



导论：大数据是国家治理能力建设的重要技术

人类发展历程表明，每一次新技术革命的深层渗透，必将引起社会结构和社会形态的变革。作为一种积极的治理资源和全新的治理技术，大数据正在悄无声息地重塑着我国的国家治理生态，提升着我国的政府治理能力，并在深层次上变革着我国的社会治理模式。但是，大数据及其技术本身也并非完美无缺。“大数据意识的缺乏、信息安全、数字鸿沟、信息孤岛、权力寻租等难题是大数据时代中国国家治理能力建设面临的基本挑战。”¹中共十八届三中全会审议并通过的《中共中央关于全面深化改革的若干重大问题决定》明确提出：“全面深化改革的总目标是完善和发展中国特色社会主义制度，推进国家治理体系和治理能力现代化。”²近年来，随着全球网络信息技术的飞速发展，人类所拥有的各类信息和数据存量也实现了爆炸式增长，各类信息源和数据广泛应用推动着人类社会开始迈入大数据时代，同时也给我国国家治理能力现代化法治化建设带来了全新的机遇和挑战。

一、研究背景：大数据技术及其产业发展的时代契机

互联网的发展是大数据发展的最大驱动力。这是因为，互联网的普及使得人们的网络行为增加并呈现出多元化发展趋势，进而导致通过互联网产生的数据发展更加迅速也更具代表性。互联网世界中的各类信息媒介，比如商品信息、社交媒体图片、文本信息、视频数据以及人与人的交互信息、位置信息、消费

¹ 唐皇凤、陶建武：《大数据时代的中国国家治理能力建设》，《探索与争鸣》2014年第10期。

² 《中共中央关于全面深化改革若干重大问题的决定》，《光明日报》2013年11月16日。

信息、交易信息等等，都已经成为大数据的最重要也是增长最快的来源。

全球最具权威的信息技术研究与咨询公司高德纳（Gartner）和国际数据公司（IDC）在 2013 年底就已经对未来的数字行业进行了准确的预测，内容涵盖信息技术行业、智能机器应用、物联网、云服务等领域。在信息技术方面，国际数据公司认为，2014 年信息技术行业总支出将增长 5.1%至 2.14 万亿美元，个人电脑营收将下降 6%。在金砖四国中，在中国经济发展的带动下，其信息技术消费较去年同期上涨 13%，按美元计算的话，尽管中国的市场份额只有美国的三分之一，但中国的信息消费增长与美国持平。对此高德纳认为，未来的信息技术产业范畴更广，或将成为个人和企业生态系统创新的助推器，包括 3D 打印、智能机器在内的数字产业革命已经开启。在物联网发展方面，国际数据公司预计，到 2020 年，物联网将产生 300 亿个自动连接的终端和 8.9 万亿美元营收。高德纳则指出，到 2015 年，分享消费者数据的智能手机应用数量将增长两倍，意味着想要获取消费者个人数据的营销人员或所有者的数目在增加。在大数据方面，国际数据公司还预测，花费在大数据技术和服务上的支出在 2014 年将超过 140 亿美元，同比上涨 30%。这部分的竞争将集中在制定“数据优化的云平台”上，也就是能够充分利用大量数据和实时数据流的云平台。而在云市场规模方面，国际数据公司预测，包括云服务和推动这些服务的云技术在内的云市场规模，到 2014 年将上涨 25%，突破 1000 亿美元。伴随着云服务公司争相实现全球扩张，数据中心数量将大幅增长，形式也将各有不同。最终，一场酣战将来自于能够创造基于云的应用与解决方案的大型硬件厂商，并最终推动这个市场的发展。³

随着近几年移动互联网、物联网、云计算等的快速发展，大数据爆炸式地增长。大数据的研究和应用越来越被人们所需求，它的应用时刻影响着人们的生活，随之催生了大数据产业这一新兴产业。正如阿里巴巴前 CEO 马云在卸任时提到：“大家还没有搞清楚 PC 时代的时候，移动互联网来了，还没搞清楚移动

³ 《新媒体时代大数据发展迎契机》，中国行业研究网，2014 年 1 月 6 日，链接地址：<http://www.chinairn.com/news/20140106/112650110.html>，最后访问日期：2016 年 8 月 28 日。



互联网的时候,大数据时代又来了。”这在某种程度上也说明了,大数据、大数据技术及其产业发展的迅猛之势,是超出了所有人的想象。同时也在某种意义上表明,大数据正在并将成为全球经济社会未来发展的一个重要的产业类型。

