

大 数 据 白 皮 书

(2018年)

中国信息通信研究院
2018年4月

CAICT 中国信通院

版权声明

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

CAICT 中国信息通信研究院

CAICT 中国信通院

前 言

随着信息技术和人类生产生活交汇融合，全球数据呈现爆发增长、海量集聚的特点，对经济发展、社会治理、国家管理、人民生活都产生了重大影响。近年来，我国的大数据在政策、技术、产业、应用等方面均获得了长足发展。

本白皮书是继《大数据白皮书（2014 年）》、《大数据白皮书（2016 年）》之后中国信息通信研究院第三次发布大数据白皮书。本白皮书在前两版的基础上，集中梳理介绍了我国大数据的最新发展态势和成果。本白皮书首先对我国大数据的发展进行了回顾与梳理，对大数据发展的总体情况进行了研判。白皮书还对大数据的技术发展、行业应用进行了梳理，探讨了利用大数据提升政府治理能力的关键问题，并对数据法律法规体系和地方大数据产业发展的新实践、新动向进行了追踪研究，力求重点介绍我国大数据发展的最新成果。随着近年来数据资产管理的概念逐渐深入人心，本白皮书专门用一章对这一问题进行了探讨。最后，结合我国大数据发展最新状况及问题，提出了进一步促进大数据发展的相关策略建议。

CAICT 中国信通院

目 录

一、大数据发展概述.....	1
二、大数据政策环境.....	4
（一）我国大数据政策回顾与大数据战略的提出.....	4
（二）国家大数据战略的内涵.....	5
三、大数据技术创新.....	8
（一）数据分析技术.....	8
（二）事务处理技术.....	11
（三）数据流通技术.....	13
四、大数据与实体经济融合应用.....	15
（一）行业应用大数据的特点.....	15
（二）行业应用大数据的深层分析.....	16
（三）行业应用大数据的关键因素.....	19
五、政务大数据发展.....	19
（一）政务大数据总体要求.....	20
（二）政务信息系统整合.....	22
（三）政务信息共享交换.....	24
（四）政务信息对外开放.....	26
六、地方大数据产业发展.....	29
（一）大数据产业发展主要模式.....	29
（二）地方大数据产业发展策略分析.....	30
（三）地方大数据产业发展成效与问题.....	32
七、数据资产管理体系.....	33
（一）数据资产管理的定位和范畴.....	33
（二）数据资产管理面临的挑战.....	35
（三）数据资产管理的发展趋势.....	37
八、数据立法重点.....	41
（一）政府数据开放.....	41
（二）个人信息保护.....	43
（三）数据流通规则.....	46
九、建议与展望.....	48
（一）制度与技术双管齐下，打破数据孤岛.....	48
（二）内部与外部多重并举，推动数据治理.....	49
（三）业务与数据加速融合，深化数据应用.....	49
（四）监管与自律同时推进，保障数据安全.....	50

图表目录

表 1	数据流通技术工具对比.....	14
表 2	部分行业代表性企业大数据应用情况.....	17
表 3	地方政府大数据产业发展定位	31
图 1	我国大数据市场产值图.....	2
图 2	事务型数据库架构演进图.....	11
图 3	我国主要政府数据开放平台上线时间.....	28
图 4	数据资产管理在大数据体系中的定位.....	34
图 5	数据资产管理体系架构.....	35

一、大数据发展概述

大数据是信息化发展的新阶段。随着信息技术和人类生产生活交汇融合,互联网快速普及,全球数据呈现爆发增长、海量集聚的特点,对经济发展、社会治理、国家管理、人民生活都产生了重大影响。在刚刚过去的 2017 年里,大数据在政策、技术、产业、应用等多个层面都取得了显著进展。

在政策层面,大数据的重要性进一步得到巩固。党的十九大提出“推动互联网、大数据、人工智能和实体经济深度融合”,习近平总书记在政治局集体学习中深刻分析了我国大数据发展的现状和趋势,对我国实施国家大数据战略提出了更高的要求。

在技术层面,以分析类技术、事务处理技术和流通类技术为代表的大数据技术得到了快速的发展。以开源为主导、多种技术和架构并存的大数据技术架构体系已经初步形成。大数据技术的计算性能进一步提升,处理时延不断降低,硬件能力得到充分挖掘,与各种数据库的融合能力继续增强。

在产业层面,我国大数据产业继续保持高速发展。权威咨询机构 Wikibon 的预测表示,大数据在 2018 年将深入渗透到各行各业(every business)¹。对于我国大数据产业的规模,目前各个研究机构均采用间接方法估算。中国信息通信研究院结合对大数据相关企业的调研测算,2017 年我国大数据产业²规模为 4700 亿元人民币,同比增长

¹ 来源:

<https://siliconangle.com/blog/2017/12/20/wikibons-2018-predictions-data-cloud-transform-ever>

² 指以数据生产、采集、存储、加工、分析、服务为主的相关经济活动,包括数据资源建设、大数据软硬件产品的开发、销售和租赁活动,以及相关信息技术服务。

30%。在这其中，大数据软硬件产品的产值约为 234 亿元人民币，同比增长 39%。而中国信息通信研究院《中国数字经济发展与就业白皮书（2018 年）》中的数据显示，2017 年我国数字经济总量达到 27.2 万亿元，同比名义增长超过 20.3%，占 GDP 比重达到 32.9%。在这其中，以大数据为代表的新一代信息技术对于数字经济的贡献功不可没。

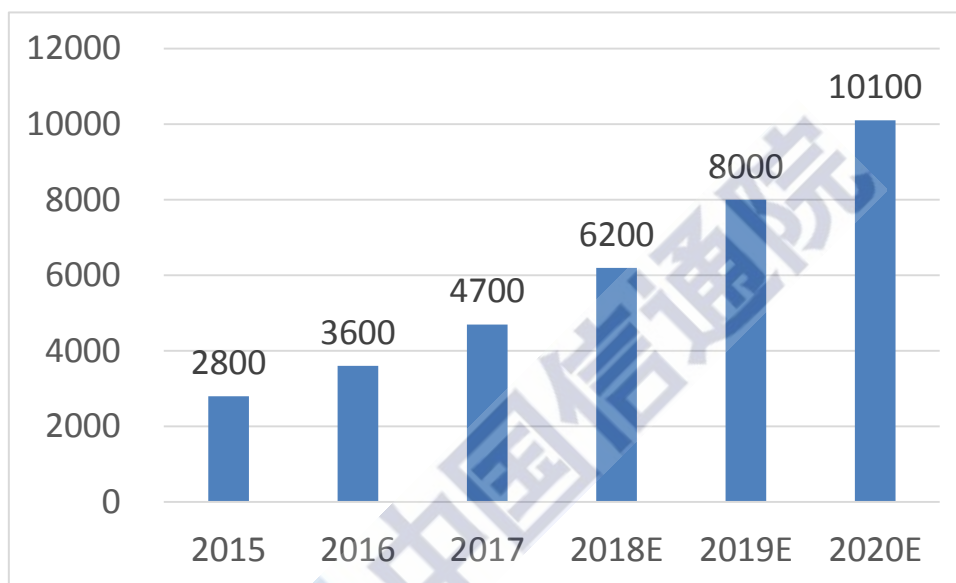


图 1 我国大数据市场产值图（单位：亿元）

在应用层面，大数据在各行业的融合应用继续深化。大数据企业正在尝到与实体经济融合发展带来的“甜头”。利用大数据可以对实体经济行业进行市场需求分析、生产流程优化、供应链与物流管理、能源管理、提供智能客户服务等，这不但大大拓展了大数据企业的目标市场，更成为众多大数据企业技术进步的重要推动力。随着融合深度的增强和市场潜力不断被挖掘，融合发展给大数据企业带来的益处和价值正在日益显现。根据中国信息通信研究院 2017 年大数据产业地图的统计，为金融、政务、电商三个行业提供大数据产品和解决方案的企业最多，分别占比 63%、57%、47%。但实践中仍然面临着缺乏

高质量数据、缺乏平台级工具、缺乏成熟商业模式等一系列问题，阻碍了实体经济行业充分利用大数据的价值。

在利用大数据提升政府治理能力方面，我国在 2017 年出台了《政务信息系统整合共享实施方案》、《政务信息资源目录编制指南（试行）》等多项政策文件推进政府数据汇聚、共享、开放，取得了诸多进展。各地纷纷将大数据作为提升政府治理能力的重要手段，通过高效采集、有效整合、深化应用政府数据和社会数据，提升政府决策和风险防范水平，提高社会治理的精准性和有效性。

在地方大数据发展实践方面，截至 2018 年 2 月底，我国各地方政府对外公布了超过 110 份大数据相关政策文件，覆盖全国 31 个省级行政区划。总体来看，我国大数据产业目前仍处于蓬勃发展阶段，各地更加注重结合当地发展特色和优势进行大数据产业发展，区域协调的发展局面正在形成。

在大数据的发展过程中，无论是政府还是企业，近年来都愈发关注数据治理和数据资产管理的重要性。2018 年 3 月，银监会出台《银行业金融机构数据治理指引》，要求银行金融机构建立自上而下、协调一致的数据治理体系。企业的数据资产管理也正在从理论走向实践，为大数据应用打下坚实的基础。为应对大数据发展带来的各种问题和需求，各国政府在立法方面也动作频频，在政府数据开放、个人信息保护和数据跨境流动方面都有了一些进展。无论是政策还是立法，都旨在实现数据价值的安全释放，提升数据管理的科学化水平。

我国要实现从“数据大国”向“数据强国”转变，还面临诸多挑

战。一是技术创新与支撑能力依然不够，我国无论是新型计算平台、分布式计算架构，还是大数据处理、分析和呈现方面与国外均存在较大差距，总体上难以满足各行各业大数据应用需求。二是信息安全和数据管理体系仍未建立，数据所有权、隐私权等相关法律法规和信息安全、开放共享的规范和标准缺乏或可操作性不强，技术安全防范和管理能力不够。三是人才队伍建设亟需加强，大数据人才远不能满足发展需要，尤其是缺乏既熟悉行业业务需求，又掌握大数据技术与管理的复合型人才。未来，需要我们继续坚持国家大数据战略，审时度势精心布局，努力开拓大数据发展新局面，更好服务我国经济社会发展和人民生活改善。

