大数据

在公共资源交易领域中的应用研究

Research on the Application of Big Data in Public Resource Trading

秦滔◎著

湘潭大学出版社

大数据

在公共资源交易领域中的应用研究

Research on the Application of Big Data in Public Resource Trading

秦滔◎著

湘潭大学出版社

版权所有　侵权必究

图书在版编目（CIP）数据

大数据在公共资源交易领域中的应用研究 / 秦滔著．
— 湘潭 ：湘潭大学出版社，2019.9
ISBN 978-7-5687-0354-3

Ⅰ．①大… Ⅱ．①秦… Ⅲ．①数据处理－应用－政府采购制度－研究－中国 Ⅳ．①F812.2-39

中国版本图书馆 CIP 数据核字（2019）第 208396 号

大数据在公共资源交易领域中的应用研究

DASHUJU ZAI GONGGONG ZIYUAN JIAOYI LINGYU ZHONG DE YINGYONG YANJIU

秦滔　著

责任编辑：王亚兰
装帧设计：何　健
出版发行：湘潭大学出版社
社　　址：湖南省湘潭大学工程训练中心
电　　话：0731-58298960 0731-58298966（传真）
邮　　编：411105
网　　址：http://press.xtu.edu.cn/
印　　刷：长沙鸿和印务有限公司
经　　销：湖南省新华书店
开　　本：710 mm×1000 mm 1/16
印　　张：15.75
字　　数：352 千字
版　　次：2019 年 9 月第 1 版
印　　次：2019 年 9 月第 1 次印刷
书　　号：ISBN 978-7-5687-0354-3
定　　价：58.00 元

作者简介：

秦滔，中共湖南省委党校、湖南行政学院教授。湖南省科技战略研究中心专家、湖南省自然辩证法研究会专家。发表多篇数字政府、大数据相关EI论文，著有《数字湖南》等著作。

前　言

新一代信息技术与经济社会各领域的深度融合，引发了数据量的爆发式增长，使得数据资源成为国家重要的战略资源和核心创新要素。大数据作为创新的重要手段，在公共资源交易领域的应用是政府治理能力现代化的重要实践，更是落实中央大数据国家战略精神的重要举措。

本书的写作源于中央有关全面深化改革会议的重要精神。笔者试图借力新一代信息技术在公共资源交易领域中的应用研究，充分利用和释放公共资源交易领域海量数据资源中蕴含的巨大价值，将汇聚、慧治、惠民与大数据发展的要求相结合，嵌入到公共资源交易全过程，打造统一规范、公开透明、阳光交易、效率公平、服务高效、监督到位的公共资源交易环境；通过总结经验、发现规律、预测趋势和辅助决策等手段，助力政府效能改革不断深化，以期达到中央对公共资源交易领域深化改革的发展期望与新要求。

本书共 10 章：第 1 章介绍大数据的内涵与特点、发展时代背景和发展历程等；第 2 章对大数据中国战略进行解读，对大数据在中国的发展现状进行阐述；第 3 章阐述公共资源交易领域中的数据利用；第 4 章是公共资源交易领域大数据在服务政府决策、深化“放管服”和智慧监管方面的应用场景与案例介绍；第 5 章着重对湖南公共资源交易领域的大数据应用痛点进行详细分析，提出痛点、难点的解决策略；第 6 章则构建了公共资源交易大数据平台架构；第 7 章阐述当前公共资源交易领域大数据分析与应用的相关技术，内容涉及大数据采集、技术、计算、分析和可视化等技术；第 8 章重点对湖南公共资源交易领域大数据应用蓝图进行设计与绘制；第 9 章重点阐述公共资源交易领域大数据应用的实践路径，构建了经济应用模型、惠民生模型、智监管模型和智优服务模型等，并细化到公共资源交易领域具体业务的大数据分析与落地方法；第 10 章从公共资源交易领

域大数据发展顶层设计、发展保障机制、数据共享与安全机制和人才机制等方面内容，对公共资源交易领域大数据发展提出合理建议。

当前，大数据在公共资源交易中的应用由点向面发展，但从系统性、学术性角度对于公共资源交易领域中的大数据及其应用方面的研究，市场中暂未有相似著作出版。因此，本书对于公共资源交易领域的专家和相关人员具有借鉴与参考价值。