2016年8月25日,由工业和信息化部、中国通信企业协会指导,中国通信企业协会通信网络运营专业委员会、北京邮电大学及中国管理科员研究院学术委员会、中国大数据技术与应用联盟共同主办的“2016中国大数据技术与应用研讨会暨联盟成立大会”在京召开。与会专家认为,随着互联网技术的飞速发展,其背后海量的数据越来越受重视,并逐渐被各行各业有效利用。例如交通、教育、医疗、电商、定位、搜索等领域,都已取得了许多成功案例,甚至大数据正在慢慢改变人们的生活习惯,我国也逐步迈向了大数据时代。⁴在会上,工业和信息化部通信发展司副司长陈家春表示:“我国数据总量正在以年均50%的速度增长,预计到2020年将占全球21%,中国正在成为真正的数据资源的大国,这为大数据产业发展提供了坚实的基础。”“据统计,2015年我国云计算带动了上下游产业规模超过了3500亿元,这些都为大数据发展提供了有力的基础支撑,同时也吸引了社会资本的关注,一种良性的大数据发展环境正在形成。”

中国本身就是全球最大的数据产地,也是最大的数据应用市场,继“人口大国”、“制造大国”之后,中国还将成为全球领先的“数据大国”。在我国重大的发展战略中,无论是“一带一路”还是“中国制造2025”,无论是“互联网+”还是“大众创业、万众创新”,都已经明显地涵盖了“数据社会化”的特征和实践。目前,我国大数据产业的发展环境逐步成熟,网络基础设施支撑能力大幅提升,大型数据中心向绿色化集约化发展,云计算产业也日渐成熟。为此,重视大数据技术,大力发展大数据产业,成为当前我国产业发展的重点方向之一。

二、研究意义:大数据是国家治理能力建设的重要推动力

“实施国家大数据战略,推进数据资源开放共享”,党的十八届五中全会的

⁴ 李景:《我国正成为数据资源大国 年均增速50%但需更开放》,中国经济网,2016年8月25日,链接地址:http://bgimg.ce.cn/cysc/tech/gd2012/201608/25/t20160825_15254068.shtml,最后访问日期:2016年10月13日。



要求，吹响了我国由数据大国向数据强国迈进的号角。这不仅对于推进“十三五”经济社会发展具有深远意义，在变革国家治理理念，革新治理手段，及优化治理过程等方面有着广泛的应用前景，对提升政府治理能力、推进简政放权也产生积极影响。为此，充分释放大数据红利对中国国家治理能力的现代化具有至关重要的意义。

首先，利用大数据能提升国家的智慧决策水平。2014年10月28日《中共中央关于全面推进依法治国若干重大问题的决定》要求“实行领导干部重大决策终身责任追究制度”，党的十八大报告指出：“坚持科学决策、民主决策、依法决策，健全决策机制和程序，发挥思想库作用，建立健全决策问责和纠错制度。”国家治理体系和治理能力的现代化法治化首先体现在人的能力现代化，人的能力来源于实践智慧，智慧的来源应该是基于大数据分析，以数据说话、以数据决策。也就是说，提高政府及其官员的决策水平，必须依靠长时期积累收集的海量数据进行科学、精准的分析而得出基本规律与核心数据，依靠数据说话、依靠数据预测、依靠数据决策、依靠数据创新。唯有如此，才能提高决策的科学性、前瞻性和准确性，满足社会和公众需求多样化，达到治理精细化，把大数据技术与思维运用到公共管理与决策中，助推治理体系和治理能力现代化。

其次，利用大数据能促进国家的智慧政府建设。“推进国家治理体系和治理能力现代化”是一项具有系统性特征的社会主义现代化建设工程，除了突出法治化的要求之外，还需要采取综合治理的思路，调动一切有利于实现国家治理体系和治理能力现代化的生产要素和管理要素，全面、有效地推进社会主义现代化国家和法治中国的建设。“大数据”就是其中的一项重要的乃至核心的生产要素和管理要素。作为治理主体之一的政府官员必须具备大数据思维、互联网思维、计算思维。习近平总书记强调，“我们必须增强忧患意识，紧紧抓住和用好新一轮科技革命和产业变革的机遇，不能等待、不能观望、不能懈怠。”抓住新一轮科技革命的机遇就必须抓住大数据。以大数据为视角，探讨大数据在政府变革、经济治理、公共服务、城市管理、社会治理等国家治理各方面发挥的



作用,展望大数据与国家治理能力现代化法治化的未来,对推进国家治理体系和治理能力现代化法治化,共同打造智慧城市、智慧政府,从而积极迈向智慧国家建设都有极大地借鉴意义。

