的报批工作；同时围绕国家对于政府数据开放共享的任务要求，加快适用于政府数据开放共享的国家标准研制。

3) 在数据管理方面，GB/T 36073-2018《数据管理能力成熟度评估模型》作为我国首个数据管理领域的国家标准已经发布。下一步急需从数据管理能力评估方法的角度开展标准化研究，落实标准在产业中的应用，推动整个行业数据管理能力的提升。

4) 在平台和工具标准方面，目前已立项《信息技术 大数据 系统通用规范》（计划号：20171082-T-469）等 5 项国家标准，标准范围覆盖大数据通用系统、大数据存储与处理系统、大数据分析系统。下一步需要围绕大数据系统功能模块，完善数据收集、数据访问等功能模块的相关标准研制；同时围绕国家标准，开展大数据系统产品的标准符合性测试评估工作。

5) 在工业大数据领域，主要开展了参考架构等基础标准的制定。下一步将

开展工业大数据平台、工业大数据系统测试、工业大数据管理、工业大数据采集与存储等方面的国家标准研制工作，通过工业大数据的标准化，全面支撑我国智能制造、工业互联网建设。

总体来说，目前我国在大数据领域在基础术语、数据资源、交换共享、数据管理、大数据系统产品、工业大数据等方面已开展了国家标准研制工作。下一步需要加强大数据已有国家标准的推广应用，开展标准试点验证；同时深入调研大数据在各个行业中的标准化需求，开展工业大数据、政务大数据等领域的标准研制，全面推进大数据标准在各个行业中的支撑引领作用。

6.3 重点标准描述

大数据标准工作组自主研制的一些重点标准已得到试验验证，并推广应用。

GB/T 36073-2018 《数据管理能力成熟度评估模型》国家标准规定了组织进行数据管理、评价的能力成熟度模型，包含了数据战略、数据治理、数据架构、数据应用、数据安全、数据质量管理、数据标准、数据生命周期管理共 8 个关键过程域，描述了每个过程域的建设目标和度量标准，可以作为组织进行数据管理工作的参考模型。适用于数据拥有方对数据进行规划、管理和能力提升方面的规范和指导，同时也适用于数据应用方案提供方进行数据相关解决方案的建设以及人员的培养等。该标准研制过程在浙江移动、天津天臣、广东佛山、鹏博士集团等单位进行了充分验证，目前该标准已经在贵州、上海以及电力、通信、金融等地方和行业进行推广应用，有力提升了地方和行业的数据管理水平。

《信息技术 大数据 系统通用规范》（计划号：20171082-T-469）国家标准规定了大数据系统的技术要求、服务要求和测试规程，适用于各类大数据系统，可作为大数据系统设计、选型、验收、检测和依据。已经在华为、阿里云、百分点、海康威视、新华三、中兴、南大通用等企业开展的试验验证和试点示范，帮助提升企业大数据产品的功能，依据该标准，国家批复成立了国家大数据系统产品质量监督检验中心。

针对大数据开放共享，开展了《信息技术 大数据 开放共享 第 1 部分：总则》（计划号：20171067-T-469）、《信息技术 大数据 开放共享 第 2 部分：政府

《数据开放共享基本要求》（计划号：20171068-T-469）、《信息技术 大数据开放共享 第3部分：开放程度评价》（计划号：20171069-T-469）3项国家标准研制工作，将进一步加大政府信息公开和数据开放力度，形成政府信息与社会信息交互融合的大数据资源，下一步将在全国各地进行试验验证和推广应用。

面向工业大数据领域，依据《智能制造 对象标识要求》（计划号：20170057-T-469）、《智能制造 制造对象标识解析体系应用指南》（计划号：20173805-T-339）标准开发了工业大数据OID数据标识管理系统，开展了应用试点，初步建立了工业大数据和标识服务平台；解决了企业数据利用率低、跨企业数据采集与共享技术不成熟、供应链上下游企业之间的数据集成共享困难等问题。

7 我国大数据标准化工作建议

7.1 完善大数据标准体系建设

当前，随着大数据技术的不断突破，产业应用的逐步深入，不断涌现出新的标准化需求。需要结合国内外大数据技术和产业现状、大数据标准化现状，以“基础统领、应用牵引”为原则，不断提炼梳理新增的大数据标准化需求，在现有大数据标准体系的基础上适时修改完善，保证大数据标准体系协调一致、布局合理，紧跟产业发展，始终保持大数据标准对大数据领域技术、产业创新发展的推动作用。

7.2 加强大数据标准的宣传推广工作

大数据领域的相关标准需要全社会的广泛参与，同时标准化的推进程度和应用深度对区域中行业产业发展的健康性和持续性有着积极的促进作用。随着多个在编大数据标准的相继发布和试验验证工作的推进，需要各级政府充分发挥引导作用，加强大数据标准的推广宣贯工作，让更多行业的相关企业和个人参与到标准的意见征集、论证应用当中，深刻理解大数据标准化的发展现状、未来规划以及对产业发展的深远意义。

7.3 加强大数据重点标准应用示范工作

标准的制定归根到底要落实到产业应用中，需要针对重点行业、领域、地区，开展大数据标准试验验证和试点示范工作。建议围绕数据开放共享、大数据系统测试、数据管理能力成熟度评估、数据管理、工业大数据、数据安全等重点领域已有的标准基础，建立标准符合性评估体系，优先在大数据综合试验区和大数据产业集聚区建立标准示范基地，开展标准应用示范工作。同时强化标准对市场培育、服务能力提升和行业管理的支撑作用。指导第三方机构，建设重点标准应用公共服务平台，开展大数据标准符合性评估、测试工作。

7.4 加强大数据标准化在数据治理领域的推进作用

数据治理的核心目标是将数据作为政府及企业的核心资产进行应用和管理，合理的数据治理能够建立规范的数据应用标准，消除数据的不一致性，提高组织内部的数据质量，推动数据广泛共享，充分发挥数据对政府及企业的业务、管理、以及战略决策的重要作用。目前我国政府、企业的数据治理能力普遍不足，需要通过标准化的手段对政府、企业的数据治理能力提供指导和规范，不断扩大标准化在数据治理领域的广泛应用，促进组织完善数据治理机制，提升数据治理能力，加强组织间的数据交换共享，提升数据价值。

7.5 加快大数据标准化人才培养

大数据人才是大数据领域发展的核心资源。2018年3月21日，教育部公布了2017年度普通高等学校本科专业备案和审批结果，283所高校获批数据科学与大数据专业，这标志着我国大数据人才培养进入了具体实施阶段。标准作为技术、产业发展的顶层支撑，也急需加快相关人才培养。建议在全国范围内组织开展大数据标准宣贯培训活动，培养掌握大数据标准技术和标准实施方法论的专业人员；鼓励和支持行业协会、高等院校、科研院所设立标准化相关研究机构，大力培育标准化科研人才；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。

7.6 加快我国大数据标准的国际化步伐

建议依托全国信标委大数据标准工作组，组织标准化核心机构和重点企业，跟踪研究大数据相关国际标准化进展，深度参与国际标准制定工作，积极贡献国际标准提案，提升自主标准的国际化水平；支持相关单位参与国际标准化工作，努力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，提升国际话语权。

8 附件：成功案例

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	案例一：地市级城市政务信息资源标准化归集、交换和应用——徐州智慧城市信息资源枢纽
2	城市运营	案例二：昆明国家经济技术开发区城市智能运营中心（IOC）
3	金融业	案例三：“数控金融”互联网金融大数据监管平台应用案例
4	农业	案例四：农业部农产品批发市场价格挖掘及可视化平台
5	制造业	案例五：海尔 COSMO Plat 空调噪音大数据智能分析
6	制造业	案例六：长安汽车智能制造技术研究所冲压质量大数据项目
7	工业	案例七：江苏省重点领域共性技术攻关项目——工业大数据元数据规范与验证技术攻关专题——工业大数据标准体系研究及重点标准编制
8	电力	案例八：国网电力大数据应用案例
9	通信	案例九：浙江移动“天盾”反欺诈系统
10	邮政	案例十：国家邮政局数据管控系统
11	科研	案例十一：重庆两江大数据科技服务双创孵化基地案例
12	卫生	案例十二：本溪大健康服务平台
13	文化	案例十三：CCDI 版权监测案例

8.1 案例一：地市级城市政务信息资源标准化归集、交换和应用——徐州智慧城市信息资源枢纽

应用领域：政务信息资源梳理、政务信息资源共享、财政资金管理、社会保障

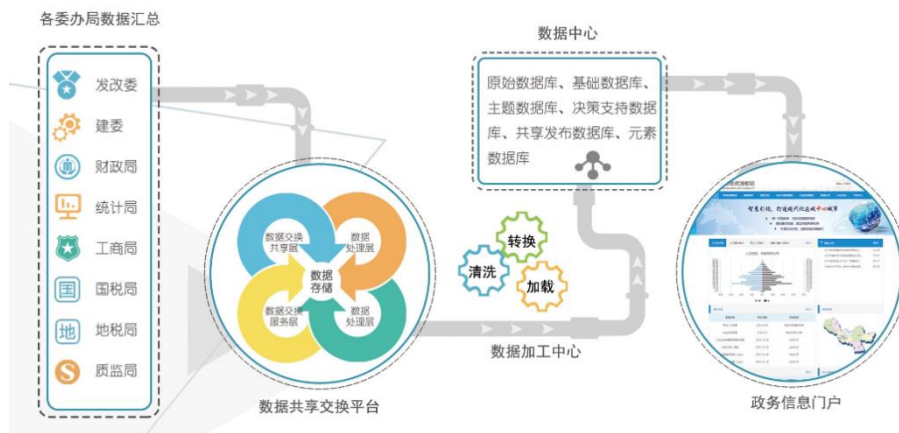
1、地市级政务信息资源标准化归集、交换、应用

徐州市政务信息资源枢纽工程以人口、法人、空间地理、宏观经济四大基础性信息资源库建设为核心，以信息资源共享交换平台建设为桥梁，以大数据技术为保障，搭建满足政务信息资源存储、传输、共享、应用、安全管理的运行支撑环境，建立统一的信息资源库和应用服务平台。该项目在地市级政务信息资源标准化归集、交换、应用方面进行了积极探索，具体体现在四个方面：

（1）统一标准平台化建设有效归集徐州政务信息数据资源

徐州市政务信息资源枢纽工程以平台化方式服务徐州市各市级政府部门及单位工作人员、信息化主管部门。通过数据交换服务、数据管理服务、数据监控

服务、信息资产管理服务、大数据分析展示服务，有效归集和使用人口基础信息库、法人基础信息库、自然资源与空间地理信息库、宏观经济基础信息库。



附图 1 案例一：整体业务逻辑图

该项目从信息资源的编目、管理、服务全流程实行统一标准和规范管理，截止 2017 年 7 月，政务信息资源枢纽已完成全市 46 家委办局的 432 类信息约 1.5 亿条数据的清洗、比对、入库工作；并累计向全市 10 家部门的 11 个业务系统提供了 369 类约 1.23 亿条数据支撑。其中人口数据 123 类约 8950.6 万条，法人数据 222 类约 3296.3 万条，空间地理数据 18 类 1478 项，宏观经济数据 35 类 3.15 万条，政务数据资源采集率达到 86%，基本实现政府部门业务数据资源共享，数据资源整合，为更深入的政府数据开放积累数据基础和标准规范。

(2) 大数据资产管理标准化应用提升了数据资源交换效率

智慧徐州信息资源枢纽项目以集中管理、高效维护为原则，制定一系列符合徐州政务信息资源资产“供”“需”双方要求的管理标准，对资源需求的申请、资源提供的流程做出清洗明确、有迹可循的监控节点；将资源的交换使用频率、资源的更新情况与各个资源的拥有部门的资产价值挂钩，有效提高各部门持有和共享本部门积累的政务信息资源的积极性，在节约沟通成本，提高绩效方面成效显著。

以人口、法人两部分数据需求为例，在信息资源枢纽建成以前，徐州有 21 个部门需要人口信息是，需要分别与公安和卫计委索取，按单次索取需要 5 个工作日计算，共需要 105 天；20 个部门需要法人登记信息，按单次索取需要 5 个工作日计算，共需要 100 天，而枢纽建成后，各部门此项工作可以在线申请，3 工作日完成数据传递，按照四库计算，每年在传输和交换基础信息方面，可以为

徐州各部门总体节约 164 天的工作量。

同时，基础信息实现集中存储和交换，按照一个部门平均需要 40 小时（5 个工作日）支持其他部门在获取自己拥有的数据资源来计算，徐州各部门可以总体节约 260（52*5）天的工作量，大大提高了各部门在信息资源积累和共享方面的工作效率。

（3）标准化管理提高了政务数据资源信息资源管理和服务水平

智慧徐州信息资源枢纽项目以人口、企业关系族谱为代表的大数据分析展示服务，形成一套包括实名认证服务、基础空间地理信息服务、数据核查服务、项检索服务（查单一结果）、专题检索服务（查一组专题数据）、数据 GIS 服务、多维指标统计服务、数据推送服务、智能信息服务、线上政务材料自动预审服务等政务大数据资源服务标准，正在利用逐步丰富的政务信息资源数据，为徐州各个部门提供服务。



附图 2 案例一：系统建设效果

（4）政务大数据标准化应用推动公共服务效能提升

通过徐州智慧城市信息枢纽项目，标准化管理在政务信息资源的大数据应用方面取得突破，获得了“真金白银”的价值回馈。例如徐州市财政局通过政务

信息资源的交换共享，综合分析人口结构、公积金缴纳信息、社保信息，在养老、优抚、精准扶贫方面优化财政支出安排，节约了上亿级别资金。徐州市人社局根据多部门人口数据比对，大量减少了因人口减员导致的养老金错误发放，同时节省了核对人口信息投入的人力物力。徐州市民政局在优抚、精准扶贫方面，同样利用多来源人口信息比对的信息资源服务，节约了大量冒领错领的资金。

2、地市级政务信息资源标准化项目的实施

徐州智慧城市信息资源枢纽项目从梳理、编目、归集、共享四个阶段，优化地市级区域政务的政务信息资源管理。项目 2015 年底启动，2017 年 1 月首次完成了徐州市政务信息资源目录的梳理工作；2017 年 2 月在全市范围内上线推广；2017 年 3 到 4 月完成了 11 家重点部门的应用需求调研工作；2017 年 6 月用大数据分析技术创造性地建立了全市人口法人族谱分析系统。是地市级区域完整利用资源目录标准化管理的实际案例，为地方政府进行政务大数据归集和共享提供了实践参考。

