

# 美国大数据政策的发展趋势

乔 健

( 中国科学技术部, 北京 100862 )

**摘 要:** 本文系统介绍了美国大数据技术及其应用的发展现状, 梳理了奥巴马政府针对大数据发展所采取的一系列政策措施, 特别是其大数据研发计划与数据开放政策, 研究了美国对大数据在隐私与安全领域影响的政策讨论, 以期为我国在大数据领域的政策制定提供借鉴。

**关键词:** 美国; 大数据; 数据开放

**中图分类号:** TP311.13    **文献标识码:** A    **DOI:** 10.3772/j.issn.1009-8623.2016.05.001

大数据技术, 是指在海量、复杂数据集合中通过分析来快速获得有价值的信息的技术。随着社会生产和生活日益互联网化, 产生的数据信息呈指数级增长。据统计, 2013 年全球产生的数据量达到了 4 泽字节 (1 泽字节 = 240 吉字节), 并且保持着大约每两年翻一番的增长趋势<sup>[1]</sup>。如此海量的数据通过分析, 可以获得很多衍生信息, 催生出大量新企业、新产品和新服务。因此, 大数据技术被世界各国所重视。据麦肯锡集团一份报告<sup>[2]</sup>的统计, 全球对于开放数据的利用, 每年可在教育、交通、零售、电力、能源、健康、金融七大领域增加 3 万亿美元经济增长, 减排 30 亿吨二氧化碳。

美国在大数据领域处于世界领先地位。2009 年, 美国国家科学技术委员会 (NSTC) 发布了《开发数字数据的威力》报告<sup>[3]</sup>, 初步提出了发展大数据的框架、建议和目标; 2012 年 3 月, 奥巴马政府宣布启动“大数据研究与开发计划”, 投入 2 亿美元进行大数据相关技术研发; 2013 年 5 月, 奥巴马政府发布行政令<sup>[4]</sup>, 加大政府数据开放力度, 以更加有效地利用宝贵的公共信息资源; 2014 年 5 月, 白宫行政办公室与总统科技顾问委员会 (PCAST) 联合发布报告《大数据: 抓住机遇, 保护价值》<sup>[5]</sup>与《大数据和隐私: 技术视角》<sup>[6]</sup>, 分

别从政策和技术的角度分析了大数据技术的发展对社会带来的影响, 特别是对隐私的影响。2016 年 5 月, 白宫又发布了《联邦大数据研发战略计划》报告<sup>[7]</sup>, 在已有基础上总结未来研发重点战略, 指导美国大数据发展进程。

## 1 美国大数据技术应用发展现状

大数据技术的关键在于从非结构化的海量数据中提取有效信息的能力。斯坦福大学与谷歌公司的一项研究<sup>[8]</sup>通过对互联网中上百万张照片进行分析归纳, 使计算机具有识别“猫”这一主题的能力。大数据技术是数学、统计学、人工智能、计算机科学等多学科互相融合的交叉学科, 通过对数据信息的自动分类、归集、预测、总结等进行分析, 实现有效信息的提取。

大数据的支撑技术还包括图像、语音识别及自然语言处理。越来越多的信息以图像和语音的形式存在, 对图像及语音信息的有效分析在三维地图、智能监控及情报安全领域得到了广泛应用; 自然语言处理则可以分析在社交媒体上呈指数增长的碎片化文字信息。对于文字、数据、图像、视频和语音等不同种类信息的联合分析, 可打破屏障、实现整合, 提高数据的价值, 这一技术称为数据融合。

作者简介: 乔健 (1984—) 男, 工学硕士, 主要研究方向为科技政策及国际科技合作。

收稿日期: 2016-05-04

大数据分析技术背后的软硬件基础设施同样至关重要。全球大部分数据存储在约 6 000 个数据中心内，而这些数据中心约有一半位于美国<sup>[9]</sup>。面对越来越复杂的数据库结构，传统的结果化查询语言（SQL）已无法进行管理，基于开源平台的 Hadoop 大数据管理软件发展迅速。美国信息产业巨头谷歌、亚马逊等公司都开发了自己的数据软件系统。数据管理分析能力已成为 IT 公司的核心竞争力。

大数据技术在各行各业催生了一批中小企业，它们大多利用可以公开获取的数据信息进行加工分析，以开发新的产品或服务。如，在农业领域，Climate Corporation 公司利用气象、地质及农产品数据库，开发出农业种植辅助决策系统及保险产品，该公司被孟山都集团以 10 亿美元收购；在医疗领域，Archimedes 公司利用电子病历数据库，开发了 IndiGO 软件系统，根据综合体检结果判断罹患心脏病、糖尿病等的健康风险，为医生提供诊断辅助；在个人金融领域，Bill Guard 公司利用美国消费金融保护局的数据库信息，分析信用卡欺诈的常见行为模式，对客户的注册信用卡提供实时保护；等等。

