

# 对大数据国家战略的几点考虑

魏 凯

中国信息通信研究院 北京 100191

## 摘要

大数据是国家发展的重要资源。对大数据资源的掌控与分析能力将成为未来国家竞争力的基础。我国大数据的发展及研究已有了良好开端,但也存在大数据资源活性不足、行业应用有待深化、关键核心技术亟待突破、产业基础薄弱、网络信息安全存在潜在隐患等问题。当前,制定国家大数据战略成为各界人士的共同呼声。国家层面的大数据战略究竟应该坚持什么思路,有哪些要素需要考虑,成为产业界和政府部门热议的话题。基于此,结合国外实践和国内问题,提出一些初步思考。

## 关键词

大数据 ; 国家战略 ; 法律规范

## *Some Considerations on the China National Big Data Strategy*

Wei Kai

China Academy of Information and Communications Technology, Beijing 100191, China

### *Abstract*

Big data is a kind of critical development resources. The ability of control and analysis on big data become the foundation of a country's competitiveness in future. Although China's big data development has a good start, there are a couple of challenges ahead, e.g., the open government and public data are insufficient, big data application is not widespread, core technologies R&D still need to be strengthened, privacy and information security are facing new risks. Currently, developing a national big data strategy is approaching a common agreement. While how to develop the strategy and what should to be considered are hot topics among the industry and government experts. Therefore, some considerations were proposed based on both other countries practice and domestic research.

### *Key words*

big data, national strategy, law and regulation

## 1 对大数据内涵的认识：新资源、新工具和新理念的综合体

大数据 (big data) 一词最早由美国航空航天局研究人员在1997年提出<sup>1</sup>，直到2011年麦肯锡公司发布报告《大数据：创新、竞争和生产力的下一个前沿》，才开始引起社会各界的广泛关注。

到目前为止，大数据仍然没有权威统一的定义，笔者认为需要从狭义和广义两个尺度认识大数据。狭义理解：大数据就是在数据采集技术快速提升、成本急剧下降的背景下，在互联网和各种信息系统上形成的体量大、格式多、速度快的数据资源，即“数据大”。广义理解：大数据不仅是“数据大”，还代表了一种新的认识世界的工具，更重要的是蕴含了一种“数据化”思维方式的变革。通过对大数据的分析，可以从中获得对现实世界的立体化的认识，这给人类提供了全新的工具，影响广泛而深远，因此有人也把大数据与火、电、计算机、互联网等并列，称为“通用目的技术”。笔者认为，当前大数据热的兴起，最重要的成果是一场“用数据说话”的思维洗礼。

就技术和应用而言，大数据是在互联网快速发展中诞生的。这个起点可追溯到2000年前后。随着全球网页规模的暴增，推动搜索引擎背后的存储和分析技术出现突变，一系列分布式技术随之出现，是今天大数据技术的源头。随后几年，海量数据驱动的业务模式及技术深入互联网营销、广告、电商等核心应用，取得巨大的商业成功，引发全社会对大数据的空前重视。

当前，大数据的技术、工具和思维理念正在从互联网向各个领域快速扩散。Gartner的调查显示，业界对大数据技术

的投资不断扩大，73%的受访者已经投资或者计划在未来24个月内投资大数据。综合IDC、Wikibon等公司的报告，仅大数据相关的数据存储、处理、分析相关的软硬件和服务，2013年产值就已达150~180亿美元量级，未来3年还将维持30%~50%的增长。在直接的产业规模之外，大家更看中大数据给各个行业带来的间接价值。大数据技术几乎可以应用到各个行业的价值创造和分配的全过程，通过降低交易成本，提高市场透明度，精确细分客户群，优化决策机制，催生新产品、新服务和新产业等方式间接创造价值巨大<sup>2</sup>。大数据不光有经济价值，还塑造了新的“数据空间”，已经成为继陆权、海权、空权、天权安全之后，另一个国家主权不可分割的组成部分<sup>[1]</sup>。

## 2 各国政府积极推动大数据发展

### 2.1 美国在推动大数据发展上最为积极，力图确立领先地位

美国依托其互联网和IT产业的领先地位，占据大数据技术发展先机，已形成明显的先发优势。Google、Facebook、Yahoo等公司的超大规模数据存储、分布式计算、深度机器学习等多项大数据关键技术已领先其他同类技术平均5~6年。这种技术领先优势借助开源模式向外扩散，带动Splunk、Tableau、Cloudera和HortonWorks等一大批专业化的大数据技术和服务公司不断涌现。大数据技术不仅在美国互联网搜索、广告、电子商务、社交网络等得到广泛应用，7成以上传统行业企业也不同程度地应用了大数据技术，特别是银行、保险、制造零售等领域已经有不少案例，国家安全、医疗、教育、科研等领域的大数据应用也在快速深化。依靠机器、算法和大数据提供某种程度的智能，从而

