大数据：
抓住机遇、保存价值

美国总统行政办公室

2014 年 5 月

“即使大数据技术重塑了我们周围的世界，今天的发言也将帮助我们持续贯彻自身的价值观念。”

“这份评估报告本质上被认为是一种对大数据作用范围的调查。在过去的 90 天中，评估小组对学术专家、产业代表、保护个人隐私的倡导者、人权团体、执法者，以及其它政府机构进行了调研。白宫科学与技术政策办公室(White House Office of Science and Technology Policy)与麻省理工大学、纽约大学、加州伯克利大学联合组织了三场大学会议。”

“在 2014 年，美国国税局通过一个名为‘Get transcript’的工具将纳税人的信息数据加以共享，纳税人可以通过它获得他们自己最近三年的纳税记录。个人纳税者可以借此下载过去的纳税申报单，这使得居民进行抵押、学生贷款、商务贷款等活动与填写纳税表变得更加便捷。”

“尽管医学技术不断变化，但健康数据仍然是我们生活中非常私密的部分。在大数据使得较之以往任何时候都更为强大的发现成为可能的同时，重新审视相关信息被所有医疗保健机构共享后的隐私保密方式也显得相当重要。医疗保健行业的领导者已经呼吁构建一个更为广泛的信用框架，使得不同来源、不同隐私保密程度的健康数据得以汇聚。这一框架需要附加《健康保险便利和责任法案》与《反基因歧视法》(“Genetic Information Non-Discrimination Act”, GINDA)中的隐私保护条款，并同时设计标准化数据结构以提高其跨平台适应性。”

“大数据正在改变世界。但是它并没有改变美国人对于保护个人隐私、确保公平或是防止歧视的坚定信仰。这份调查报告旨在鼓励使用数据以推动社会进步，特别是在市场与现有的机构并未以其他方式来支持这样的进步的领域，与此同时，我们也需要相应的框架、结构与研究，来帮助保护我们的核心价值观念。”

亲爱的总统先生：

我们正生活在社会、经济与技术革命之中。我们如何通信、交际、度过闲暇时光以及开展业务已经转移到了互联网上。互联网又渗透进入我们的手机，进入蔓延在我们家园和城市中的设备中，进入推动工业经济的工厂中。其导致的数据爆炸和挖掘正改变着我们的世界。

今年一月，你要求我们进行为期 90 天的调查，检验大数据将如何改变我们生活和工作的方式，改变政府、公民、企业家和消费者之间的关系。这次调查的重点在于公共和私营部门如何在将风险最小化的同时，将大数据的价值最大化。它也为大数据确定了发展我们的经济，改善健康和教育以及使我们国家更加安全和节能的机会。

虽然大数据毫无疑问地加大了政府权力累积未经核实的事实的可能性，但它也提供了增强公众责任、隐私和权利的方案。如果正确实施，大数据将成为历史前进的推动力，帮助我们国家保持长期以来成为我国特点的公民和经济活力。

大数据技术将变革生活中的每一个领域。它们使之成为可能的知识发现提出了我们为隐私保护构架的框架如何在大数据生态系统中应用的重大问题。大数据也引发了其他问题。这个报告的一个重大发现在于，大数据分析有一定可能使长久存在的公民权利保护黯然失色，特别在于个人信息如何利用于住房、信贷、就业、健康、教育及市场领域上。美国人与数据的关系将扩展他们的机会和潜力，而不是缩减。

我们正在建设我们将继承的未来。美国比世界上的任何其他国家更适合确保数据革命能够持续地为个人发展和社会良好效力。我们欣然提交本报告，建议我们如何在保护诸如隐私、公正、自决等基本价值的同时拥抱大数据技术。我们致力于这一倡议和改进。我们今天在这发起的对话将帮助我们在大数据重塑我们周围世界的同时，坚持我们自身的价值。