目　录

第 1 章　大数据时代来临 …… (1)
1.1　大数据的内涵与特点 …… (1)
1.2　大数据发展的时代背景 …… (4)
1.3　大数据发展历程 …… (7)
第 2 章　大数据在中国 …… (12)
2.1　大数据中国战略 …… (12)
2.2　大数据在中国的发展现状 …… (17)
第 3 章　公共资源交易领域中的数据利用 …… (23)
3.1　公共资源交易的相关概念 …… (23)
3.2　大数据在公共资源交易中的价值 …… (27)
3.3　全国公共资源交易中心网站数据应用情况 …… (30)
3.4　全国公共资源交易中心数据应用情况 …… (33)
3.5　湖南省公共资源交易平台数据应用分析 …… (35)
第 4 章　公共资源交易领域大数据应用场景与案例 …… (38)
4.1　服务政府决策等方面的应用案例 …… (39)
4.2　深化“放管服”方面的应用案例 …… (45)
4.3　大数据综合应用
——努力构建贵州省公共资源智慧交易新格局 …… (50)
第 5 章　湖南公共资源交易领域大数据应用痛点分析与解决策略 …… (55)
5.1　湖南省公共资源交易数据应用痛点分析 …… (56)
5.2　公共资源大数据应用问题解决策略 …… (65)

第 6 章　湖南公共资源交易大数据平台建设架构 …… (76)
6.1　湖南省公共资源交易大数据平台建设目标 …… (77)
6.2　湖南省公共资源交易大数据交换平台 …… (79)
6.3　湖南省公共资源交易大数据交换平台功能架构 …… (80)
6.4　湖南省公共资源交易数据湖设计 …… (87)

第 7 章　湖南公共资源交易领域大数据应用建议技术 …… (97)
7.1　公共资源交易领域大数据技术参考架构 …… (99)
7.2　数据采集技术 …… (101)
7.3　数据存储技术 …… (106)
7.4　大数据计算技术 …… (113)
7.5　大数据分析技术 …… (117)
7.6　大数据可视化技术 …… (120)

第 8 章　湖南公共资源交易领域大数据应用蓝图设计 …… (122)
8.1　以“稳”为目标的湖南省公共资源交易大数据应用蓝图 …… (124)
8.2　以“管”为目标的湖南省公共资源交易大数据应用蓝图 …… (127)
8.3　以“服”为目标的湖南省公共资源交易大数据应用蓝图 …… (128)
8.4　以“惠”为目标的湖南省公共资源交易大数据应用蓝图 …… (130)

第 9 章　湖南公共资源交易领域大数据应用实践路径 …… (132)
9.1　湖南省公共资源交易大数据应用实践原则 …… (133)
9.2　湖南省公共资源交易大数据应用项数据可能来源分析 …… (133)
9.3　湖南省公共资源交易大数据应用分析步骤范例 …… (137)
9.4　湖南省公共资源交易大数据统计分析 …… (139)
9.5　湖南省公共资源交易大数据应用 …… (194)
9.6　湖南省公共资源交易大数据综合应用 …… (212)

第 10 章　湖南省公共资源交易大数据发展建议 …… (240)
10.1　优化大数据应用环境 …… (240)
10.2　夯实大数据应用基石 …… (242)

参考文献 …… (245)

第 1 章　大数据时代来临

大数据已经成为社会发展的重要资源和信息资产。在当今信息时代，人类社会生存与发展所依赖的基础由原有的自然资源和能源两种，增加为自然资源、能源和数据三种。我们生活在一个充满“数据”的时代，阅读、购物、看病、旅游、使用微博或微信，人类的行为和机器的行为都在不断产生新的数据，数据的产生如此之快，数据的汇聚如此之广泛，对我们的影响如此深刻，远远超出了人类的想象。随着信息技术和人类生产、生活的深度融合，互联网快速普及，全球数据呈现爆发式增长、海量集聚的特点，数据资源的开发利用，对经济发展、社会进步、国家治理和人民生活产生重大影响。习近平同志强调，大数据发展日新月异，我们应该审时度势、精心谋划、超前布局、力争主动，深入了解大数据发展现状和趋势及其对经济社会发展的影响，分析我国大数据发展取得的成绩和存在的问题，推动实施国家大数据战略，加快完善数字基础设施，推进数据资源整合和开放共享，保障数据安全，加快建设数字中国，更好服务我国经济社会发展和人民生活改善。因此，在这个数据的时代，大数据将为我们提供新的发展机遇和创新发展的能力。