二、大数据政策环境

在刚刚过去的 2017 年里，大数据从政策层面备受关注。在党的十九大报告“贯彻新发展理念，建设现代化经济体系”一章中，专门提到“推动互联网、大数据、人工智能和实体经济深度融合”，高屋建瓴地指出了我国大数据发展重点方向。2017 年 12 月 8 日，十九届中共中央政治局就实施国家大数据战略进行了集体学习，习近平总书记深刻分析了我国大数据发展的现状和趋势，对我国实施国家大数据战略提出了五个方面的要求。本章将对国家大数据政策进行梳理，并对国家大数据战略的内涵进行分析。

（一）我国大数据政策回顾与大数据战略的提出

2014 年，大数据首次写入政府工作报告，而这一年也成为实际

意义上的“中国大数据政策元年”。从这一年起，“大数据”逐渐成为各级政府关注的热点，政府数据开放共享、数据流通与交易、利用大数据保障和改善民生等概念逐渐深入人心。

2015 年 8 月 31 日，国务院正式印发了《促进大数据发展的行动纲要》（以下简称“《行动纲要》”），成为我国发展大数据产业的战略性指导文件。《行动纲要》作为我国推进大数据发展的战略性、指导性文件，充分体现了国家层面对大数据发展的顶层设计和统筹布局，为我国大数据应用、产业和技术的发展提供了行动指南。

2016 年，《中华人民共和国国民经济和社会发展第十三个五年规划纲要》（以下简称“《十三五规划纲要》”）正式公布。“十三五规划纲要”的第二十七章题目为“实施国家大数据战略”。这也是“国家大数据战略”首次被公开提出。《十三五规划纲要》对“国家大数据战略”的阐释，成为各级政府在制订大数据发展规划和配套措施时的重要指导，对我国大数据的发展具有深远意义。

2016 年底，工业和信息化部正式发布《大数据产业发展规划（2016-2020 年）》。《大数据产业发展规划》以大数据产业发展中的关键问题为出发点和落脚点，明确了“十三五”时期大数据产业发展的指导思想、发展目标、重点任务、重点工程及保障措施等内容，成为大数据产业发展的行动纲领。农业林业、环境保护、国土资源、水利、交通运输、医疗健康、能源等主管部门纷纷出台了各自行业的大数据相关发展规划，大数据的政策布局逐渐得以完善。

（二）国家大数据战略的内涵

全面准确的理解国家大数据战略的内涵与意义，才能形成广泛的社会共识、充分的调动社会资源、完成构建国家大数据体系的各项任务。全面深入了解大数据及其相关技术的发展脉络和历史轨迹，可以引导我们准确深刻的把握大数据与国家总体目标相关性和内生性。2017年12月8日，中共中央政治局就实施国家大数据战略进行第二次集体学习，习近平总书记在主持学习时，深刻分析了我国大数据发展的现状和趋势，对我国实施国家大数据战略提出了五个方面的要求，一是推动大数据技术产业创新发展；二是构建以数据为关键要素的数字经济；三是运用大数据提升国家治理现代化水平；四是运用大数据促进保障和改善民生；五是切实保障国家数据安全与完善数据产权保护制度。我们认为，上述五大要求构成了国家大数据战略的“五大内涵”。

一是推动大数据技术产业创新发展。总书记指出，我们要瞄准世界科技前沿，集中优势资源突破大数据核心技术，加快构建自主可控的大数据产业链、价值链和生态系统。近年来，我国在大数据技术产业方面取得了不少突破。2014-2016年，百度、阿里和腾讯先后拿下国际上知名的 **Sort Benchmark** 大赛冠军。这个竞赛全面比拼分布式系统软件架构能力，包括如海量数据分布式存储、计算任务切片调度等方面的能力。而这一赛事 2014 年之前的冠军均被微软、Yahoo、亚马逊等包揽。这从一个侧面反映了我国产业界在大数据处理技术水平的快速提升，但是在互联网与大数据技术的创新与发展方面，同世界先进水平相比还有很大距离。

二是构建以数据为关键要素的数字经济。总书记提出，要坚持以供给侧结构性改革为主线，加快发展数字经济，推动实体经济和数字经济融合发展，推动互联网、大数据、人工智能同实体经济深度融合，继续做好信息化和工业化深度融合这篇大文章，推动制造业加速向数字化、网络化、智能化发展。2016年，我国数字经济总量达22.6万亿元，占GDP比重达30.3%。数字经济已经成为带动中国经济增长的核心动力。工业互联网、分享经济、网络零售、移动支付等领域的快速发展，既为大数据的发展提供了重要应用场景，也对大数据产业的技术水平提升起到了促进作用。

三是要运用大数据提升国家治理现代化水平。总书记强调，要建立健全大数据辅助科学决策和社会治理的机制，推进政府管理和社会治理模式创新，实现政府决策科学化、社会治理精准化、公共服务高效化。要实现这一目标，不但要重点推进政府数据本身的开放共享，还应当将各级政府的平台与社会多方数据平台进行互联与共享，并通过大数据管理工具和方法，全面提升国家治理现代化水平。

四是要用大数据促进保障和改善民生。总书记指出，大数据在保障和改善民生方面大有作为。要坚持问题导向，抓住民生领域的突出矛盾和问题，强化民生服务，弥补民生短板。民生大数据应用一向是大数据的重点行业应用，医疗、教育、社保、交通等行业的大数据应用在2017年也不断取得突破。大数据在流行病预测、个性化医疗、智能交通、治安管理等更广泛的社会场景中，将为增进民生福祉创造更大的技术红利。

五是要切实保障国家数据安全。总书记强调，要加强关键信息基础设施安全保护，强化国家关键数据资源保护能力，增强数据安全预警和溯源能力。要加强政策、监管、法律的统筹协调，加快法规制度建设。目前，关键数据基础设施的公权力属性、数据的生成、数据的权属、数据的开放、数据的流通、数据的交易、数据的保护、数据的治理以及法律责任等问题，都亟需得到法律的确认。

以上五个角度共同构成了国家大数据战略的主要内涵。大数据是信息化发展的新阶段，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。

三、大数据技术创新

如今，大数据技术体系纷繁复杂，但其中有诸多技术格外受到关注。随着社交网络的流行导致大量非结构化数据出现，传统处理方法难以应对，数据处理系统和分析技术开始不断发展。从 2005 年 Hadoop 的诞生开始，形成了数据分析技术体系这一热点。伴随着数据量的急剧增长和核心系统对吞吐量以及时效性的要求提升，传统数据库需要向分布式转型，形成了事务处理技术体系这一热点。然而，时代的发展使得单个企业、甚至单个行业的数据都难以满足要求，数据融合的价值更加显现，形成了数据流通技术体系这一热点。本章将对数据分析、事务处理、数据流通这三类典型的技术体系的最新进展进行介绍。

（一）数据分析技术

从数据在信息系统中的生命周期看，数据分析技术生态主要有 5

个发展方向，包括数据采集与传输、数据存储与管理、计算处理、查询与分析、可视化展现。在数据采集与传输领域渐渐形成了 Sqoop、Flume、Kafka 等一系列开源技术，兼顾离线和实时数据的采集和传输。在存储层，HDFS 已经成为了大数据磁盘存储的事实标准，针对关系型以外的数据模型，开源社区形成了 K-V (key-value)、列式、文档、图这四类 NoSQL 数据库体系，Redis、HBase、Cassandra、MongoDB、Neo4j 等数据库是各个领域的领先者。计算处理引擎方面，Spark 已经取代 MapReduce 成为了大数据平台统一的计算平台，在实时计算领域 Flink 是 Spark Streaming 强力的竞争者。在数据查询和分析领域形成了丰富的 SQL on Hadoop 的解决方案，Hive、HAWQ、Impala、Presto、Spark SQL 等技术与传统的大规模并行处理（massively parallel processor, MPP）数据库竞争激烈，Hive 还是这个领域当之无愧的王者。在数据可视化领域，敏捷商业智能（business intelligence, BI）分析工具 Tableau、QlikView 通过简单的拖拽来实现数据的复杂展示，是目前最受欢迎的可视化展现方式。

相比传统的数据库和 MPP 数据库，Hadoop 最初的优势来源于良好的扩展性和对大规模数据的支持，但失去了传统数据库对数据精细化的操作，包括压缩、索引、数据的分配裁剪以及对 SQL 的支持度。经过 10 多年的发展，数据分析的技术体系渐渐在完善自己的不足，也融合了很多传统数据库和 MPP 数据库的优点，从技术的演进来看，大数据技术正在发生以下变化：

- 1) 更快

Spark 已经替代 MapReduce 成为了大数据生态的计算框架，以内存计算带来计算性能的大幅提高，尤其是 Spark2.0 增加了更多了优化器，计算性能进一步增强。

2) 流处理的加强

Spark 提供一套底层计算引擎来支持批量、SQL 分析、机器学习、实时和图处理等多种能力，但其本质还是小批的架构，在流处理要求越来越高的现在，Spark Streaming 受到 Flink 激烈的竞争。

3) 硬件的变化和硬件能力的充分挖掘

大数据技术体系本质是数据管理系统的一种，受到底层硬件和上层应用的影响。当前硬件的芯片的发展从 CPU 的单核到多核演变转化为向 GPU、FPGA、ASIC 等多种类型芯片共存演变。而存储中大量使用 SSD 来代替 SATA 盘，NVRAM 有可能替换 DRAM 成为主存。大数据技术势必需要拥抱这些变化，充分兼容和利用这些硬件的特性。

4) SQL 的支持

从 Hive 诞生起，Hadoop 生态就在积极向 SQL 靠拢，主要从兼容标准 SQL 语法和性能等角度来不断优化，层出不穷的 SQL on Hadoop 技术参考了很多传统数据库的技术。而 Greenplum 等 MPP 数据库技术本身从数据库继承而来，在支持 SQL 和数据精细化操作方面有很大的优势。