目录

1、大数据与个人

什么是大数据？

自从古代有过第一次计数和农作物产量记录以来，数据收集和分析便成为社会功能改进的根本手段。17、18 世纪的微积分、概率论和统计学所提供的基础性工作，为科学家提供了一系列新工具，用来准确预测星辰运动、确定公众犯罪率、结婚率和自杀率。这些工具常常带来惊人的进步。在 19 世纪，约翰 • 斯诺(John Snow)博士运用近代早期的数据科学绘制了伦敦霍乱爆发的“群聚”地图。霍乱在过去被普遍认为是由“有害”空气导致的，斯诺通过调查被污染的公共水井进而确定了“霍乱”的元凶，并同时奠定了疾病细菌理论的基础。[1]

从数据中撷取洞见以提振经济行为，这也是美国工业的惯常做法。弗雷德里克 • 温斯洛 •泰勒(Frederick Winslow Taylor)在宾夕法尼亚州的米德瓦尔钢铁厂采用秒表和笔记板来分析生产力，这大大增加了车间产量，也铸就了他的信念，即数据科学可以为生活中每一个方面都带来革命性影响。[2] 1911 年，泰勒撰写了《科学管理原理》，以回应西奥多 • 罗斯福(Theodore Roosevelt)总统有关提升“国家效能”的倡议：

> 从我们单个人的行动到大型企业的工作，科学管理的基本原理可以应用到一切类型的人类行为中……无论何时，只要正确运用这些原理，必定会产生真正令人惊讶的成果。[3]

今天，数据比以往任何时候都更加深入地与我们的生活交织在一起。我们期待着用数据解决各种问题、改善福利，以及推动经济繁荣。数据的搜集、存储与分析技术不断提升，这种提升看上去正处于一种无限的向上轨迹之中。它们的加速是因为处理器能力的增强、计算与存储成本的降低，以及在各类设备中嵌入传感器的技术的增长。2011 年，新生成的和复制的信息量估计超过了 1.8 ZB（泽字节）；[4] 而在 2013 年，这一数字估计可达 4 ZB。[5]

---

[1] Scott Crosier, *John Snow: The London Cholera Epidemic of 1854*, Center for Spatially Integrated Social Science, University of California, Santa Barbara, 2007, http://www.csiss.org/classics/content/8.

[2] Simon Head, *The New Ruthless Economy: Work and Power in the Digital Age*, (Oxford University Press, 2005).

[3] Frederick Taylor, *The Principles of Scientific Management* (Harper & Brothers, 1911), p. 7, http://www.eldritchpress.org/fwt/ti.html.

[4] John Gantz and David Reinsel, *Extracting Value from Chaos*, IDC, 2011, http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf.

[5] Mary Meeker and Liang Yu, *Internet Trends*, Kleiner Perkins Caulfield Byers, 2013, http://www.slideshare.net/kleinerperkins/kpcb-internet-trends-2013.

什么是泽字节（ZB 或 Zettabyte）

一泽字节等于 10^{21} 字节，或相应的信息单元。想想看，一个字节等于文本中的一个字符。1ZB 相当于存储 323 兆份列夫•托尔斯泰所著的 1250 页的《战争与和平》所需的容量。[6]或者想象一下，假定每一个美国人每秒钟拍一张照片并连续拍 1 个月，所有这些照片存储进来容量就相当于 1ZB。

世界上每天大约有 5 亿张照片上传或分享，另外每分钟还有超过 200 小时的视频上传、分享。但是，即使是人们自己产生的信息，即从语音通话、电子邮件、文本到上传的图片、视频、音乐等全方位交流产生的信息，与每天产生的与他们相关的电子记录等数字化信息相比，在数量上都是相形见拙的。

这些趋势还将持续下去。我们只是处在所谓的“物联网”(“Internet of Things”)的相当初级的阶段。在物联网中，我们的各种应用设备、运输工具以及持续增长的“可穿戴”技术产品将可以彼此交换信息。技术的进步将促成创建、捕捉、管理与存储信息的成本降至 2005 年的六分之一。自 2005 年以来，人们在硬件、软件、人才与服务方面的商业投资增长了近 50%，达到 4 万亿美元。

“物联网”

“物联网”这个术语用来描述具有可交换信息能力的设备网络。这些设备通常嵌入了传感器，并通过有线或无线网络连接后进行彼此间的信息交换。它们可能包括你的温控器、汽车，甚至是你咽下去的“小药片”，医生可以用它来监控你的肠胃以及消化道的健康状况。这些连接的设备通过互联网传输、编制和分析数据。