1.1　大数据的内涵与特点

1. 大数据的内涵

大数据，在业界和学术界并未有统一定义。常见的大数据定义有：

大数据是指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合，是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

——麦肯锡

大数据是指无法在一定时间内用常规软件工具对其内容进行抓取、管理和处理的数据集。

——维基百科

大数据是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

——Gartner

大数据指不用随机分析法（抽样调查）这样的捷径，而采用所有数据进行分析处理。

——维克托·迈尔-舍恩伯格和肯尼斯·库克耶编写的《大数据时代》

笔者认为，大数据可以从三个层面来定义或理解：一是大数据是海量数据或巨量数据的集合。从业界与学术界对大数据的定义可见一斑。这一数据集合的数量已超过人类社会五千年来所有数据的总量。二是大数据是海量数据处理、加工、分析与利用的技术工具集合。这一工具集合包括了我们已知的对数据采集、传输、存储、管理、分析和呈现的多种类信息技术。三是大数据是人类思考决策的新思维。人类利用杂合数据源和新的分析技术，挖掘数据背后真实的事物规律，获得决策力、洞察发现力和流程优化能力的新思维。

因此，本书对于大数据的定义如下：大数据是人类社会海量信息资源的数字化集合，其规模超过传统数据软件处理能力范围，需要新的软件工具对数据进行获取、存储、管理和分析，以新方式呈现事物规律，提高人类决策力和洞察发现力的新思维方式。

（1）究竟多大的数据量才可以称之为大数据？

对于数量而言，人类往往是通过计量单位来衡量数量的多少或大小。例如，对于物体的质量以克和千克来衡量，又如两个天体间的距离通常是以千米、亿公里和光年来衡量。在计算机系统中，我们定义信息存储的最小单位为 bit，按照进率 1024（2 的十次方）来计算，其衡量单位由小到大为 bit、Byte、KB、MB、GB、TB、PB、EB、ZB、YB、BB、NB、DB。

1 Byte = 8 bit

1 KB = 1 024 Bytes = 8 192 bit

1 MB = 1 024 KB = 1 048 576 Bytes

1 GB = 1 024 MB = 1 048 576 KB

1 TB = 1 024 GB = 1 048 576 MB

1 PB = 1 024 TB = 1 048 576 GB

1 EB = 1 024 PB = 1 048 576 TB

1 ZB = 1 024 EB = 1 048 576 PB

1 YB = 1 024 ZB = 1 048 576 EB

1 BB = 1 024 YB = 1 048 576 ZB

1 NB = 1 024 BB = 1 048 576 YB

1 DB = 1 024 NB = 1 048 576 BB

多大容量的数据才算大数据，大数据的规模并没有具体的衡量标准，需要根据数据所有者和使用者的情况来确定。对于个人而言，TB 级别的数据可称为大数据；对于数据业务量不大的企业，如普通的政府部门，省级公共资源交易中心、各级政务大厅而言，TB 级别的数据是大数据；对于大的互联网企业，如百度、阿里巴巴而言，EB 级别的数据称为大数据。

数据的规模是衡量大数据的重要指标，但不是唯一的，还需要从另外两个维度来衡量：一是从时间序列累积大量的数据，二是在深度上更加细化数据。比如一份现在看起来很小的数据，但是纵向积累久了也可以变成大数据，横向与其他数据关联起来也可能形成大数据。而一份很大的数据如果没有关联性、没有价值，也不是大数据。

随着云计算、物联网和移动互联网的应用，数据量迅猛增长。数据来源有两种：一种与人有关，比如政府、企业等为人们服务时产生的数据；另一种与物有关，在 5G 移动泛在、万物互联时代，物联网应用的浪潮将带动数据量爆发式增长。数据量的爆发式增长，一方面是信息终端的大面积普及，信息源大量增加；另一方面是基于云计算的互联网信息平台快速增长，数据向平台大规模集中。