再次,利用大数据能提升国家的智慧服务能力。“公共服务型”政府是我国市场化改革、政府职能转变的必然选择。“治理理论”的主张,诸如治理主体多元化、治理对象参与性、治理手段多样化等,与大数据及其治理不谋而合。首先,大数据能助推国家公共服务向精细化个性化转变。姜奇平指出:大数据逐渐成为信息社会经济、社会和文化各个方面活动的原型结构,其最大的改变就是把人类“实现目的”变成专业化的工作。⁵大数据时代,个体需求是社会生产国家服务所关注的核心,而让数据挖掘更加深入和精细化,实现对民众需求的精确把握,从而合理配置公共服务资源,才能引导政府提供更加个性化和人性的公共服务。其次,大数据能助推国家公共服务从封闭低效型向协同高效节约型转变。从世界趋势看,最大限度开放数据是必然发展趋势,其带来的积极效应将推动政府信息资源的进一步公开,而不断成熟的数据分析技术能使公开的政务信息得以有效利用并发挥更大价值。政府在人口、教育、治安、就业、交通、社保、卫生、工商、税收、民政等方面拥有天然的数据优势,通过这些数据的深度挖掘,政府管理者可准确预测到民众和企业的服务需求,进而提供以需求为导向的公共服务,提高工作效率,提高服务质量和公众满意度,降低公共服务成本,提升公共服务能力。

⁵ 姜奇平:《大数据与信息社会的意义结构》,《互联网周刊》2012年第12期。

第一章 走进大数据时代

2012 年 1 月在瑞士举行的达沃斯世界经济论坛上，一份名为《大数据，大影响》（Big Data Impact）的报告引起热议。这份报告认为，大数据就像货币和黄金一样，是一种新型的经济资产。多家投资机构甚至据此判断，“大数据”将会成为一条全新投资主线。

第一节 大数据的潜力

2014 年 2 月，科学记者史蒂夫·洛尔（Steve Lohr）在《纽约时报》撰文写道：科学、体育、广告、公共健康等各个不同领域，都越来越趋向基于数据的发现和决策。

“这是一场革命，我们现在做的只是冰山一角，但是由于庞大的数据新来源而带来的量化方法，将横扫学界、商界和政界，所有领域都将被触及。”哈佛大学定量社会研究所主任加里·金（Gary King）在接受《纽约时报》记者史蒂夫采访时说。⁶麻省理工大学管理学院的经济学家埃里克·布吕诺尔夫松（Erik Brynjolfsson）接受采访时说，要充分理解大数据的潜在影响，必须通过显微镜来观察。四个世纪前发明的显微镜使人们前所未有地在细胞层面观察和测量事

⁶ 郭晓科：《大数据》，清华大学出版社，2013 年版，第 1 页。



物。这是度量方面革命性的举措。数据测量就是现代人的显微镜。例如，谷歌（Google）的搜索、脸谱（Facebook）的帖子和推特（Twitter）的信息，使得细节化、即时化的测量行为和情绪成为了可能。布吕诺尔夫松说，在商业和经济等其他领域，决策将越来越依赖于数据和分析，而非经验和直觉。

有很多事件可以说明以数据为先的思考方式所带来的好处。最著名的还是2003年迈克尔·刘易斯（Michael Lewis）所写的著作《点球成金》（Money ball），讲述了低预算的奥克兰运动家队如何通过整理后的数据和晦涩的棒球统计资料来发现被低估的运动员。繁重的数据分析不仅在棒球领域成为一项标准，而且也被应用于其他运动，如在去年由英国足球运动员布拉德·皮特主演的电影版《点球成金》之前，就已经使用了数据分析。沃尔玛、科尔之类的零售商也通过分析销售情况、定价，以及经济、人口统计、天气等数据，决定不同店铺的产品类别及降价促销的时机。

全球国际快递（UPS）等运输公司通过采集卡车运输次数和路线来调整运输的路线。Match.com 等在线相亲服务网站，经常查看在线用户的个人情况、反应情况和沟通情况，以提高男女约会安排的匹配率。而美国的警察部门，以纽约为首，使用了计算机化的人像绘图，并通过分析逮捕记录、发工资日、体育赛事、降雨和节日等变量来试图预测可能发生犯罪的“热点区域”，从而可以提前在这些地区部署警力。吕诺尔夫松和其他两名同事2011年发表的研究表明，由数据引导的管理方式正在美国企业界蔓延，而且开始出现回报。他们研究了179家大型企业，发现那些正在使用“基于数据的决策方式”的企业所获得的利润比其他企业高出5至6个百分点。