关于“大数据”有许多种定义，这种差别取决于你是一位计算机科学家，还是一位金融分析师，抑或是一位为风险投资人推销一个概念的企业家。多数定义都反映了那种不断增长的捕捉、聚合与处理数据的技术能力，而这个数据集在数量、速率与种类上持续扩大。换言之，“现在，数据可以更快获取，有着更大的广度和深度，并且包含了以前做不到的新的观测和度量类型。”[7]更确切地说，大数据集是“庞大的、多样化的、复杂的、纵深的和/或分布式的，它由各类仪器设备、传感器、网上交易、电子邮件、视频、点击流，以及现在与未来所有可以利用的其他数字化信号源产生”。[8]

就大数据而言，真正重要的是它能做什么。先且不论我们如何把大数据界定为一种技术现象，大数据分析那多元而广阔的潜在用途将面临一些关键性的问题，即我们的法

[6] “2016: The Year of the Zettabyte,” Daily Infographic, March 23, 2013, http://dailyinfographic.com/2016-the- year-of-the-zettabyte-infographic.

[7] Liran Einav and Jonathan Levin, “The Data Revolution and Economic Analysis,” Working Paper, No. 19035, *National Bureau of Economic Research*, 2013, http://www.nber.org/papers/w19035; Viktor Mayer-Schonberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, (Houghton Mifflin Harcourt, 2013).

[8] National Science Foundation, Solicitation 12-499: *Core Techniques and Technologies for Advancing Big Data Science & Engineering* (BIGDATA), 2012, http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf.

律、伦理与社会规范在大数据时代是不是有足够的能力保护个人隐私和其它价值。前所未有的计算能力与持续的改进能力为我们的生活带来了可能是先前从未预料到的发现、创新与进步。但是，这些能力对于普通消费者来说，多数都是既不可见，也无法得到的，因此，它们在持有大数据的一方与有意无意地提供着数据的一方之间，形成了一种非对称的力量。

部分挑战也在于如何理解大数据发挥作用的许多不同的应用场景。大数据可以被看成一种资产、一种公共资源，或者一种个体身份的表达；[9]它的应用或许可以驱动未来的美国经济，也可以是我们所珍视的自由的一种威胁。大数据可能是所有这些事情。就这为期 3 个月的研究而言，评估组并不旨在对大数据的作用做出全面的解答。大数据技术和支撑它的产业都在不断地创新和变化中。相反，我们的研究集中在探讨个人与数据的搜集、利用方之间那些至关重要的问题。

这份评估报告的范围

今年 1 月 17 日， 奥巴马总统在司法部就改革美国信号情报工作的演讲中，责成他的顾问约翰•波德斯塔(John Podesta)就大数据技术正在或将要对经济、社会与政府行为的范围内发生的影响做出全面评价。为此，波德斯塔召集了商务部部长佩尼•普利茨克(Penny Pritzker)、能源部部长欧内斯特•莫尼兹(Ernest Moniz)、总统科学顾问约翰•霍尔德伦(John Holdren)、总统经济学顾问杰弗里•泽恩斯(Jeffrey Zients)，以及其他高级政府官员。总统科学和技术顾问委员会(President’s Council of Advisors for Science & Technology, PCAST)组织了一项平行报告，以评估基础技术。他们的成果支持了本报告中的许多技术性判断。

这份评估报告本质上被认为是一种对大数据作用范围的调查。在过去的 90 天中，评估小组对学术专家、产业代表、保护个人隐私的倡导者、人权团体、执法者，以及其它政府机构进行了调研。白宫科学与技术政策办公室(White House Office of Science and Technology Policy)与麻省理工大学、纽约大学、加州伯克利大学联合组织了三场大学会议。白宫科学与技术政策办公室也发放了“咨询请求”，就大数据和个人隐私问题寻求公众意见，并得到了超过 70 起回复。此外，白宫网站平台就公众对于大数据的各类使用及不同类型的大数据技术所持的态度，做了一项粗略的调查。在本报告附录中，可以看到工作组各项工作的列表。

大数据有什么不同？

[9] Harvard Professor of Science & Technology Studies Sheila Jasanoff argues that framing the policy implications of big data is difficult precisely because it manifests in multiple contexts that each call up different operative concerns, including big data as property (who owns it); big data as common pool resources (who manages it and on what principles); and big data as identity (it is us ourselves, and thus its management raises constitutional questions about rights).