(2) 大数据与云计算、人工智能关系

大数据与云计算、物联网和人工智能等新一代信息技术之间相互影响、相互促进、相互融合。例如，运满满通过车联网设备和信息平台，每天获取 3 TB 至 4 TB 的数据，运用先进的大数据算法模型，实现了智能车货匹配、智能实时调度等。云计算是硬件资源的虚拟化，而大数据是海量数据的高效处理。简单来说，云计算是大数据的基础，有了云计算才能大量集中数据从而产生大数据。同时，大数据也支撑了云计算应用创新，带动云计算的发展。人工智能的核心在于大数据支撑。围棋人工智能程序“阿尔法狗”打败柯洁，离不开大数据的支持。大数据技术能够通过数据采集和分析等方式，从海量数据中快速获得有价值的信

息，为深度学习人工智能算法提供坚实的素材基础。反过来，人工智能技术也促进了大数据技术的进步。两者相辅相成，任何一方技术的突破都会促进另外一方的发展。

2. 大数据的特点

维克托·迈尔-舍恩伯格和肯尼斯·库克耶的《大数据时代》将大数据的特征归纳为4个“V”：Volume（大量），Variety（多样），Value（价值）和Velocity（速度）。

一是数据体量巨大。截至目前，人类生产的所有印刷材料的数据量是200 PB（1 PB=210 TB），而历史上全人类说过的所有的话的数据量大约是5 EB（1 EB=210 PB）。

二是数据类型繁多。通常数据被分为结构化数据和非结构化数据。结构化数据常以文本或表格的形式呈现，非结构化数据则包括网络日志、音频、视频、图片和地理位置信息等，人类当前数据中95%是非结构化数据，并且此类数据仍保持高速增长。这些多类型数据对数据的处理能力提出了更高要求。数据类型繁多不仅仅是指数据格式多样性：文字、影音、网页、串流等结构性和非结构性的数据。另外，还涵盖来源多样性：数据可能来源于汽车，可能是沃尔玛每周统计的数据，可能是交警处罚信息等。

三是价值密度低，商业价值高。价值密度的高低与数据总量的大小成反比。以视频为例，一部1小时的视频，在连续不间断的监控中，有用的数据可能仅有一二秒。如何通过强大的机器算法更迅速地完成数据的价值“提纯”，成为目前大数据背景下亟待解决的难题。

四是处理速度快。这一点也是和传统的数据挖掘技术有着本质的不同。例如，Google Flu Trends对于流感预测，通过实时收集数据进行运算，如果运算速度比较慢，处理不了那么多数据，那么在流感爆发的时候，都还没有计算出这个区域有没有流感潜伏，这就失去了时效性。所以处理速度很重要。

大数据的意义就在于，可供人类分析和使用的数据量在大增，通过对大数据的整合、交换、分析和挖掘，可以发现新的知识、创造新的价值，带来“大知识”“大利润”“大科技”和“大发展”。

1.2 大数据发展的时代背景

随着互联网的兴起和网络技术的飞速发展，越来越多的人与物理设备接入和

使用网络，全球网络中的在线人数与物理设备数量急剧增加。网络中的人，以及接入的设备产生大量的数据。例如，人类用手机会产生数字信息，工业互联网设备产生工业数字信息，接入网络的交通摄像头产生视频信息，气象传感器产生气象信息，我们的生活被数字信息包围。同时，智能设备的普及、存储设备性能的提高和网络带宽的增加，以及计算机的运算能力不断提升，为大数据的采集、存储、传输、处理和分析提供了基础。

1. 数据大量产生

从人类有了计算机开始，信息技术经历了三个阶段：第一阶段是计算机的发明与应用期，这个阶段数据的处理量非常小，计算机的运算能力相对较弱；第二个阶段是互联网的发明与应用期，这一阶段互联网只是为数不多的企业、政府、高校和个人在使用，传输的数据也不多；第三阶段是互联网的发展与普及期，这个阶段一直延续至今，并且仍将持续下去。互联网的普及深刻地改变着整个社会，是人类始料未及的。其他所带来的是大量数据的高速产生与广泛利用，并使数据成为人类社会发展的最重要资源之一。整体来看，数据大量产生是由于上网人数与上网设备的急剧增加。