1  
[http://www.foreignpolicy.com/articles/2012/10/08/big\\_data](http://www.foreignpolicy.com/articles/2012/10/08/big_data)

2  
<http://www.drc.gov.cn/n/20140219/1-224-2878793.htm>

取得新的增长,正成为美国各行业信息化演变的主要方向,“第二次机器时代”的转折点正在来临<sup>3</sup>。

然而即便在领先的美国,大数据的发展初期,很多问题仍无法完全依靠市场自发解决,突出表现在以下几点:首先是技术问题,数据的采集、高效存储、深度分析等基础技术还不能适应发展需求,医疗、国防、交通、科研等方面的大数据技术水平仍然薄弱;其次是数据,商业机构往往面临数据短缺,而美国政府掌握的数据仅次于制造业,商业化利用价值非常高,但缺少开放;再次是法律制度,美国一些隐私保护组织呼吁加强保护,但以互联网公司为代表的商业机构则主张宽松对待,双方在收紧与放松之间展开拉锯战;最后是人才,麦肯锡的一份报告预计:到2018年,美国在“具备深入分析能力”的大数据专业人才缺口将达到14万~18万人。

美国政府意识到必须有所作为,于是2012年在全球率先推出“大数据行动计划(big data initiative)”,强化顶层设计,统筹推进技术研发、数据开放、率先应用和负面问题应对<sup>4</sup>。

- 加大技术研究:资助DARPA、NSF、NASA、NIH等机构开展核心技术研究。例如,在大数据计划资助下,加州大学伯克利分校开发的开源软件平台“伯克利数据分析软件栈(Berkeley data analytics stack),其中的内存计算软件Spark性能比大名鼎鼎的Hadoop提高近百倍,有望取代后者,成为大数据技术平台的主力<sup>5</sup>。

- 数据开放方面:在全球较早建立了统一的数据门户(data.gov),已开放37万个数据集和1 209个数据工具。奥巴马在2013年5月进一步要求,政府必须实现新增和经处理数据的开放和机器可读,激发大数据创新活力。

- 政府应用方面:美国政府是大数据的积极使用者。2013年曝光的“棱镜门事件”显示出美国国家安全部门大数据应用的强大实力,其应用范围之广、水平之高、规模之大都远远超过人们的想象。2012-2013年,美国国家安全局(NSA)、联邦调查局(FBI)及中央情报局(CIA)等联邦政府机构还大量采购亚马逊的云服务,以支撑其大数据应用。

- 关注负面影响:随着应用的深入,美国政府对大数据带来的负面影响也更加重视,美国白宫2014年5月发布的《大数据:抓住机遇,守护价值》报告中提醒,在发挥正面价值的同时,应该警惕大数据应用对隐私、公平等长远价值带来的负面影响。

## 2.2 英、日、澳、新等国家迅速跟进,努力缩小差距

英国将大数据列为战略性技术,给予高度关注。英国政府紧随美国之后,推出一系列支持大数据发展举措。首先是给予研发资金支持。2013年1月,英国政府向航天、医药等8类高新技术领域注资6亿英镑研发,其中大数据技术获得1.89亿英镑的资金,是获得资金最多的领域。其次是促进政府和公共领域的大数据应用。据测算,通过合理、高效使用大数据技术,英国政府每年可节省约330亿英镑,相当于英国每人每年节省约500英镑。为了在医疗领域更好地应用大数据,2013年5月,英国政府和李嘉诚基金会联合投资设立全球首个综合运用大数据技术的医药卫生科研机构,将透过高通量生物数据,与业界共同界定药物标靶,处理目前在新药开发过程中关键的瓶颈,之后还将汇集遗传学、流行病学、临床、化学和计算机科学等领域的顶尖人才,集中分析庞大的医疗数据。

日本政府把大数据作为提升日本竞争

3 美国麻省理工学院教授 Andrew McAfee 和 Erik Brynjolfsson 在 2014 年出版的新书《The Second Machine Age: Work, Progress and Prosperity In a Time of Brilliant Technologies》中提出的概念。