这章首先界定一下大数据真正新颖和不同的是什么，它得益于总统科学和技术顾问委员会（PCAST）的工作。PCAST 写作了一份平行而独立的报告《大数据与个人隐私：一种技术的视角》。[10]

“3V”：数量（Volume）、类别（Variety）、速度（Velocity）

为了本项研究，评估小组聚焦的是那些数量巨大、类别繁多且高速运行的数据，而传统的数据采集与分析模式已经难以应对了。我们将其特征通俗地称为“3V”。数据采集、存储与处理成本的下降，连同像传感器、相机、地理位置及其它观测技术提供的新的数据来源，意味着我们生活在一个数据采集几乎无处不在的世界中。采集与处理的数据量是空前的。从基于网络的应用、可穿戴技术与先进传感器到监测生命体征、能源使用状况与慢跑者跑步速度的监测仪，由此带来的数据爆炸将推进人们对于高性能计算技术的需求，并推动针对最复杂数据的管理能力的提升。

不仅是数据的数量正在快速增长，它的格式也越发多样，来源也越发广泛。就像总统科学和技术顾问委员会的报告中所说的，有些数据是“天生数字化的”(“born digital”)，意思是说它就是特别创造出来用于计算机和数据处理系统的。这些例子存在于电子邮件、网页浏览，或 GPS 定位之中。其它数据是“天生模拟的”(“born analog”)，这是说它从物理世界中发散出来，但可以不断被转化成数字格式。模拟数据的例子包括手机、相机或摄像设备录制的语音或可视信息，或者还有通过可穿戴设备监测到的身体活动数据，如心率或排汗量。[11]“数据融合”(“data fusion”)能够将分散的数据源整合在一起，随着这种能力的提升，大数据可以带来一些远见卓识。

大数据来源是什么？

数据的来源与格式，连同其类别与复杂程度，都处于持续增长之中。部分数据来源如下：公众网络；社交媒体；移动应用程序；联邦、州和地方记录与数据库；聚集商业交易与公共记录中的个人数据而形成的商业数据库；地理空间数据；各类调查；通过扫描并借助光学字符识别转化而成电子形式的传统离线文献。更多具有上网功能的设备与传感器的出现扩大了从物理实体，包括通过传感器和射频识别(radio-frequency identification, RFID)芯片采集数据的能力。而个人定位数据则来自 GPS 芯片、移动设备蜂窝信号基站的三角测量、无线网络映射，以及个人支付行为。[12]

---

[10] President’s Council of Advisors on Science & Technology, *Big Data and Privacy: A Technological Perspective*, The White House, May 1, 2014.

[11] The distinction between data that is “born analog” and data that is “born digital” is explored at length in the PCAST report, *Big Data and Privacy*, p 18-22.

[12] See, e.g., Kapow Software, *Intelligence by Variety - Where to Find and Access Big Data*, http://www.kapowsoftware.com/resources/infographics/intelligence-by-variety-where-to-find-and-access-big- data.php; James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKin- sey Global Institute, 2011, http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

不仅如此，数据采集与分析的执行速度越来越接近即时时间，这意味对于一个人就其周边环境或生活所做的决定产生即时的影响而言，大数据分析有着越来越大的潜力。高速数据的例子包括记录使用者在线与网页互动活动的点击流数据，即时追踪定位的移动设备获得的 GPS 数据，以及得到广泛分享的社交媒体数据。客户与公司希望通过分析这种数据使其即刻获益的要求越来越高。事实上，如果手机定位应用不能即时准确地确认手机位置，它根本就不会有什么用处，并且，在确保我们的汽车安全运行的计算机系统中，实时操作就至为关键了。

新机会、新挑战

大数据技术能够将大量的数据集以从前不可能的方式分析出有价值的东西。的确，部分大数据所能产生的卓见是研究者过去从未敢想过的。但是，有关大数据的技术能力已然达到了成熟与普及的水平，它要求我们思考如何努力在大数据提供的机遇与这些技术所带来的社会、伦理问题之间做出平衡。