大数据的预测能力正在被发掘，并在公共健康、经济发展和预测等领域显示出了成功的希望。研究人员发现，就在某一地区医院的急诊室里流感病人增加的约两周前，谷歌搜索里对“流感症状”和“流感治疗”的强暴请求、高峰（急诊室的报告通常比实际接诊情况晚两周左右）。据国际数据资讯公司（Global Pulse）估测，数据数量一直在增加，每年增长50%。这个速度不仅是指数据流的增长，而且还包括全新的数据种类的增多。如今全球有数不清的数据感应器，

应用于工业设备、汽车、电子量表、集装箱等。它们可以测量并传递地点、移动、振动、温度、湿度，甚至空气中的化学变化。将这些传递沟通的感应器与计算机智能连接起来，你就能看到物联网和工业互联网的崛起。

如果说数据是新型经济资产，那么赋予数据生产力的则是互联网。

2015年5月，全球知名咨询公司麦肯锡（McKinsey & Company）的研究部门麦肯锡全球研究所（McKinsey Global Institute）在对13个国家（法国、美国、英国、德国、日本、意大利、加圣大、俄罗斯、瑞典、韩国、中国、巴西、印度）调研后发布了题为《互联网的价值：网络对经济增长、就业及繁荣的影响》（Internet matters.: the net's sweeping impact on growth, jobs, and prosperity）的调查报告。报告从7个方面论述了互联网对经济增长的贡献（Matthieu et al., 2015）：

（1）互联网体量巨大，发展迅猛。全球网民数量已达30亿，并仍在以每年2000万的速度增长。互联网经济平均已占13个国家GDP的10.4%，超过农业和能源，成为经济的巨大推动力。

（2）各国互联网经济发展水平差异大，在瑞典和英国这样的发达国家，互联网经济占GDP的比重是15%，而在13个国家中有9个，这一数字仍在6%以下。全球互联网经济的发展潜力不容小觑。

（3）互联网经济对GDP贡献巨大。报告中称，从2000年到2014年的15年间，数据覆盖的13个国家互联网经济占GDP增长的11%，并且影响在持续扩大；在互联网产业发达的国家（瑞典、德国、英国、法国、美国、韩国、加拿大、意大利、日本），15年间互联网对GDP增长的贡献率达14%，15年间的最后5年（2010~2014年），这一数字达25%；即使在互联网产业发展中国家（中国、印度、巴西、俄罗斯），互联网经济对GDP增长的贡献也达到了5%。

（4）互联网产业的发达程度与提升生活质量关系紧密。相关研究显示，过去的15年中，互联网发展给被调查的13个国家所带来的人均GDP增长约为500美金，相当于19世纪工业革命历经50年取得的成绩。

（5）相较于传统行业，互联网更能催生就业机会。法国的一项研究显示，



过去的 15 年,互联网每摧毁 1 个就业岗位,便新创造 2.4 个就业岗位。麦肯锡全球调查数据支持了这一论断,并把这一数字更新为 2.6 个。

(6) 互联网促使经济发展现代化。互联网的使用为中小企业提升了 10% 的生产力。倚力于互联网技术的中小企业,增长速度和产品出口量都是同类企业的 2 倍。

(7) 互联网创造令人惊异的消费者剩余。月均每用户的消费者剩余从德国的 15 欧元到英国的 25 欧元。仅 2014 年一年,法国创造的消费者剩余为 70 亿欧元,美国为 520 亿欧元。

在同一份报告中,麦肯锡全球研究所指出使用互联网或物联网收集市场和消费者信息的公司,很快会发现它们的数据库已被浩瀚的信息充塞。传统的数据处理工具已无法实现数据收集、储存、检索、共享、分析和视觉化的功能。倘若不加整理,互联网终有一天会成为克利夫,斯托尔(Cliff Stoll)曾担心的“塞满垃圾信息的荒野”——“大数据”的意义正在于寻找管理海量数据的科学路径。报告指出长于“大数据时代”的人士将迎来各种各样的机遇。《纽约时报》的分析文章认为美国至少还需要 14 万至 19 万具备“深厚分析”专业技能的人才,以及 150 万熟悉数据的经理级人才(Steve, 2015)。麦肯锡预计下一个 10 年美国健康医疗业的年均增长约 1%,亦即超过 3000 亿美元的潜在商业价值;欧洲发达国家的公共事业年均增长约为 0.5%,也就是 2.55 千亿欧元的市场。在这份预测大数据带来大市场的报告发布一个月之后,2015 年 6 月,麦肯锡全球研究所又密集地发布了另一份报告:《大数据:创新、竞争和生产力的下,一个前沿》(Big data: The next frontier for innovation, competition, and productivity),在这份报告中,“大数据”概念得到了清晰阐释。