## 2 奥巴马政府的大数据相关政策

一方面，奥巴马政府视数据为宝贵的社会财富，政府部门产生或收集的数据属于公共所有，在不触犯隐私及公共安全的情况下应向社会免费开放，这使得私营部门可以基于此开发新产品、新服务和新技术，促进中小企业发展，增强全社会经济活力；另一方面，各联邦机构大力投入，加强大数据相关技术研发，抢占技术前沿领域的战略制高点。

### 2.1 开放数据计划

2009 年，奥巴马上任伊始就发表了《透明和开放的政府》备忘录<sup>[10]</sup>，要求联邦各部门采取具体措施，贯彻政府部门透明、参与和合作的原则。2013 年，奥巴马发布行政命令，落实开放数据计划。在这项行政命令中，奥巴马要求所有政府部门的公共数据库在非特殊情况下要向公众免费开放，并且要与行业接轨，使用便于机器处理的通用数据库格式。此项行政命令由白宫预算管理办公室与白宫首席信息官共同牵头执行，纳入各联邦部门绩效考核

指标。政府数据库通过统一的数据平台 data.gov 发布。目前，该平台上已有超过 18 万个政府数据库向公众开放。

为了使开发数据计划取得预期的效果，美国政府大力推进建立公私合作的数据创新生态系统。为此，白宫在开源程序平台 GitHub 上启动了一项在线开放数据工程，欢迎社会各界对开放数据计划献计献策，如程序员可以针对政府开放数据库编写应用程序接口（API）软件，供全社会共享。此外，纽约大学政策实验室（GovLab）推出了“开放数据 500 强”计划，从各行各业中选出利用政府开放数据获得商业成功的 500 个最佳案例，总结商业模式并进行分享。

### 2.2 “我的数据”计划

与开放数据计划不同，“我的数据”计划并不是将数据向全社会开放，而是将个人医疗、能源、金融、教育等信息完整地以标准数据格式提供给本人，从而促进数据分析利用，形成数据市场，为个人提供便利并促进创新。

“我的数据”计划最早开始于 2010 年美国退伍军人部启动的“蓝按钮”计划，该计划的最初目的是为美国的退伍军人提供标准化健康医疗信息下载服务，后来该计划推广至全民，目前有超过 1.5 亿美国人可享受“蓝按钮”服务，一键式下载体检、药物史、处方、疫苗等完整医疗数据。有超过 500 家公司为“蓝按钮”计划提供后续数据处理服务，开发衍生市场。仿照“蓝按钮”计划，美国能源部于 2012 年启动了“绿按钮”计划，使个人可下载住宅能耗详情；2014 年，美国税务局又推出“我的税单”服务，方便个人进行纳税并据此办理个人信贷。目前“我的数据”计划方兴未艾，个人数据市场的形成将进一步促进大数据创新及应用。

### 2.3 大数据研究与开发计划

2012 年，奥巴马政府启动大数据研究与开发计划，由美国国家科学基金会、国立卫生研究院、国防部、国防部高级研究计划局（DARPA）、能源部、地质调查局 6 家联邦机构共同开展，共安排 2 亿美元预算，深入研发大数据获取、存储及分析等相关技术。DARPA 在其 XDATA 计划中开发针对不完整、缺损数据库的分析软件；国立卫生研究院重点分析生物大数据；能源部加强气象数据的分析、管理和

可视化；国家科学基金会的大数据项目使得人类基因组测序的成本降低到了原来的 1/40。

美国大数据计划的跨部门协调由 NSTC 下设的“网络与信息技术研发计划”（NITRD）分委员会负责，并专门设立了跨部门高级别协调工作组。近年来，大数据研发计划的参与部门增加到了 15 个，研发预算也逐年增长，2016 年财年达到 2.9 亿美元。

为了更好地促进大数据研发计划的技术转化，2015 年，美国国家科学基金会投入 500 万美元，在全美建立四家大数据区域创新中心，覆盖 50 个州，吸引超过 250 家大学、企业及城市政府等组织参与，以加强不同部门之间的协作，深入研究蓬勃发展的数据科学与技术，解决区域及全国性挑战。

### 3 美国对大数据隐私相关问题的政策思考

历史上的技术进步通常伴随着新的社会问题以及相关的政策讨论，如 1890 年便携式照相机的发明导致了人们对于隐私的普遍担忧，最终于 1934 年修改了美国民法，对侵犯隐私行为作出新的界定。大数据技术在大大增强人们获取未知信息能力的同时，也给隐私和安全带来了新的挑战。白宫报告《大数据：抓住机遇，保护价值》对这一问题进行了系统性思考，并提出了政策性建议。

