

基于大数据环境下创新型 档案管理与服务研究

谭萍著

吉林人民出版社

前 言

信息时代，大数据技术冲击着整个世界，变革着人们的生活、工作和思维方式。在大数据环境下，档案管理工作也受到了很大地影响和冲击，档案部门的档案管理能力和水平面临全新的考验。对大数据环境下的档案管理模式变化进行研究，有利于保护各类档案文化资源、优化档案业务流程、挖掘档案增值信息、加强档案管理机构的职能建设，达到促进档案事业发展的目的。

大数据指的是需要创新处理模式才能具有更强的决策力、洞察力和流程优化能力的海量、高增长率和多样化的信息资产。因此，档案也可以归入大数据的行列，但反之，大数据可不仅局限于档案，档案仅为大数据的一种而已。在大数据环境下，伴随着互联网技术的飞速发展、各类社会媒体的普遍应用，数据呈现出数量大、种类多、变化迅速、价值总量高的特点，而档案信息资源作为最重要的社会信息资源之一，也呈现出大数据的特点。在大数据环境下，档案信息资源的数量急剧增长，种类愈发繁杂，数字化、信息化水平不断提升，使用传统的管理手段已经难以处理新形态的档案信息资源，对大数据环境下的创新型档案管理与服务的研究显得愈发重要。

《基于大数据环境下创新型档案管理与服务研究》正是在这样的时代背景下，围绕档案管理与服务这一关键内容进行的研究。本著作共分为八章，从大数据的理论知识 and 档案的基础知识入手，对传统档案工作与管理进行了简单阐述，重点讲述了大数据环境下档案管理方面的创新内容，探析了在大数据环境下的档案整合、挖掘、开发与利用等问题，研究了大数据环境下档案信息服务的创新性内容，以及档案编研工作的意识创新与实践创新路径，最后对大数据环境下档案工作的终端——智慧档案馆进行了建设实例分析与改进策略总结，提高了本书的实用性与应用性。本书既可以作为学习档案管理的学生的辅导书，又可以作为研究大数据的人员及对档案工作有研究人员的参考书。

本书在撰写过程中得到了国内外许多专家学者的支持和帮助，在此致以衷心的感谢。另外，由于时间仓促，书中难免存在疏漏之处，恳请广大读者批评指正。

目 录

第一章 大数据概述	1
第一节 大数据的定义	1
第二节 大数据的发展趋势	6
第三节 大数据的研究现状与未来展望	9
第四节 大数据的应用与需求	14
第二章 档案管理工作的基础认知	22
第一节 档案的发展沿革	22
第二节 档案的分类	27
第三节 档案工作的内容与性质	31
第四节 档案工作的要求及意义	38
第三章 传统文件档案及其管理	54
第一节 文书档案及其管理	54
第二节 人事档案及其管理	60
第三节 会计档案及其管理	67
第四节 科技档案及其管理	78
第四章 电子文件管理及纸质档案的数字化	87
第一节 电子文件的计算机管理	87
第二节 电子文件管理软件及其应用	104
第三节 纸质档案的数字化	109

第五章 大数据环境下档案管理探析	119
第一节 档案信息资源的整合	119
第二节 档案信息资源的挖掘	135
第三节 档案信息资源的开发与利用	148
第六章 大数据环境下档案信息服务创新性研究	162
第一节 档案信息服务研究现状	162
第二节 档案信息服务研究的主要内容	165
第三节 档案信息服务模式面临的机遇与挑战	175
第四节 档案信息服务应用实例分析	178
第七章 大数据环境下档案创新性编研	186
第一节 基于大数据环境下档案编研工作现状	186
第二节 基于大数据环境下档案编研工作的意识创新	188
第三节 基于大数据环境下档案编研工作的实践创新	191
第八章 基于大数据环境下档案工作的终端——智慧档案馆	197
第一节 智慧档案馆概述	197
第二节 我国部分城市智慧档案馆的建设	202
第三节 智慧档案馆风险分析	205
第四节 建设智慧档案馆存在的问题及策略	209
参考文献	214
后 记	216