第二节 什么是大数据？

“大数据”一词由英文“Big Data”翻译而来，过去常说的“信息爆炸”、“海量数据”等已经不足以描述这个新事物。麦肯锡全球研究所报告《大数据：创新、竞争和生产力的下一个前沿》对“大数据”定义如下（James，2011）：大数据是指大小超出了传统数据库软件工具的抓取、存储、管理和分析能力的数据库。这个定义有意地带有主观性，对于“究竟多大才算是大数据”，其标准是可以调整的，即：我们不以超过多少 TB（1000GB）为大数据的标准。我们假设随着时间的推移和技术的进步，大数据的“量”仍会增加。还应注意，该定义可以因部门的不同而有所差异，这取决于什么类型的软件工具是通用的，以及某个特定行业的数据集通常的大小。因此，今天众多行业的大数据范围可以从几十 TB 到数千 TB。

作为特指的大数据，按 EMC 的界定，其中的“大”是指大型数据集，一般在 10TB 规模左右；多用户把多个数据集放在一起，形成 PB 级的数据量；同时这些数据来自多种数据源，以实时、迭代的方式来实现。大数据通常与 Hadoop、NoSQL、数据分析与挖掘、数据仓库、商业智能以及开源云计算架构等诸多热点话题联系在一起。

IBM 负责软件和硬件两大集团的高级副总裁 Steve Mills 在 IBM 201110D 大会上说：“分析不再是一个工具，而是一项必要的能力，能让业务流程智慧运转的能力。企业必须将对信息的洞察力转化为行动，不是为了获得竞争优势，而是因为它已经变成生存的根本。”IBM 公司把大数据概括成了三个 V，即大量化（Volume）、多样化（Variety）和快速化（Velocity），并向客户推出了“大数据解决方案”服务。IBM 公司所概括的这三个大数据的特点也反映了大数据所潜藏的价值（Value），或许可以认为，这四个 V 就是大数据的基本特征。

“大数据”的首要特征是数据量大。基于电脑的数据储存和运算是以字节



(byte) 为单位的, 1KB (Kilobyte) = 1024B, 又称千字节; 更高级的数量单位分别是 1MB (Megabyte, 兆字节)、1GB (Gigabyte, 吉字节)、1TB (Trillionbyte, 太字节)、1PB (Petabyte, 拍字节)、1EB (Exabyte, 艾字节)、1ZB (Zettabyte, 泽它字节) 和 1YB (Yottabyte, 尧它字节), 每个单位之间的运算关系是乘以 1024。截至 2015 年, 美国几乎所有部门中每一个雇员数量在 1000 人以上的企业所存储的数据平均值至少为 400TB, 是美国零售商沃尔玛 1999 年的数据仓库的四倍。很多经济部门中, 每个企业平均存储数据超过 2PB。欧洲的组织 2015 年存储容量总计接近 25EB, 大约为整个美国存储容量 (35EB 以上) 的 70%。全球企业 2015 年在硬盘上存储了超过 15EB 的新数据, 消费者在 PC 和笔记本电脑等设备上存储了超过 12EB 新数据, 而 1EB 数据就相当于美国国会图书馆中存储数据的 4000 多倍 (James, 2015)。数据容量增长的速度大大超过了硬件技术的发展速度, 以至于引发了数据存储和处理的危机。大量的数据会被处理掉, 比如医疗卫生提供商会处理掉他们所产生的 90% 的数据 (手术过程中产生的几乎所有实时视频图像)。

然而, 大数据不只是大。海量数据引发的危机并不单纯是数据量的爆炸性增长, 还牵涉到数据类型的改变, 即多样化 (Variety)。原来的数据都可以用二维表结构存储在数据库中, 如常用的 Excel 软件所处理的数据, 我们称之为结构化数据。但是现在, 更多互联网多媒体应用的出现, 使诸如图片、声音和视频等非结构化数据占到了很大比重。有统计显示, 全世界结构化数据增长率大概是 32%, 而非结构化数据增长率则是 63%, 预计至 2020 年, 非结构化数据占有比例将达到互联网整个数据量的 85% 以上。用于产生智慧的大数据, 往往是这些非结构化数据。⁷ Informatica 中国区首席产品顾问但彬认为: “大数据” 包含了 “海量数据” 的含义, 而且在内容上超越了海量数据, 简而言之, “大数据” 是海量数据 + 复杂类型的数据。但彬进一步指出: 大数据包括交易和交互数据集在内的所有数据集, 其规模或复杂程度超出了常用技术按照合理的成本和时限