8.2 案例二：昆明国家经济技术开发区城市智能运营中心（IOC）

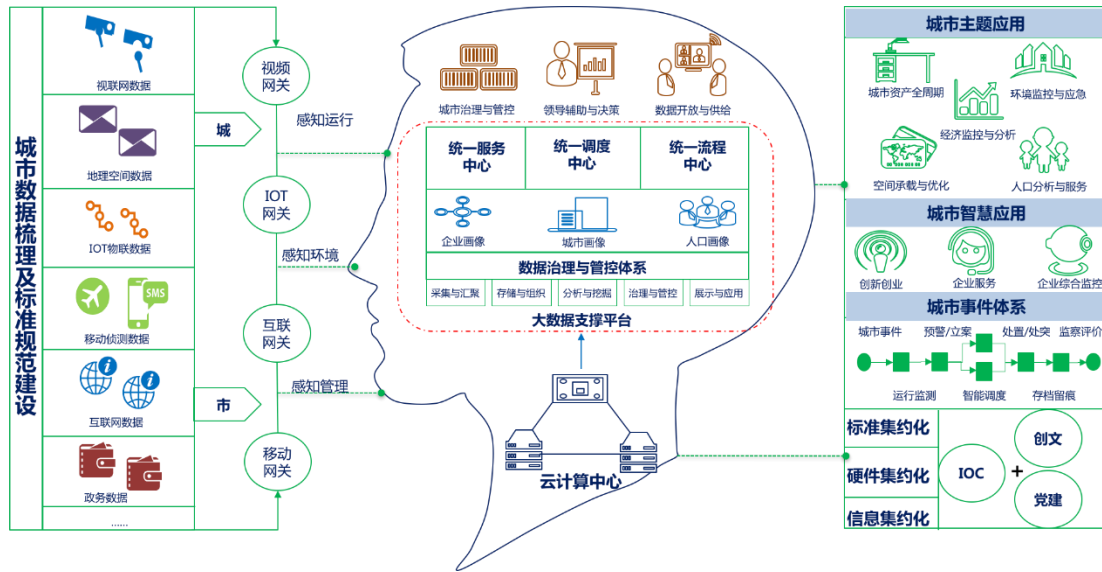
应用领域：智慧城市

1、建设目标

为实现昆明经济技术开发区“十三五”期间提质发展创新区的总体战略目标，围绕经开区管理的重大问题和经济转型升级的薄弱环节，以云计算、大数据、“互联网+”等先进技术手段与经开区管理融合为抓手，基于东方国信城市智能运营中心（IOC），以构建经开区城市智能中心为主攻方向，加速经开区城市管理、产业升级、民生服务与信息化的深度融合，实现跨越式发展。

围绕以上定位，从城市系统性的视角出发，基于大数据技术的城市智能中心（IOC）在信息世界整体层面即为城市“大脑”，政府专网、互联网、物联网（IOT）等泛在网络为神经，各传感器、摄像头、政务管理的专业系统以及移动应用为功能器官，城市各类空间、人口、法人、宏观经济、构筑物以及各专业系统的数据为运行信号。通过城市动态连接、感知、分析、反应城市各个角落的状态，通过

预警预报、自动控制、数据决策等形式，影射与服务实体城市管理需要。



附图 3 案例二：IOC 总体设计图

2、运行情况

昆明国家经济技术开发区城市智能运营中心于 2017 年 5 月投入试运行，已接入经开区 900 路摄像头，城市空间数据 50 多万条，物联网 IOT 数据以 4 家典型企业 5 个设备为试点累计入库数据 2000 多万条；政务数据 1800 多万条，涉及经开区 40 多部门；5 类非结构化数据 20 多万条，指标 1000 多项，形成 40 多项服务，主题应用 16 个，试运行期间每天收录城市事件数据 60 多条。累计存储数据达 10TB。

依托 IOC 的“创文”工作的快速落地，凸显了 IOC 大数据支撑智能化城市管理的能力，有效促进了各部门数据交换和共享的主观意识，为未来 IOC+智慧社区、IOC+智慧城市治理等奠定了基础。

3、标准规范建设

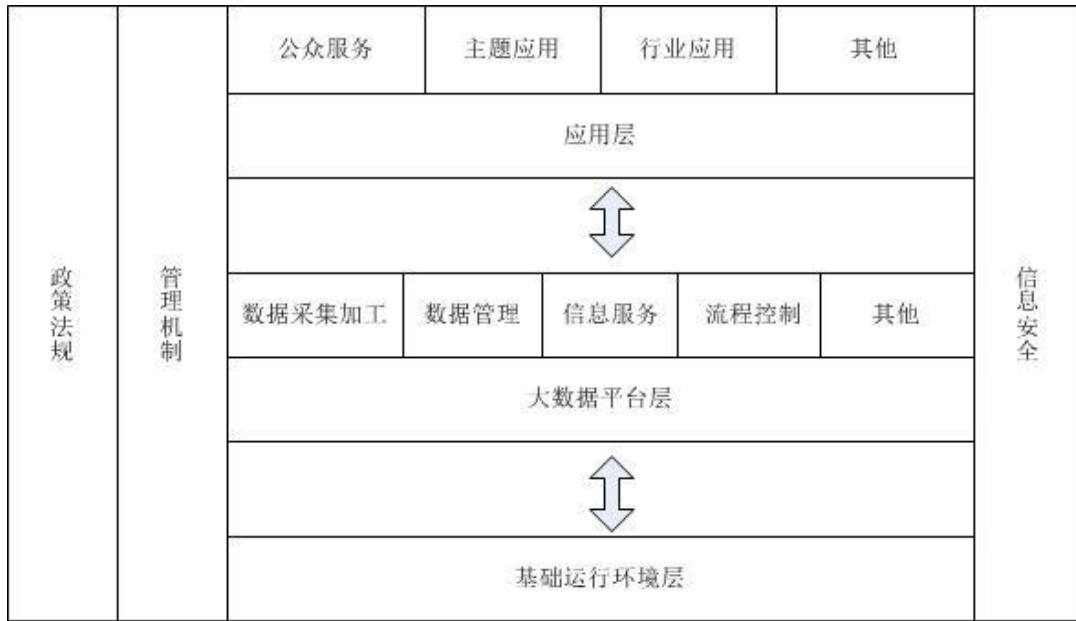
(1) 标准体系设计

昆明国家经济技术开发区城市智能运营中心建设过程中坚持贯彻大数据标准体系框架，并在大数据标准体系基础上结合地方实际情况扩展了相关标准规范，从而支撑 IOC 的搭建。

1) IOC 标准化技术参考模型

IOC 标准化参考模型是制定昆明经济开发区 IOC 标准体系的基础，是从系统工程的角度抽象概括出的 IOC 技术框架。基于城市智能运营中心（IOC）产品架

构，抽象概括出的 IOC 标准技术参考模型见下图。



附图 4 案例二：IOC 标准化技术参考模型图

IOC 标准技术参考模型由基础运行环境层、大数据平台层、应用层组成，信息安全、政策法规和管理机制贯穿于各个层面中。

基础运行环境层位于整个技术体系结构的底层，为 IOC 提供必要的网络、软硬件基础运行环境，以及高效、可靠的信息传输服务通道，是各类 IOC 应用的最终承载者。

大数据平台层提供基础功能服务、信息交换服务、事务处理服务、流程控制服务等各种通用服务，能有效地简化 IOC 的设计和实现。

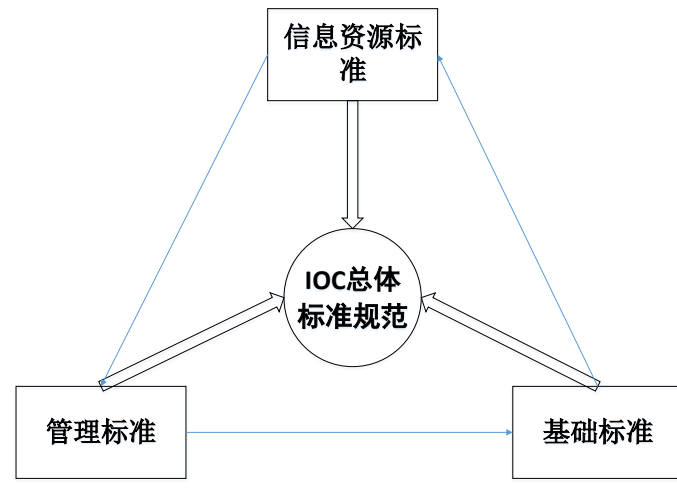
应用层是 IOC 最终的用户表现层，是 IOC 面向最终用户的层面。主要包括各类主题应用和服务。

信息安全在各层面上为 IOC 提供机密性、完整性、可用性、鉴别、抗抵赖等安全服务。

政策法规和管理机制是推动 IOC 建设运转的中枢，涉及基础层、大数据平台层、应用服务各个层面的技术和运营管理。包括制定标准和贯彻标准中所遵循的标准化方针、政策、组织制度等。

2) IOC 标准规范体系框架

本项目重点建设的是 IOC 大数据平台层和应用层的内容，依据 IOC 标准化技术参考模型，制定本项目 IOC 标准体系框架，见下图。



附图 5 案例二：IOC 标准规范体系框架

IOC 标准规范体系框架由 4 个部分组成：

- a) 总体标准：包括 IOC 建设总体性、框架性的标准和规范。
- b) 基础标准：主要包括 IOC 数据的管理、共享、交换、应用相关的标准和规范。
- c) 信息资源标准：包括数据分类与编码、数据加工、数据质量控制等方面的标准和规范。
- d) 管理标准：主要包括 IOC 数据安全管理的标准规范。

IOC 标准规范体系表如下：

附表 2 IOC 标准规范体系表

分类		序号	标准名称
总体标准		1	《IOC 标准规范总体设计》
信息资源标准	分类标准	2	《城市数据分类与编码规范》
	数据加工	3	《数据整合与处理规范》
	数据质量控制	4	《数据质量校核规范》
		5	《数据治理与管控流程规范》
基础标准	数据管理	6	《城市管理及运营指标体系规范》
	数据共享交换与服务	7	《数据采集接口规范》
		8	《数据分析与应用接口规范》

		9	《应用系统接入接口规范》
		10	《数据共享与服务接口规范》
管理标准	信息安全	11	《数据安全管理办法》

(2) 实施思路

I0C 标准规范的建设是 I0C 整个建设内容之一，时间紧、任务重，既要做到数据、平台、应用系统建设达到预期效果又要遵循一定的规范约束，保持先进性和扩展性，必须合理安排标准、数据、平台、应用系统建设等工作。

1) 标准急用先行、平台成熟优先

对于影响 I0C 数据中心数据采集、加工、存储的标准，即《数据采集接口规范》、《数据安全管理办法》、《数据整合与处理规范》、《城市管理及运营指标体系规范》优先编制。

核心平台即 I0C 大数据平台采用经过多方 POC 测试和央采软件协议供货采购目录中的产品，从而降低风险。

2) 示范推动、逐步扩展、动态维护

项目建设中涉及人口、经济、节能、安监环保等 16 个面向领导、部门、企业和公众等不同终端的主题应用系统，标准的建设很难在短时间内一次性建设完成同时支撑多个应用，因此，在实施过程中对重要的标准除了优先编制外，还采用迭代更新的方式，例如，以“创文”为契机，结合业务需求，从指标数据的采集、处理、分类、管理、展示等全过程首先形成该领域的一系列标准规范，凸显成效后扩展其他领域。

4、标准应用案例

在 I0C 标准体系的指引下构建了昆明国家经济技术开发区 I0C 典型应用案例：

(1) 对园区资产进行全生命周期管理；

在城市资产全生命周期管理领域，以“城市资产”为核心进行数据的收集整理、整合、分析和统计利用、决策建议等。



附图 6 案例二：系统截图 - 主题应用和资产生命周期管理

(2) 对园区空间布局进行承载优化;

综合运用城市规划、国土资源、城市建设、地下管线、产业布局、人口分布等城市大数据，实现城市资源、产业、人口等资源的空间布局分析、资源集约利用分析、资源承载力分析等功能，实现城市规划更科学、空间布局更合理、产业升级转型更科学、人力资源布局更合理。

(3) 对园区提供更及时的园区环境监测与应急指挥;

基于经开区行政区划图，叠加交通运营、犯罪、治安、社区、教育、卫生等数据，分析城市交通拥堵指数、城市犯罪率、城市治安状况及社区治理情况等。



附图 7 案例二：系统截图-环境监测

(4) 通过企业画像，对于园区工业经济进行全方位监控与宏观分析；综合运用工业和商业企业的工商、税务及进出口等大数据，按工业类型实现产业发展趋势、投资收益、供应链、能耗与产出等分析，实现经开区经济形势分析、经济运行监控、经济运行分析等的“一张图”展示。

(5) 通过人口画像，提供更精准的民生服务与公共管理；通过对各个部门关于人口数据的汇总与分析，以“人口”为核心，在社保、民政、公安、教育、计生、住房、交通、残联等领域，采集户籍人口、长期居住人口、流动人口数据并进行整理、整合、分析与挖掘。

(6) 为园区搭建“以市场为导向”的创新创业平台；本着“塑造一个人，培养一批人，带动一群人，影响一代人”理念，遵循以市场为导向，聚焦科技创新应用，建立创新创业平台，汇聚创新要素，培育创新创业队伍，培育创新环境。

(7) 为园区企业构建“一站式”服务平台；构建起政府和企业沟通的信息平台，成为政府为企业提供高效快捷服务的绿色通道，为企业自身管理提供基础办公平台，为企业生产性服务提供工业设计、物流、人才、交易、供需、行业舆情等信息服务，成为企业成长的助力平台。

(8) 为园区“节能减排、环保安全”组建工业企业综合监控平台；

面向工业企业的节能、减排、环保、安监等业务应用，提高工业企业综合监控平台，实现工业企业的节能、减排、环保、安监等监控设备的接入、组建园区工业企业监控的物联网（IOT），实现工业企业的大数据采集、大数据分析等，为企业的节能、降耗、环保和安全提供优化方案和大数据服务。



附图 8 案例二：系统截图-工业企业综合监控平台

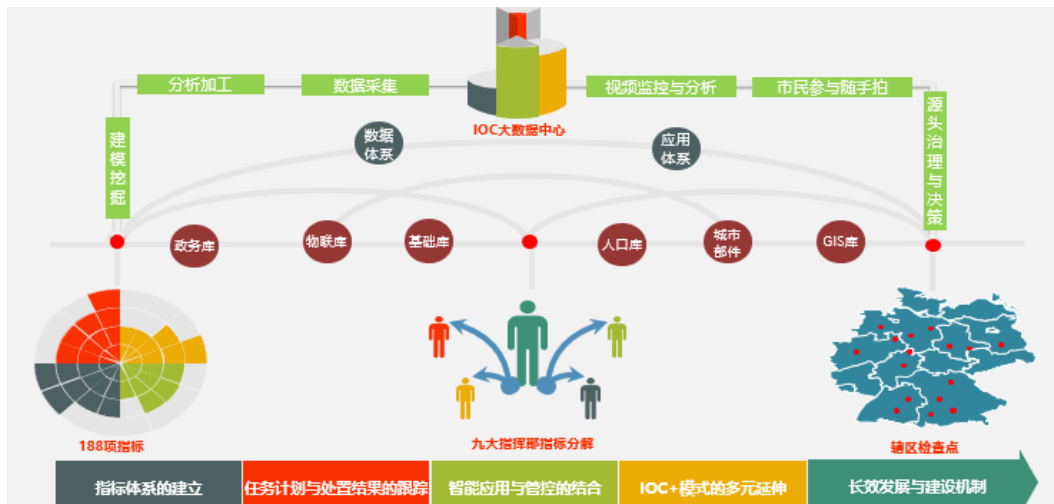
(9) 搭建智能的统一指挥和调度中心

将各城市事件串成一个整体，统一指挥和调度，提升区域“综合性”问题的解决能力和各部门“联动”水平。在智能事件调度平台的基础上结合“创文”工作形成了“创文”大数据指挥中心，昆明经开区通过“政府+市民”参与的数字化管理以及大数据中心统一指挥，实现了“信息收集全面及时，情况处置快速高效”。王喜良市长表示，大数据指挥中心在“创文”工作中发挥了很好的作用，希望各级相关部门健全优化城市管理长效机制，以“创文”工作推动城市管理水平迈上新台阶。