大数据应用的威力与机遇

若使用得当，大数据分析能够提高经济生产率，改善客户与政府服务体验、挫败恐怖分子并且拯救生命。例如：

> 大数据与不断发展的“物联网”使得人们将产业经济与信息经济进行整合成为可能。喷气式发动机和运货卡车现在能够装配许多传感器以监控上百个数据点，并且在需要维护时自动报警。[13]这就使得整个系统更加扁平化，减少了维护成本，并同时增强了安全性。
>
> 医疗保险和医疗补助服务中心(The Centers for Medicare and Medicaid Services, CMS)已经开始在要求支付前用预测分析软件来标示看似报销欺诈的凭据。欺诈预防系统有助于实时甄别高风险医疗保健提供者的欺诈、浪费与滥用行为，它已经终止、阻止或确认了 1.15 亿美元的欺诈性支付，在该程序上头一年花的每 1 美元带来了 3 美元的成本节约。[14]
>
> 在阿富汗战争最激烈的那几年，美国国防高级研究计划局(Defense Advanced Research Projects Agency, DARPA)派遣了数据科学家团队和可视化技术团队到战地。在一个名为 Nexus 7 的计划中，这些团队被直接派进作战部队，用他们的工具帮助指挥官解决特定的作战计划。在其中一个地区，Nexus 7 的工程师将卫星数据和监

---

[13] Salesforce.com, “Collaboration helps GE Aviation bring its best inventions to life,” http://www.salesforce.com/customers/stories/ge.jsp; Armand Gatdula, “Fleet Tracking Devices will be In- stalled in 22,000 UPS Trucks to Cut Costs and Improve Driver Efficiency in 2010,” FieldLogix.com blog, July 20, 2010, http://www.fieldtechnologies.com/gps-tracking-systems-installed-in-ups-trucks-driver-efficiency.

[14] The Patient Protection and Affordable Care Act provides additional resources for fraud prevention. Centers for Medicare and Medicaid Services, “Fraud Prevention Toolkit,” http://www.cms.gov/Outreach-andEducation/Outreach/Partnerships/FraudPreventionToolkit.html.

测仪数据融合，观察交通工具是如何在道路网中流动，这使其更容易定位并摧毁简易爆炸装置。

有一个大数据研究综合了通过监测器采集的数百万个来自新生儿重症监护病房的数据样本，以确定哪些新生儿有可能感染了潜在的致命性传染病。通过分析所有数据（不只是医生在他们的巡视中标记的），该项目能够识别像体温升高、心率加快这样的因素，以此作为有可能发生了某种感染的早期预警信号。这些早期感染信号并不是经验丰富、工作细致的医生通过传统方式能够了解到的。[15]

大数据技术也具有其它极大的前景，它可以用来更好地管理电网间的调配需求、改进能效、为发展中国家提高农业生产力，以及预测传染病的传播等许多其它的应用领域。

大海捞针

现在，计算能力要做到“大海捞针”不仅是可能的，而且依然成为现实。过去，搜索多个大数据集既需要合理组织数据，也需要提出特定的研究问题，依赖选择对的查询以返回正确的结果。大数据分析令数据科学家积聚了海量数据，包括非结构化数据，并且使他们能够找出异常点与数据模式。在这种发现的模式中，为了找到针，你得有个大海；为了获得确定的洞见，你需要一定量的数据。而在其中所涉及的巨大数据量内，就隐含了对于个人隐私的关键性挑战。

例如，Broad 研究院的基因研究人员发现，海量的基因数据集在识别遗传变异对疾病的意义中有着关键的作用。在这个研究中，当样本数量是 3,500 时，和精神分裂症有关的遗传变异无法检测出来；当使用 10,000 个样本时，也只能有细微的识别；但是当样本达到 35,000 时，统计学上的意义便突然显示出来。正如一个研究人员所观察到的一样，“当达到某个拐点时，一切都变了。”[16]对于更多数据的获取，尤其是像基因数据等私人敏感的数据，对于研究者来说将会是一个巨大的挑战，这一情况是由以限制其访问的隐私法为主的各种因素造成的。