⁷ 曹磊、陈薇娜等: 《大数据: 数字世界的智慧基因》, 文汇报, 2011 年 11 月 8 日, 第 12 版。

捕捉、管理及处理这些数据集的能力。

简单来说,大数据由三项主要技术趋势汇聚组成。一是海量交易数据:在从 ERP 应用程序到数据仓库应用程序的在线交易处理(OLTP)与分析系统中,传统的关系数据以及非结构化和半结构化信息仍在继续增长。随着企业将更多的数据和业务流程移向公共和私有云,这一局面变得更加复杂。二是海量交互数据:这一新生力量由源于 Facebook、Twitter、LinkedIn 及其他来源的社交媒体数据构成。它包括了呼叫详细记录(CDR)、设备和传感器信息、GPS 和地理定位映射数据、通过管理文件传输(Manage File Transfer)协议传送的海量图像文件、Web 文本和点击流数据、科学信息、电子邮件,等等。三是海量数据处理:大数据的涌现已经催生出了设计用于数据密集型处理的架构,例如具有开放源码、在商品硬件群中运行的 Apache Hadoop。Hadoop 是一种以可靠、高效、可伸缩的方式对大量数据进行分布式处理的软件框架。它的可靠性在于提前假设计算元素和存储会失败,因此它维护多个工作数据副本,确保能够针对失败的节点重新分布处理;高效性则表现在它以并行的方式工作,通过并行处理加快处理速度。Hadoop 还是可伸缩的,能够处理 PB 级数据。此外,由于 Hadoop 依赖于社区服务器,因此它的成本比较低,任何人都可以使用。对于企业来说,难题在于以具备成本效益的方式快速可靠地从 Hadoop 中存取数据。脸谱是 Hadoop 最知名的用户之一。通过 Hadoop,类似脸谱的社交网站和类似国内淘宝网的商业网站实现了“推荐你可能认识的人”、“可能想读的书”、“可能喜欢的商品”等服务。

第三节 大数据与云计算

我们再来了解一下与大数据密切相关的“云计算”概念。《互联网周刊》主编姜奇平认为,大数据并不像某些人说的,是云计算之“后”的又一浪,而就



是云计算本身，因为两者都是数据的大规模集聚与定制化分布的结合。

云计算的基本原理是，使计算分布在大量的分布式计算机上，而非本地计算机或远程服务器中。云计算意味着只需要一台笔记本或者一个手机等互联网终端，就可以通过网络服务来实现我们需要的一切，包括浏览文档、图片、视频，甚至运行超级计算这样的任务。云计算的应用包含这样的一种思想：把力量联合起来，给其中的每一个成员使用。⁸

终端用户通过网络浏览器或手机应用程序使用云计算，而这些商业软件和数据被存储于距离遥远的服务器上。支持者声称云计算使得企业的程序运行得更有效，更迅速。通过提高管理性能和缩减维护费用，并且对 IT 进行强化使其能更迅速地对外资源做出调整，以满足流动性和不可预见的商业需求。

云计算的概念可以回溯到 20 世纪 60 年代，约翰·麦卡锡（John McCarthy）曾预言“有朝一日，计算可能会变成公共设施”。几乎所有当今云计算的特征（弹性供给、作为公共设施提供、在线、无穷供给），道格拉斯·帕克希尔（Douglas Parkhill）在他 1996 年出版的著作《效用计算的挑战》（The Challenge of the Computer Utility）中都已涉及。另外，他在书中还提到了云计算，电力工业与公共使用，私人、政府与社区使用形式的对比等内容。其他学者表示云计算的起源可以提前到 20 世纪 50 年代。当时，计算机科学家赫伯·格劳希（Herb Grosch）假设整个世界将在由大约 15 个巨大的数据中心运作的终端上运作。

沃尔玛或者谷歌这样的大公司，很早就开始使用云服务，但是成本极其昂贵。今天的商业硬件、云建筑及开源软件正在将大数据带入普及民用的领域。即使是在车库中创业的公司也可以用较低的价格租用云服务了（Edd, 2015）。云计算应用的快速增长得益于高效网络的普遍适应性、低成本的电脑、存储设备以及硬件虚拟化的广泛应用等因素。