随着信息化进程的不断深入，企业、政府和个人的数据不断增长，尤其企业与政府的数据成倍增长。例如，搜索引擎服务终端百度的数据量 1 000 PB，每天响应 138 个国家数十亿次请求，每日新增 10 TB；2019 年一季度微信用户数量达 11 亿，使用超过 4 400 台服务器存储用户产生的信息，压缩后的数据 100 PB，每天新增 200～300 TB；智慧城市中 1080P 高清网络摄像机一个月产生 1.8 TB 数据，大城市 50 万个摄像头，一个月产生 3 PB 的数据量；2018 年双十一当天，淘宝总计产生了超过 10 亿条包裹的数据。又如，截至 2018 年 12 月，我国网民规模为 8.29 亿人次，互联网普及率达 59.6%。美国互联网用户规模接近 2.5 亿，非洲以及南美地区的第三世界国家互联网网民数也增长迅速。截至 2018 年 12 月，我国手机网民规模达 8.17 亿人次，全年新增手机网民 6 433 万人次；我国网民使用手机上网的比例达 98.6%，较 2017 年底提升 1.1 个百分点；网民使用电视上网的比例达 31.1%，较 2017 年底提升 2.9 个百分点；使用台式电脑上网的比例为 48.0%，较 2017 年底下降 5 个百分点（数据来源：《CNNIC：2019 年第 43 次中国互联网络发展状况统计报告》）。《全球联网和物联网设备预测更新》指出，截至 2018 年底，全球联网设备数量达到 220 亿台。该报告预测，到 2025 年将有 386 亿台联网设备，到 2030 年将有 500 亿台联网设备，如图 1.1 和图 1.2 所示。

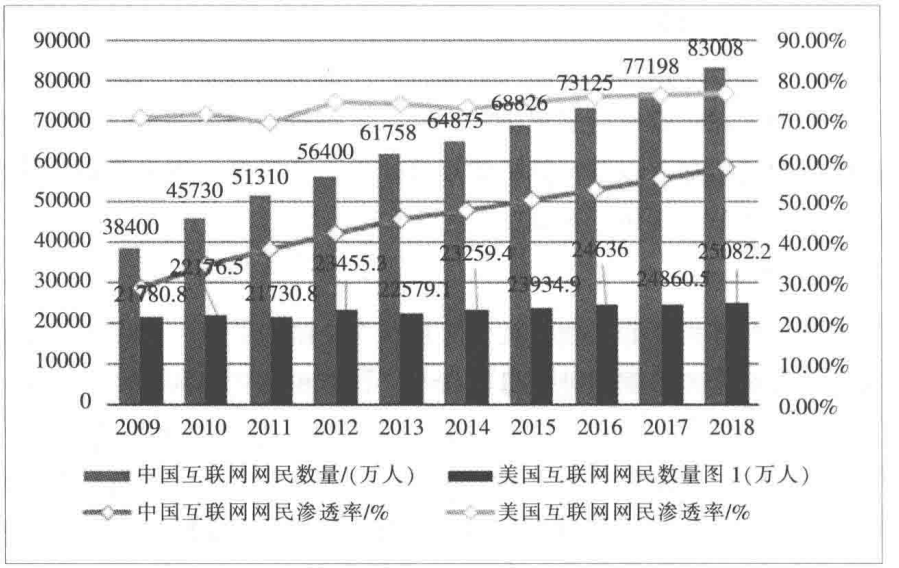


图 1.1　2009～2018 年中美互联网网民数量及渗透率对比

（数据来源：赛迪资讯）

当前数据的主要来源有：

① 物联网传感器和视频监控设备时时刻刻都在收集海量数据；

② 计算机产生的数据；

③ 用户越来越多的线上行为产生了大量非结构化数据。

2. 数据处理能力大幅提高

从 1946 年计算机发明、1965 年摩尔定律、1988 年普适计算、1989 年数据挖掘、2004 年社交媒体出现、2005 年物联网出现，至 2012 年大数据元年，我们所看到的是计算机处理能力越来越强大，获取数据计算能力的方式越来越便捷和便宜，计算机系统对于数据的存取和加工处理能力越来越强，挖掘数据的能力成倍增加，深度与广度不断加大，数据产生的主体越来越多样化、产生的速度一日千里，数据呈爆炸式的增长，数据的价值由量变变为了质变。AI、区块链、大数据、云计算和 5G 等技术的发展与突破，带来了算力的巨大提升和释放，技术之间的相互取长补短实现了大数据作为生产资料、算力作为生产力、区块链作为生产关系的科技矩阵。