第一章 大数据概述

第一节 大数据的定义

一、大数据的概念及其内涵

大数据，也就是海量数据，一般指所含的数据集规模巨大，现在大众的软件工具无法在合理的时间进行采集、存储、分析管理的数据信息。因其在各个行业的广泛应用，使之关注热度历年来居高不下。作为人们获得新的认知、理念和创造价值的源泉，大数据的数据来源可以囊括我们在日常生活中普遍可以见到的上传到网页上的图像、视频、录音，高速公路上车辆与收费记录、日常监控录像，医院的治疗病例、高端的基因测序，天文学中通过望远镜收集的信息数据等。

大数据最早出现在何时，何地？通过对大量文献资料追踪溯源，笔者发现这个词出现的最早时间在 1980 年的美国，著名的未来学家托夫勒在其所著的《第三次浪潮》中，将大数据热情地颂为“第三次浪潮的华彩乐章”。2008 年 9 月，《自然》杂志推出了名为“大数据”的封面专栏。2009 年“大数据”才开始成为互联网技术行业中的热门词汇，被世人推崇讨论。从 1980 年到 2019 年，尽管大数据的发展已有 30 多年的时间，但目前对于大数据仍没有一个统一的、完整的、科学的定义。

（一）狭义的大数据概念

受早期研究者将数据作为一种工具的思想的影响，很多研究机构和学者将大数据作为一种辅助工具或者从其体量特征来进行定义。

高德纳（Gartner）咨询管理公司数据分析师 Merv Adrian 认为，大数据超过了在正常的时间内和常用硬件环境下，常规的软件工具计算、分析相关数据的能力。

作为大数据研究讨论先驱者的咨询公司麦肯锡，2011 年在其大数据的研究报告

Big Data: The next frontier for innovation, competition, and productivity 中根据大数据的数据规模来对其诠释, 它给出的定义是: 大数据指的是规模已经超出了传统的数据库软件工具收集、存储、管理和分析能力的数据集。需要指出的是, 麦肯锡在其报告中强调, 并不是超过某一个特定的数据容量才能定义为大数据, 因为随着技术的不断进步, 其数据集容量也会不断地增长, 行业的不同也会使大数据的定义不同。

电子商务行业的巨人亚马逊的专业大数据专家 John Rauser 将大数据定义为超过了一台计算机的设备、软件等的处理能力的数据量。

日本野村综合研究所的著名学者城田真琴和周自恒在其专著《大数据的冲击》中通过对大数据的起源进行探讨, 对大数据作出如下定义: 大数据指的是运用现有的一般技术难以进行管理的大量数据的集合^①。

简以概之, 对于大数据的狭义理解, 研究者大多从微观的视角出发, 将大数据理解为当前的技术环境难以处理的一种数据集或者能力; 而在宏观方面, 研究者目前则还没有提出一种明确的看法, 但多数都提出了对于大数据的宏观理解, 需要注意其不同行业领域的差异及随着技术进步, 其数据容量不断增长的特点。

(二) 广义的大数据概念

广义的大数据概念是以对大数据进行分析管理, 挖掘数据背后所蕴含的巨大价值为视角, 对大数据的概念进行定义。

维基百科对大数据给出的定义是: 或称为巨量数据、大资料, 指的是所涉及的数据量规模巨大到无法通过当前的技术软件和工具在一定的时间内进行截取、管理、处理, 并整理成为需求者所需要的信息。

肯尼思·库克耶, 被誉为“大数据时代的预言家”的维克托·迈尔-舍恩伯格在其专著《大数据时代: 生活、工作与思维的大变革》中将大数据定义为: 大数据是人们获得新的认知、创造新的价值的源泉; 大数据还为改变市场、组织机构, 以及政府与公民关系服务^②。他们还认为大数据是人们在大规模数据的基础上可以做到的事情, 而这些事情在小规模的数据基础上是无法完成的。

IBM 组织则是从大数据的特征出发对其进行定义, 它认为大数据具有 3V 特征, 即: 规模性 (Volume)、多样性 (Variety) 和高速性 (Velocity), 故大数据是指具有容量难以估计、种类难以计数且增长速度非常快的数据。

① 城田真琴. 大数据的冲击 [M]. 周自恒, 译. 北京: 人民邮电出版社, 2016: 13.

② 维克托·迈尔-舍恩伯格, 肯尼思·库克耶. 大数据时代: 生活、工作与思维的大变革 [M]. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2012: 12.