IOC+“创文”应用中 188 项指标的构建和管理，是《城市管理及运营指标体系规范》的一个典型应用。



附图 9 案例二：系统截图-统一指挥和调度中心



附图 10 案例二：IOC+“创文”体系

8.3 案例三：“数控金融”互联网金融大数据监管平台应用案例

应用领域：金融

1、项目参考的大数据领域、互联网金融领域的国家、行业标准

项目充分参考和结合多项已发布和在研的相关国家标准和国际标准，如《金融服务 信息安全指南》《信息技术 大数据 术语》《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》《信息技术安全技术 信息安全管理体系实施指南》等。

2、相关大数据标准在具体项目的实施应用

2015年6月至8月，国务院连续下发了针对大数据行业的三个指导性文件（国务院办公厅《运用大数据加强对市场主体服务和监管的若干意见》、国务院《积极推进“互联网+”行动的指导意见》、国务院《促进大数据发展行动纲要》），要求各级政府部门要充分运用大数据先进理念、技术和资源，加强对市场主体的服务和监管，推进简政放权和政府职能转变，增强政府服务职能和监管的针对性和有效性，提高政府治理能力。2015年7月18日，人民银行等十部门发布《关于促进互联网金融健康发展的指导意见》（银发〔2015〕221号，以下简称《指导意见》）。《指导意见》按照“鼓励创新、防范风险、趋利避害、健康发展”的总体要求，提出了一系列鼓励创新、支持互联网金融稳步发展的政策措施，积极鼓励互联网金融平台、产品和服务创新，鼓励从业机构相互合作，拓宽从业机构融资渠道，坚持简政放权和落实、完善财税政策，推动信用基础设施建设和配套服务体系建设。并提出了“依法监管、适度监管、分类监管、协同监管、创新监管”的原则，

勤智数码科技股份有限公司研发的“勤智·数控金融”互联网金融监管大数据平台，运用大数据、人工智能、区块链等技术，为金融监管部门的业务管理提供全面的信息化管理支持，提高其行政管理效率和履职能力。科学合理界定了各互联网金融业态的业务边界及准入条件，落实监管责任，明确风险底线，促进地区互联网金融产业健康发展，更好地服务实体经济。并实现了制度建设和技术保障的有效结合，以及对金融监管部门业务管理活动进行全方位、全过程风险监控，实现线上线下、业务层面与监管层面的实时交互对接，提高其业务管理的信息化水平、办事效率和履职能力，坚决守住不发生行业风险和区域性金融风险的底线。

（1）“勤智·数控金融”在互联网金融监管标准化上的创新：

全面：平台作为RegTech监管科技的实际应用，以大数据为基础，以云计算为手段，对接区块链技术，利用人工智能实现各种互金业态的全面监管；

高效：实时对接监管企业平台数据和第三方数据，实时建模分析，实时预警，

实现高效、快速、自动地处理；

专业：针对互金业态的风险预警、防控模型，汇聚银行高管、海归博士、金融才俊等专家团队，经过实际验证；

安全：自下而上的信息安全体系，区块链技术的首次应用，有效保障资产安全，实现数据不可逆、不可伪造，奠定证据保全的基石；

准确：征信要素完备，有效防范信用风险，融合政府、企业、第三方机构、网络等大数据搭建信用体系。

可控：完全自主知识产权的产品，包括从底层大数据平台，风险模型，前沿技术，到顶层应用的全方位自主知识产权产品。

（2）互联网金融监管的标准化将进一步提升监管水平

针对互联网金融监管存在的条块分割，信息孤岛，质量堪忧，手段缺乏的困境。“勤智·数控金融”将为互联网金融监管机构提供全面、专业、规范的工具和体系。

“系统贯通”：通过建设“勤智·数控金融”平台，推动金融监管部门业务管理向信息化、数据化、自流程化与融合化发展，并实现对金融监管部门已有的业务监管系统的整合，提高其针对金融业务的日常监管和风险管理能力。同时通过大数据手段和思维，对监管权力的运行进行风险控制，将权力、制度、行为和流程数据化，通过数据管住人、事、物，以数据为重要核心，完善权力监督的技术体系。

“数据整合”：通过“勤智·数控金融”平台业务管理系统的建设，将制度建设和技术保障有效结合，对金融监管部门的业务管理活动进行全方位、全过程风险监控，实现线上线下、业务层面与监管层面的实时交互对接，提高金融监管部门业务管理的信息化水平、办事效率和履职能力，坚决守住不发生行业风险和区域性金融风险的底线。

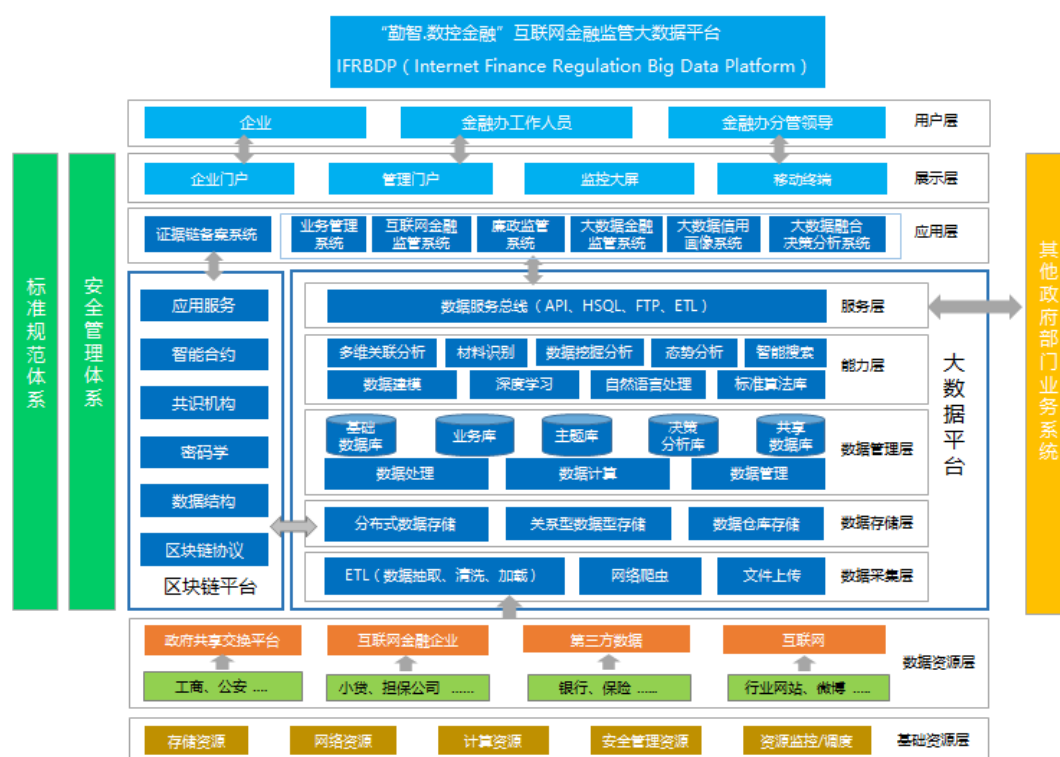
“多维监测”：通过第三方等多维度的数据监测，“数控金融”平台实现实时汇聚分析来自于 40 多个维度的第三方数据，如企业接入数据、政府公开数据、运营商数据、互联网公开数据、电商平台数据、各类租赁的 O2O 平台等。实现多维度、多业态、多场景、全方位的监测。

“丰富手段”：采用大数据、云计算、分布式存储和计算、人工智能、区块

链等技术，积极探索新兴信息技术在金融监管部门业务监管中的运用，适应互联网和大数据时代对互联网金融业务管理的需要，提高地区金融监管机构的履职能力。

3、互联网金融监管标准化管理将显著提高监管效率

“勤智·数控金融”互联网金融监管大数据平台已运行超过 15054 小时，拥有超过 200 个分析模型，超过 400 张报表，具备 300 多个流程上百个角色，已经产生超过 7000 多项数据项。“数控金融”互联网金融监管大数据平台基于 SOA 体系进行系统架构，由十一个层面，两大体系构成：十一个层面由基础资源层、数据资源层、数据采集层、数据存储层、数据管理层、能力层、服务层、应用层、展示层、用户层、区块链平台组成；两大体系包括标准规范体系和安全管理体系。



附图 11 案例三：勤智数控金融业务整体架构图

8.4 案例四：农业部农产品批发市场价格挖掘及可视化平台

应用领域：农业

1、主要攻关内容分为两大部分：分析挖掘模型建立研究和可视化展示

项目背景：我国农业数据的系统采集最早可追溯到 1950 年左右，随着物联网、互联网、移动互联网技术的应用，农业数据的产生速度及采集频率加快，但农业数据普遍存在频次低、及时性差的问题，数据多为年度数据且更新滞后 1-2 年。目前农业部信息中心拥有采集积累的 20 年批发市场价格数据，是国内农产品高频数据中，时间序列最长的数据之一，具有时间序列长、频次高、监测点覆盖区域广的特点，是市场稀缺数据，价值巨大。但是农业部信息中心对于 20 年历史价格数据的挖掘分析建模及数据应用较少，数据价值挖掘需求迫切。

项目目标：一方面围绕农业部农产品批发市场 20 年数据资源，充分考虑聚合其他农业环节数据，最大程度挖掘批发市场数据应用价值。挖掘我国农产品价格的波动趋势以及农产品间价格的相关性等重要特征，认识批发市场农产品价格变化规律，完善农产品批发市场价格指导机制，保障全国农产品的供需平衡及市场的稳定发展。另一方面通过搭建可视化展示平台，研究数据展示效果。

2、项目针对技术标准与标准应用的效果

技术标准的研究，使得项目通过对于批发市场大数据平台涉及的数据进行数据清洗、数据管理、数据开放、指标口径、分类目录、交换接口、访问接口、数据质量、技术产品、安全保密等流程的处理，各个环节均均遵循统一的技术标准，并依托大量的数据建模，研究制定挖掘分析与可视化展示关键标准的制定和实施，为构建涉农大数据开发应用的标准体系提供支持。

技术路线与技术标准研究：(1) 利用大数据技术，对历史数据进行数据清洗和梳理，得到符合数据应用要求的干净数据，明确批发市场数据资产的数量和质量。清洗后的数据可以应用于数据开放共享、建模及应用开发。(2) 对历史数据进行系统的分析，描绘不同类型农产品价格周期、传导机制，通过价格数据建模，对全国的重点生鲜农产品价格实现预警。(3) 对原有的数据展示进行可视化专题开发，采用 B/S 架构，可视化与后端实现解藕和分离，全面采用 HTML5 与 WebGL 技术、react、vue 技术以 3D 或 2D 中国地图板块为主视觉区，可以对象形式操

作图形元素，具备良好的事件交互机制，吸附各农产品批发市场价格信息的可视化功能，基于绿色、生态、有机、科技等关键词，进行农产品批发市场价格数据的专题页面与可视化呈现设计。

技术标准的研究驱动的应用效果：通过数据清洗，全面梳理了价格数据资产的数量和质量。通过建立价格波动周期模型、价格预测模型、农产品价格传导时滞模型、农产品价格相关性模型等价格模型，形成了一套统一的技术标准应用场景，挖掘出我国农产品价格的波动趋势，了解到批发市场农产品价格的变化规律，辅助农业部更及时更准确的进行价格监测预警，防止农产品价格暴涨暴跌的频发，保障农产品的供需平衡。

3、案例实施

项目开展情况：农业部农产品批发市场大数据挖掘与可视化项目已完成，正常运营中。

项目指导意义：

一方面，本次项目的研究成果可为以下 4 种相关部门或机构提供参考：

（1）农业部的垂直部门，例如种植业司、畜牧业司等，在对生产环节做监控分析时，也需要下游的批发市场流通及消费数据作为参考。

（2）发改委、商务部均在监测农产品的价格，也有拓宽数据源的需求，使用同类数据进行数据比对、参考、数据关联分析等。

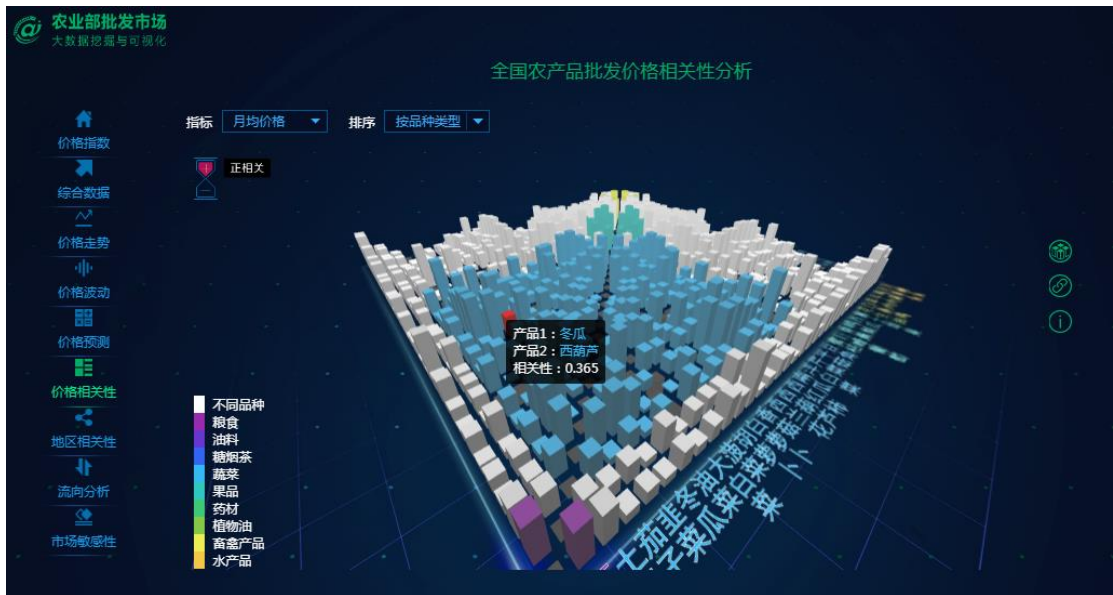
（3）产地政府，需要及时给农民提供价格行情服务，为当地农民提供种植及销售决策建议，以便当地农民合理安排农产品的生产，政府也可以根据未来市场需求趋势，及时调整种养殖政策。