5) 深度学习的支持

深度学习框架出现后，和大数据的计算平台形成了新的竞争局面，以 Spark 为首的计算平台开始积极探索如何支持深度学习能力，

TensorFlow on Spark 等解决方案的出现实现了 TensorFlow 与 Spark 的无缝连接，更好地解决了两者数据传递的问题。

（二）事务处理技术

随着移动互联网的快速发展，智能终端数量呈现爆炸式增长，银行和支付机构传统的柜台式交易模式逐渐被终端直接交易模式替代。以金融场景为例，移动支付以及普惠金融的快速发展，为银行业、支付机构和金融监管机构带来了海量高频的线上小额资金支付行为，生产业务系统面临大规模并发事务处理要求的挑战。

传统事务技术模式以集中式数据库的单点架构为主，通过提高单机的性能上限适应业务的扩展。而随着摩尔定律的失效（底层硬件的变化），单机性能扩展的模式走到了尽头，而数据交易规模的急速增长（上层应用的变化）要求数据库系统具备大规模并发事务处理的能力。大数据分析系统经过 10 多年的实践，积累了丰富的分布式架构的经验，Paxos、Raft 等一致性协议的诞生为事务系统的分布式铺平了道路。新一代分布式数据库技术在这些因素的推动下应运而生。

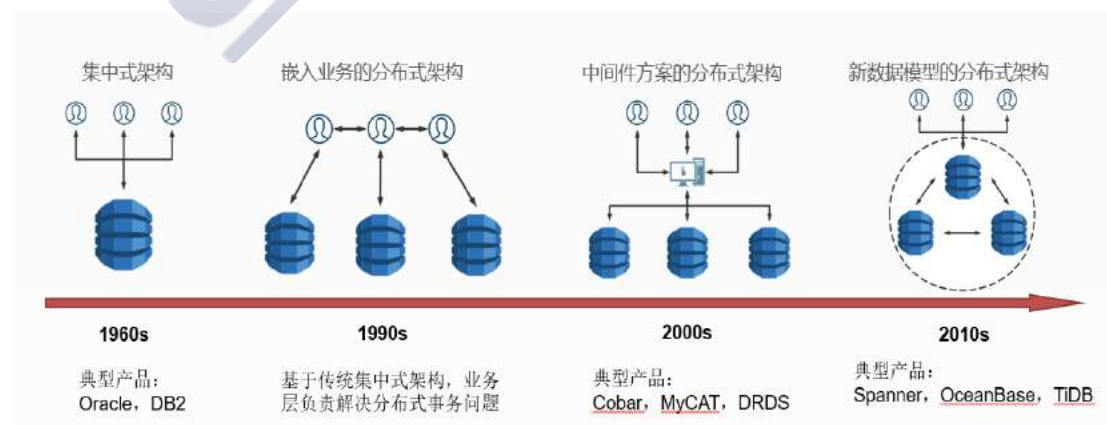


图 2 事务型数据库架构演进图

如图 2 所示，经过多年发展，当前分布式事务架构正处在快速演进阶段，综合学术界以及产业界工作成果，目前主要分为三类：

- 1) 基于原有单机事务处理关系数据库的分布式架构改造：利用原有单机事务处理数据库的成熟度优势，通过在独立应用层面建立起数据分片和数据路由的规则，建立起一套复合型的分布式事务处理数据库的架构。
- 2) 基于新的分布式事务数据库的工程思路的突破。通过全新设计关系数据库的核心存储和计算层，将分布式计算和分布式存储的设计思路和架构直接植入数据库的引擎设计中，提供对业务透明和非侵入式的数据管理和操作/处理能力。
- 3) 基于新的分布式关系数据模型理论的突破。通过设计全新的分布式关系数据管理模型，从数据组织和管理的最核心理论层面，构造出完全不同于传统单机事务数据库的架构，从数据库的数据模型的根源上解决分布式关系数据库的架构。

分布式事务数据库进入到各行各业面临诸多挑战，其一是多种技术路线，目前没有统一的定义和认识；其二是除了互联网公司有大规模使用外，其他行业的实践刚刚开始，需求较为模糊，采购、使用、运维的过程缺少可供参考的经验，需要较长时间的摸索；其三缺少可行的评价指标、测试方法和测试工具来全方位比较当前的产品，规范市场，促进产品的进步。故应用上述技术进行交易类业务进行服务时，应充分考虑“可持续发展”、“透明开放”、“代价可控”三原则，遵循“知识传递先行”、“测试评估体系建立”、“实施阶段规划”

三步骤，并认识到“应用过度适配和改造”、“可用性管理策略不更新”、“外围设施不匹配”三个误区。

大数据事务处理类技术体系的快速演进正在消除日益增长的数字社会需求同旧式的信息架构缺陷，未来人类行为方式、经济格局以及商业模式将会随大数据事务处理类技术体系的成熟而发生重大变革。

（三）数据流通技术

数据流通是释放数据价值的关键环节。然而，数据流通也伴随着权属、质量、合规性、安全性等诸多问题，这些问题成为了制约数据流通的瓶颈。为了解决这些问题，大数据从业者从诸多方面进行了探索。目前来看，从技术角度的探索是卓有成效和富有潜力的。

从概念上讲，基础的数据流通只存在数据供方和数据需方这两类角色，数据从供方通过一定手段传递给需方。然而，由于数据权属和安全的需要，不能简单地将数据直接进行传送。数据流通的过程中需要完成数据确权、控制信息计算、个性化安全加密等一系列信息生产和再造，形成闭合环路。

安全多方计算和区块链是近年来常用的两种技术框架。由于创造价值的往往是对数据进行的加工分析等运算的结果而非数据本身，因此对数据需方来说，本身不触碰数据、但可以完成对数据的加工分析操作，也是可以接受的。安全多方计算这个技术框架就实现了这一点。其围绕数据安全计算，通过独特的分布式计算技术和密码技术，有区分的、定制化的提供安全性服务，使得各参与方在无需对外提供原始

数据的前提下实现了对与其数据有关的函数的计算，解决了一组互不信任的参与方之间保护隐私的协同计算问题。区块链技术中多个计算节点共同参与和记录，相互验证信息有效性，既进行了数据信息防伪，又提供了数据流通的可追溯路径。业务平台中授权和业务流程的解耦对数据流通中的溯源、数据交易、智能合约的引入有了实质性的进展。

除了以上两种技术框架外，近年来还涌现出多种数据流通的技术工具，这里将其列表总结如下。

表1 数据流通技术工具对比³

技术工具	同态加密	零知识证明	群签名	环签名	差分隐私
原理概述	对原始数据进行加密，使得加密数据和原始数据进行相同处理时，结果相同	证明者向验证者证明一个声明的有效性，而不会泄露除了有效性之外任何信息	允许群体中的任意成员以匿名方式代表整个群体对消息进行签名，并可公开验证	一种简化的群签名，环签名中只有环成员没有管理者，不需要环成员间的合作	通过添加噪音来达到隐私保护效果
技术特点	可在不解密的情况下对密文进行计算和分析	证明者无需任何事件相关数据，就能向验证者证明事件的真实可靠	能为签名者提供较好的匿名性，同时在必要时又通过可信管理方追溯签署者身份	不需要分配指定的密钥，无法撤销签名者的匿名性	具有严谨的统计学模型，能够提供可量化的隐私保证
适用领域	云计算、电子商务、物联网等	电子商务、金融、银行、电子货币等	公共资源管理、电子商务、金融等	云存储、电子货币等	电子商务、物联网等
成熟度	全同态加密理论上可行，商用化程度还需提高	通用场景的零知识证明理论较为成熟，性能优化后逐渐商用	广泛应用在网络安全中，需要提高计算效率	建立更好的安全性模型，与群签名、CPK结合，优势互补	还需研究复杂数据的差分隐私保护和有效控制连续数据的累计误差

³ 来源：《数据流通关键技术白皮书（1.0版）》，大数据发展促进委员会，2018年。

四、大数据与实体经济融合应用

党的十九大报告中指出，要加快大数据与实体经济的深度融合。经过几年的发展，各行各业对于大数据应用的重要性基本得到统一，但受限于各种各样的因素，各行业的大数据应用水平还有较大差异。本节将以部分行业为例，分析各行业大数据发展现状及原因，并给出行业大数据应用发展的路径。

（一）行业应用大数据的特点

近年来，在全球经济数字化浪潮的带动下，我国大数据与实体经济的融合应用不断拓展。大数据企业正在尝到与实体经济融合发展带来的“甜头”。利用大数据可以对实体经济行业进行市场需求分析、生产流程优化、供应链与物流管理、能源管理、提供智能客户服务等，这不但大大拓展了大数据企业的目标市场，更成为众多大数据企业技术进步的重要推动力。随着融合深度的增强和市场潜力不断被挖掘，融合发展给大数据企业带来的益处和价值正在日益显现。

然而总体来看，目前我国在大数据与实体经济融合领域整体上还处于发展初期。相对于发达国家，在融合行业数量、融合应用深度、融合业务规模、融合发展均衡性等方面还有一定差距。这一阶段主要特点如下：

一是业务类型不均衡：大数据融合应用主要集中在外围业务上，而在核心业务方面的渗透程度还有待提高。据调查显示⁴，在应用大数据的行业企业中，营销分析、客户分析和内部运营管理是应用最广

⁴数据来自《中国大数据发展调查报告》，大数据发展促进委员会 2018 年 4 月发布

泛的三个领域。61.7%的企业将大数据应用于营销分析，50.2%的企业将大数据应用于客户分析，将近50%的企业将大数据应用于内部运营管理。相比之下大数据分析在产品设计、产品生产、企业供应链管理等核心业务的应用比例还有待提升，大规模应用尚未展开。