国际数据公司（IDC）则在 IBM 的基础上，根据自己的研究，将 3V 发展为 4V，即数据规模巨大（Volume）、数据的类型多种多样（Variety）、数据的产生、处理、分析速度加快（Velocity）、数据的价值难以估测（Value）。所以，IDC 认为，大数据指的是具有规模海量、类型多样、数据的产生、处理、分析速度加快，需要超出典型的数据库软件进行管理，且能够给使用者带来巨大价值的数据集。

通过对大数据的定义进行梳理可以发现，大多数研究机构和学者是从数据的规模量，以及对于数据的处理方式来对大数据进行定义的，且多是从自身的研究视角出发，因此对于大数据的定义，可谓是仁者见仁，智者见智。

本书在参照了学术领域各个研究机构和行业企业对大数据的定义的基础上，将大数据定义为在信息爆炸时代所产生的巨量数据或海量数据，并由此引发的一系列技术及认知观念的变革。它不仅仅是一种数据分析、管理以及处理方式，也是一种知识发现的逻辑，通过将事物量化成数据，对事物进行数据化的研究分析。大数据具有客观性、可靠性，既是一种认识事物的新途径，又是一种创新发现的新方法。

二、大数据的应用

大约从 2009 年开始，“大数据”成为互联网信息技术行业的流行词汇。美国互联网数据中心指出，互联网上的数据每年将增长 50%，每两年翻一番，而目前世界上 90% 以上的数据是最近几年才产生的。此外，数据又并非单纯指人们在互联网上发布的信息，全世界的工业设备、汽车、电表上有着无数的数码传感器，随时测量和传递着有关位置、运动、震动、温度、湿度乃至空气中化学物质的变化，这也产生了海量的数据信息。

大量数据的充斥不只影响企业界。贾斯汀·格里莫将数学与政治科学联系起来，对博客文章、国会演讲和新闻稿进行计算机自动化分析等，希望借此洞察政治观点是如何进行传播的。在科学、体育、广告和公共卫生等其他许多领域中，也有着类似的情况。这意味着大数据应用已经朝着数据驱动型的发现和决策的方向发生转变。

在公共卫生、经济发展和经济预测等领域中，大数据的预见能力也正在被开发。在一次调查中，研究者发现“流感症状”和“流感治疗”等词汇在谷歌上的搜索查询量增加，而在几个星期以后，到某个地区医院急诊室就诊的流感病人数量就有所增加，因此，Google 流感趋势（Google Flu Trends）就可以利用搜索关键词预测流感疫情。

除此之外，洛杉矶警察局和加利福尼亚大学合作，利用大数据预测犯罪的发生；

统计学家内特·西尔弗(Nate Silver)利用大数据预测2012年美国选举结果;麻省理工学院利用手机定位数据和交通数据建立城市规划;梅西百货根据需求和库存的情况,基于SAS的系统对多达7300万种货品进行实时调价;医疗行业早就遇到了海量数据和非结构化数据的挑战,近年来很多国家都在积极推进医疗信息化发展,使医疗机构可以做大数据分析。

三、大数据的价值及其战略意义

(一) 大数据的价值

现在的社会是一个高速发展的社会,科技发达,信息通畅,人们之间的交流越来越密切,生活也越来越方便,大数据就是这个高科技时代的产物。阿里巴巴创办人马云曾提到,未来的时代将不是IT时代,而是DT(Data Technology,数据科技)的时代,可见大数据的重要性。

有人把数据比喻为蕴藏能量的煤矿。煤炭按照性质有烟煤、无烟煤、褐煤等,而露天煤矿、深山煤矿的挖掘成本又不一样。与此类似,大数据并不在“大”,而在于“有用”。价值含量、挖掘成本比数量更为重要。对于很多行业而言,如何利用这些大规模数据是赢得竞争的关键。

大数据的价值体现在三个方面:(1)为大量消费者提供产品或服务的企业可以利用大数据进行精准营销;(2)做小而美模式的中小微企业可以利用大数据做服务转型;(3)面对互联网的压力,必须转型的传统企业需要与时俱进,充分利用大数据的价值。

不过,“大数据”在经济发展中具有巨大作用并不代表其能取代一切对于社会问题的理性思考,科学发展的逻辑不能被湮没在海量数据中。著名经济学家路德维希·冯·米塞斯曾提醒过:“就今日言,有很多人忙碌于资料之无益累积,以致对问题之说明与解决,丧失了其对特殊的经济意义的了解。”这确实是需要警惕的。