数据处理能力主要由计算机系统提供，这其中可以是单个计算机，也可以是云计算，更可能是超级计算机。现在，单个计算机运算 1024 M 位的指数仅需 31 秒；

2018 年 3 月，NVIDIA 推出最新小型超算——“NVIDIADGX2”，它采用 16 块 TeslaV100 计算卡并行计算，能够提供最高达 2PFLOPs（2000TPLOPs）的深度计算能力。而云计算则更为强大，通过网络“云”将巨大的数据计算处理程序分解成无数个小程序，然后，通过多部服务器组成的系统处理和分析这些小程序得到结果并返回给用户。利用这项技术，可以在很短的时间内（几秒钟）完成对数以万计的数据的处理，从而实现强大的网络服务。超级计算机则能提供强悍算力，具有很强的计算和处理数据的能力，中国的“神威·太湖之光”，持续性能为 9.3 亿亿次/秒，峰值性能可以达到 12.5 亿亿次/秒。2019 年全球排名第一的超级计算机 Summit（美国）其数据处理能力的峰值速度为 20 亿亿次/秒。

1.3　大数据发展历程

大数据发展至今，大体上可以分为三个阶段：

（1）萌芽（至 1980 年）：随着数据挖掘理论和数据库技术的成熟，一些商业智能工具和知识管理技术开始出现。1980 年，托夫勒的《第三次浪潮》一书中将“大数据”称颂为“第三次浪潮的华彩乐章”。这是人类第一次对大数据的出现所做的描述与对未来数据价值的预测。

（2）出现阶段（1980～2008 年）：1997 年，迈克尔·考克斯和大卫·埃尔斯沃斯描述“模拟飞机周围的气流——是不能被处理和可视化的。数据集通常之大，超出了主存储器、本地磁盘，甚至远程磁盘的承载能力，称之为‘大数据问题’”。2002 年在“9·11”袭击事件后，美国政府前国家安全顾问约翰·波因德克斯特领导国防部整合现有政府的数据集，组建了一个用于筛选通信、犯罪、教育、金融、医疗和旅行等记录来识别可疑人的大数据库。2007～2008 年，随着社交网络的激增，技术博客和专业人士为“大数据”概念注入了新的生机。《连线》的克里斯·安德森认为：当时处于一个“理论终结时代”“当前世界范围内已有的一些其他工具将被大量数据和应用算法所取代”“应该深入参与大数据计算的开发和部署工作，因为它将直接有利于许多任务的实现”。2008 年 9 月，《自然》杂志推出名为“大数据”的封面专栏。

（3）热门阶段（2009 年至今）：2009 年，美国政府通过启动 Data.gov 网站；2010 年，肯尼斯库克尔发表大数据专题报告——《数据，无所不在的数据》；2011 年，麦肯锡全球研究院发布《大数据：下一个创新、竞争和生产力的前沿》，称“大数据时代已经到来”。2012 年，美国政府发布大数据国家战略，欧

洲主要国家相续发布其大数据国家战略，至此，大数据成为政府、企业和大众所关注的热点，持续至今。2013 年是中国的大数据元年，人们对于大数据的热情不断增长，为大数据发展提供了支撑，中国政府又于 2015 年发布大数据国家，大数据以势不可挡的姿态进入人们的思想意识，并在社会的各个领域探索与落地实践。

1. 大数据发展历程事件回顾（2005～2018 年）[①]