二是地域分布不均衡：大数据融合应用在地区之间发展不均衡，各地大数据应用发展程度差距较大。受经济发达程度、人才聚集程度和技术发展水平影响，大数据应用的产学研力量仍主要分布在北京、上海、广东、浙江等东部发达地区。相关的数据显示⁵，中西部地区的大数据应用虽然市场需求较大，但发展水平仍较低。

三是行业分布不均衡：大数据融合应用主要集中在部分行业中，如前所述，大数据与金融、政务、电信等行业的融合效果较好，而在其它众多行业的融合效果则有待深化。在下文中将着重对此现象的原因进行深入分析。

（二）行业应用大数据的深层分析

企业和行业大数据应用体系其实就是在生产业务系统之外构建统一的企业级数据仓库。回顾各个领先行业企业级数据仓库建设路径，从技术架构上大都经历了从传统数据库或者数据仓库的架构到 MPP 数据库架构再到 Hadoop 的架构体系。除技术架构外，企业级数据仓库的建设还包括数据模型、数据管理体系以及数据应用体系的建设，整个企业级数据仓库最终实施效果依赖于企业内部专业而有力度的组织机构来推动。以下以金融、电信、能源、交通、互联网等几个行

⁵数据来自《中国大数据产业发展评估报告（2017年）》

业为例，选取代表企业对其行业大数据应用情况进行简要梳理。

表 2 部分行业代表性企业大数据应用情况

	大型国有银行 (以中国工商银行为例)	运营商 (以中国联通为例)	电网 (以国家电网为例)	石化 (以中石化为例)	大型互联网 (以阿里巴巴为例)
数据平台	2000 年开始建立数据仓库。	2011 年开始建立数据仓库。	2014 年开始大数据平台试点。	2016 年打造云计算大数据平台。	2004 年开始建立数据仓库。
数据管理体系	2007 年建立了全行统一的数据体系。2013 年搭建 Hadoop 信息库。	2012 年开始整合全国数据建立大数据中心。	2009 年启动国网 SG-CIM 模型和主数据标准研究。	2016 年启动 ERP 系统主数据管理平台。	2010 年引入 Hadoop&Hive 平台进行新一代的数据平台的构建。
数据应用	2014 年研发了流数据平台，具备实时大数据应用能力。	2015 年开始运营数据，与各行业结合做大数据应用。	2015 年开始大数据应用的建设推广。	2015 年宣布借助阿里等企业打造数据商业服务新模式。	2008 年，阿里巴巴建立了数据平台，将交易核心过程进行重组，成立用户中心开展数据应用。
组织机构	1996 年即成立软件开发中心专注电子化建设。	2017 年成立专门的大数据运营公司。	2013 年成立专业的大数据团队，负责服务内部和外部客户。	早期即建立信息化管理部，无专门数据管理部门。	2008 年确定了数据和云计算两个重要战略，建立全球顶尖团队，搭建全新技术架构。

从上表中不难看出，各行业（以代表性企业为例）在企业级数据仓库建设方面进展不一。从技术角度来看数据仓库建设较早的行业经历了从传统的数据仓库过渡到 Hadoop 和 MPP 数据库架构的过程，而数据仓库建设较晚的行业由于后发优势直接使用 Hadoop 或者 MPP 数据库来了构建企业级数据仓库。

从数据管理的角度来看，国有大型银行、电信领域是最早建立统

一数据标准和数据模型的行业，尤其是国有大型银行有上百套业务系统，数据标准化、数据建模、数据治理的复杂度高，实践的难度最大，有很多可供参考的经验。

在数据应用方面，互联网公司、运营商和国有银行的进展也相对较快，这些行业都有较强的精细化经营的诉求，尤其是互联网行业，数据平台直接融入到业务之中，无论是精准广告、内容推荐、用户标签、风险控制都高度依赖于数据分析体系的支持。互联网公司和电信运营商都在经历由内向外的路径，即不仅服务支持企业内部经营分析，而且能够将数据价值释放到社会和其他行业。从组织架构的情况来看，企业级数据仓库建设既需要大数据相关技术人员的支持，也需要数据管理团队的推动，前者负责平台搭建、运维，后者负责数据标准、数据模型、数据治理、数据生命周期的实施。相比之下，电网、石化部门的数据仓库建设起步较晚，一方面在于这些行业所处的经营的阶段不同，另一方面因为前期信息的任务主要在于支撑各类业务系统的运行。但随着这些实体经济巨头企业对数据管理和应用的重视，这些行业与大数据的融合步伐将会不断加快。借助后发优势和企业规模效应，我们有理由相信这些行业的大数据应用前景将非常广阔。

对比一些传统企业和部分政府机构，大部分在数据应用方面往往还处于初级阶段：统一的数据仓库还未创设，一体化的大数据平台还未搭建，数据治理或管理体系尚处于初级阶段，数据管理的专门机构未设置。对此，需要这些行业和企业建立科学的数据治理和资产管理体系，提升与大数据融合的能力。

（三）行业应用大数据的关键因素

根据上文的分析，我们可以总结出行业应用大数据的关键因素。这些因素对于行业利用大数据提升业务能力具有一定的指导效果。

一是要建立一体化的大数据平台。大数据应用效果较好的行业，通常都建立了生产业务和数据分析深度融合的系统。通过一体化大数据平台，数据的汇聚和共享得以实现，从而提升了数据价值。

二是要形成良好的数据管理体系。大数据应用效果较好的行业，通常都已经开展了成熟的数据治理和数据资产管理实践。数据的共享和集成水平比较高，标准化的数据管控体系得以建立，数据的质量、安全得以保证。

三是形成了平民化的数据应用。大数据应用效果较好的行业，通常都建立了与需求深度耦合又简单易用的数据应用工具。这使得大数据的使用者从企业数据专家扩展到了普通业务人员，从而真正实现了“人人产生数据、人人应用数据”。

四是组建了强有力的数据管理部门。数据管理职能应该有专门的部门实施，因此应成立专门数据管理领导小组和数据管理（处理）部门，将数据的监管职责赋予数据管理部门，由数据管理部门集中管理监控数据，各有关职责部门配合。

五、政务大数据发展

大数据是提升政府治理能力的重要方式之一。我国政府多年的信息化发展积累了海量的政务数据，如何健全和完善政务数据的应用机制、厘清政务存量数据、将数据进行共享开放、从数据中挖掘价值，

最终用于政府治理，切实解决人民群众在同政府打交道时的实际困难、社会企业对于政务数据的迫切需求和提升政府工作效率成为政务大数据深层次应用的最主要问题。

（一）政务大数据总体要求

国家大数据战略实施以来，我国政府出台了多项顶层设计，为大数据产业的快速成长提供良好的发展环境。2015 年《促进大数据发展行动纲要》的发布吹响了我国大数据发展的号角。2016 年以来，关于电子政务、政务信息、政务系统相关文件频发，循序渐进、有条不紊的指导政务大数据的有序发展。特别是 2017 年起，“加快国务院部门和地方政府信息系统互联互通，形成全国统一政务服务平台”、“深入推进“互联网+”行动和国家大数据战略”等要求陆续提出，为政府信息化建设提供了新的商业机遇和建设方向。

2016 年 4 月 14 日，发布了《推进“互联网+政务服务”开展信息惠民试点实施方案》。2016 年 9 月 19 日发布了《政务信息资源共享管理暂行办法》。2016 年 12 月 27 日，发布了《国务院关于印发“十三五”国家信息化规划的通知》。2017 年 1 月 12 日，《“互联网+政务服务”技术体系建设指南》。总体说来，各指导性文件逐步明确了四个方面的内容，一是在政务信息共享原则方面，提出以共享为原则，不共享为例外；需求导向，无偿使用；统一标准，统筹建设；建立机制，保障安全。二是在政务信息资源分类方面，提出将现有的政务信息按照重要程度和等级分类，划分为无条件共享、有条件共享和不予共享三类。三是在平台建设方面，提出共享平台是管理国家政

务信息资源目录、支撑各政务部门开展政务信息资源共享交换的国家关键信息基础设施，包括共享平台（内网）和共享平台（外网）两部分。四是在分工职责方面，提出了国家发展改革委、国家网信办组织编制信息共享工作评价办法，国家网信办负责组织建立政务信息资源共享网络安全管理制度，国家发展改革委、财政部、国家网信办建立国家政务信息化项目建设投资和运维经费协商机制。

经过了多个指导性政策文件的发布与宣贯，政务信息系统和资源整合也逐步走向了落地的进程。2017年5月18日，发布了《政务信息系统整合共享实施方案》提出，2017年底完成国务院部门系统整合清理工作，2018年6月实现各个部门整合后的政务信息系统统一接入国家数据共享交换平台。具体提出十项主要任务和方法，包括上下联动，开展“互联网+政务服务”试点；一体化服务，规范网上政务服务平台体系建设；完善标准，加快构建政务信息共享标准体系；构建目录，开展政务信息资源目录编制和全国大普查；强化协同，推进全国政务信息共享网站建设；推动开放，加快公共数据开放网站建设；促进共享，推进接入即统一数据共享交换平台；设施共建，提升国家统一电子政务网络支撑能力；推进整合，加快部门内部信息系统整合共享；审、清结合，加快消除“僵尸”信息系统。2017年7月，发改委和网信办发布了《政务信息资源目录编制指南（试行）的通知》提出政务信息资源目录按照三个维度进行划分，从资源维度将政务信息分为基础信息、主体信息和部门信息，涉及人口基础、法人信息、自然资源、社会信用、公共服务、健康保障、社会保障、安全生产和

其他信息；从涉密维度分为涉密信息和非涉密信息；从共享维度即按照无条件共享、有条件共享和不可共享三类。提出将目录元数据按照信息资源分类、信息资源名称、信息资源代码、信息资源提供方式、信息资源摘要、信息资源格式、具体信息（名称、数据类型）、共享属性（共享类型、共享条件、共享方式）、开放属性（是否开放、开放条件）、更新周期、发布日期、关联资源代码进行划分。规定了数据资源目录的编写流程，分为四个阶段，一是前期准备，包括组织准备、目录划分、资源调查，二是目录编制与报送，三是目录汇总与管理，包括审核汇总和管理维护，四是目录更新。