在这个快速发展的智能硬件时代,困扰应用开发者的一个重要问题就是如何在功率、覆盖范围、传输速率和成本之间找到那个微妙的平衡点。利用相关数据和分析,企业组织可以降低成本、提高效率、开发新产品、做出更明智的业务决策等等。通过结合大数据和高性能的分析,下面这些对企业有益的情况都可能会发生:(1)及时解析故障、问题和缺陷的根源,每年可能为企业节省数十亿美元。(2)为成千上万的快递车辆规划实时交通路线,避免拥堵。(3)分析所有SKU,以利润最大化为目标来定价和清理库存。(4)根据客户的购买习惯,为其推送他可能感兴趣的优惠信息。

(5) 从大量客户中快速识别出金牌客户。(6) 使用点击流数据分析和数据挖掘来规避欺诈行为。

(二) 大数据的战略意义

大数据技术的战略意义不在于掌握庞大的数据信息，而在于对这些含有意义的数据进行专业化处理。换言之，如果把大数据比作一种产业，那么这种产业实现盈利的关键在于提高对数据的“加工能力”，通过“加工”实现数据的“增值”。

随着云时代的来临，大数据吸引了越来越多人的关注。《著云台》的分析师团队认为，大数据通常用来形容一个公司创造的大量非结构化和半结构化数据，这些数据在下载至关系数据库用于分析时会花费过多时间和金钱。大数据分析常和云计算联系到一起，因为实时的大型数据集分析需要像 Map Reduce（一种编程模型，用于大规模数据集的并行运算）一样的框架来向数十、数百，甚至数千的计算机分配工作。

四、大数据与传统数据库

大数据分析相比于传统的数据仓库应用，具有数据量大、查询分析复杂等特点。传统数据库是 GB/TB 级（计算机储存容量单位，1PB=1024TB、1TB=1024GB）高质量、较干净、强结构化、自上而下、重交易、确定解的数据仓库。大数据是 PB 级以上的，有噪声、有冗余、非结构化、自下而上、重交互、满意解的数据仓库。大数据出现后，大数据对数据库高并发读写要求、对海量数据的高效存储和访问需求、对数据库高可扩展性和高可用性的需求，使传统 SQL（关系型数据库）逐渐没有用武之地，NoSQL（非关系型数据库）这一模式变得非常流行。互联网巨头对于 NoSQL 数据模式应用非常广泛，如谷歌的 Big Table、Facebook（社交网站）的 Cassandra（开源分布式 NoSQL 数据库系统）、甲骨文公司的 NoSQL 及亚马逊的 Dynamo(Key-value 模式下的存储平台）等。虽然大数据的存储和处理已有了成熟的解决方案，但是在系统软件中占较大比重的操作系统却没有太大变化，一些重要的技术问题还没有解决，如操作系统对新兴计算资源的直接抽象的调度、分布式文件系统下的统一数据视图、全数据中心范围内能耗管理、大数据下的安全性等，相关技术还不成熟，需要进一步研发。

五、大数据与 Web（互联网）

大多数研究大数据的商业公司都有明确的商业目的，即更好地支撑 Web 服务，如谷歌搜索引擎服务、Facebook、SNS 网站、新浪微博网站等。在大数据驱动下，

Web 服务可以提供更加流畅的网页交互体验，更加快速的社会资讯获取，让日常工作和生活更加便捷，实现更加深入的人、机、物融合。

回顾一下 Web 的发展，Web1.0 时代的 Web 内容主要由网站服务商提供，Web2.0 时代，用户大量参与 Web 内容的生产，如博客和微博内容。Web3.0 时代，人、机、物将共同参与，使 Web 形成对真实世界的全面映射。

第二节 大数据的发展趋势

一、大数据对社会发展的影响

大数据用户行为反映其真实需求。一切行为皆有前兆，未来的不确定性，是人类产生恐惧的根源之一，而大数据技术在一定程度上可以帮助我们预测未来。简单来说，大数据可以帮助人们从各种各样的数据中快速获取有价值的信息能力。大数据时代，软件价值体现在其带来的数据规模、流量与活性；公司价值在于其拥有数据的规模、活性以及收集、运用数据的能力，这些决定了公司的核心竞争力。从国家层面看，国家数据主权体现在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。