2005 年，Hadoop 项目诞生。Hadoop 最初只是雅虎公司用来解决网页搜索问题的一个项目，后来因其技术的高效性，被 Apache Software Foundation 公司引入并成为开源应用。Hadoop 本身不是一个产品，而是由多个软件产品组成的一个生态系统，这些软件产品共同实现全面功能和灵活的大数据分析。从技术上看，Hadoop 由两项关键服务构成：采用 Hadoop 分布式文件系统（HDFS）的可靠数据存储服务，以及利用一种叫作 MapReduce 技术的高性能并行数据处理服务。这两项服务的共同目标，是提供一个使结构化和复杂数据的快速、可靠分析变为现实的基础。

2008 年，“大数据”得到部分认可。业界组织计算社区联盟（Computing Community Consortium）发表了一份有影响力的白皮书——《大数据计算：在商务、科学和社会领域创建革命性突破》，它使人们的思维不止局限于数据处理的机器，提出了大数据真正重要的是新用途和新见解，而非数据本身。此组织可以说是最早提出大数据概念的机构。

2009 年，印度政府建立生物识别数据库。联合国全球脉冲项目已研究了对如何利用手机和社交网站的数据源来分析预测从螺旋价格到疾病暴发的问题。

2009 年，美国政府启动 Data.gov 网站，向公众提供各种各样的政府数据。该网站超过 4.45 万点击量的数据集被用于保证一些网站和智能手机应用程序来跟踪从航班到产品召回再到特定区域内失业率的信息，这一行动激发了从肯尼亚到英国范围内政府相继推出类似举措。

2010 年，肯尼斯·库克尔发表大数据专题报告——《数据，无所不在的数据》。报告中提道：世界上有着无法想象的巨量数字信息，并以极快的速度增长。

① 今日头条，https://www.toutiao.com/a6629481307321664013/?tt_from=android_share&utm_campaign=client_share&timestamp=1562990510&app=news_article&utm_medium=toutiao_android&req_id=20190713120149010152019225718BA9B&group_id=6629481307321664013。

从经济界到科学界，从政府部门到艺术领域，很多方面都已经感受到了这种巨量信息的影响。科学家和计算机工程师已经为这个现象创造了一个新词汇“大数据”。库克尔也因此成为最早洞见大数据时代趋势的数据科学家之一。

2011 年，IBM 的沃森超级计算机每秒可扫描并分析 4 TB（约 2 亿页文字量）的数据量。

2011 年，全球知名咨询公司麦肯锡全球研究院发布《大数据：创新、竞争和生产力的下一个新领域》，大数据开始备受关注，这也是专业机构第一次全方位介绍和展望大数据。报告指出，大数据已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来。报告还提到，“大数据”源于数据生产和收集能力以及速度的大幅提升——由于越来越多的人、设备和传感器通过数字网络连接起来，产生、传送、分享和访问数据的能力也得到彻底变革。

2011 年，工信部发布的物联网“十二五”规划中，把信息处理技术作为 4 项关键技术创新工程之一被提出来，其中包括海量数据存储、数据挖掘和图像视频智能分析，这些都是大数据的重要组成部分。

2012 年，在瑞士达沃斯召开的世界经济论坛上，大数据是主题之一，会上发布的报告《大数据，大影响》（Big Data，Big Impact）宣称，数据已经成为一种新的经济资产类别，就像货币和黄金一样。

2012 年，美国奥巴马政府在白宫网站发布了《大数据研究和发展倡议》，这一倡议标志着大数据已经成为重要的时代特征。2012 年 3 月 22 日，奥巴马政府宣布将花费 2 亿美元投资大数据领域，这是大数据技术从商业行为上升到国家科技战略的分水岭，在次日的电话会议中，政府将数据定义为“未来的新石油”，大数据技术领域的竞争，事关国家安全和未来。国家层面的竞争力将部分体现为一国拥有数据的规模、活性以及解释、运用的能力；国家数字主权体现了对数据的占有和控制。数字主权将是继边防、海防、空防之后，另一个大国博弈的空间。

2012 年，美国软件公司 Splunk 于 4 月 19 日在纳斯达克成功上市，成为第一家大数据处理的上市公司。Splunk 成功上市促进了资本市场对大数据的关注，同时也促使 IT 厂商加快大数据布局。

2012 年，联合国在纽约发布了一份关于大数据政务的白皮书，总结了各国政府如何利用大数据更好地服务和保护人民。这份白皮书举例说明在一个数据生