（4）销地政府，为保障菜篮子供应充足、安全，销地政府对当地的农产品价格稳定及供应充足监测比较关注，同时还关注本地区农产品的消费结构，来源地份额，来源地农产品生产情况等。

另一方面，本项目在农产品价格大数据清洗、数据脱敏、数据挖掘、数据结果可视化等环节形成了一套成熟的研究方法体系，并在农产品价格的区域传导性、价格敏感性、品种间价格相关性方面形成了一些结论性成果，可为地方农业局或农业大数据企业在农业数据的研究和探索方面提供借鉴和参考。



附图 12 案例四：系统截图--价格指数



附图 13 案例四：系统截图—价格相关性分析

8.5 案例五：海尔 COSMO Plat 空调噪音大数据智能分析

应用领域：家电行业、智能制造

应用背景：海尔胶州空调互联工厂部署有国内唯一的分贝检测设备，当空调测试分贝大于标准分贝时，系统判断为不合格并将结果输出至 COSMO Plat-IM (MES) 系统，但此设备无法识别空调运行中的异音，如摩擦音、共振音、口哨音等。此外，每天快节奏、高强度的空调装配流水线工作导致检测工人听取噪音

时间过长，易产生疲劳和误判，偶尔有不合格品流到下线，影响产线整体检验的可靠性。因此，急需找到新式噪音识别方法，解决企业当前痛点。

核心方案：基于标准化思路的核心问题研究：COSMOplat 是海尔自主研发、自主创新的共创共赢工业互联网平台，通过整合平台上的软件及硬件资源，与美林数据共同开发了空调噪音智能检测系统，有效地解决了无法准确、可靠识别异音的痛点。解决方案包括非结构化音频数据实时采集与存储、分析建模与智能识别、结果输出与可视化展现三大部分，核心过程如下：

阶段 1：模型搭建的标准化研究

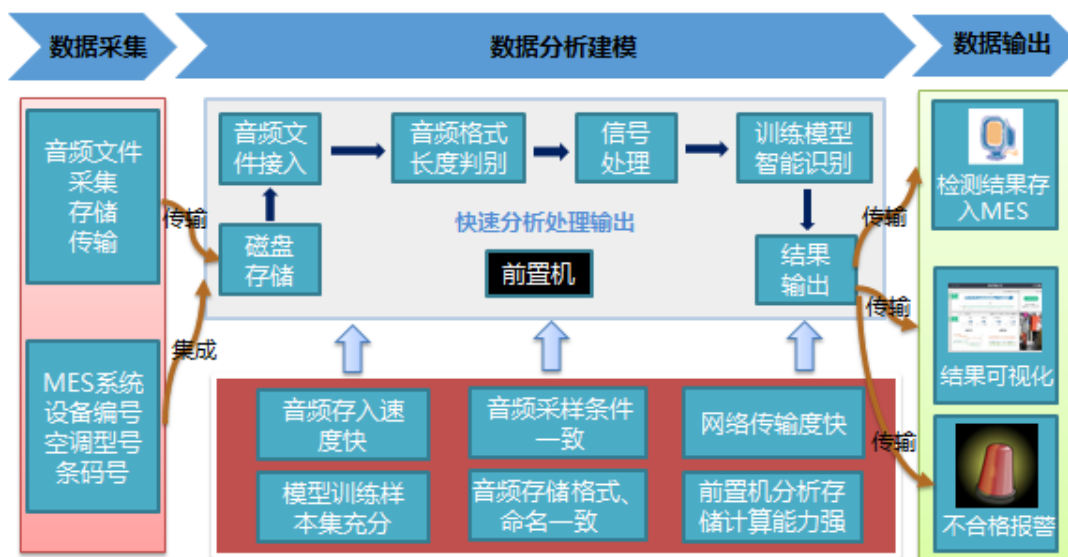
针对生产线采集的大量历史检测音频，利用端点检测技术对产品运转过程中起、停机阶段的音频区段进行智能切割，利用数字滤波技术自动对音频进行降噪。通过特征自动提取与样本标定，利用机器学习技术构建智能分类模型，模拟人工判断行为，构建标准化的模型研究思路。

阶段 2：参数调优的标准化思路

智能分类模型需通过大量音频数据进行模型训练与优化，并验证其准确性。算法专家利用历史音频对模型进行验证与参数调优，通过不断扩充训练样本及模型自学习，确保识别准确率满足生产线质检精度要求，最终形成一套基于标准化思路的调优方法。

阶段 3：上线实施，技术标准研究成果的应用

构建音频采集系统，实现产品分贝检测产线对音频的实时同步采集与型号关联。智能识别模型自动完成音频文件的接入、特征提取、智能判别等工作，输出对应产品条码号的实时判别结果，对异音自动报警，并针对识别结果对产品异音原因进行智能分类，辅助返修排故。系统将智能检验结果实时反馈至企业 COSMOplat 工业互联网平台，支持产线质量问题在线统计与分析。



附图 14 案例五：系统核心思路

项目实施过程中参考了《信息技术大数据术语》《非结构化数据管理系统技术要求》《信息安全技术数据库管理系统安全技术要求》等大数据相关标准，并与海尔工业智能研究院有限公司一起，结合项目具体实施过程中的现场问题和解决过程为《信息技术大数据存储与处理系统功能测试规范》《非结构化数据访问接口规范》《实时数据库通用接口规范》等在研以及拟研制的大数据相关国家标准反馈了标准立项诉求和标准内容建议。

实际效果：海尔 COSMOPlat 空调噪音大数据智能分析项目通过传感器、分贝检测系统、业务系统、模型算法的集成与交互，在企业解放人力、减少误判、提高检验可靠性等方面均有了极大提升。此项智能检测系统的实施充分利用了设备端的嵌入式智能计算技术，以分布式信息处理的方式实现了设备端的智能和自治，通过服务器、业务系统间的交互协作，实现了检测系统整体的智能化。项目的实施为海尔集团在旗下其他分厂生产线部署基于声音检测的空调状态智能识别系统积累了丰富经验，为行业内公司在产线智能化改造与转型升级等方面做出了示范。

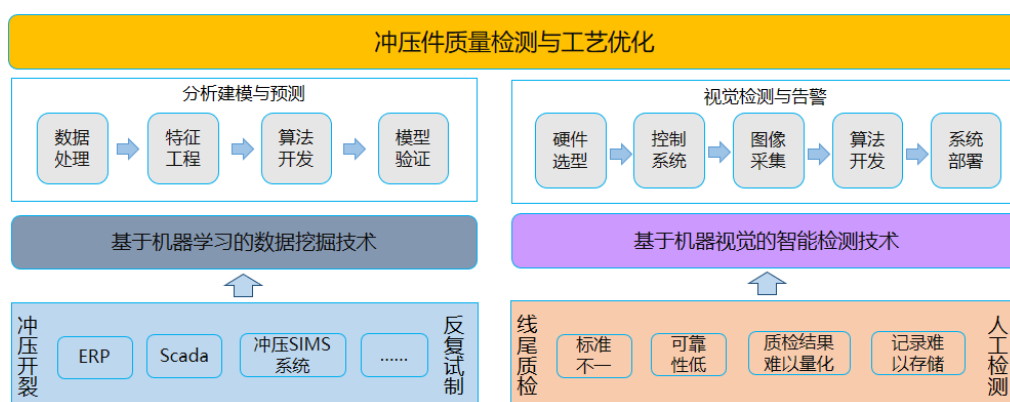
8.6 案例六：长安汽车智能制造技术研究所冲压质量大数据项目

应用领域：汽车工业、智能制造

长安汽车作为中国知名汽车制造企业，中国品牌汽车产销累计已突破 1000 万辆，并连续 10 年位于中国品牌汽车销量前茅。近些年，长安总体战略也是从传统汽车制造企业转向智能制造服务型企业，重点发展科技、智能制造、服务三大方面，并在重庆本部成立了智能化中心。长安汽车某工厂冲压车间共建有三条冲压生产线，主要负责生产侧围、翼子板、车门、引擎盖等轮廓尺寸较大且具有空间曲面形状的乘用车车身覆盖件。

目前在冲压生产过程中，一方面由于冲压设备性能、板材材料性能、生产加工过程参数等波动，部分侧围在拉伸工序中易产生局部开裂现象，需反复进行参数调整与试制；另一方面，在冲压产线线尾，需对冲压件外观质量进行统一检测，现有检测方式为人工手动检测，需在有限生产节拍时间内，快速分拣出带有开裂、刮伤、滑移线、凹凸包等表面缺陷的冲压件，检测标准不统一、稳定性不高、质检数据难以有效量化和存储，不利于企业数据资源收集、质量问题分析与追溯。

美林数据通过建设大数据存储与处理平台，实现了工厂冲压车间的所有设备、模具、材料、生产制造过程数据、质量检验数据的集成、存储与统一管控。平台建设过程参考了《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》《非结构化数据访问接口规范》等标准。



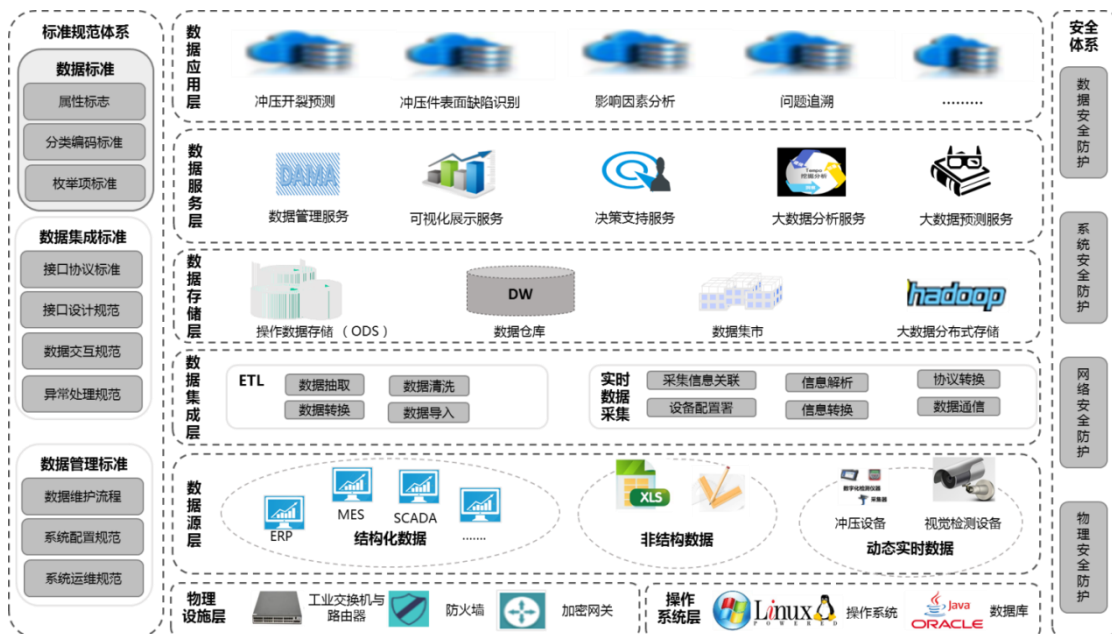
附图 15 案例六：解决方案

依据大数据存储与处理平台，借助基于机器学习的数据挖掘、基于机器视觉的智能检测技术，实现了冲压侧围件开裂预测与产品件表面缺陷的智能识别检验。

➤ 依据冲压设备加工参数、板材参数、模具性能参数及维修记录等，通过数据挖掘机器学习算法，建立冲压工艺侧围开裂智能预测模型。通过样本积累与模型训练调优，准确预测冲压侧围件的开裂风险。最后，确定了冲压制造过程影响因素间的相关性，制定了生产过程参数组合控制策略，为冲压制造过程工艺优化和质量把控提供支持。

➤ 基于机器视觉的冲压件缺陷智能识别检测，立足生产线现有条件，设计图像采集系统，通过图像实时采集与智能分析，快速识别冲压件是否存在表面缺陷，并自动将所有检测图像及过程处理数据存储至大数据平台。通过质检数据、生产过程工艺参数、产品设计参数间的关联，借助大数据分析技术，形成冲压产品质量问题分析管理的闭环连接，实现冲压产品质量的精确控制和优化提升。

➤ 项目实施总体框架如下：



附图 16 案例六：项目实施总体框架图

在项目具体实施过程中，美林数据通过与长安汽车项目团队进行业务研讨及评定，综合制造业底层数据采集、平台数据规范设计、数据质量检测、大数据分析总体技术要求等方面内容，编写了企业级《制造业大数据分析业务指南》。此外，根据项目实施经验，为工业大数据领域相关标准、应用规范的立项、研制、

发布等提供了信息反馈和经验积累，后续将协助大数据标准工作组确定并推出相关标准内容。

8.7 案例七：江苏省重点领域共性技术攻关项目---工业大数据元数据规范与验证技术攻关专题---工业大数据标准体系研究及重点标准编制

应用领域：工业

1、主要攻关内容分为两大部分：工业大数据标准体系研究和重点标准编制。

(1) 工业大数据标准体系研究

全面分析智能制造发展的要求和工业大数据产业发展趋势，研究智能制造企业的研发、生产、运营、营销和管理方式，分析国内外工业大数据服务产业标准的现状、趋势和需求，梳理相关国际标准、国家标准、行业标准、团体标准、企业标准，提出符合针对智能制造相关技术研发和业务应用需求，研究工业大数据标准化需求，以指导成体系成系统的标准制定工作，形成工业大数据发展趋势的标准体系框架和标准体系表。