（一）泛互联网化

泛互联网化是收集用户数据的唯一一种低成本方式，能够为人们提供数据规模和数据活性。泛互联网化带来软件使用的三个变化：跨平台化、碎片化和门户化。

1. 跨平台化

应用软件深度整合网络浏览器功能，移动终端（手机、平板电脑）拥有相同的体验和协同的功能。

2. 门户化

用户无须启用其他软件即可完成绝大多数的工作并满足沟通需求，对于个性化的用户需求，可以直接调用第三方应用或者插件来满足需求。

3. 碎片化

把原来大型臃肿的软件拆分成多个独立的功能组件，用户可以按需下载使用。

这三个特征的核心意义分别在于收集用户行为数据，提高客户黏性及降低软件总体拥有成本，改变商业模式。

（二）行业垂直整合，越靠近消费端的企业越有发言权

开源软件加速了基础软件的同质化趋势，而软、硬件一体化的趋势，进一步弱化了产业链上游的发言权。大数据产业结构发展趋势有两个维度：第一维度是大数据产业链，围绕数据的采集、整理、分析和反馈；第二维度是垂直的行业，如媒体、零售、金融服务、医疗和电信等行业。

从这两个维度来看，大数据有三类商业模式：（1）大数据价值链环节，专注于价值链的高附加值环节。（2）垂直产业的大数据整合，利用大数据提高垂直产业效率。（3）大数据使用者，提供大数据基础设置、技术和工具。

（三）数据成为资产

未来企业的竞争将是拥有数据规模和活性的竞争，将是对数据解释和运用的竞争。大数据成为企业和社会关注的重要战略资源，并已成为大家争相抢夺的新焦点。因而，企业必须要提前制定大数据营销战略计划，抢占市场先机。围绕数据，可以演绎出六种新的商业模式：租售数据模式、租售信息模式、数据媒体模式、数据使用模式、数据空间运营模式和大数据技术提供商。

1. 租售数据模式

简单来说，租售数据模式就是租 / 售广泛收集的、精心过滤的、时效性强的数据。

2. 租售信息模式

一般聚焦于某个行业，广泛收集相关数据、深度提取整合信息，以庞大的数据中心加上专用传播渠道，为某行业服务的模式。此处，信息指的是经过加工处理，承载一定行业特征的数据集合。

3. 数据媒体模式

数据媒体模式将产业定位于媒体上，利用数据挖掘技术帮助客户开拓精准营销，企业收入来自客户增值部分的分成，这就是数据媒体模式。这类企业成长非常快，一般擅长数据挖掘分析技术帮助一些数据大户如银行、运营商等开展新的业务。

4. 数据使用模式

数据使用模式将产业定位在某一具体行业通过大量数据支持对数据进行挖掘分析后预测相关主体的行为，以开展业务，这就是数据使用模式。数据使用模式依据大数据技术开展高收益、低风险的业务，为企业创造新的盈利模式。未来将会有更多的数据使能型的业务模式出现，它们将具备创新业务的特质。这一模式的关键成功因素是维护数据的真实性和完整性，并适时进行风险分析。数据越完善，风险越低，越有利于保证企业的高收益。

（四）数据管理成为核心竞争力

数据管理成为核心竞争力，直接影响财务表现。当“数据资产是企业核心资产”的概念深入人心之后，企业对于数据管理便有了更清晰的认识，将数据管理作为企业核心竞争力，持续发展，战略性规划与运用数据资产，成为企业数据管理的核心。数据资产管理效率与主营业务收入增长率、销售收入增长率呈显著正相关；此外，对于具有互联网思维的企业而言，数据资产竞争力所占比重为 36.8%，数据资产的管理效果将直接影响企业的财务表现。

（五）数据泄露事件频发

未来几年数据泄露事件的增长率也许会达到 100%，除非数据在其源头就能够得到安全保障。可以说，在未来，每个财富 500 强企业都会面临数据攻击，无论他们是否已经做好安全防范工作。现在所有企业，无论规模大小，都需要重新审视安全的定义。在财富 500 强企业中，超过 50% 将会设置首席信息安全官这一职位。企业需要采取新的措施来确保自身以及客户的数据安全，所有数据在创建之初便需要获得安全保障，而并非在数据保存的最后一个环节，仅仅加强后者的安全措施已被证明于事无补。