4 [http://www.whitehouse.gov/sites/default/files/microsites/ostp/big\\_data\\_press\\_release\\_final\\_2.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf)

5 取代而非补充, Spark Summit 2014 精彩回顾. <http://www.csdn.net/article/2014-07-17/2820713>

力的关键。日本政府认为,提升日本竞争力,大数据应用不可或缺。日本在新一轮IT振兴计划中把发展大数据作为国家战略的重要内容,新的ICT战略重点关注大数据应用技术。日本总务省2012年7月推出了新的综合战略“活力ICT日本”,将重点关注大数据应用,并将其作为2013年6个主要任务之一,聚焦大数据应用所需的社会化媒体等智能技术开发以及在新医疗技术开发、缓解交通拥堵等公共领域的应用。

此外,澳大利亚、新加坡等国也非常重视大数据发展。2013年8月初,澳大利亚出台公共服务大数据政策,提出了大数据分析的实践指南,希望通过大数据分析系统提升公共服务质量,增加服务种类,为公共服务提供更好的政策指导。在新加坡,多个国际领先企业在该国设立大数据技术研发中心,加速数据分析技术的商业应用。2014年初,新加坡资讯通信发展管理局(IDA)还聘请了首任首席数据科学家,专门推进政府数据的开放和价值开发。

### 3 我国大数据发展面临的突出问题

我国各界高度关注大数据的发展,积极推进相关技术研发和应用。目前,百度、腾讯、阿里巴巴等骨干互联网企业已建立了世界上规模最大的大数据平台,单集群规模达到上万台,在分布式系统、超大规模数据仓库、深度学习等关键技术上有所突破。大数据驱动的互联网、金融、流通、新媒体等跨界业务创新加速发展。但相对而言,我国大数据发展的信息化基础薄弱,大数据前沿技术原创能力不足,关键产品和服务的供应较为落后,大数据技术应用的扩散速度较慢,急需推动促进发展。

对我国来说,数据资源不丰富、技术差距大和法律法规不完善是当前大数据发

展面临的主要问题<sup>[2]</sup>。

一是我国数据源不够丰富,数据开放程度较低。丰富的高质量数据资源是大数据产业发展的前提。近几年在互联网产业及金融、电信信息化快速发展的带动下,我国数据资源总量有了快速增长,已达到全球的13%,但其他行业受信息化水平制约,数据储量仍不丰富。已有数据资源还存在标准化、准确性、完整性低,利用价值不高的情况。同时,我国政府、企业和行业信息化系统建设中受到各种因素制约,形成了众多“信息孤岛”。再加上体制机制、利益关系等还没有理顺,数据开放程度严重滞后于大数据发展需求。

二是我国大数据技术存在水平不高、技术扩散不畅的问题。我国大数据技术的发展模式也与全球类似,互联网企业具备快速将国际先进的开源大数据技术整合到自身系统中的能力,并构建了单集群上百万节点的大型系统,但仍缺乏原创技术,对开源社区的贡献不足,进而对前沿技术路线的影响比较微弱。同时,由于本土开源社区等产业组织发育滞后,国内领先企业在大数据方面的技术创新也难以向社会扩散。以往我国在数据库技术上的相对落后局面仍然没有根本改变,商用大数据产品供给能力还比较薄弱,使得依赖供应商的传统行业大数据应用水平受到限制。

三是大数据相关的法律法规有待进一步完善。随着大数据挖掘分析越来越精准、应用领域不断扩展,个人隐私保护和数据安全变得非常紧迫。在隐私保护方面,现有的法律体系面临着两个方面的挑战:一是法律保护的个人隐私,主要体现在个人可识别信息(personally identifiable information, PII),但随着技术的推进,以往并非PII的数据也可能会成为PII,使得保护范围变得模糊;二是以往建立在“目的明确、事先同意、使用限制”等原则之上的

个人信息保护制度,在大数据场景下变得越来越难以操作。而我国个人信息保护、数据跨境流动等方面的法律法规尚不健全,这成为制约大数据产业健康发展的重要原因之一。需要结合我国法治建设的实际情况,探索通过行业自律等方式弥补法律体系不完善的弊端。

## 4 对我国大数据国家战略的思考

针对大数据发展面临的问题,需要加强统筹规划,围绕经济社会发展的实际需求,以市场为导向、以企业为主体、以加强公共数据资源开放共享和推动政府和公共服务领域的大数据应用为抓手,保障大数据安全,营造发展环境,引领大数据在关键行业应用,以突破关键技术为核心,拉动培育自主可控产业体系,推动大数据为经济社会可持续发展做出积极贡献。本着这样的思路,建议国家大数据战略需要考虑数据资源、行业应用、技术产品、法律法规和安全保障5个要素。