大数据之下的数据集群与数据的关系可能会出乎人的预料，但同时也很深刻。同时，即使有海量的数据，大数据分析的结果也不一定完美。图像识别并不能识别这个图像是否重要。相关性仍然不等于因果性。利用大数据技术找到的相关性，或许不能为对结果、行为的预测以及其他个人判断提供恰当的基础。与一般数据一样，在大数据中，解释始终是重要的。

完美个性化的福利和后果

融合大量不同类型的数据并实时处理他们，就有可能在消费者开口之前，就提供给他们正确的信息、产品或者服务。少量数据能够被结合在一起，从而创造出某个人的清晰的行为图谱，进而预测他们的偏好与行为。这些详细的私人档案和个性化的经历在消

[15] IBM, “Smarter Healthcare in Canada: Redefining Value and Success,” July 2012, http://www.ibm.com/smarterplanet/global/files/ca__en_us__health care__ca_brochure.pdf.

[16] Manolis Kellis, “Importance of Access to Large Populations,” *Big Data Privacy Workshop: Advancing the State of the Art in Technology and Practice*, Cambridge, MA, March 3, 2014, http://web.mit.edu/bigdatapriv/ppt/ManolisKellis_PrivacyBigData_CSAIL-WH.pptx.

费者市场上很有用，它能够向确定的一类人推送产品与服务，他们中的一员可能是一位酷爱编织的专业会计，也可能是一位喜欢恐怖电影的家庭主厨。

不幸的是，“完美的个性化”(“perfect personalization”)也会在定价、服务与机会方面造成微妙的或是不明显的歧视。例如，一项研究表明，涉及到黑人常用名（例如，“杰梅因(Jermaine)”）的网络搜索比涉及到白人常用名（例如，“杰弗里(Geoffrey)”）的搜索结果中更容易出现含有“逮捕”意味的广告。这项研究无法确定为什么种族偏见的结果会产生，因为在算法上，广告显示的生成是一个基于多变量的综合决策过程。[17]显然，不同的群体通过不同的信息服务所产生的结果，有可能对个人造成真实的伤害，这种伤害可能发生在他们求职、买房甚至只是简单的搜索信息的时候。

还有一处值得关注：大数据技术能够从意识形态或文化上把人隔离开来，就像泡沫过滤器一样，有效地防止他们接触到一些对他们的偏见与假设构成挑战的信息。[18]一些公司正在搜集并处理大量急剧增长的数据，并煞费苦心地挖掘个人资料与他们的喜好。然而，公众对这些活动的范围与规模的认知是有限的，消费者是很少有机会来控制这些被搜集并且反复使用的数据文件。

模糊与再识别

数据整合等技术在使大数据分析功能日益强大的同时，也为对目前个人隐私的保护带来了严峻挑战。当数据开始连接到个人或设备时，一些隐私保护技术将设法去除这种链接，或者将个人身份信息“模糊化”(“de-identify”)——但是一些同样有效的技术也可以把这些碎片化的链接复原，并重新确定相应的个人或设备信息。同样，整合不同的数据可能会导致一些分析师所说的“马赛克效应”(“mosaic effect”)，即个人身份信息甚至可以从不包括其个人识别码的数据库中得到或者推断出，只要明确包括其爱好等倾向在内的行为图谱即可。

许多技术人员认为，数据的模糊化处理作为保护个人隐私的一种手段，其作用也只是有限的。[19]事实上，对数据进行收集与模糊化处理是基于相关公司不恢复数据的承诺与对应的安保措施的基础上的。对数据进行加密、删除独特标识符、打乱数据使其无法识别个人，或者在其个人资料的控制上给予使用者更多的权限是目前采用的几种技术解决方案。但是有目的的模糊化处理可能使数据丧失其实用性与确保其出处及相应责任的能力。此外，它很难预测再识别技术将如何演变以应对看似匿名的数据。这将导致大量的不确定性，个人该怎样控制他或她的数据？他或她该怎样反对建立在海量数据之上的决策？

数据的保持

---

[17] Latanya Sweeney, “Discrimination in Online Ad Delivery,” 2013, http://dataprivacylab.org/projects/onlineads/1071-1.pdf.