（六）数据生态系统复合化程度加强

大数据的世界不只是一个单一的、巨大的计算机网络，而是一个由大量活动构件与多元参与者元素所构成的生态系统，是由终端设备提供商、基础设施提供商、网络服务提供商、网络接入服务提供商、数据服务使能者、数据服务提供商、触点服务、数据服务零售商等众多参与者共同构建的生态系统。而今，数据生态系统的基本雏形已然形成，接下来将进行系统内部角色的细分（市场的细分）、系统机制的调整（商业模式的创新）、系统结构的调整（竞争环境的调整）等，从而使数据生态系统复合化程度逐渐加深。

二、云平台数据更加完善

企业越来越希望能将自己的各类应用程序及基础设施转移到云平台上。就像其他信息技术系统那样，大数据的分析工具和数据库也将走向云计算。

云计算能为大数据带来哪些变化呢？

（1）云计算为大数据提供了可以弹性扩展、相对便宜的存储空间和计算资源，使中小企业也可以像亚马逊一样通过云计算来完成大数据分析。自 2013 年开始，大数据技术已开始和云计算技术紧密结合，预计未来两者关系将更为密切。除此之

外，物联网、移动互联网等也将一齐助力大数据革命，让大数据营销发挥出更大的影响力。

从历史上，传统的 IDC 就是这种模式，互联网巨头都在提供此类服务。但近期网盘势头强劲，从大数据角度来看，各家纷纷嗅到大数据商机，开始抢占个人、企业的数据资源。海外的 Dropbox，国内微盘都是此类公司的代表。这类公司的想象空间在于可以成长为数据聚合平台，盈利模式将趋于多元化。

从数据量上来看，非结构化数据是结构化数据的 5 倍以上，任何一个种类的非结构化数据处理，都可以重现现有结构化数据的辉煌。语音数据处理领域、视频数据处理领域、语义识别领域、图像数据处理领域都可能出现大型的、高速成长的公司。

(2) 云计算的 IT 资源庞大、分布较为广泛，是异构系统较多的企业及时准确处理数据的有力方式，甚至是唯一的方式。

当然，大数据要走向云计算，还有赖于数据通信带宽的提高和云资源池的建设，以确保原始数据能迁移到云环境以及资源可以随需求弹性扩展。

第三节 大数据的研究现状与未来展望

一、国内外大数据研究发展现状

(一) 大数据已深耕于经济领域且创造了巨大的经济价值

在此，我们以美国为例并谈其大数据产业创造的巨大的价值，具体表现在以下几个方面。

1. 大数据在欧洲公共管理部门得到深入应用

大数据为 OECD 组织中的欧洲国家公共管理部门创造了 1500 ~ 3000 亿欧元或更高的潜在经济价值，这些经济价值主要通过政府公共管理机构开支的减少、转移支付的下降及税收的增加来实现。

2. 大数据提高了美国的征税效率

由于迅速发现异常的能力日益增加，政府税务部门可以缩小“税收缺口”，即纳税人应付税款与其自愿缴税额之间的差额，并且防止出现违规情况。大多数税务机构实行“自愿缴税与追讨欠税并举”的模式。在这种模式下，它们接受纳税人的纳税申报单并办理退税，并对一部分纳税申报单进行抽查，以找出有意或无意欠税的情况。

3. 大数据使美国医疗服务质量得到提高

对于医疗服务的提供方和支付方来说，在减少医疗成本的同时不断提高医疗质量和效率仍然是一个难以实现的目标，而这也是改善民生的重大机遇。2010年，全美医疗支出占国内生产总值的17.9%，比2000年增长13.8%。而且，某些慢性疾病（如糖尿病）的患病率正在增加，将消耗更多的医疗资源。

对这些疾病和其他相关健康服务的管理将深刻地影响国家的福祉，大数据可以在这方面发挥作用。为在广大人群中取得最有效的医疗效果，更多地使用电子健康记录（电子健康档案），并与新的分析工具相结合，将有助于信息的挖掘和整合。研究人员可以利用信息判断发展趋势，并依据真实的医疗服务质量开展医疗评估。

4. 大数据使美国的交通更加便利

通过完善信息和自动驾驶功能，大数据有可能在许多方面彻底改变交通的面貌。开车的人越多，交通堵塞就越容易发生，其后果是浪费能源，造成全球气候变暖，耗费时间和金钱。手持设备、车辆和道路上的分布式传感器则可以提供实时交通信息。这些信息，再加上更好的自动驾驶功能，可以使驾驶更安全，减少交通堵塞。智能汽车日益互联的新型交通生态系统有可能彻底改变道路使用方式。