### 4.1 促进大数据资源开放与流通

首先是加快推动公共数据向社会开放。坚持分级、分类、分阶段开放的原则,着力推进地理、人口、法人、金融、税收、医疗、社保、农业、交通、统计等公共数据对外开放力度,提升公共数据再利用水平。推进政府数据建设,健全政府数据管理制度,规范数据采集,保证信息数据质量。引导企业应用公共数据,加大对民生类公共数据应用的政策倾斜和支持力度。鼓励有条件的地区开展试点建设,探索开放模式,形成示范效应。加强安全保障,保守国家秘密,维护社会伦理,保护个人隐私和商业秘密。

其次是促进商业化数据流通交易。支持企业间点对点数据交换与共享,探索建立大数据交易市场,开展大数据交易试点,支持各行业数据的汇聚与交易,通过市场化机制促进数据资源的有效流通。依托数据流通与交易实践,引导形成数据流通交易规则体系,建立数据流通交易中的安全与隐私保护标准,确保数据资源的有序流通。

### 4.2 深化行业大数据应用

在应用方面,政府首当其冲,应该带头利用大数据技术提升政府治理与社会服务能力。要培育政府部门应用大数据提升政府治理能力的理念,重点推动政府在国家统计、市场监管、舆情监测、环境保护、公共安全、国家安全等方面的大数据应用,鼓励政府以外包服务方式采购大数据分析基础设施和服务。支持医疗卫生、食品安全、教育科研、城市交通、社会保障等众多公共领域的企事业单位开展大数据应用,推进各公共服务领域的大数据应用示范建设,鼓励公共领域机构开展大数据创新,以提升公共服务能力并惠及民众生活。

而在商业化应用方面,政府应该退后一步,把工作重点放在完善制度、创造环境方面,支持互联网、电信等数据资源丰富、IT能力较强的企业深化内部应用,稳妥有序地将自身数据资源和服务能力对外开放。一是鼓励互联网企业建设大数据分析平台,支撑内部应用,并将大数据能力向其他行业进行辐射,推动互联网企业以在线云服务方式向传统行业企业开放大数据分析平台能力。二是支持电信企业在保障数据安全和个人隐私的前提下,充分利用电信行业丰富的数据资源,提升电信运营企业在客户管理、市场营销、产品开发、网络优化等方面的大数据应用水平,鼓励

电信企业对外提供数据资源与分析能力，为各类企业、行业单位和政府部门提供大数据服务，促进社会化大数据产业生态快速、有序发展。

### 4.3 突破大数据关键技术产品

首先，要攻克大数据核心关键技术。搭建国际化的学术交流和人才流动平台，产、学、研、用联合开展大数据理论研究。攻克面向EB级超大规模数据集的新型数据存储、表示、管理和查询关键技术；面向文本、图像、视频大数据和物联网感知数据，研发自然语言分析、信息提取、知识表现、自动化推理、机器学习等核心技术，突破基于超大规模深度神经网络的深度学习相关理论与技术，研发数据安全与隐私保护理论与技术。

其次，要大力培育自主可控的大数据产品与服务。整合国内互联网、制造企业和科研机构力量，共同研发支持超大吞吐量的新一代融合并行分布式计算平台、支持结构化和非结构化数据的大规模数据仓库、基于深度神经网络的超大规模数据分析引擎等平台级软件，培育自主开源软件社区，构造生态链。大力发展面向垂直领域的数据分析、数据资产管理、商业智能应用软件和在线服务。

### 4.4 完善大数据法律法规

在个人隐私保护方面，要严格保护大数据应用中的个人信息。落实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，探索制定大数据环境下数据收集、开放、交换、应用等环节的个人信息保护指南，明确大数据应用相关各方的个人信息保护义务和责任。

在个人和企业数据资产保护方面，要探

索建立数据产权保护制度。探索制定信息资产管理制度，建立信息资源基本权利体系，明确各类市场主体所积累的信息资产所有权归属，建立规范化管理和使用机制，保护信息所有者、信息主体及公众合法利益。

在国家数据主权方面，要建立数据跨境流动管理制度。制定跨境数据流动分级分类管理制度，涉及国家秘密、国家安全以及经济安全的数据严格禁止跨境，对政府和公共部门数据跨境流动，实施安全风险评估机制。建立外资安全审查机制，强化对外资的数据跨境流动管理。