如果说“云计算”是大数据时代必备的技术平台的话，那么“众包”（crowd sourcing）业务就是大数据时代全新的生产组织模式。这是《连线》（Wired）杂

⁸ 万川梅：《云计算应用技术》，西南交通大学出版社，2013 年版，第 9 页

志 2006 年发明的专业术语,用来描述一种新的商业模式,即企业利用互联网来将工作分配出去、发现创意或解决技术问题。通过互联网控制,这些组织可以利用志愿员工大军的创意和能力——这些志愿员工具备完成任务的技能,愿意利用业余时间工作,满足于对其服务收取小额报酬,或者暂时并无报酬,仅仅满足于未来获得更多报酬的前景。尤其对于软件业和服务业,这提供了一种组织劳动力的全新方式。⁹

“众包”模式使得科学发现不再是专业学者埋头于实验室的苦差事,而是全球科学家、学生和感兴趣的民众都可以参与的大众活动。谷歌公司在这个领域做了许多工作,他们开发了 Google.org——这是一个利用谷歌在信息技术处理数据方面特长建立的全球公众都能够参与的科学研究平台。从 2008 年 11 月起,Google.org 启动了名为“流感趋势”的项目,使用一种复杂的算法,对关于流感的网络搜索进行跟踪,从而对流感在人群当中传播的方式作出估计。其后,Google.org 还组织了地球引擎项目,将大量的卫星图像和数据开放给公众,让每个人都可以对气候影响下的水源变化和沙漠化进行跟踪研究。这些项目都在寻求一种“长尾效应”,用来解决过去一直无法展开研究的科学难题(曹磊等,2014)。

第四节 大数据的价值

大数据的核心价值在于“预测”,不管是宏观经济领域,还是医学领域,甚至包括城市管理领域,都可以看到依据大数据进行“预测”所获得的明显成效。这是由于“发现”和“挖掘”了大数据的“潜在价值”,使得人们既能从“已知”发现问题、总结经验、归纳规律,又能从“未知”的多样可能性中发现问题、

⁹ 张利彬、钟复平、涂慧:《众包问题研究综述》,《科技进步与对策》2012 年第 6 期,第 154-160 页。



把握方向。¹⁰

麦肯锡全球研究所报告《大数据：创新、竞争和生产力的下一个前沿》(James et al., 2011)指出，对企业而言，大数据的价值体现在两个方面：分析使用和二次开发。大数据的分析使用可以揭示之前由于分析成本太高而忽略的信息，如消费者的同伴影响、购物交易信息分析、社交网络信息和地理数据等。过去的10年中，大数据分析方面已经积累了一些开发新产品和新数据的成功经验，比如数据推送改变了Facebook一类社交网站的商业化模式，其8亿用户潜在的个人信息和商业价值都能被转换成各类广告用途，与广告主对接，哪怕用户的一句“最近胃疼”的状态更新，也会吸引胃药广告和保险广告的“轮番轰炸”。

对个人而言，智能手机的普及为开发大数据应用搭载了最好的平台。苹果公司2011年秋天发布了iPhone手机上的问答软件Siri。这款软件源起于五角大楼的一项研究，之后却成了硅谷的一项创新产品。苹果公司2010年买入Siri技术，并不断加入更多数据。如今，使用苹果手机的人们提供了数以百万计的问题，Siri正在成长为一个越来越熟练称职的个人助理，提供事件提醒、天气预报、就餐建议等服务，回答的问题也愈加广泛。

在种类繁多的大数据中，有一类是探测人们所在位置时产生的定位信息。GPS技术的不断发展使得我们可以在几十米的距离内定位像手机那么小的装置，同时我们也看到了个人位置数据被用来创造新的商业和创新企业模式，这种模式几乎涉及全球每个人的生活。

个人定位数据涉及通信、零售、媒介等多个行业。这个领域蕴含着创造新价值的巨大潜力，麦肯锡全球研究所预测服务提供商将获得超过1000亿美元的收入，为消费者和终端用户创造的价值将达7000亿美元。早期个人定位数据的来源是个人信用卡和借记卡付款，消费者在销售点终端(POS)的固定地点刷卡，与个人身份识别数据相关联。随着手机用户增加，使用基站信号定位越来越普遍。现在，许多智能手机配备WIFI网络功能，这是用来收集定位数据的

¹⁰ 张弛：《大数据思维范畴探究》，《华中科技大学学报（社会科学版）》2015年第2期，第120-125页。



另一个来源。Skyhook 一类的服务商将不同 WIFI 网络的物理位置绘制在地图中，移动设备通过联入搜索到的 WIFI 网络，广播自己的位置信息。智能手机技术使个人位置数据更准确也更容易获得，尤其对移动设备应用程序的开发人员来说。此外，正在开发的新技术甚至可以收集在 GPS 信号极弱的建筑内的个人定位数据。