5. 大数据使美国的教育质量得到提升

大数据可以对美国教育及其在全球经济中的竞争力产生深远影响。例如，通过深入地跟踪和分析学生的在线学习活动，研究人员能够确定学生的学习方式和提高学习效率的方法。这种分析可以针对成千上万的学生进行，而不是孤立地进行小型研究。课程的教学方法，无论是在线的，还是传统的，都可以根据大规模分析所收集到的信息进行改进。

（二）欧美等发达地区大数据国家战略

1. 美国已经布局大数据产业

美国政府将大数据视为强化美国竞争力的关键因素之一，把大数据研究和生产计划提高到国家战略层面。2012年3月，美国政府宣布投资2亿美元启动“大数据研究和发展计划”，这是继1993年美国宣布“信息高速公路”计划后的又一次重大科技发展部署。美国政府认为大数据是“未来的新石油与矿产”，将“大数据研究”上升到国家战略层面，对未来的科技与经济发展必将带来深远影响。

以美国科学与技术政策办公室（OSTP）为首，国土安全部、美国国家科学基金会、国防部、美国国家安全局、能源部等已经开始与民间企业或大学一起进行多项大数据相关的研究开发。奥巴马指出，通过提高从大型复杂的数字数据集中提取知识和

观点的能力，承诺帮助加快在科学与工程中的步伐，改变教学研究、加强国家安全。

据悉，美国国防部已经在积极部署大数据行动，利用海量数据挖掘高价值情报，提高快速响应能力，实现决策自动化。而美国中央情报局通过运用大数据技术，将分析搜集数据的时间由 163 天缩减到 27 分钟。

2012 年 5 月，美国数字政府战略发布，提出要通过协调化的方式，所有部门共同提高收集、储存、保留、管理、分析和共享海量数据所需核心技术的先进性，并形成合力；扩大大数据技术开发和应用所需人才的供给。以信息和客户为中心，改变联邦政府工作方式，为美国民众提供更优质的公共服务。

2018 年，纽约大学计算机系教授塞恩告诉记者，根据前期计划，美国希望利用大数据技术在科研教学、环境保护、工程技术、国土安全、生物医药等多个领域取得突破。其中具体的研发计划涉及美国国家科学基金会、国家卫生研究院、国防部、能源部、国防部高级研究局、地质勘探局等 6 个联邦部门和机构。

2. 欧盟及日本等国相继出台大数据战略

继美国率先开启大数据国家战略先河之后，欧盟、日本及韩国等国家也相继出台了相应的战略举措。数据规模及运用数据的能力将成为影响综合国力的重要因素，对数据的占有和控制也将成为国家间争夺的焦点。

法国政府为促进大数据领域的发展，以培养新兴企业、软件制造商、工程师、信息系统设计师等为目标，开展了一系列的投资计划。法国政府在其发布的《数字化路线图》中表示，大力支持“大数据”在内的战略性高新技术，法国软件编辑联盟也号召政府部门和私人企业共同合作，投入 3 亿欧元资金用于推动大数据领域的发展。

法国生产振兴部部长 Arnaud Montebourg、数字经济部副部长 Fleur Pellerin 和投资委员 Louis Gallois 在第二届巴黎大数据大会结束后的第二天共同宣布了将投入 1150 万欧元用于支持 7 个未来投资项目。这足以证明法国政府对于大数据领域发展的重视。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”。众所周知，法国在数学和统计学领域具有独一无二的优势。

为了提高信息通信领域的国际竞争力，培育新产业，同时应用信息通信技术应对抗灾救灾和核电站事故等社会性问题，日本总务省于 2012 年 7 月新发布“活跃 ICT 日本”新综合战略，针对大数据推广的现状、发展动向、面临的问题等进行探讨，以期对解决社会问题作出贡献。

2013 年 6 月，安倍内阁正式公布了新 IT 战略——“创建最尖端 IT 国家宣言”。

“宣言”全面阐述了2013～2020年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”。

（三）我国大数据国家战略

伴随着大数据时代的来临，世界各国对数据的重视达到了前所未有的高度。套上大数据的光环后，那些原本存放在服务器上平淡无奇的旧数据一夜之间身价倍增。按照世界经济论坛报告的看法，“大数据为新财富，价值堪比石油”。正如大数据之父维克托所预测，“虽然数据还没有被列入企业的资产负债表，但这只是一个时间问题”。