(2) 重点标准编制。

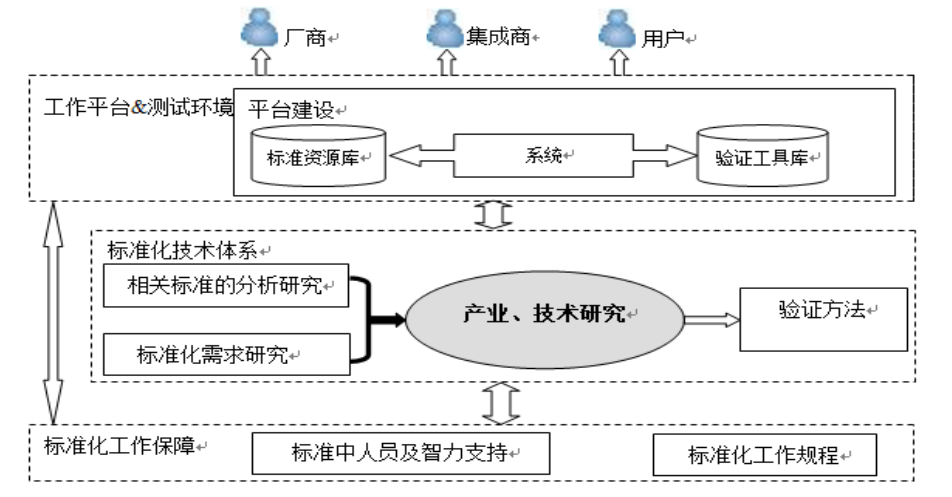
结合产业需求，针对现阶段工业领域急需解决的问题，采用急用先行的原则，结合工业大数据标准体系框架，研制《工业大数据产品核心元数据规范》、《工业大数据 OID 标识分配与注册解析规范》、《工业大数据 OID 对象标识符编码与存储规范》等重点关键的工业大数据领域重点标准。

2、本案例先进性和技术路线

(1) 总体设计

1) 各项研究任务协调配合、有机互动、整体推进

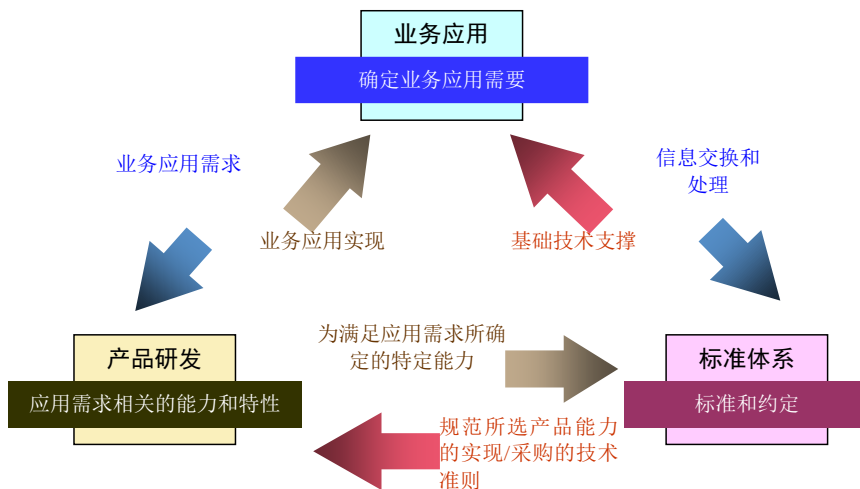
本项目研究内容包括相关技术、相关标准体系和文本的研制工作。各项工作相互配合，相互促进。



附图 17 案例七工作流程图

2) 标准研究与标准验证、应用验证等内容协调互动

工业大数据相关业务、产品和标准化三者之间是相互促进、相互制约的，标准的需求主要来自业务和系统的需求。这三者之间相互依托，层层递进，协同完成项目目标。



附图 18 案例七逻辑结构图

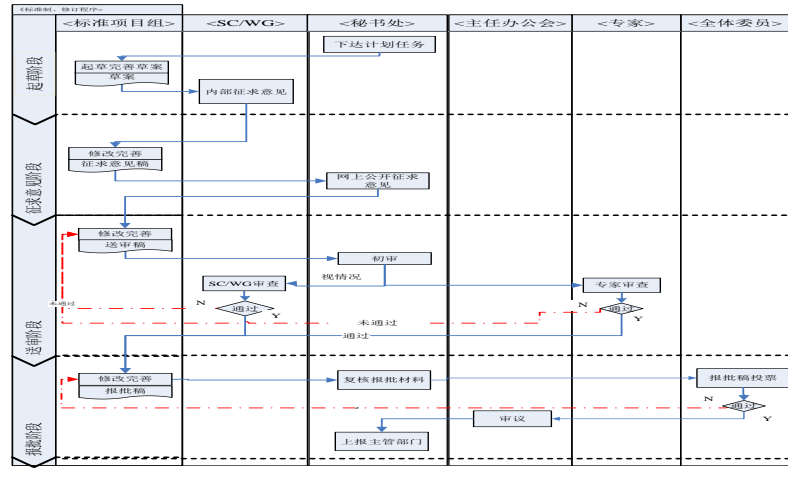
3) 相关标准研制

工业大数据相关服务标准的制定将征集用户需求，并针对尚未规范化的、紧迫需要规范和统一的内容，适时推出标准内容。

以现有技术为基础，提炼适合系统的技术及参数要求，结合技术发展趋势，联合相关生产厂商、学术机构、高等院校和行业协会及用户等各方组成标准起草组共同研究标准内容。

(2) 项目实施

按照标准化法规定，标准在研制过程中主要经历起草、征求意见、送审和报批等四个阶段。



附图 19 案例七项目实施阶段说明

3、案例实施

(1) 江苏中堃数据技术有限公司基于公关专题，结合工业大数据重点在线缆行业的能耗分析和预警预测、营销分析和预警预测突破。其中通过能耗横向对比分析，纵向时间序列分析，发现企业耗能设备，并结合入厂时间，单位产量能耗等指标，帮助企业能耗下降 5%。通过设备能源利用率分析，为企业提升 10% 左右的产品产量。营销分析预测是通过分析同行业的相关数据，为决策者提供有力的决策支撑，辅助决策者做出最优的选择。

(2) 在新能源领域，江苏中天科技软件技术有限公司在相关标准的指导下构建智慧能源管理平台，实现了集中数据存储和大数据分析评估，可以输出多维度报表；支持平台电站资产评估分析，投融资决策；充分考虑了大数据的采集和处理稳定性和安全性。

(3) 徐工信息公司结合工业大数据重点服务标准的研制和标准试点验证系统建设，在工程机械智能制造新模式探索中积极应用该标准系统，同步开展产品核心元数据规范和 OID 标识管理规范标准验证试点工作，同时扩展行业推广应用。

(4) 江苏蓝创智能科技股份有限公司将在橡胶机械行业智能制造新模式中推广应用系列标准，验证标准在该行业中产品数据描述匹配能力，提高企业有效维护和管理产品数据的能力，提升工业产品生产溯源查询和产品数据共享能力。

(5) 徐工集团牵头承担了标准试点验证项目，积极参与了标准草案和验证系统方案的编写及讨论，并在子企业进行试点，开展验证系统对接研发和验证系统的应用与验证，输出工程机械行业的标准验证效果和相应的《应用指南》。

(6) 苏州洞察云信息技术有限公司——星云测试精准测试系统为锐捷网络的大型核心交换机程序提供可视化、智能化全生命周期测试解决方案，实现以极低的硬件消耗采集设备内部进程高速运行的代码逻辑数据，提供达到航天级别的测试数据和智能测试分析结果。

8.8 案例八：国网电力大数据应用案例

应用领域：国家电网

从电力行业不同业务领域来看，在电网生产领域，积极推进大数据技术在智能电网发、输、变、配、调、用六大环节的广泛应用能够提高供电效能、促进经济运行、增强电网安全；在经营管理领域，积极推进大数据技术在电网规划、配网运行、运营监测和人财物集约化管理等方面的广泛应用，以促进经营管理模式创新发展；在优质服务领域，积极推进大数据技术在智能电表增值服务、需求侧管理等方面的广泛应用，以促进优质服务能力提升和新型业务形态发展。

1、配变重过载预警分析的技术标准的研究和应用

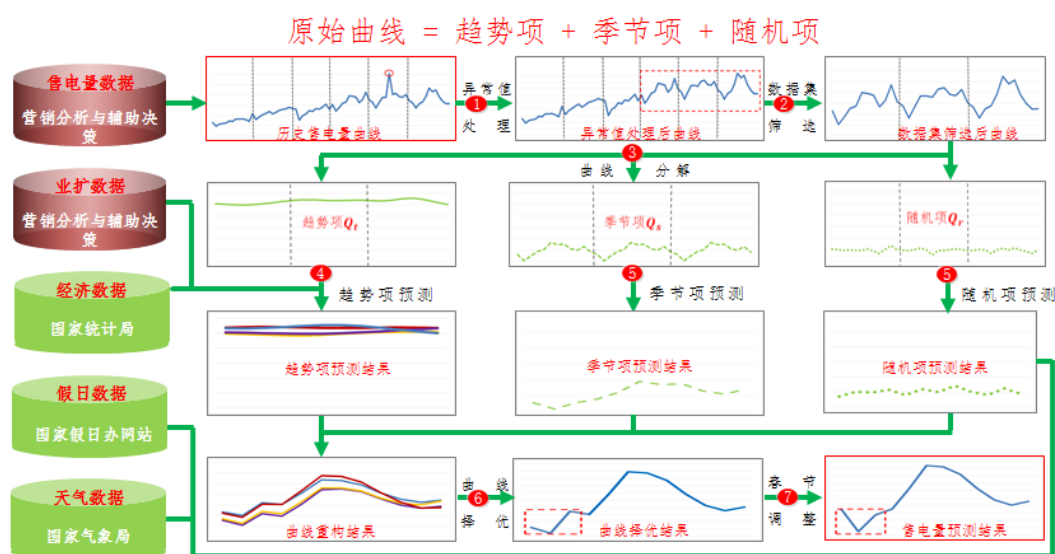
随着经济的持续发展，用电负荷相应快速增长。由于地区发展不均衡性、配网投资合理性、配网结构复杂性、突发事件偶然性等内外部原因，配变重过载时有发生。国网公司在不断加大配电网的建设和改造力度，并新建大量配电线路和配电台区，但每年仍有大量配变重过载现象发生。供电可靠性下降、客户投诉上升，电网公司缺乏有效预警机制，无法充分发挥事前引导作用，重过载预警需求较大。

当前业务部门开展重过载预警，多以人工经验或是简单阈值预警为主，配变重过载预判准确率较低，只维持在 50%左右，未综合考虑影响配变重过载发生的外部环境、经济因素及配变自身属性因素。多家网省公司都开展了变压器重过载动态分析的工作，分析配网变压器重过载产生的因素，研究各因素对配变发生重过载影响的重要程度。

美林数据基于大数据挖掘技术的配变重过载预警分析，结合技术标准的研究工作，是通过选取有可能影响迎峰度夏期间配变发生重过载的信息数据，如配变的历史负荷数据、配变所属区域数据、设备信息、客户信息数据、气温数据等，综合考虑电力负荷的周期性波动，同时结合气温、日照等气象因素对迎峰度夏期间配变负荷的影响，运用逻辑回归模型，构建重过载预警中期和短期模型，对配变重过载现象进行预测预警分析。充分利用了大数据分析、存储等技术标准相关内容，此项基于大数据挖掘的配变重过载预测分析技术，能够更加准确、及时的预判配变重过载情况，短期预警模型准确度可达到 88%左右，中期重过载预警模型准确度可达到 70%左右。通过模型预测的中期和短期重过载预警模型可最大程度降低配变重过载水平，减轻配变重过载带来的不良影响，提升配网供电能力。

2、售电量精准预测分析的技术标准的研究和应用

随着电力体制改革和智能电网建设的不断深入，售电量已成为考核电力企业的一个重要指标，月度售电量预测对于国家电网公司合理地确定销售电量总定额、分解售电量销售指标、制定有序用电方案、指导发电厂和输配电网的合理运行、推动电力市场的发展和建设都具有十分重要的意义。此外随着售电侧的放开，售电量预测的准确与否直接和售电公司的利润挂钩，售电量预测的重要性便更加凸显。本应用主要预测省公司、地市公司 10 大行业（8 大行业、居民及全行业）未来月度售电量，且结合最新的因素及售电量数据，通过自学习的方式得到最新预测结果，对公司电网规划、有序用电具有重要的指导意义。



附图 20 案例八：模型构建原理

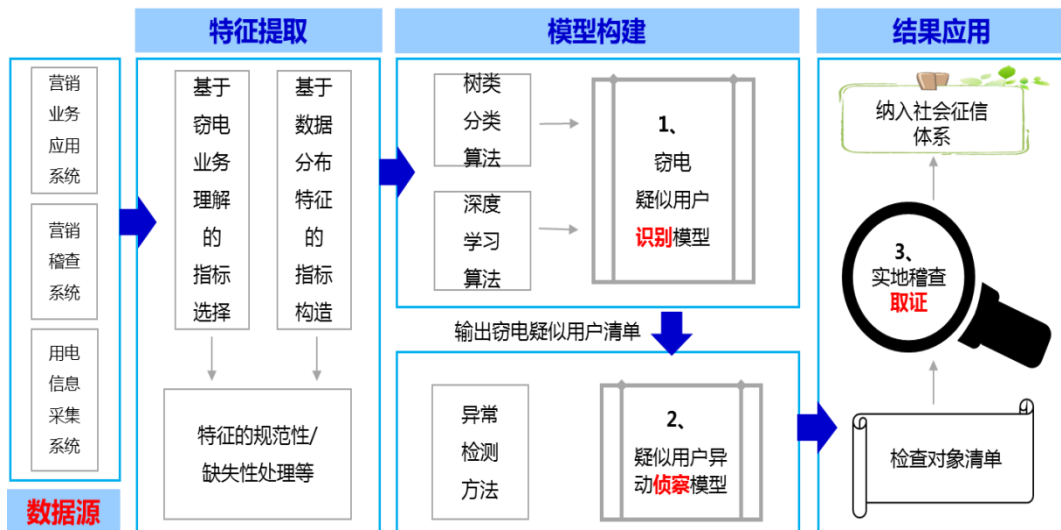
采用的关键算法：线性回归、L1/2 稀疏迭代回归算法、BP 神经网络、回升状态网络 ESN、支撑向量机 SVM、逻辑回归等。

目前已在多个省级电力公司应用，效果较好，准确度达到了 98%以上。

3、反窃电分析的技术标准的研究和应用

随着经济快速发展，社会用电量逐年增长，同时窃电行为也日益频繁，给电力企业造成了大量的经济损失。窃电是导致线损率升高的一项重要因素，目前电力企业反窃电主要通过例行用电检查和应用用采数据构建人工规则来判断，此方法检查窃电用户具有随机性，且人工规则得出的疑似窃电清单数量往往数以万计，远远超出稽查力度而难以实施。

陕西省、福建省电力公司在 2015 年开展了基于大数据与计算智能的反窃电研究，该项研究以电能表和采集终端中的电能计量数据、事件记录、用户及终端档案信息等数据为基础，利用各类规则对异常信息进行综合判断、分析，并结合大数据挖掘技术实现海量数据准实时处理，对现场计量异常情况、窃电行为进行在线监测，发现疑似窃电用户并输出疑似窃电用户清单，同时支持动态产生异常事件告警，实现对现场窃电行为的在线诊断及窃电行为分析的全过程管理。