（二）政务信息系统整合

国家及地方政府对于政务信息整合的指导和要求可以分为三个方面，即统一数据资源整合和基础设施建设、整合和升级信息系统、完善数据资源标准体系建设。

1、数据资源整合和基础设施建设

统筹规划、协同推进。整合分散的数据中心资源，充分利用现有政府和社会数据中心资源，运用云计算技术，整合规模小、效率低、能耗高的分散数据中心，构建形成布局合理、规模适度、保障有力、绿色集约的政务数据中心体系。统筹发挥各部门已建数据中心的作用，严格控制部门新建数据中心。加快完善国家基础信息资源体系，加快建设完善国家人口基础信息库、法人单位信息资源库、自然资源和空间地理基础信息库等基础信息资源。依托现有相关信息系统，逐步完善健康、社保、就业、能源、信用、统计、质量、国土、农业、城乡

建设、企业登记监管等重要领域信息资源。

目前，各地政府新规划基础设施建设均在充分整合、利用现有各级数据中心的基础上，集约化、合理化的建设绿色、环保、需求导向的现代数据中心成为各级政务应用提供基础设施环境成为大数据提升政府治理能力的重要任务之一。在数据中心建设基础上，充分利用云化技术提升物理资源利用率，为各政府部门提供专业化的云计算服务。同时，考虑各级数据中心、分散网络节点的互联互通，在提升网络带宽和传输速率的基础上，合理的利用网络资源。在政务数据整合方面，纷纷制定适用于本地政务数据的政务数据资源目录，进行集中存储和统一管理。整合实有人口、法人、空间地理等基础数据库建设，加强内部共享和动态更新，提高数据准确性。

2、信息系统整合和升级

整合各类政府信息平台和信息系统。严格控制新建平台，依托现有平台资源，在地市级以上（含地市级）政府集中构建统一的互联网政务数据服务平台和信息惠民服务平台，在基层街道、社区统一应用，并逐步向农村特别是农村社区延伸。除国务院另有规定外，原则上不再审批有关部门、地市级以下（不含地市级）政府新建孤立的信息平台和信息系统。通过规划建设，逐步形成统一的互联网政务数据服务平台，实现基础信息集中采集、多方利用，实现公共服务和社会信息服务的全人群覆盖、全天候受理和“一站式”办理。

目前，各地政府均不同程度的根据业务特点开展了信息系统的整合和创新。实践证明，通过大数据推动各政府部门业务协同、流程再

造、决策支撑，是有效提升政府治理能力的重要手段之一。推动网络化政务服务，实现“数据多走路，群众少跑腿”成为各地政府推行大数据创新的重要目标，创新、丰富的网上办事大厅、移动应用、微信办公等方式成为了政府治理大数据创新应用的主要手段。

3、数据资源标准体系建设

突出重点、攻坚克难，推进政府大数据标准制定工作，重点制定数据流通标准、数据安全与隐私保护标准以及面向政府大数据平台架构与评测的标准。重点规定元数据、数据开放、数据共享、数据交换、数据质量等内容，安全与隐私保护标准要重点规定数据安全和隐私保护等内容，面向政府大数据平台架构与评测的标准要重点规定平台架构、评测方法等内容。

目前，各地政府均不同程度的建立了政府内部、政府和企业、政府和公众的数据整合流通标准和规则，包括数据开放、数据共享、数据交换等一系列标准，解决政府内部数据共享、政府数据对外开放、政府和企业数据交换等问题。制定数据安全和隐私的标准，形成阶段性政府数据安全使用的标准和隐私保护的基本条款，并向完善的安全和隐私保护标准的目标迈进。建立政府大数据平台架构体系和评测标准，梳理政府大数据平台架构的通用特点，形成通用架构标准和基本的评测标准，并最终根据不同政府职能和业务对通用架构进行细分，形成涵盖多个政府业务的架构体系，并制定相应的评测标准。

（三）政务信息共享交换

国家和地方政府对于数据共享交换的总体要求是打通政府间数

据壁垒和隔阂，实现政府数据的跨部门流动和互通，能够有效发挥政府数据的关联分析能力，建立“用数据说话、用数据决策、用数据管理、用数据创新”的政府管理机制，实现基于数据的科学分析和科学决策，构建适应信息时代的国家治理体系，推进国家治理能力现代化。通过统筹完善，逐步推动政府数据资源共享，制定政府数据资源共享管理办法，整合政府部门公共数据资源，促进互联互通，提高共享能力，提升政府数据的一致性和准确性。明确各部门数据共享的范围边界和使用方式，跨部门数据资源共享共用格局基本形成。充分利用统一的国家电子政务网络，构建国家、省市、乡镇等多级政府数据共享交换平台。

目前，各地政府不同程度的制定了数据共享交换办法。明确政府数据共享的类型、范围、共享义务主体、共享权利主体、共享责任和共享绩效考核评估办法。各级政府部门依据政府数据共享办法制定本部门政府数据共享的具体目录，依据政府数据共享目录向其他政府部门提供政府数据共享服务；明确政府数据共享使用的方式，按照全公开使用、半公开使用、不公开使用等不同级别，界定对政府数据共享使用的数据公开范围，同时规定政府数据共享使用人的义务和责任。

各级政府在地方大数据规划中也对数据共享交换计划进行了明确规定，明确政府数据共享的年度目标、双年度目标以及中长期目标，确定各政府部门为实现政府数据共享达标所应采取的具体措施和工作安排，明确政府数据共享的具体程序和工作流程，明确政府数据共享的负责人员、责任部门以及究责措施。

为推动政府信息共享交换工作落实，多数地方政府制定了政府数据共享绩效考核管理办法，建立政府数据共享评估指标体系，对各级政府部门提供政府数据共享服务的情况进行评估考核；依托政府数据共享平台统计和反馈功能，自动、逐项评价共享数据的数量、质量、类型和使用程度等情况；引入第三方评估评级机构，对各级政府部门的政府数据共享计划及其执行情况进行评估评级，将评估评级结果纳入政府部门信息化工作考核报告，与电子政务项目立项申报关联起来，严格执行激励约束措施，推动共享数据滚动更新，提高共享数据数量质量，确保政府数据共享取得实绩。

（四）政务信息对外开放

政府数据资源是量体大、集中度高、辐射范围广、与社会公众关联紧密、开发利用价值高、积聚带动效应明显的大数据资源。推进落实政府数据开放建设工程，逐步实现政府数据依托两大平台向社会开放，是建立健全数据驱动型增长新模式，推动经济社会全面发展，促进治理能力现代化的重要抓手。

坚持政府数据以开放为原则、以不开放为例外，按照“试点先行，制度保障，平台搭建”的总体思路。首先有条件的省市区域进行开展政府数据开放示范试点，以点带面、以局部辐射全局，按照全面规划、布局合理的原则，逐步向其他区域扩散。同步建设国家政府数据统一开放平台，建成面向互联网、实现跨地区跨部门跨行业政府数据异构存储的国家政府数据统一开放平台，以“增量先行、存量补进”为原则，分步实现各民生保障服务相关领域的政府数据集向社会开放。

建立政府数据正负面清单。形成包括立法信息、基础信息、宏观经济信息、社会管理信息、公共服务信息、司法信息、重要行业信息、市镇公共信息等在内的政府数据开放目录，率先将信用、交通、医疗、卫生、就业、社保、地理、文化、教育、科技、资源、环境、金融、统计、气象、企业登记等重要领域政府数据纳入开放清单。同时，最终实现除国家及商业机密、涉及个人的数据和信息、执法记录等法定不能开放的数据之外的所有数据，均当无保留全部向社会开放。

完善健全数据开放制度和机制。制定政府数据开放短期、中期长期计划和工作流程，提出相关具体要求，吸引更多政府和社会数据纳入开放范围。明确要求政府数据应当采用开放、机器可读的数据格式和标准，通用、可扩展的元数据，确保政府数据能够支持开放下游的任何信息处理和传播活动。建立数据开放责任制度，明确企业和社会公众对政府数据享有获取权。建立完善数据安全审查制度，制定政府数据安全审查管理办法，明确政府数据开放前后及开放过程中应当采取的安全保密审查程序和流程。制定完善政府数据开放绩效考核评估制度。依托平台统计和反馈功能，自动、逐项评价开放数据的数量、质量、类型和使用程度等情况，并形成建立和惩罚机制，提高各政府部门和社会机构的数据开放热情。

目前，我国各地政府数据开放进程都已起步。从地区来看，已有十余个省市依托各自的数据开放平台或专门网站开放了一批数据。如北京、上海、浙江、福建、贵州等试点地区，以及佛山、青岛、武汉、长沙等地。截至2018年1月中旬，北京市数据开放平台已开放42个

政务部门 18 个领域的 748 个数据集，上海市已开放 42 个政务部门 12 个领域的 1564 个数据集，浙江省已开放 39 个政府部门 8 个领域的 292 个数据集，贵州省已开放 58 个部门 13 个领域的 470 个数据集，福建省数据开放平台对既有开放数据和数据查询网站进行了整合。从行业来看，司法、信用、气象、林业等部门通过专门网站提供数据的浏览下载。