如今，导航设备、基站跟踪、智能手机是大多数定位数据的来源。导航设备频繁更新位置数据导致数据量飙升；全球庞大的手机用户群产生的个人定位数据相当惊人；而智能手机用户使用的各种应用程序要求定位跟踪，也使得定位数据快速增长。来自麦肯锡的数据表明，2009 年全球范围生成的个人定位数据超过 1 拍字节，并且以每年约 20% 的速度增长。日益增长的手机用户群每天都在生成庞大的数据，亚洲已成为产生个人定位数据的领先地区。2010 年，中国的手机用户世界第一，达 8 亿部；印度排第二，超过 6.5 亿部；北美大约是 3 亿部，位列第三。

个人定位数据主要应用在以下 3 方面：一是个人使用的定位服务，包括智能路由、汽车车载智能通讯、智能手机移动定位等；二是个人定位数据组织性的使用，比如地理定位广告、电子收费、保险定价和应急响应；三是聚合定位数据在宏观层面上的使用，包括城市规划和零售业务智能。

定位数据体现了互联网技术与移动终端的不断融合；而“众包”这种大数据时代全新的生产组织模式，则影响着这个世界传统的思维逻辑。从 20 世纪 90 年代互联网搜索引擎投入民用以来，科学家就一直试图通过互联网在时时更新的线索中发现下一个全球流行疾病的线索。这个数据挖掘项目很长一段时间中都处在科学实验的阶段，直到 2008 年 11 月谷歌推出了流感趋势监测系统（Google Flu Trends），这是一个运用类似数据“众包”处理方案（Crowd sourcing）预测流感暴发情况的在线数据处理平台。谷歌的科研团队发现，每周世界上都有上千万的互联网用户在线搜索健康信息。在流感高发期，以“流感”为关键词的在线检索会大大增加；过敏高发期，以“过敏”为关键词的在线检索增加；到了夏天，这一高频检索词则换成了“晒伤”……他们意识到也许已经是感染



流感病毒的人群才会上网检索相关信息，在流行病疫情和在线检索行为之间或许存在某种关联性。尽管不是所有搜索流感信息的人都真的得了流感，但所有检索信息叠加就会呈现出一种趋势。谷歌的科研团队将通过分析网上检索信息得到的趋势与传统流感监测系统相比较，发现检索高频区正是流感暴发区。他们认为通过计算在线检索行为的频率，可以预测世界上不同国家和地区的流感暴发程度，并在 2009 年 2 月的科学杂志《自然》发表了这一题为《使用搜索引擎查询数据监测流感疫情》的研究成果（Jeremy et al., 2009）。2014 年，杜佳斯和她的团队将这一成果的数据与前去相关暴发地医院就医的数字进行比较，发现紧随网上增长的检索行为，短期内医院就医人数出现增加（Andrea et al., 2014）。

继“谷歌流感趋势”之后，还有美国的“健康地图”（Health Map）和日本的“发现病毒”（Bio Caster）等在线公共卫生数据挖掘平台投入使用。“健康地图”通过不间断地扫描博客、推特、官方监控数据、新闻网站和 RSS 链接及用户上传的信息，用十种语言发布监测结果，美国卫生和公共服务部借助与它的合作发布季节性流感趋势及 H1N1 病毒在美国的传播情况。约翰·布朗斯坦（John Brownstein）是哈佛大学医学院、波士顿儿童医院流行病学家，也是“健康地图”的联合创建人之一。他认为数据挖掘的意义在于将监测数据转化为有效的健康护理，“我们更需要看到（数据处理）产生的影响”。在他看来，学界对使用社交网站挖掘数据的信心在不断增加。2012 年 1 月，布朗斯坦和他的同事通过“健康地图”实现了对 2010 年海地地区霍乱疾病的动态监控，比当地的医疗工作者通过传统数据检测发布的报告提前了 2 周（Trop et al., 2012）。布朗斯坦表示，“随着我们使用数据挖掘技术发布论文的增多，相信会有越来越多的研究者会接受这种方法的合法性”。鲁米·川那拉（Rumi Chunara）是哈佛医学院研究生物传感器的工程师。他认为非正式数据是传统监测方式的有益补充，尤其是在拥有较少健康资源但手机拥有率较高的地区和人群中。

“发现病毒”的开发者奈杰尔·科利尔（Nigel Collier）是位于东京的日本国立信息情报研究所的计算机专家。他认为对于自动化工作的软件来说，一大