[18] Cynthia Dwork and Deirdre Mulligan, “It's Not Privacy, and It's Not Fair,” 66 *Stan. L. Rev. Online* 35 (2013).

[19] See PCAST report, *Big Data and Privacy*; Harvard Law Petrie-Flom Center, *Online Symposium on the Law*, Ethics & Science of Re-identification Demonstrations, http://blogs.law.harvard.edu/billofhealth/2013/05/13/online-symposium-on-the-law-ethics-science-of-re-identification-demonstrations/.

在过去，对于个人信息的自然控制的保存技术经常可以保证足够的隐私。数据可以被摧毁，对话可以被遗忘，记录可以被消除。但在数字世界，信息可以被获取、拷贝、分享、精确的翻译并且无限期的保存。从前存储大量数据的成本巨大，现在这些数据可以储存在一粒米大小的芯片里，既简单又实惠。结果是数据一旦被创造出来，就可以在许多情况下永恒的有效。此外，电子数据经常涉及到复杂多样的人群，使得个人的控制难以实现。比如，谁是一张照片的拥有者？是照片的拍摄者，还是照片里表现的人，是第一个邮寄它的人，抑或是邮寄它的地址？这些新科技的发展基本改变了一个人和与他/她相关的数据的关系。

数据自由的分享和复制的确比任何时候都要更多。个人、政府、企业、人际网络、同事、上台控制私人数据的其他政党，他们间的特殊责任仍在彼此区分。然而，技术发展的目标是明确的：越来越多的私人数据会产生，并在他人的控制下保存。保证数据的安全是当务之急。出于这个原因，“公众-个人合作社的各种模型”(“models for public-private cooperation”)，例如在 2014 年 2 月成立的行政网络安全结构(Administration’s Cybersecurity Framework)，是确保该基础设施的安全与可调整性的重要部分，而正是这套基础设施，正在为世界上许许多多的数据库提供服务。[20]

证明我们的价值

无论大数据所带来的问题是多么的严重与重要，政府依然会支持相关电子经济的发展并提供免费的数据流来激发大数据的创造力。科技的进步总是会产生如何权衡我们的隐私与社会价值之间的关系的问题。美国在公共领域内，在国会上，在法庭里，均遭受过这个的问题所引发的争议的挑战。而在历史长河之中，无论科技如何变化，我们一直坚定地保护宪法赋予公民的权力。

奥巴马总统上任伊始，政府就号召公众与私营部门善加利用数据的力量，使其提高生产力，改善生活质量，服务大众社会。这也就意味着，这项研究并不仅仅涉及大数据科技的可行性，还包括了大数据是如何可能挑战一般美国人的价值观与美国当下的法律框架。这份报告集中叙述了联邦政府如何在大数据科技改变消费者与公民的世界观的同时，确保我们价值观的延续与法律的与时俱进。

去年，关于隐私方面的公共争议主要集中于政府，尤其是在情报机构如何收集、储存，并应用数据这一方面。这份报告在很大程度上搁置了由信号情报领域的大数据使用而引发的问题，对这一问题的详细处理可以参加总统在 1 月份发布的政策指南。相应地，这份调查报告也同样考虑到了政府通过收集与使用这些大型数据库给公众带来了便利。公众的信任要求政府合理地运行与工作，并要求较之个人，政府必须以一个更严格的标准来收集与使用个人信息。正如奥巴马总统所明确指出的，“对于一个领导者而言，仅仅说‘相信我们，我们不会滥用我们所收集到的数据’是不够的。”[21]

[20] President Barack Obama, *International Strategy for Cyberspace*, The White House, May 2011, http://www.whitehouse.gov/the-press-office/2014/02/12/launch-cybersecurity-framework.

[21] President Barack Obama, Remarks on the Administration’s Review of Signals Intelligence, January 17, 2014, http://www.whitehouse.gov/the-press-office/2014/01/17/remarks-president-review-signals-intelligence.