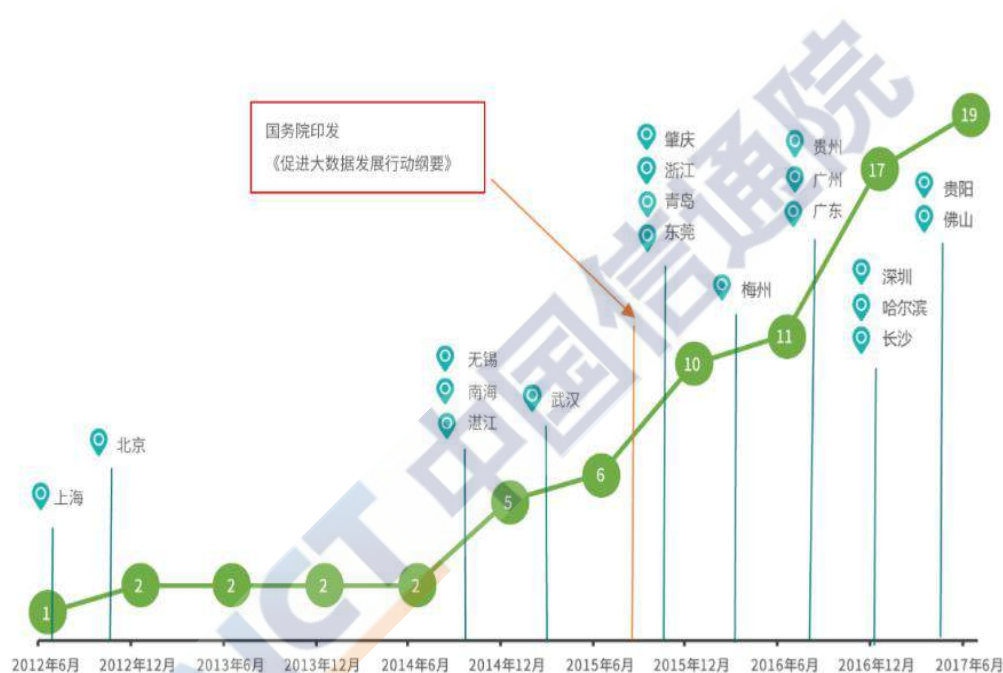


图3 我国主要政府数据开放平台上线时间⁶

数据服务方面，我国大部分数据开放平台不具备公众互动功能，社会参与有限。使用登记方面，我国试点地区将仅对大规模、连续利用数据服务的机构和个人实施网络实名登记，从而在促进与规范间进行平衡，但对“大规模”的判定还需明确。数据评价方面，我国一些地区也采用了数据评价做法，未来将继续加强数据评价与数据撤回、数据完善、考核评估间的衔接。

⁶ 来源：《中国地方政府数据开放平台报告》

六、地方大数据产业发展

大数据产业对于推动地方经济发展具有重要推动作用。一直以来，各地政府纷纷把大数据产业作为发展大数据的核心工作。截至 2018 年 2 月底，地方政府对外公布了超过 110 份大数据相关政策文件，覆盖 31 个省级行政区域。总体来看，我国大数据产业目前仍处于蓬勃发展阶段，逐步形成区域协调发展局面。

（一）大数据产业发展主要模式

地方政府结合自身经济基础、产业结构特点与人力资源条件等要素，积极寻求发展具有本地特点的大数据产业，形成了不同的发展模式，优化了我国大数据产业结构。

（1）以北京、广东、江苏为代表的引领型发展模式

北京、广东、江苏凭借强大的经济、科技与人力资源实力，在关键技术、先进产品、产业生态体系构建方面，制定了明确的发展目标。北京提出建设“全国大数据和云计算创新中心、应用中心和产业高地”，江苏提出“争创全国领先、特色明显的国家大数据综合试验区”，广东提出“打造全国数据应用先导区和大数据创业创新集聚区，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”。

（2）以苏州、南宁为代表的落实型发展模式

苏州、南宁等地强化大数据工作落实力度，从国内外大数据发展背景、本地现状与基础、发展路径与策略、基础设施建设、行业应用、产业创新、产业生态打造等方面，制定了详细深入的发展规划。南宁

市政府对大数据产业的发展模式、商业模式以及相关重大工程给出了详细说明，全面体现实现“规划与计划相结合，继承与创新相结合”的工作思路。

（3）以赶超发达地区产业为目标的追赶型发展规划

部分省份与城市在制定大数据发展规划时，鉴于自身产业基础条件的限制，重点采取跟随策略，根据《纲要》中提出的要求，逐一进行落实。借助大数据产业发展浪潮，推动当地电子信息产业发展，为当地经济扩张寻求新的增长点，提高本地经济活力，优化了整体的产业结构。

（二）地方大数据产业发展策略分析

通过分析已经对外公布的 55 份大数据发展规划或行动计划等政策文件，多数地方政府明确了定位、规划了产业目标，以此来指引本地大数据发展的各项工作。

1. 大数据产业发展定位

在已发布大数据政策的地方政府中，有 20 个省级或地市级政府明确提出了大数据产业发展定位，涉及面向全球、面向全国、面向区域等三个层面，包括人才、创新创业、数据资源聚集、应用、产业中心等不同类型。

表 3 地方政府大数据产业发展定位

类型	面向全球	面向全国	面向区域
人才发展中心		湖北、四川	宜春
创业创新中心	广东	北京、青岛、宁波、贵安新区、四川	淮南
资源聚集中心	珠三角、广东	贵州、湖北、青海、武汉、天津滨海新区、四川	沈阳、兰州、淮南
应用示范中心	珠三角、广东、深圳	北京、贵州、浙江、湖北、青岛、宁波、淮南、贵安新区、吕梁、厦门、河南、四川	青海、台州、宜春
产业中心 / 高地	珠三角、广东、深圳	北京、重庆、浙江、湖北、新疆、武汉、沈阳、宁波、天津滨海新区、贵安新区、吕梁、湖州、福建、宜春、河南、长沙、四川	青岛、合肥、淮南、台州、钦州

从已发布的大数据发展规划文件来看，多个省市以发展面向全国的大数据产业中心或高地为目标，部分省市在多方面发展大数据，以广东为例，提出了“用 5 年左右时间，打造全国数据应用先导区 and 大数据创业创新集聚区，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”。中部、西部的一些省市也积极面向全国发展大数据，例如，贵州提出到 2020 年，“大数据、云计算应用和服务水平居国内领先地位，产业体系健全，成为西部地区重要的、全国有影响力的战略性新兴产业基地”。

2. 大数据产业创新特点分析

基于传统产业分类以及大数据应用创新特点，按照工业大数据、农业农村大数据、服务业大数据以及新兴产业大数据等类别进行分析，其中，服务业又可细分为金融、电子商务、交通等多个领域。

工业仍是各地最为关注的大数据产业创新领域。55 个省级和地

市级政府部门中，有 39 个提到了工业大数据，占比达到 71%。大数据应用于工业研发设计、生产制造、经营管理、市场营销、售后服务等产品全生命周期、产业链全流程各环节，成为推进智能制造的重要手段。

农业农村是地方政府关注的另一大热点大数据应用领域。有 33 个省级政府部门或地市级政府部门提到了农村农业大数据，占总数的 60%。重点应用在农产品质量安全追溯、农产品产销信息监测预警、农业自然灾害预测预报、动物疫病和植物病虫害监测预警等方面。

大数据在服务业拥有广阔应用空间。互联网金融、数据服务、数据材料、数据制药等新业态成为新热点，各地积极培育新技术、新产品、新业态和新模式。

从目前各地发布的大数据产业政策文件看，电力、教育、文化创意等领域的大数据产业创新受到的关注度不高，今后需要加大在这些领域的推动力度。

（三）地方大数据产业发展成效与问题

初步形成协同发展的大数据产业生态体系。加速培育和引进大数据骨干企业，大力扶持特色鲜明的创新型中小企业，形成协同发展的大数据产业生态体系。试验区采取有针对性措施，引进和培育大数据企业。上海涌现出宝信软件、银联智策等行业应用龙头企业，星环科技、华院数据等技术型企业在数据挖掘、大数据平台、数据安全等领域快速成长。重庆成功推动阿里巴巴西部创新中心、中科曙光-美国 VMWare 合资公司、浪潮大数据等项目落户重庆，猪八戒网公司估值

已超百亿元，成为国内最大的工业设计众包平台，中科云丛已成为国内知名的人脸识别大数据企业。沈阳注重龙头企业引进，中兴、浪潮、360、华为、SAS、国信优易等 60 余家企业落户沈阳，在浑南、和平、铁西形成大数据企业集聚态势。这些骨干企业对推动当地经济发展、提升大数据发展、促进高新技术产业就业，均起到重要作用。

产业发展层次较低，产业布局亟需完善。除了京津冀、珠三角、上海三个综合试验区之外，其他综合试验区大数据产业整体发展层次较低，大数据产业链布局亟需完善。同时，中西部试验区由于产业基础较为薄弱、科研教育资源少、经济发展水平较低等原因，大数据龙头企业都相对较少，导致细分行业对个别企业依存度过高，存在一定的风险。贵州省在电子设备制造、IDC 产业领域取得有目共睹进展，但在以海量计算、人工智能等为代表的高端产业领域，与国内先进省份仍有较大区域。

七、数据资产管理体系

近年来，数据治理和数据资产管理的重要性愈发凸显。有效的数据资产管理是大数据与实体经济深度融合的必经之路。数据成为资产的概念逐渐深入人心，甚至有人建议将数据计入资产负债表。数据资产管理将从“理论”走向“实践”，将影响数据的存量和增量，提升数据的质量和 value，保障数据的安全，为大数据应用及人们未来的便捷生活打下夯实的基础。

（一）数据资产管理的定位和范畴

数据资产管理在大数据技术体系中的定位如图 4 所示，它位于应用和底层平台中间。数据资产管理包括两个重要方面，一是数据资产管理的核心活动职能，二是确保这些活动职能落地实施的保障措施，包括组织架构、制度体系。数据资产管理在大数据应用体系中，处于承上启下的重要地位。对上支持以价值挖掘为导向的数据应用开发，对下依托大数据平台实现数据全生命周期的管理。



图 4 数据资产管理在大数据体系中的定位⁷

目前，数据资产管理已经形成了一套科学的管理范畴。根据 DAMA 等机构的总结，数据资产管理主要包含 9 个活动职能和 2 个保障措施，9 个活动职能指的是数据标准管理、数据模型管理、元数据管理、主数据管理、数据质量管理、数据生命周期管理、数据安全、数据资产价值评估和数据资产运营流通，2 个保障措施包括组织架构和制度体系。