附图 21 案例八：数据处理分析逻辑图

电力企业通过美林数据提供的大数据分析技术实现了将同期线损率降低 1% 的目标，以平均电价 0.6 元/千瓦时计算，该项应用成果直接提升效益约 5.58 亿元。

4、设备采购供应商信用评价分析的技术标准的研究

我国电力企业主要从事发电、输电、配电的系统规划、建设、生产、运营等业务。电力装备采购，是电力系统实现外延式扩张和内涵式发展的基础，做好物资采购供应工作，是实现电力企业社会效益及企业效益最大化的基础。我国电力企业物资采购供应商数量多，规模差异大，既有世界知名跨国公司，也不乏国内中小规模企业，供应商资信情况不一，信用风险定量评估困难。

美林数据以逻辑回归方法构建模型，模型以企业违约次数、供应商主营业务收入、供应商资产负债率、供应商的行业地位、供应商项目是否得到政策支持为判别指标，并通过相关检验，证明了所建模型的有效性。通过计算供应商不违约的概率，据此得到供应商的信用得分，在此基础上利用聚类分析进行分类，确定各个类别的上下界限，最终得到供应商信用等级评判标准。

利用聚类分析对供应商信用得分进行分类，得到供应商信用等级评判标准，从而完成对设备采购供应商的定量评价，为设备的采购、供应商的选择提供了科学的决策支撑。模型基于供应商相关信息，利用数据挖掘技术，对供应商信用等级进行了定量分析，避免了主观性判别带来的弊端，为提高采购管理水平和供应商管理水平，强化对供应商的管控能力提供了科学支撑。

5、设备资产经济寿命分析的技术标准的研究

设备资产全寿命周期管理是安全管理、效能管理和全寿命周期成本管理在资产管理方面的有机结合，是使设备资产的寿命周期费用最小的一种管理理念。设备大修计划是基于设备资产状态评价结果制定的，本质是为了延长设备使用寿命、保障供电安全、提高设备效率。大修项目影响设备总成本曲线变化趋势，导致盈亏平衡点变化，结合同类型资产经济绩效刻画，对大修设备的成本、价值进行预测，得到其综合经济效益（利润），并与同类型设备经济效率的阈值进行比较分析，从而实现对大修项目辅助决策分析。

6、客户欠费风险分析的技术标准的研究

一直以来电力企业都是采用先用电后缴费的市场规则，同时由于催费措施落后，缺少有效的欠费回收风险分析手段，未形成差异化风险防控策略，使得部分电费回收周期长，最终导致电费回收成为困扰电力企业的一大问题。陕西省电力公司此前通过人工方式来识别电费回收风险，但是这种方法无法精准定位高风险

客户，并且由于风险防范时间点相对滞后，没有形成事前的标准化预警机制，难以及时制定差异化的防控措施。陕西省电力公司在 2016 年利用海量用电客户历史数据信息建立基于客户信用的客户欠费风险分析与预测模型，以高效准确识别欠费高风险客户，并制定差异化的催费措施，提高电费回收率，加强电费回收管理。

美林数据构建的客户欠费风险分析与预测模型通过分析欠费客户的基本属性特征及历史行为特征，围绕欠费情况、缴费行为等业务视角，利用机器学习算法建立客户欠费风险预测模型，提供了有效的客户电费风险预警机制。该电力企业通过实施客户欠费风险分析，精准定位了欠费高风险客户群，通过欠费风险预测模型，输出欠费高风险客户 12.5 万户，在提高电费回收率的同时大大降低了电力公司人力成本。

7、变压器预测性检修分析的技术标准的研究

近年来，我国对于变压器的检修都是采用状态检修和预防性检修相结合的方法，但随着用电方的数量飞速增长，供电方设备数量急剧增加，所要用到的电力元件也越来越多，使得整个电力传输系统变得更加复杂，所要解决的问题难度也越来越大，要想做到对已故障或即将发生故障的元件进行及时检测和维修，以及持续、稳定地将发电系统产生的电力传输到用户方，并极大地减少电力和维修费用的损耗，困难变得越来越大。

因此，美林数据通过对设备运行历史海量数据的挖掘展开预测性检修的研究，以较高的准确率预测出设备运行的未来状态，预判设备发生故障的可能性，从而达到基于设备的状态来指导检修的目的。预测性检修的分析结果对于指导检修计划编制、合理安排电网运行方式、优化计划停电策略等发挥了举足轻重的作用。

以上七项国网电力大数据分析应用案例在业务场景分析、数据资源收集与管理、分析挖掘模型建立、部署实施、运维保障等各阶段过程中参考了《信息技术大数据术语》、《信息技术 数据元素值格式记法》、《信息安全技术 数据库管理系统安全评估准则》、《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》等标准，并结合在国网电力系统进行大数据技术应用项目的实施经验和国网业务特点，后期将与国家电网相关机构，协助信标委大数据标准工作组为《信息技术 大数据 工业应用术语》、《信息技术 数据元素值表示--格式记法》、《信

息技术 大数据参考架构》、《信息技术 大数据 分析总体技术要求》、《信息安全技术 隐私保护评估方法》、《信息技术 大数据工业应用参考架构》六项标准的立项研制、内容编写等提供行业具体应用实施情况反馈和详细建议。

8.9 案例九：浙江移动“天盾”反欺诈系统

应用领域：通信领域、反欺诈

1、项目背景

近年来电信诈骗违法犯罪愈演愈烈，严重侵害人民群众利益，打击电信诈骗已经迫在眉睫。当前我省公安在各地陆续建立反欺诈中心，加强对电信诈骗的打击。面对当前诈骗方式不断翻新的情况，需要社会各界通力协作，尤其是采用高新技术对电信诈骗实现精准打击。

浙江移动公司提供的“天盾”大数据反通信欺诈系统，利用大数据及时发现疑似受害人，可以及时对用户进行劝阻，该系统自 2016 年 8 月正式上线以来，累计避免群众损失超过 4600 万元，效果显著。

2、系统的标准化研究方案及反欺诈标准应用体系建设

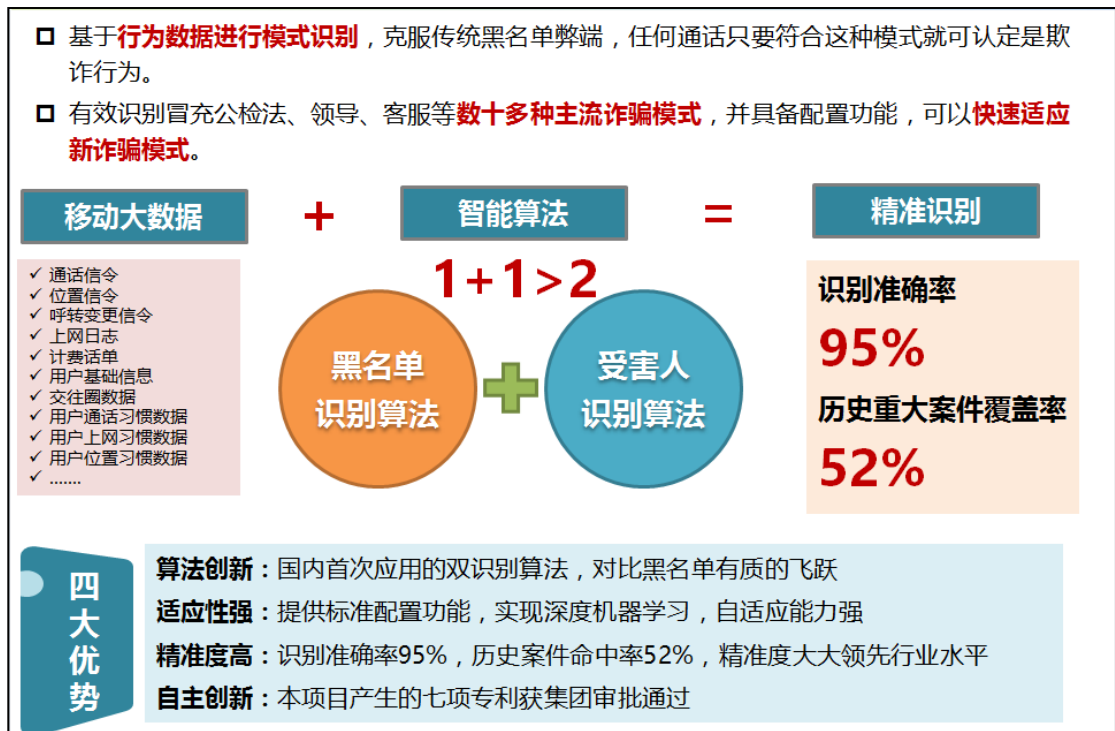
本项目系中国移动大数据在电信反欺诈领域的重点创新与成功实践。系统平台遵循《信息技术大数据系统通用规范》《信息技术大数据存储与处理系统技术要求》等基础性国标。并在架构设计、模型算法上做出自主创新，结合大数据和机器学习技术，构建双识别模型和流处理系统，搭建“天盾”大数据反欺诈系统，有效识别十多种主流诈骗模式，从识别到实施拦截仅用 2 分钟。

数据规模方面，完整 O/B/M 三域数据整合，数据量超 5PB；数据处理方面，每秒过滤 1 万次通话，每日处理通话信令量超 3 亿，设计容量超 10 亿；团队建设方面，自建数据建模团队，自主掌控社交网络、时间序列、流处理等先进技术；技术方面，通过大数据和机器学习，革新传统建模手段，基于社交网络采用随机森林算法，基于行为数据采用时间序列分析，全面实现深度机器学习，基于流处理先进技术，实现诈骗实时识别干预。同时与公安、银行开展进行深度合作，构建协同的反通信欺诈生态体系，建立反欺诈联盟，打通内外部数据合作通道，实现数据传输反馈一体化，对于电信诈骗行为进行实时有效干预。

本项目将应用和场景中提炼的需求，以及架构上的创新，向在研国标《信息技术基于大数据参考架构下的接口框架》《信息技术大数据存储与处理系统功能测试规范》等进行重要反哺。中国移动通信集团浙江有限公司也是《数据管理能力成熟度评估模型》的试点单位之一，为国标的建设做出了突出贡献。

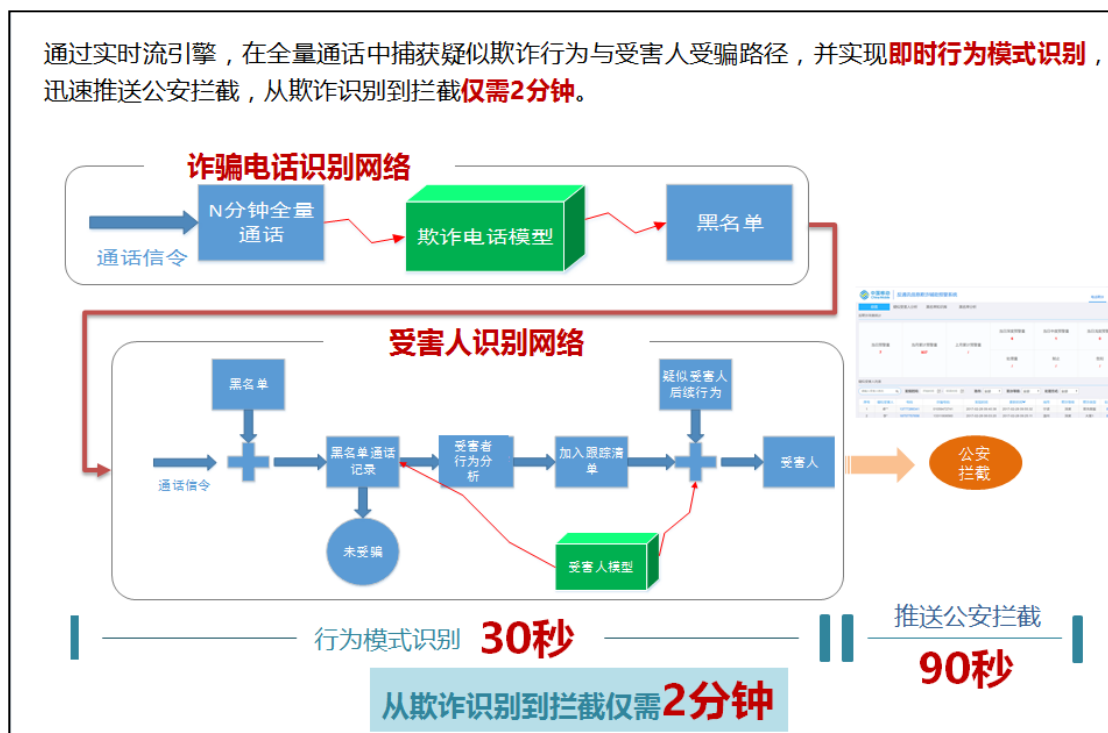
3、反欺诈技术标准研究的创新点

(1) 自主研发行业领先的双识别智能算法:国内首次应用的双识别算法，对比黑名单有质的飞跃;提供标准配置功能，实现深度机器学习，自适应能力强;识别准确率 95%，历史案件覆盖率 52%，精准度大大领先行业水平。



附图 22 案例九：业务逻辑介绍

(2) 强劲的实时事中识别干预能力:通过实时流引擎，在全量通话中捕获疑似欺诈行为与受害人受骗路径，并实现即时行为模式识别，迅速推送公安拦截，从欺诈识别到拦截仅需 2 分钟。



附图 23 案例九：系统业务流程图

(3) 互联网化运营开发模式快速响应:实现敏捷开发和迭代，36 天完成上线，按日迭代，版本发布周期从传统的 30 天缩减到 7 天，支撑的案件类型从最初的 2 类扩展到近 20 类。

(4) 专利成果突出:本项目共涉及相关八项专利，创新成果突出。

4、反欺诈标准研究的应用成果及对地方的指导意义

(1) 自 2016 年 6 月上线以来，累积电话实时识别 35703 次，道接阻止诈骗案件 11144 宗，直接挽回经济损失超过 6035 万元。

(2) 自 2016 年 6 月上线以来，使大部分诈骗手段显著失效，沉重打了诈骗分子，浙江公安部门反馈的电话欺诈案件逐月显著减少，从最高峰的 7 月 2962 件，下降到 12 月的 1799 件，下降比例达 40%。

该项目对社会安全、经济秩序管理等有重要影响，已经获得了社会、相关部门的广泛认可和好评。对各地方的公共管理、安防监控、社会监管，及其它行业（如金融、互联网等）也有重要指导意义。

8.10案例十：国家邮政局数据管控系统

应用领域： 邮政

按照国家对政府简政放权、加强事中事后监管、加强行业安全监管等相关要求，为有效解决三级邮政管理部门监管人少、事多、覆盖面大、监管基础能力薄弱、信息化基础较差等问题，依靠信息化手段创新监管方式、提升监管能力，促进企业落实安全主体责任，提高政府履职能力，国家邮政局向国家发展改革委申报了“邮政业安全生产监管信息化工程”的建设，并被列入了国家安全生产监管信息化工程（一期）的建设内容。