### 4.5 保障大数据安全

首先，要构建大数据安全管理制度体系。加强大数据安全顶层设计，积极推动数据信息保护立法，加快大数据安全人才队伍建设。探索建立面向政府信息采集和管控、敏感数据管理、数据交换标准和规则、个人隐私等领域的大数据安全保障制度，明确数据采集、使用、开放等环节涉及信息安全的范围、要求和责任，确保国家利益、社会安全、商业秘密、个人隐私等信息不受侵犯。

其次，加强大数据安全保障能力建设。加强大数据安全基础设施和软硬件建设，改变传统的端级防护、单点布防的安全解决方案，提高针对大数据网络攻击、数据存储应用系统漏洞等安全防护能力。加强数据匿名保护、数据水印、数据溯源、角色挖掘、风险自适应的访问控制等大数据安全保护关键技术研究，从技术角度降低大数据泄露、滥用的风险，保障大数据信息的合法、合理利用。

## 5 总结与展望

大数据的发展涉及资源、技术、产业、

政策和法律等方方面面,不能完全依靠市场,需要政府统筹谋划,协同推动。无论是从美、日、欧等国家和地区的政策实践看,还是从国内发展需求看,各方都认为有必要尽快出台我国的大数据国家战略。

种种迹象表明,大数据战略的出台日益临近。在2014年《政府工作报告》中就提出“设立新兴产业创业创新平台,在新一代移动通信、集成电路、大数据、先进制造、新能源、新材料等方面赶超先进,引领未来产业发展”。2015年《政府工作报告》又提出了“推动移动互联网、云计算、大数据、物联网等与现代制造业结合”。工业和信息化部2015年工作重点中提出,要推动出台大数据应用与发展指导性文件<sup>6</sup>,并制定工业大数据创新发展指导意见<sup>7</sup>。2015年国家发展和改革委员会在高技术工作年度工作会议上,进一步明确提出要启动“国家大数据战略及行动计划”<sup>8</sup>。相信随着相关

政策的出台和落地,数据割据、技术薄弱、法规缺失等问题将得到改善,大数据在我国的应用发展也将迈上新的台阶。

## 参考文献

- [1] 沈国麟. 大数据时代的数据主权和国家数据战略. 南京社会科学, 2014(6): 113~119  
Shen G L. Data sovereignty and national strategy in big data era. Nanjing Journal of Social Sciences, 2014(6): 113~119
- [2] 工业和信息化部电信研究院. 大数据白皮书(2014). [http://www.catr.cn/kxyj/qwfb/bps/201405/t20140512\\_1017466.html](http://www.catr.cn/kxyj/qwfb/bps/201405/t20140512_1017466.html), 2014  
China Academy of Telecommunication Research of MIIT. Big data white paper (2014). [http://www.catr.cn/kxyj/qwfb/bps/201405/t20140512\\_1017466.html](http://www.catr.cn/kxyj/qwfb/bps/201405/t20140512_1017466.html), 2014

6  
<http://www.miit.gov.cn/n11293472/n11293832/n11293907/n11368223/16342761.html>

7  
<http://www.miit.gov.cn/n11293472/n11293832/n11293907/n11368223/16342761.html>

8  
<http://www.dqfgw.gov.cn/fgwweb/WN010018/16681.htm>

### 作者简介



**魏凯**,男,中国信息通信研究院(原工业和信息化部电信研究院)通信标准研究所移动与大数据研究部副主任、高级工程师,国际电信联盟(ITU-T)标准化局第16研究组Q21联合报告人,数据中心联盟大数据组组长,中关村大数据交易产业联盟副秘书长。主要从事互联网技术、标准与产业政策研究工作。主持和参与科学技术部、国家发展和改革委员会、工业和信息化部科研和产业化项目。近年来参与国务院《关于促进云计算创新发展培育信息产业新业态的意见》、工业和信息化部等5个部委《关于数据中心建设布局的指导意见》等重要政策文件的起草。发表论文30多篇,申请发明专利1项,主持完成通信行业标准10多项、ITU-T国际建议书2项,3次获得省部级科学技术奖。

收稿日期: 2015-05-03; 修回日期: 2015-05-10

论文引用格式: 魏凯. 对大数据国家战略的几点考虑. 大数据, 2015010

Wei K. Some considerations on the China national big data strategy. Big Data Research, 2015010