#### 3.1 美国隐私伦理与法律框架的建立

美国对隐私伦理的最根本表述是其宪法第四修正案，即“人民的人身、住宅、文件和财产不受无理搜查和扣押，不得侵犯”。但这一原则无法预测未来的技术发展会给隐私权利带来怎样的挑战，其现代隐私法律框架体系是在实践中逐步讨论确立的。

1973 年，美国确立了个人信息数据操作原则。当时，信息统计系统越来越多地用于卫生、教育和共同福利等事务，引发人们对数据安全的担忧。在此背景下，法律界提出了公平信息操作原则（FIPP），其核心是：个人有权知晓何种个人信息被收集、作何使用；有权反对某种使用方式；有权修改错误信息；收集机构有义务保证信息可靠及安全等。这一原则后来成为美国一系列立法的基石，如个人金融领域的《公平信用报告法案》（FCRA）、医疗领域的《医疗保险可靠性法案》（HIPAA）和教育领域的《儿童网上隐私保护法案》（COPPA）等。

需要指出的是，美国法律体系中没有对个人信

息保护的总体立法，而是在实践中根据需要针对不同主体和应用领域进行具体规定。这样做的出发点是为制度创新留有余地，但事实上也造成了不同主体适用不同法律，同时存在法律监管“灰色地带”的混乱局面。

#### 3.2 大数据给隐私带来的挑战

使用大数据技术，通过整合从前所认为的低敏感度信息，可以得到确定当事人身份、描述其行为模式的高敏感度信息，这种现象被称为“马赛克效应”。具体来说，在实践中大数据技术的应用带来了以下方面的风险和挑战：

（1）对于元数据的保护亟待加强。元数据指的是描述数据各种属性的信息，即“关于数据的数据”。例如，在通信领域，通信内容（语音通话、短信、电子邮件等）受到《电子通信隐私法案》（ECPA）的保护，但通话时长、电子邮件地址、拨打号码等“元数据”缺乏法律保护，很容易被获取。在大数据技术的分析下，以上元数据可以揭示相当多的信息，使个人隐私面临严重威胁。

（2）消费者信息不对称，面临歧视风险。目前互联网对于个人相关信息的收集普遍采用“通知—同意”模式，即以消费者同意列出的各项隐私协议作为使用服务的前提。但在实践中，这项看似公平的举措形同虚设，因为个人很少仔细阅读隐私协议，对于个人信息后续使用情况也无从得知。实际上，目前已经形成了庞大的个人数据产业链，诸多公司利用大数据技术分析个人的消费习惯、支付能力、健康状况或失业风险，将结果出售给零售商、保险公司、银行等机构。目前这一新兴领域尚无明确监管法案，消费者信息面临滥用威胁，特定群体很容易由于数据分析结果而受到歧视。

（3）大数据教育具有隐私风险。目前，大数据正在被应用于美国的教育领域。互联网在线教育可以对学生进行针对性测评，个性化调整教学进度、内容与方法，并进行系统性分析研究。但这种教学方式将收集到关于学生的大量学习数据，通过分析可以揭示学生的学习能力、智力水平、性格特征等大量敏感信息，如不慎泄露，将对学生们的未来发展造成不利影响。目前，对这一领域的监管还是空白。

（4）大数据医疗面临政策障碍。医疗也是大

数据的重点应用领域，通过对大量电子病历的统计分析，可以在很大程度上帮助医生得到诊断决策和治疗建议。为了使统计结果更加准确，往往需要收集患者生活习惯、饮食特点甚至经济状况等非传统领域的“医疗数据”，美国现行的《医疗保险可靠性法案》对以上信息的收集有诸多限制。如何平衡隐私和使用需要，成为美国大数据医疗领域的博弈热点。

### 3.3 对于大数据隐私相关问题的政策性建议

目前，美国对于大数据安全与隐私问题的立法还处于讨论阶段，总体思路是既要规避大数据技术应用可能带来的种种风险，又要为未来的技术与商业模式创新留有足够的发展空间。针对这一问题，白宫报告提出了以下主要建议：

（1）关注大数据结果的使用，而非数据的收集和分析过程。随着互联网与物联网的发展，越来越多的信息被数字化，全球数据量呈指数级增加。目前，数据通过云计算被存储与分析的体系基本形成，这就决定了数据一旦形成，就难以被彻底删除。未来大数据技术的发展难以预测，随着各学科成果的交汇，现在看似无价值的数据可能在未来提供新的信息。以上规律决定了限制数据收集和分析的政策导向既不现实，又会阻碍技术进步与创新。而大数据技术的负面影响主要体现在对分析结果的滥用上，应该加强对大数据应用的政策监管。