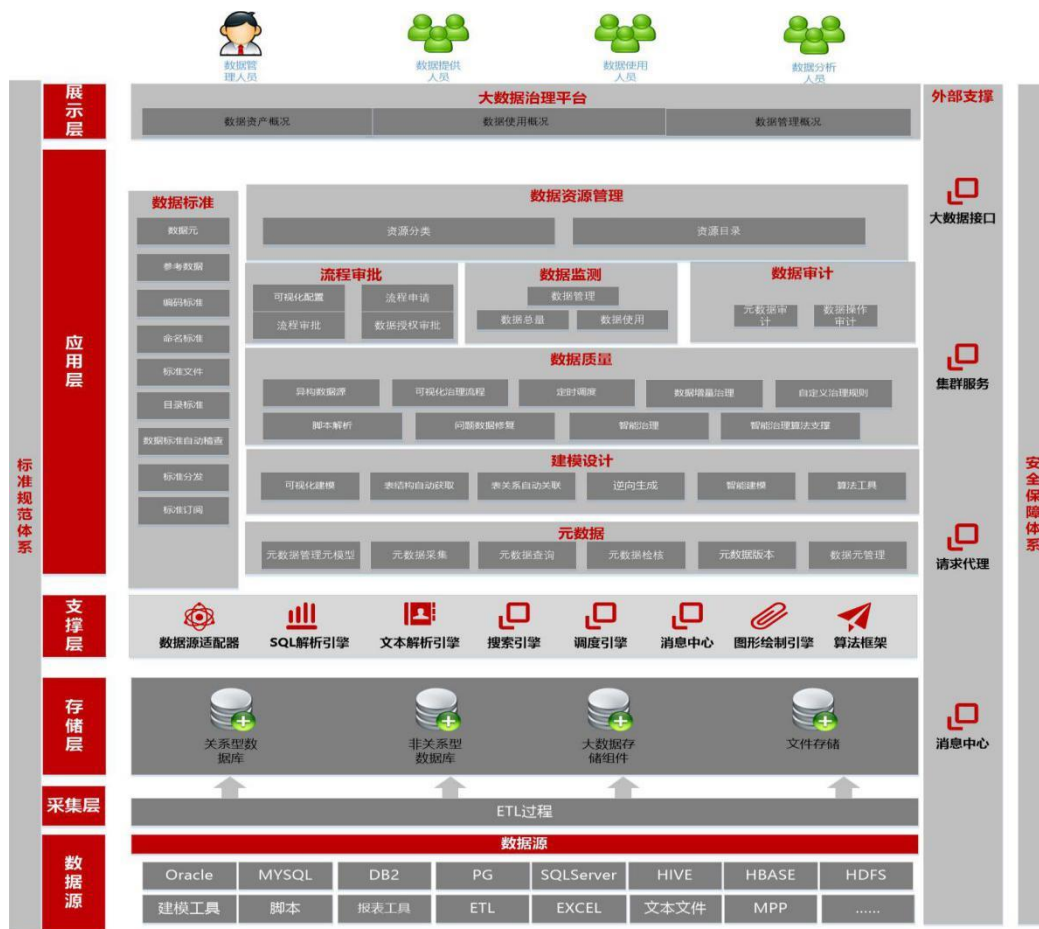
安全生产监管信息化工程（一期）国家邮政局建设部分项目于2016年6月获得批复并正式启动建设。数据资源建设是工程的重要建设内容之一，主要实现邮政行业相关的数据资源整合利用。

1、数据资源标准研究的工作目标

建设国家邮政局数据资源管控系统，实现对国家邮政局数据资源的规范性、可用性、有效性和安全性进行统一管理。并借助数据资产可视化技术，实现对国家邮政局数据资源规模及动态的统一展现，为国家邮政局相关创新型业务开展提供数据服务支撑。

2、数据资源标准应用的建设内容

国家邮政局数据管控系统实现了国家邮政局数据整体态势呈现，数据管理，数据安全以及数据服务支撑等能力。具体涵盖数据资源管理、元数据管理、数据质量管理、数据状态监测、数据审计、流程审批、数据规范管理等一系列功能和技术标准的应用场景。



附图 24 案例十：系统总体架构图

3、国家邮政局数据管控系统主要应用的关键技术包括以下几个方面

(1) 目录联动技术和标准的研究

产品创新性设计上下级目录采用接口协议（如：WebService）进行联动，通过自主研发的目录联动技术实现上级目录变动（如：增删改）实时传递到下级，对数量多、分类广、分散在不同机构的信息资源特征进行描述。目录标准创建时，只需配置好相关连接信息，通过 Webservice，调用其他系统信息到目录标准管理进行自动编目，无需进行二次创建和配置，有效节省配置时间，显著提高工作效率，极大地提高数据治理效率与质量。新增了 Webservice 消息接收，解析和转发的需求。20 个业务系统的授权，待办消息通过 webservice 接口接收和转发，其中待办消息需要进行解析并存储到数据底盘中。

(2) 数据质量自动校验技术和标准的研究

产品采用业界领先的配置化度量规则和校验方法，系统内置治理流程及多种常用检核规则，并率先设计通过规则管理页面自定义添加，结合治理规则组件、

算法组件实现自动化检核，并智能化生成数据质量评估报告。

（3）3D 可视化呈现技术

产品采用 WebGL3D 绘图标准将 JavaScript 和 OpenGL ES 2.0 结合在一起，通过增加 OpenGL ES 2.0 的一个 JavaScript 绑定实现硬件 3D 加速渲染，顺利实现复杂导航的创建和数据视觉化，进而实现复杂 3D 结构的网站页面设计。

（4）业界领先的数据安全加固技术和标准的研究

产品采用自主研发的高可靠数据安全加固技术，针对最细粒度数据单元利用数据及资源级别访问安全控制，进行多样化共享设置，并通过安全出口使用自定义或第三方安全算法进行加密，结合 SSL 传输通道加密，高效实现字段级加密及脱敏处理。

（5）数据源无缝集成技术

产品通过自创的数据源适配技术实现各种常用数据源（包括 mysql、oracle、db2、sybase、sqlserver、达梦等）之间的数据适配，可靠屏蔽数据源之间的差异，高效实现数据源的无缝集成。

4、核心功能

- （1）构建了国家邮政信息资源目录，实现了数据资源的分类管理；
- （2）构建了标准的数据服务支撑平台支撑数据应用需求；
- （3）建立了元数据标准，对元数据全生命周期统一管理；
- （4）实现了自动化的数据质量管理功能对数据资源使用情况的动态监测；
- （5）通过对数据和操作的审计，大大提升了邮政数据平台的安全性，实现事后追溯；



附图 25 案例十：资源管控系统

5、相关标准的研究成果和应用价值

- (1) 构建了国家邮政信息资源目录体系，实现了数据资源的分类管理；
- (2) 构建了标准的数据服务支撑平台，支撑数据应用需求；
- (3) 建立了标注的规范库，同时构建了元数据标准，实现了对元数据全生命周期的统一规范的管理；
- (4) 实现了自动化的数据质量管理功能；
- (5) 实现了对数据资源使用情况的动态监测；
- (6) 通过数据审计功能，实现了对数据的审计和操作的审计，大大提升了数据平台的安全性，当出现问题时可以实现事后追溯。

6、大数据标准的研制和应用及建设效果

国家邮政局数据管控系统建设过程中坚持贯彻大数据标准体系框架，并在大数据标准体系基础上结合国家邮政数据资产管控规范要求，对国家邮政局现有应用系统数据进行全面调研、梳理和分析，对国家邮政局现有数据资源进行重新规划，制定《国家邮政局数据资源目录》，完善《国家邮政局数据库设计规范》和《国家邮政局资源管理规范》。从现有应用系统中识别和抽取主数据资源，对全部应用系统数据资源进行编目。同时，对现有数据交换与共享平台（ESB）进行优化完善，实现相关系统与国家邮政局内部各系统间、国家邮政局与共建单位之间的数据交换与共享，完善《国家邮政局数据交换规范》。最终以数据资源管控系统为工具，以不断完善的邮政行业数据标准与规范为依据，以数据管控组织与

策略为支撑，对数据资源从模型设计到共享服务的全生命周期进行管控，逐步形成国家邮政局的数据管控体系。

国家邮政局数据管控系统于 2017 年 10 月份投入试运行，系统共接入 10 大类基础数据信息，对接了 28 个业务系统，实现了主数据的同步和数据交换，上线后三个月内完成数据交换 856 万次。系统主要完成了 4 个数据资源的规划包括数据资源目录规划、主数据规划、共享交换数据规划和数据标准规范的规划；完成了 3 个数据管理功能包括元数据管理、数据质量管理和数据审计管理。

在大数据标准化体系的指导下，国家邮政数据管控系统的建设，为各个行业构建大数据资产管控系统提供了技术指导意见，为国家政务信息共享平台在地方项目的落地提供了参考标准和依据。

8.11 案例十一：重庆两江大数据科技服务双创孵化基地案例

应用领域：科学研究

项目概要说明：重庆两江大数据科技服务双创孵化基地建设内容包括三部分：一是打造一个大数据众创空间+孵化器；二是建设一个大数据双创服务平台——“重庆两江新区大数据协同创新平台”；三是建立两江大数据人才培养基地，培养一大批支撑重庆大数据产业发展的高端人才。本项目聚合了重庆大数据领域高端资源，有效组织重庆多家研究机构共同开展大数据标准研究，参与了多项大数据领域标准编制，包括《信用信息征集规范第 1 部分：总则》《信用标准体系总体架构》（报批中），《信息技术大数据面向应用的基础计算平台基本性能要求》（编写中）等。目前，项目已成功通过国家发改委 2017 年项目中期检查，成果显著。同时，本项目已获批重庆市大数据创新应用试点项目（大数据产业公共服务平台类）。已成为重庆地区大数据产业公共服务的桥头堡，将在今后对国内外大数据的标准化发展进行前瞻性研究，同时发挥其在标准化、专业事业、评估方面的咨询作用，积极参与制定全市大数据收集、管理、开放、应用等标准规范建设。

1、 大数据众创空间+孵化器

依托市级大数据众创空间——重科“数聚邦”，打造“线上+线下”的孵化

体系。在现有“数据帮”的基础上，扩建面积到1万平方米，培养孵化大数据双创企业80家。建设成为集“信息发布、产品展示、数据开放、创新创业资源对接、信用管理”于一体的在线创业服务平台。

2、大数据标准化科技服务协同创新平台

平台致力于充分整合市内外大数据技术创新要素，完善大数据产业技术创新链条，促进创新资源的优化配置、开放共享及金融等各类资源的结合，面向我市大数据领域科技型中小企业提供大数据产业共性技术服务。平台坚持“标准引导→业务转变→数据应用变革”的大数据驱动路线，通过建立完善标准化体制机制，优化新型标准体系，夯实标准化技术基础，增强标准化服务能力，充分发挥“标准化+”效应，推动重庆科技创新发展。大数据标准化科技服务协同创新平台建设包括标准化服务平台、大数据智能应用工具集、大数据基础信息库、大数据产品设计与测试。

（1）打造重庆大数据技术标准创新平台

加快培育标准化服务机构。以重庆两江新区大数据科技服务协同创新平台为载体，建成长江上游地区的标准创新孵化平台、标准技术服务公共平台和标准化人才培养平台。推动标准化服务提供主体和方式的多元化。支持各类标准化科研机构开展标准化理论、方法、规划、政策研究，提升标准化科研水平。加快标准化科研机构改革，激发科研人员创新活力，提升服务产业和企业能力。

（2）大数据智能应用工具集

面向有大数据技术研发、应用需求但是又缺乏技术、软硬件设施支撑的大数据科技型中小企业，提供功能丰富、方便快捷的大数据智能应用工具系统。

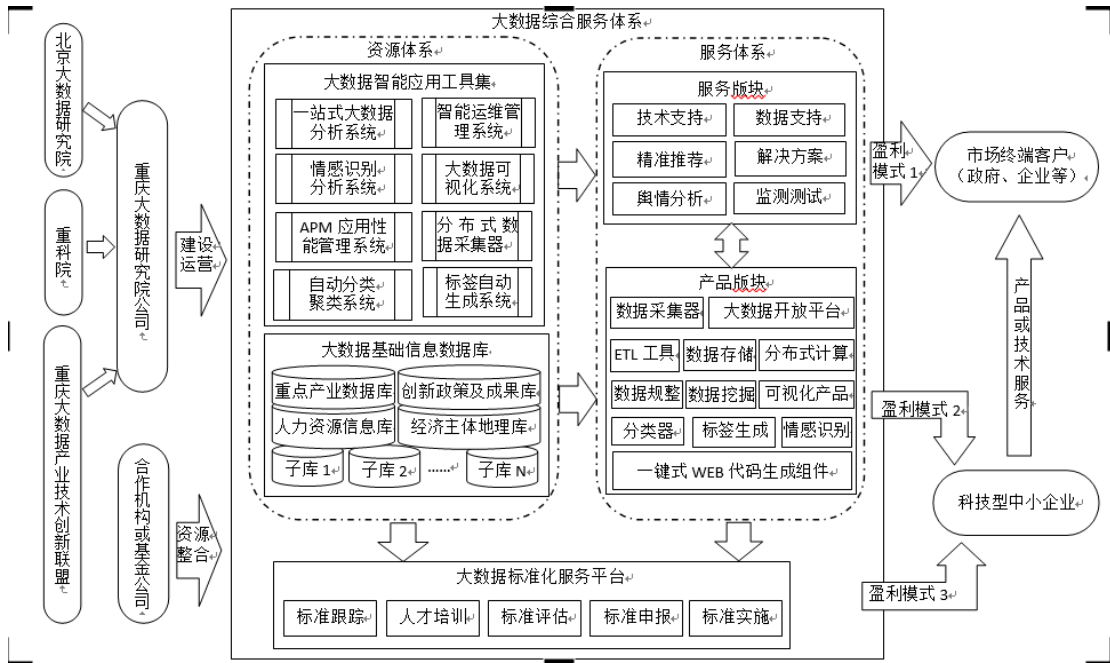
（3）大数据基础信息数据库

面向有大数据研究、数据交互及应用需求但是又缺乏数据源的大数据科技型中小企业提供基础信息数据库，建设内容为“3+N+1”数据库模式，“3”即重点企业企业数据库、人力资源信息库和两江创新创业数据库3个基础数据库。“N”即在以上3个基础数据库的基础上，将共享数据库持续延伸扩展到N个其他数据库。“1”即通过搭建1个公共服务数据交换平台，向用户提供最佳性能、灵活快捷的应用交付服务，实现共享、简单、快捷、灵活的应用。

（4）大数据产品设计与测试

面向有大数据产品开发需求但是又缺乏技术、软硬件设施支撑的大数据科技型中小企业，提供大数据产品设计、产品测试、认证标准检索与管理咨询服务，主要包括需求分析审查、设计审查、代码审查、单元测试、功能测试、性能测试、可恢复性测试、资源消耗测试、并发测试、健壮性测试、安全测试、安装配置测试、可移植性测试、文档测试以及验收测试等内容，预防和发现产品缺陷。