⁷ 来源：《数据资产管理实践白皮书（1.0 版）》，大数据发展促进委员会，2017 年。



图5 数据资产管理体系架构

（二）数据资产管理面临的挑战

企业日常经营活动中积累的大量数据，除了支持业务流程运转之外，越来越多地被用于帮助企业提升管理决策效率、实现价值挖掘和业务创新。企业日常经营决策过程的背后，实质是数据的生产、传递和利用的过程，风险控制、产品定价、绩效考核等管理决策过程需要大量高质量数据支撑。提升数据质量、降低成本已经成为行业企业热点关注话题。如果不能对数据进行有效梳理及精细化管理，其价值就得不到很好体现，严重影响数据价值发挥，甚至会给运营管理带来负面作用。此外，日益全面、严格的监管措施和信息披露要求，也对企业数据提出了前所未有的挑战，主要体现在以下几个方面：

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

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

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

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

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

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

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

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

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

（三）数据资产管理的发展趋势

随着数据资产管理生态系统的不断发展，现有的实践体系也在迅速发展，可以从数据对象、数据采集、处理架构、组织职能、管理手段和应用范围六个方面来预测其发展趋势。

数据对象纷繁复杂。目前，企业数据管理的主要数据对象仍然是结构化的文本数据。未来，随着网络爬虫、视频处理、语音识别、自然语言处理、图像处理、人脸识别等相关技术逐渐成熟并被产业界进一步深度应用，城市数据、视频数据、语音数据、图形图像数据等将被越来越多的进行管理和应用。预计到2020年，66%的企业将采用高级分类处理方案来采集、保存并处理非结构化的数据，以提高分析效率。

数据采集途径丰富。随着传感器、5G及NB-IoT的发展，数据采集及传输途径也将得以扩充。由社交媒体和机器人过程自动化(RPA)等转型技术创建的新数据通道将为数据治理和数据质量组织带来机遇和挑战。这些渠道的数据，其规模、数量、速度和变化(SVVV)等

特征与主数据管理和数据治理的传统领域的特征显著不同。数据采集的变化和传统数据管理架构产生了“差异”，这样的“差异”要求管理组织采用不同的方法来管理数据质量和标准，以满足相关数字业务流程所要求的灵活性。

处理架构更新换代。由于越来越多的文件、文本、日志等半结构化、非结构化数据加入形成“数据湖”，数据的处理架构也在发生变化。支持主流大数据分析平台的处理架构以及批处理、流计算等技术正在被应用于数据资产管理。预计到 2020 年，主流的分析架构都将包含基于目标进行优化的解决方案，其中三分之一的产品会将关系型及非关系型数据的处理结合在一起。数据处理的底层架构将全面采取包括分布式文件系统、MPP 数据库、传统数据仓库、流计算引擎、交互式计算引擎、离线计算引擎在内的“计算&存储混搭架构”，并逐渐由传统的“ETL”数据集成过程向“ELT”转变。以 Hadoop、Spark 等分布式技术和组件为核心的数据处理架构，能够支持批量和实时的数据加载以及灵活的业务需求，将是持续焦点。

组织职能升级变迁。当前主流管理制度体系中，数据管理职能由 IT 部门来负责，业务部门配合 IT 部门执行数据管理并提出需求。未来，随着数据分析与业务融合越来越深入，业务部门将成为数据应用的主角，在数据资产管理中扮演越来越重要的角色。据 Gartner 预测，未来 50% 的全球性组织将聘用首席数据官(CDO, Chief Data Officer)，在数据高度监管的银行金融或医疗健康领域，此类人才需求量更大。

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

应用范围不断扩大。数据的应用范围将由传统的内部应用为主发展为支撑内部和服务外部并重。数据资产的意义价值也从对内强化能力扩展到了对外合作开放上，从而实现数据资产保值到增值的跨越。内部应用一般包括管理优化、研判决策、风险规避、业务拓展、管控成本等。由原来的只应用于领导决策场景扩展到全员业务分析使用。外部运营包括智能推荐、精准营销、分析报告以及风险防控等。从使用对象来看，数据资产的使用者不仅包括企业决策人员，还包括运维用户、业务管理人员、数据分析人员、数据科学家等各种角色。

数据资产管理的主体可以从企业向更广泛的概念推广。可以说，构建科学的数据资产管理体系是个人层面、企业层面、社会层面、国家层面乃至国际层面都关注的热点话题之一。**个人层面**，个人数据利用与数据安全保护之间需要合理的平衡，降低组织运营与合规方面的风险。大数据商业化应用中涉及的用户数据处理需要对用户隐私进行脱敏加密，以实现可控的隐私保护目标。**企业层面**，数据资产价值的估算可以帮助企业更准确的掌握信息化投资收益，也是数据交易流通

的前提之一。好的数据资产管理策略能有效规避风险，节约投入成本。

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

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

国际层面，数据资产管理知识体系涉及管理、技术等多个学科，是一个非常复杂的系统工程，涉及很多技术难点和管理内容需要结合多方力量达成一致标准，国际化的共识机制是数据资产管理有效执行的重要前提和保障。

八、数据立法重点

为应对大数据发展带来的各种问题和需求，各国政府动作频频，通过修改原有法律法规、制定新的法律政策等方式，从国家、企业和个人三个维度出发，完善政府数据开放、数据流通规则以及个人信息保护等方面的制度规定，为大数据的持续健康发展提供法律上的保障。

（一）政府数据开放

从国家层面而言，政府数据作为公共资源，在不危害国家安全、不侵犯商业秘密和个人信息的前提下，将其最大限度地开放给社会进行增值利用，有利于增加政府透明度，激发社会创新活力，提高公共服务水平，促进经济转型升级，提升政府治理能力。

美国打响了政府数据开放的第一枪，2009年1月，美国总统奥巴马发布备忘录——《开放和透明的政府指令》，数据门户网站data.gov上线，要求建立更加开放、参与、合作的政府。随之，全球开放数据运动相继展开，2011年，巴西、印尼、墨西哥、挪威、南非、菲律宾、英国和美国签署了《开放数据声明》，“开放政府合作伙伴”（Open Government Partnership，简称 OGP）宣告诞生。根据相关决议，新加入 OGP 的国家必须签署《开放政府宣言》，制定国家数据开放行动计划并征求公众意见，并且就行动进展定期提交独立报告。2015年经济合作与发展组织 OECD 发布的《开放公共部门信息的政府措施评估》报告显示，制定政府数据开放战略成为各国普遍趋势。主要经济体相继发布国家战略或政策积极推动政府数据开放，

并跟随数据开放的形势和要求不断修正和调整相关政策以适应新形势的发展。

在我国政府职能转变和产业转型发展的背景下，在《政府信息公开条例》的基础上，面对大数据时代的道理，我国政府正在逐步从日趋成熟的“政府信息公开”向蹒跚起步的“政府数据开放”探索前进。近年来，党中央、国务院在公共信息资源开放的有关工作中不断进行战略部署，陆续发布了一系列战略规划和政策文件，推动公共信息资源开放。今年，中央网信办、发展改革委、工业和信息化部联合印发《公共信息资源开放试点工作方案》，确定在北京、上海、浙江、福建、贵州等5省份开展公共信息资源开放试点工作。而在此前，国家已陆续发布一系列文件推进公共信息资源开放工作。如《关于推进公共信息资源开放的若干意见》是我国第一份从国家层面专门针对公共信息资源开放工作进行指导规范的文件；《关于运用大数据加强对市场主体服务和监管的若干意见》要求有序开放政府数据和全社会信息资源，提升经济社会运行效率和公共服务整体效能；《促进大数据发展行动纲要》、《“十三五”国家信息化规划》要求到2018年形成公共数据资源开放共享的法规制度和政策体系，建成国家政府数据统一共享交换和开放平台；《2016年推进简政放权放管结合优化服务改革工作要点》、《关于促进云计算创新发展培育信息产业新业态的意见》、《关于积极推进“互联网+”行动的指导意见》等文件也作出了相应要求。2017年4月，贵阳市发布全国首部专门的政府数据共享开放地方性法规《贵阳市政府数据共享开放条例》，对政府数据

共享开放的管理机制、部门职责、平台建设维护、采集汇聚、共享开放范围边界及使用具体要求、监督管理、安全保障等措施进行了规范。

政府数据开放涉及多层级、多部门利益，并且是一个涉及安全、经济和技术多个领域的综合性问题。结合国内外实践经验，要切实推动政府数据开放有效开展，必须从国家层面制定政府数据开放战略或政策，对中央、地方统一部署政府数据开放任务和要求；建立连接中央、地方的统一的政府数据开放网站，不断开放高价值数据；加强配套措施的建设，包括明确数据开放原则，制定政府数据资产目录、政府数据开放清单，建立政府数据开放评估体系，加强政府数据开放标准体系建设，加强安全保障体系建设等。

（二）个人信息保护

当前，大数据技术及其相关产业的发展正前所未有的改变着个人信息的收集和使用方式，对现有的个人信息保护制度带来新的问题和挑战。近年来，云计算、物联网、移动互联网等新技术新业务的快速发展，给现有的个人信息保护法律制度带来了新的挑战，各国修订、立法活动更加频繁。

进入 21 世纪以来，全球个人信息保护立法持续升温，截止 2016 年 12 月，全球共有 115 个国家和地区制定了专门的个人信息保护法，确立了个人信息收集、使用以及跨境传输等基本规则。当前世界各国正试图通过修订或增加法律法规中关于个人信息的保护范围以及强化保护力度以应对大数据背景下给个人信息保护工作带来的巨大挑战。欧盟于 2016 年出台了《一般数据保护条例》（GDPR），增加了

数据主体的被遗忘权和删除权，引入强制数据泄露通告、专设数据保护官员等条款，同时包含了更严厉的违规处罚。非欧盟成员国的公司（包括免费服务）只要满足下列两个条件之一：（1）为了向欧盟境内可识别的自然人提供商品和服务而收集、处理他们的信息；（2）为了监控欧盟境内可识别的自然人的活动而收集、处理他们的信息，该公司就受到 GDPR 的管辖。这个条例将对中国企业的移动应用安全，以及数据收集、处理和交易产生重大影响。