（2）发展隐私保护第三方中介机构，加强信息处理透明度。目前，个人与机构在数据利用领域信息高度不对称，现行的“通知—同意”模式无法保护消费者的隐私与权利。应该加强第三方监管，大力发展监督数据使用的第三方中介机构，根据消费者的需求对各种软件和服务进行隐私保护评级，保护个人对于数据的知情权与维护权。各机构应该对大数据的算法和应用保持透明。

（3）加快大数据隐私保护技术的研发与标准制定。与大数据的分析利用相比，数据的隐私保护技术研发投入尚显不足。数据加密、数据访问及处理可追溯、数据标签、安全多方计算等领域具有很大的应用潜力，从根本上提高网络安全与隐私保护水平是有希望的。除加大研发投入外，更重要的是制定行业标准，将数据的分析利用纳入可监管、可追查、可控制的安全轨道。

## 4 结语

作为前沿技术，大数据正日益得到世界各国的重视，其对社会影响的深度和广度早已超出了单纯的技术范畴，带来了如雨后春笋般的新产品、新服务与新商业模式，也带来了对现有政策框架的挑战。我国于2015年发布了《促进大数据发展行动纲要》，全面推进大数据发展和应用。在坚持创新驱动发展、大力推进自主创新的过程中，我国应具有国际视野，关注美国等发达国家在大数据领域的研发动态与政策趋势，在加强基础研究、稳步推进政府数据开放、推动政策法规建设等方面积极借鉴国际经验，加快建设数据强国进程。■

### 参考文献：

- [1] Mary Meeker, Liang Wu. Internet Trends[R]. California: KPCB, 2013.
- [2] James Manyika, Michael Chui, Diana Farrell, et al. Open Data: Unlocking Innovation and Performance with Liquid Information[R]. Chicago: McKinsey Global Institute, 2013.
- [3] NSTC. Harnessing the Power of Digital Data for Science and Society[R]. Washington DC, 2009.
- [4] The White House. Executive Order: Making Open and Machine Readable the New Default for Government Information[R]. Washington DC, 2013.
- [5] EOP. Big Data: Seizing Opportunities Preserving Values[R]. Washington DC, 2014.
- [6] PCAST. Big Data and Privacy: A Technological Perspective [R]. Washington DC, 2014.
- [7] NSTC. Federal Big Data Research and Development Strategic Plan[R]. Washington DC, 2016.
- [8] Le Q V. Building High-Level Features Using Large Scale Unsupervised Learning[A]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Vancouver: IEEE, 2013: 8 595-8 598.
- [9] Charles McLellan. The 21st Century Data Center: An Overview[EB/OL]. (2013-4-2) [2016-04-15]. <http://www.zdnet.com/article/the-21st-century-data-center-an-overview/>.
- [10] The White House. Transparent and Open Government[R]. Washington D C, 2009.

(下转第34页)

## Knowledge and Technology Transfer Based on Partnerships: Introduction and Enlightenment of KTP Program in UK

LI Zhen-xing<sup>1</sup>, WANG Xi<sup>2</sup>

(1. Chinese Academy of Science and Technology for Development, Beijing 100038;

2. Torch High Technology Industry Development Center, Ministry of Science and Technology of China,  
Beijing 100045)

**Abstract:** This paper introduces the capital source, funding mode, applicable standards, procedure for examination and approval, organization, implementation and assessment process of Knowledge Transfer Partnership (KTP) program, then analyzes KTP's implementation effect. Insights are given as following: the innovation capability of the company could be promoted by funding of related S&T program; good design of the benefit mechanism is the basis of success of the S&T program; people is the core of the knowledge and technology transfer process; full participation of the adviser is helpful to the successful implementation of the program; innovation of the funding mode is helpful to increase the efficiency of Input-Output of public fund. The fund idea, and mode and experience of KTP could be used for the reference of design and implementation of S&T programs in China.

**Key words:** UK; KTP program; S&T program; technology transfer

---

(上接第 4 页)

## Introduction and Enlightenment of Development Trends of Big Data Policies in U.S.

QIAO Jian

(Ministry of Science and Technology of the People's Republic of China, Beijing, 100862)

**Abstract:** This paper systematically introduces the current status of the Big Data technology and its application in the United States, and combs the policies and measures of the Obama Administration to develop the Big Data technology, especially the big data research and development initiative and the open data policy. The policy debate on the privacy and security issues of the Big Data is also studied, with the hope of shedding lights on the Big Data policy establishment in China.

**Key words:** U.S.; big data; open data