现在许多国家将大数据视为国家战略，并且在实施上，也已经进入到企业战略层面。

争夺新一轮技术革命制高点的战役已经打响，中国政府在美国提出《大数据研究和发展计划》的2012年也批复了“十二五国家政务信息化建设工程规划”，专门有人口、法人、空间、宏观经济和文化等五大资源库的五大建设工程。我国的开放、共享和智能的大数据的时代已经来临。

2012年8月，国务院印发了《关于促进信息消费扩大内需的若干意见》，推动商业企业加快信息基础设施演进升级，增强信息产品供给能力，形成行业联盟，制定行业标准，构建大数据产业链，促进创新链与产业链有效嫁接。

同时，构建大数据研究平台，整合创新资源，实施“专项计划”，突破关键技术。大力推进国家发改委和中科院基础研究大数据服务平台应用示范项目，广东率先启动大数据战略，推动政府转型，上海也启动大数据研发三年行动计划。

当前，在政府部门数据对外开放，由企业系统分析大数据进行投资经营方面，上海无疑是先行一步。2014年5月15日，上海市推动各级政府部门将数据对外开放，并鼓励社会对其进行加工和运用。

根据上海市经信委印发的《2014年度上海市政府数据资源向社会开放工作计划》，目前已确定190项数据内容作为2014年重点开放领域，涵盖28个市级部门，涉及公共安全、公共服务、交通服务、教育科技、产业发展、金融服务、能源环境、健康卫生、文化娱乐等11个领域。

其中市场监管类数据和交通数据资源的开放成为重点，随着这些与市民息息相关的信息查询的开放，企业运用大数据在上海“掘金”的时代来临，企业投资和上海民生相关的产业，如交通运输、餐饮等，可以不再“盲人摸象”。

在立足国家战略和产业政策推动大数据收集和分析技术快速发展的同时，我们

也逐步认识到避免数据垄断和保护数据安全的重要性，开展了相关法律法规的探讨和研究。

2017年12月，习近平总书记在中共中央政治局第二次集体学习时强调，要推动大数据技术产业创新发展。我国网络购物、移动支付、共享经济等数字经济新业态新模式蓬勃发展，走在了世界前列。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。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。要运用大数据促进保障和改善民生，坚持以人民为中心的发展思想，推进“互联网+教育”“互联网+医疗”“互联网+文化”等，让百姓少跑腿、数据多跑路，不断提升公共服务均等化、普惠化、便捷化水平。

毫无疑问，中国的底子太薄了。但是，大数据是信息化时代的“石油”。开发大数据资源的能力将影响未来国家的核心竞争力。我国不能幻想走别人修好的道路，更不能靠等，只能依靠自身加速前行，提升将数据转化为信息和知识的速度和技术，而这种转化速度和技术体现了大数据技术能力的高低。

二、大数据的未来展望

（一）促进我国科研事业进一步发展

应用大数据技术，对于我国社会的稳定发展以及社会经济水平的提升有着非常重要的意义。此外，在我国现代化发展过程中，部分科学研究事业需要对大量的研究数据进行有效分析，而传统的数据分析手段难以获得良好的分析效果。在科学研究事业的发展过程中，通过应用大数据技术，能够对复杂的数据信息进行有效地采集和分析处理，从而在海量复杂数据中获得自身所需数据。较之于传统科研工作的人工数据处理模式，应用大数据技术能够促进数据采集的合理性与准确性得到进一步提升，且数据分析结果的准确性和可靠性也进一步提高，从而能获得自身所需的研究成果，对于我国科研事业的进一步发展也有着积极的意义。可见，大数据技术对于我国科研工作中项目数据的获取能够提供良好的信息保障，对于我国科研事业的进一步发展有着积极的意义。

（二）完善社会经济市场

近年来，我国大数据技术得到了迅速发展，在社会经济市场的各个领域中也得到了普及，并且成为当下各个行业生产经营发展的重要趋势。此外，大数据技术的应用还催生了云计算技术、物联网技术以及网络通讯技术等多项技术，对于我国社会经济