整体来看，我国个人信息保护以分散立法为主，尚未制定专门统一的个人信息保护法。个人信息保护立法体系由法律、行政法规、部门规章、地方性法规和规章、各类规范性文件等组成，形成了多层次、多领域、内容分散、体系庞杂的个人信息保护法律体系。近年来，在法律规定中直接对“个人信息保护”进行规定的趋势日趋明显。2012 年，《全国人民代表大会常务委员会关于加强网络信息保护的決定》首次以法律文件的形式对个人信息保护的要求做了明确规定，为各项工作的开展提供了上位法依据；2013 年，工业和信息化部出台了《电信和互联网用户个人信息保护规定》具体规定了电信业务经营者、互联网信息服务提供者收集、使用用户个人信息的规则和信息安全保障措施等要求；2016 年，《网络安全法》将个人信息保护纳入网络安全保护的范畴，其第四章“网络信息安全”对个人信息保护问题作了专章规定。此外，《征信业管理条例》（2013 年）《刑法修正案（九）》（2015 年）以及相关司法解释，如《最高人民法院、最高人民检察院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》（2017）《最高人民法院关于审理利用信息网络侵害人身权益民事纠纷案件适用法律若干问题的规定》（2014）等，也分别从行业监管、刑事、民事等方面进一步补充健全了我国的个人信息保护法律体系。上述法律规定相继发布，标志着我国个人信息保护法律体系框架基本形成，初步建立起了以《网络安全法》、《人大决定》为核心，包括《刑法》、《征信业管理条例》、《电信和互联网用户个人信息保护规定》等法律、行政法规和部门规章在内的个人信息保护立法体系。从内容上来看，涵盖了个人信息保护领域的大部分要素和制度。我国个人信息保护以分散立法为主，尚未制定专门统一的个人信息保护法。

制定国家专门统一的个人信息保护法是我国未来发展的必然选择。从国际社会来看，制定个人信息保护法符合国际个人信息保护立法发展的大势所趋。从国内来看，制定统一的个人信息保护法是我国个人信息保护和社会经济法发展的现实需求。首先，能够回应社会各界呼声，近年来不断发生的个人信息泄露事件引发高度关注，制定一部统一的个人信息保护法成为社会各界的一致呼声。其次，有利于从国家层面明确重大问题和基本制度，通过制定统一的个人信息保护法，明确公民个人对其信息享有的基本权利，企业收集和使用个人信息的基本规范，并对数据跨境流动规则等基本问题作出回应。最后，有利于促进我国互联网产业做大做强。制定个人信息保护法不仅能加强我国公民个人权利的保护，更能为我国互联网企业开拓海外市场提供良好的信誉保证。

（三）数据流通规则

随着大数据应用日益渗透到各行各业中，数据所蕴含着的巨大商业价值也越来越为人们所重视，数据日益成为重要的企业资产和国家战略资源。数据资源通过交易流通，能释放更大的价值，提升生产效率，推进产业创新。通过市场化的手段来促进数据流通成为一种趋势，数据流通市场应运而生。

2014年5月27日，美国联邦贸易委员会（FTC）公布《数据经纪商：呼吁透明度与问责制度》报告，该报告显示，美国“数据经纪商”行业正在迅速发展壮大，海量的个人数据被用来进行以营销为目的的消费者数据分析，数据收集和分析几乎覆盖全美消费者，所提供的服务包括出售市场营销产品、开展风险消减服务、提供查询服务等。报告分别针对这三类数据经纪商提出了法律规制方面的建议。虽然数据经纪商提供的服务不同，但从FTC针对三类数据经纪商的规制建议来看，目的都是为了保障消费者对个人数据被收集和使用时的知情权和访问权，在个人数据被不当利用时还可赋予用户退出等相关权利。2017年1月，欧盟委员会发布了《打造欧洲数字经济 数据所有权》的白皮书。白皮书指出，欧盟现有的数据相关法律框架并不能有效地促进数据的运作，建议未来从法律角度对数据所有权问题进行回应。报告基于数据所有权，提出创造一种非排他性、灵活和可扩展的所有权，这种非排他性的所有权可以由对其进行系统性操作的参与者来申明。

当前我国并没有针对数据流通的专门性立法，立法中与数据流通

相关的内容一般是针对涉禁止违反信息传播、禁止侵害商业秘密、个人信息保护、数据安全以及平台责任方面。针对上述内容，《民法总则》《网络安全法》《著作权法》《保守国家秘密法》《全国人民代表大会常务委员会关于维护互联网安全的决定》《全国人民代表大会常务委员会关于加强网络信息保护的决定》《消费者权益保护法》《国家安全法》《刑法》《反恐怖主义法》《反不正当竞争法》等法律，《互联网信息服务管理办法》《著作权法实施条例》《电信条例》《国际联网安全保护管理办法》等行政法规，《电信和互联网用户个人信息保护规定》《网络借贷信息中介机构业务活动管理暂行办法》《互联网新闻信息服务管理规定》等部门规章，《最高人民法院最高人民检察院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》等六部司法解释，《互联网新闻信息服务新技术新应用安全评估管理规定》等规范性文件作出了规定。

数据资源的流动性和可获取性是大数据应用和产业发展的基础。为促进数据的开放和社会化利用，规范数据流通行为，应当建立数据流通负面清单制度，禁止危害国家安全、侵犯个人信息及企业商业秘密等数据的流通。此外，数据流通主体（企业、数据交易平台等）应对可能侵犯个人信息、商业秘密、国家安全等的数据流通进行风险评估，及时终止不当流通行为，并向国家数据流通主管部门报告。还应当完善第三方平台监管机制，建立数据交易机构资质审核和准入机制，加强事前准入、事中监测和事后处置等监管机制和手段。可以通过建立具有可操作性的数据流通规则及相关标准，鼓励地方和行业组织先

试先行，尝试创设相关规则 and 标准，对流通过程中的数据质量、数据分类、数据安全等问题作出规范。

九、建议与展望

大数据从概念产生到应用成熟，中间横亘着一道又一道的障碍。能否突破这些障碍，关系到大数据能否发挥实效，真正成为引领信息技术变革、助力数字经济发展、提升政府治理能力和公共服务水平的关键因素。在《大数据白皮书（2016 年）》中，我们提出了避免盲目跟风、推动数据共享、强调供需对接、完善法律制度、突出地方特色等五点建议。这些问题有些得到了改善，有些仍然是大数据发展过程中较大的问题。站在当下，我们提出针对大数据发展的如下几点建议：

（一）制度与技术双管齐下，打破数据孤岛

数据流通不畅一直是制约大数据发展的关键障碍。人人都想要别人的数据，但都不愿意把自己的数据给别人。与此同时，以前信息系统建设都从一个个“烟囱”开始，数据缺乏互通的技术基础。从国家层面到企业内部，情况大同小异。麦肯锡 2016 年底的一份报告显示，大数据在很多领域没有达到预期效果，很重要的原因就是数据割裂。为解决这一问题，需要制度手段与技术手段双管齐下。这些年，推动数据开放共享的政策举措在一直在加强，然而效果与预期还有差距，未来，如果同态加密、差分隐私、多方安全计算、零知识证明技术如能进一步取得突破，数据共享和流通将有望再前进一大步。

（二）内部与外部多重并举，推动数据治理

据调查,数据分析工作,往往有 80%的时间和精力都耗费在搜集、清洗和加工数据上。数据质量不过关,会让数据分析效果大打折扣,甚至让分析结果谬以千里。很多单位大数据应用效果不佳,多半问题出在数据管理上。大家都同意把数据当做资产,甚至认为有朝一日会计入资产负债表。然而,数据资产管理不像大数据分析挖掘那么光鲜亮丽,就像城市的“下水道工程”,短期只有投入没有产出。但长期来说又不得不做,是战略层面的事情,否则返工的成本巨大。以后,随着每个企业都将成为数据驱动的企业,数据资产管理这样基础性的操作要尽早完成。同时,全行业的数据治理也应提上日程。例如,金融行业的《银行金融机构数据治理指引》就将整个行业的数据治理进行了顶层设计,为行业数据融通奠定了坚实基础,也为其它行业的数据治理开了一个好头。

（三）业务与数据加速融合，深化数据应用

如前所述,虽然大数据的应用取得了长足进展,但行业与大数据融合的不平衡问题还很严重。目前,大数据在互联网、金融、电信等领域产生了实实在在的效益,医疗、工业领域也正在加速。但总体上只能说刚刚走出了半步,大多数是“平行替代”或“补课”,还远远没能达到“深度融合”的阶段。例如,在金融和电信行业,往往只是采用 Hadoop 等工具来重构原来的昂贵的数据仓库。而政务、医疗、工业等领域的大数据应用,则大多是“补课”:即在业务系统之外,新建原来缺失的数据平台。客观地说,目前这样的阶段对于很多行业

来说是“必经之路”。在这一阶段，需要鼓励大数据技术企业不断提升大数据平台和应用的可用性和操作便捷程度，优先支持面向传统企业的产品、服务和解决方案的开发，简化大数据底层繁琐复杂的技术，便捷大数据应用的部署。随着这些“替代”或“补课”的深入推进，业务与数据将加深融合，数据驱动的新模式、新业态更值得期待。

（四）监管与自律同时推进，保障数据安全

数据安全是大数据发展的底线。如前所述，我国大数据的安全保障能力还不够强，安全体系建设还未完成。一方面需要强化数据法律的建设，加强重要基础设施和关键领域的法律监管，尤其在个人信息保护方面需要“重拳整治”。另一方面需要强调行业自律，由于数据不可避免的出现“寡头现象”，部分大企业所拥有的数据涉及到众多用户的信息安全，这就需要企业强化自律。从政府角度，需要主动适应并努力引领新变化，加强政策、监管与法律的统筹协调，动态优化政策法规体系，积极构建大数据健康发展的有利环境。

CAICT 中国信通院

中国信息通信研究院

地址：北京市海淀区花园北路 52 号

邮政编码：100191

联系电话：010-62300249

传真：010-62304980

网址：www.caict.ac.cn