3、大数据标准化人才培养基地



附图 26 案例十一：重庆两江大数据科技服务双创孵化基地运营模式

与重庆互联网学院、北京大数据研究院、重庆大学、重庆理工大学、重庆电讯职业学院、重庆工程学院共建大数据标准化人才培养基地。面向有技术人才水平提升需求的大数据科技型中小企业，提供专业的大数据标准化教学培训课程体系，全面提升企业家的标准化战略创新能力、企业管理层和员工的标准化技能、标准化专业技术人员的专业水平、政府公务人员和社会公众的标准化知识，为重庆标准化改革发展提供人才保障。

8.12案例十二：本溪大健康服务平台

应用领域：保险业、健康管理、物联网、大数据、云计算

1、医疗大数据标准化管理的探索和突破，包括在保险领域的探索，以及医疗平台服务标准的建设。

(1) 医疗大数据标准化管理在保险领域的探索

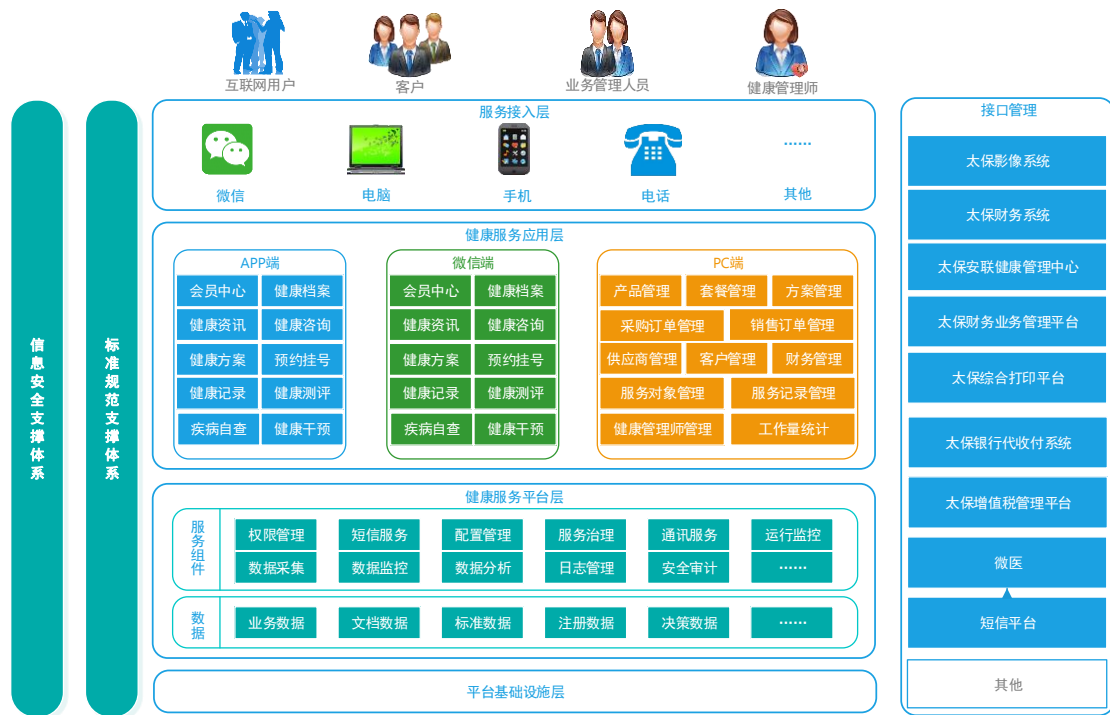
为了全面贯彻落实国家层面《关于促进健康服务业发展的若干意见》（国发〔2013〕40号）和国务院《关于创新重点领域投融资机制鼓励社会投资的指导意见》（国发〔2014〕60号），《全国医疗卫生服务体系规划纲要（2015—2020年）》，国务院《关于城市公立医院综合改革试点的指导意见》，《政府和社会资本合作项目政府采购管理办法》（财库〔2014〕215号）文件精神，推进健康产业发展，深化医改、改善民生，优化医疗资源配置和健康服务升级，提升全民健康水平，本溪进行面向保险行业的在线健康服务平台项目建设。

在线健康服务平台是以医疗健康信息平台为基础，采用“互联网+健康”模式，运用云计算、物联网、互联网、健康感知设备等新技术，全面整合医疗卫生资源和体系，为居民提供全面健康服务和便捷医疗服务的综合服务平台。本项目立足于未来健康产品专业化管理，并整合自身自有服务资源和第三方服务资源，将现有整合资源服务与独立服务项目整合打包成在线健康产品，打造保险行业健康产品与客户粘性最强、与服务结合最紧、最能彰显特色的在线健康产品，在利用标准化进行医疗健康资源整合方面进行创新。

(2) 健康大数据服务标准对医疗体系的优化升级

在线健康服务平台对线上线下医疗健康资源进行标准化梳理，从技术上打通线上和线下的通道，并针对不同的人群，进行定位分析，依据统一规范提供标准化针对性的服务：针对慢病人群提供慢病服务；针对妇幼人群提供妇幼孕育服务；针对商业健康保险人群提供商业健康保险服务、健康档案服务；针对网络医疗人群提供预约服务、咨询服务、支付服务、境外就医、医保政策咨询。在线健康服务平台，利用各种先进的技术，为居民提供超预期的医疗健康服务体验。同时，也综合利用公立医疗健康资源、社会医疗健康资源，构建一个信息流、商流、资金流高效协同的健康产业联盟，利用信息化平台让各类健康资源拥有一个统一的

规范的协同体系，让联盟在标准的引领和规范下实现持续健康发展。



附图 27 案例十二：总体架构

2、本案例先进性和技术路线

(1) 健康信息资源的标准化管理。

在线健康服务平台基于电子病历、健康档案、全员人口信息三大库，聚合居民在家庭、体检中心、基层医院、各级专科医院中的行为数据和医疗记录的全过程档案，形成全链条健康大数据中心，实现对健康信息标准化建模，并规范和安全地提供妇幼和慢病健康大数据的分析和运营服务，同时对数据的聚合与应用的全过程进行痕迹化管理和实时监控。

(2) 在基于“互联网+”模式的大健康服务平台建设上进行标准化创新

本溪市随着人口健康信息平台的建设，形成了本溪各医疗卫生机构的信息共享和业务合作的基础。以人口健康信息平台为基础，推动了基于“互联网+”模式的大健康服务平台的建设。

本溪市大健康服务平台运营和建设包括两部分：

一是政府和企业合作探索新型运营健康运营服务模式，由政府给予政策支持，通过企业投入软硬件基础设施和运营，政府以购买服务形式，构建新的产业联盟。

二是分阶段开展六大应用服务：

- 1) 慢病服务。以慢病管理为切入点，实现居民健康的全周期智能化管理。
- 2) 智能预约服务。推进预约为入口的便民服务建设，增加覆盖，建设全业务的预约。
- 3) 妇幼孕育服务。妇幼孕育服务针对的孕产期的妇女和不满 6 周岁的儿童及其家人提供的全周期的资讯、咨询、检查预约、各项事务办理服务。为孕产妇和新生儿提供事前、事后、事中的各类便民服务。
- 4) 商业健康保险服务。帮助居民处理商业健康保险的购买、实施结算、审核等服务，以信息化手段提供全流程服务。
- 5) 网上咨询服务。提供居民疾病咨询、用药咨询、检查检验结果咨询、导医咨询等服务。
- 6) 支付服务。多种窗口和形式支持方便、快捷、随时随地完成快速缴费的平台。

通过本项目的建设，本溪市在推进健康产业发展，深化医改、改善民生，优化医疗资源配置和健康服务升级方面取得了明显进步，健康医疗资源整合标准化管理为提升全民健康水平提供了明确的方向，也为辽宁其他城市医疗健康资源的整合和全民健康事业的发展提供了经验参考。

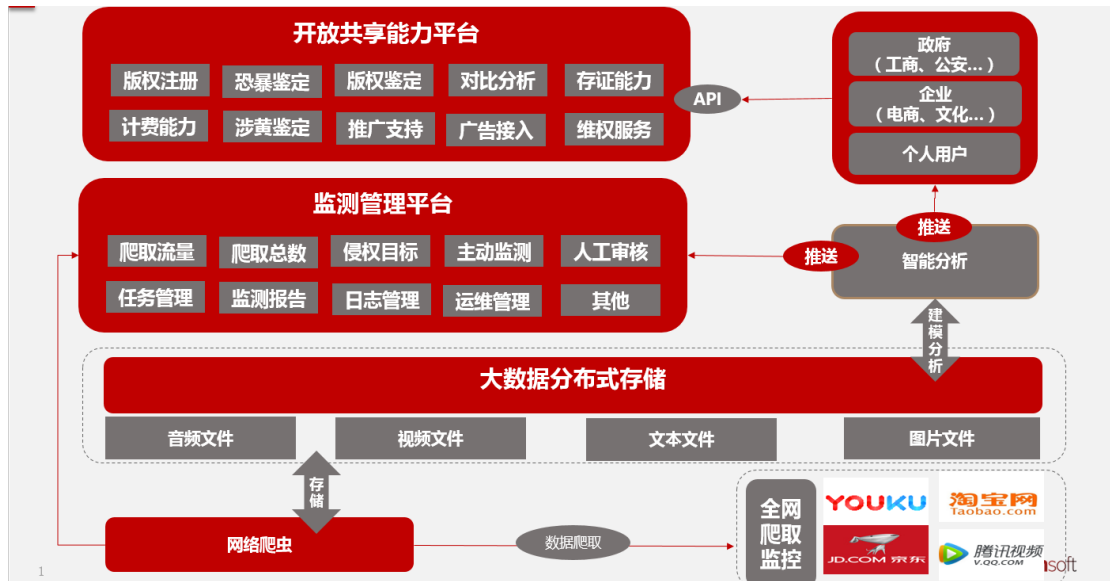
8.13 案例十三：CCDI 版权监测案例

应用领域：广电、知识产权

CCDI 项目是国家新闻出版广电总局和贵州省政府根据党中央、国务院关于实施“互联网+”行动计划，推进传统媒体与新兴媒体融合发展和《促进大数据发展行动纲要》，重点谋划和全力推进的新闻出版广电行业战略性新兴产业项目。

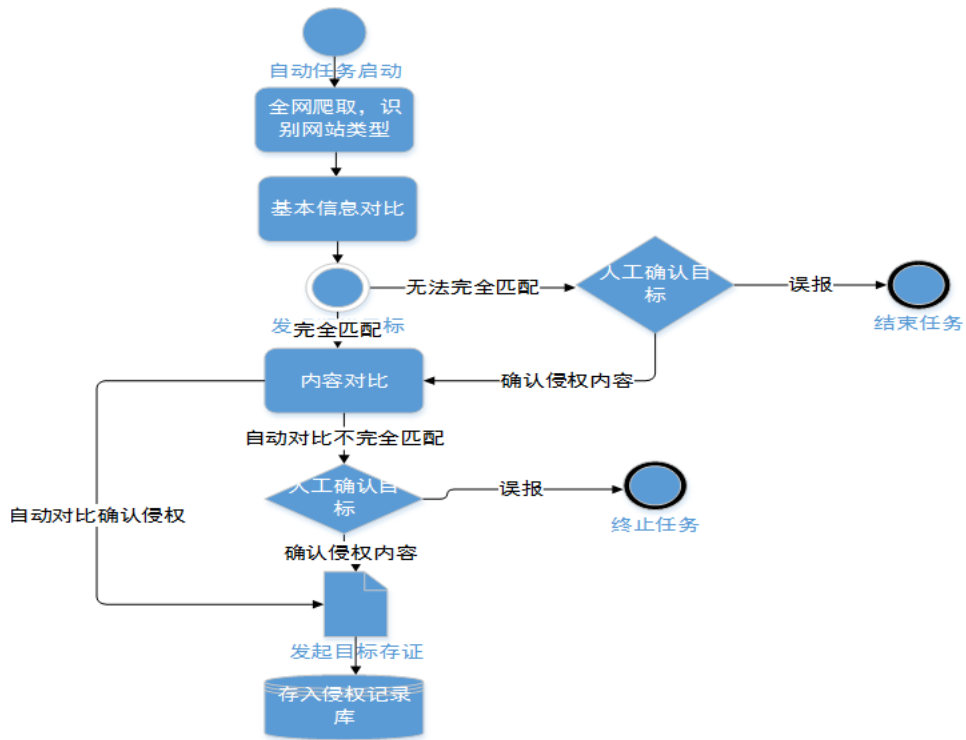
版权云大数据监测平台旨在通过对全网（互联网+广电网）媒体（包含文字，图片，视频，音频）侵权内容的爬取，借助分布式存储计算平台构架，将海量的泛互联网内容进行实时比对，由系统自动发现网络版权违规行业，并由系统自动完成相关存证工作，最后，将版权信息、商标专利、版权服务等数据可视化分析，实现从后端到前端的一体化服务打通，为中国互联网版权监测与版权相关服务生态系统的打造提供了良好的基础。

版权监测平台的总体架构分为四层，底层为数据获取层，主要通过网络爬虫以及合作商的接口开放等方式获取正版版权相关信息；上一层为数据存储层，将文件、图片、音频、视频等不同类型的文件存储到分布式存储计算平台中；再上一层，通过监测管理平台的主动监测与人工审核机制，利用大数据技术实现版权鉴定与识别；最上层通过开放能力平台，实现对政府、对企业和对个用户的版权相关服务。



附图 28 案例十三：平台架构图

在获取到版权监测对象的不同类型数据后，系统自动调用比对接口实现对内容版权的真实性研判，一旦确认侵权，则通过存证接口，实现侵权的取证与存证，为后续的一系列相关服务提供支撑。版权监测流程如下：



附图 29 案例十三：版权监测流程图

版权监测大数据平台建设过程中始终坚持以标准规范引领系统建设的原则，过程中参考了《数字产品版权保护服务平台技术要求和测评方法》等标准，结合项目建设经验，总结并编撰了《版权数据互联网采集与清洗规范》《基于源内容对象嵌入关联标识符编码规范》《数据安全管理办法》《视频数据共享交换接口规范》《小文件数据存储与检索策略规范》等标准与规范。

通过版权监测大数据平台建设，利用先进的大数据技术，对全网媒体进行监测，打击网络盗版，维护知识产权的作用，从而更好地促进了我国文化产业发展，进一步增强文化自觉，使我国更加深入参与全球文明对话，增强国际话语权，提高国家文化软实力。



附图 30 案例十三：系统截图-版权云监测平台的视窗

版权监测大数据平台已于 2017 年 5 月上线试运行，目前已采集数据超过 2P，经比对的数据超过 10 亿条，系统记录的疑似侵权行为数据超过 20 万，存证的侵权网站超过 500 个。