这份调查报告开阔了我们对于大数据问题的视野，它将大数据的应用范围远远扩大到情报领域之外。这种新的技术不仅仅只是在考察个人隐私，无论这种隐私是被定义成不被干涉，或者是掌握某人身份等其他权利。在这份调查报告中，一些影响最为深远的挑战主要集中在以下方面：大数据的分析有可能导致新型的不公平待遇，尤其是对于弱势群体；也可能产生不透明的决策制定环境，以至于个人自治完全迷失于在一堆无法理解的算法之中。

这些都不是不能解决的问题，但是它们都需要更加深入与严肃的思考。谨记历史学家梅尔文•克兰兹伯格(Melvin Kranzberg)的科技第一定律是非常重要的：“技术既无好坏，亦非中立。”[22]科技可以被用来服务群众，但也可以伤害个人。不管科技多么先进，美国公众都保留着一种力量，即他们能够通过制定政策与法律来管理新技术的使用，进而在某种程度上保护基本的价值观。

大数据正在改变世界。但是它并没有改变美国人对于保护个人隐私、确保公平或是防止歧视的坚定信仰。这份调查报告旨在鼓励使用数据以推动社会进步，特别是在市场与现有的机构并未以其他方式来支持这样的进步的领域，与此同时，我们也需要相应的框架、结构与研究，来帮助保护我们的核心价值观念。

[22] Melvin Kranzberg, “Technology and History: Kranzberg's Laws," 27.3 *Technology and Culture*, (1986) p. 544-560.

2、奥巴马政府在数据开放与隐私保护问题上的发展路径

回顾美国历史，技术与隐私法都处于不断交替发展之中。在营造创新环境、促进经济繁荣的同时，美国一直在全球范围内扮演着保护个人隐私的领导角色。

宪法第四修正案保护了“人民的人身、住宅、文件和财产不受无理搜查和扣押的权利”。对实在空间与有形资产的保护体现了尊重、重视人身安全与个人尊严的意识，公民良好的社会行为与民主社会的正常运行依赖于此二者[23]。 在美国，一个保护隐私利益的法律框架已经建立起来，并覆盖了宪法、联邦、各州等各个层面。“隐私权”不是一个狭隘的概念，而是由一系列概念组成的，它们针对侵害公民隐私权的各种行为，形成了各个样式的有针对性的保护措施。

在美国，数据收集与将数据造福大众有着同样长的历史。宪法第二章第一款授权进行十年一度的人口普查，以分配美国众议院议席。在实践中，人口普查从来没有仅仅只进行简单的人数计算，而是收集一些更为具体的以公共利益为目的的人口统计信息[24]。

自从奥巴马总统执政以来，联邦政府采取了史无前例的政策措施，将更多的它所拥有的数据向公众、公司与创新者开放。从 2009 年开始，奥巴马政府将大量资料库向公众开放，并且将许多数据公布在美国政府的中央信息交换库——Data.gov 网站上。这种将政府的信息数据当作一种资产并加以披露，使其易于获取与使用的做法，换句话说，就是信息的公开化。这大大加强了社会民主程度、开拓了经济发展机会、改善了公众生活质量。

为了挖掘源于公开数据的信息价值，需要开发分析处理信息数据的工具。因此奥巴马政府也在关于信息运算、分析、存储与加密的数据基础科学上，进行了重大投资。

奥巴马政府在做出这些投资的同时，也承认信息的收集、运用以及共享所面临严峻的挑战。联邦研究基金(Federal research)为此划拨经费用于技术开发，以及因运用大规模数据网络而产生的伦理问题研究。美国在隐私保护问题上长期引领历史潮流，奥巴马政府在 2012 年发布了一份关于消费者隐私保护的突破性蓝图，其中包括保护消费者账单

---

23. See, e.g., *City of Ontario v. Quon*, 560 U.S. 746, 755-56 (2010) (“The [Fourth] Amendment guarantees the privacy, dignity, and security of persons against certain arbitrary and invasive acts by officers of the Government.”); *Kyllo v. United States*, 533 U.S. 27, 31 (2001) (“‘At the very core’ of the Fourth Amendment ‘stands the right of a man to retreat into his own home and there be free from unreasonable governmental intrusion.’”); *Olmstead v. United States*, 277 U.S. 438, 478 (1928) (Brandeis, J., dissenting) (“They [the Framers] sought to protect Americans in their beliefs, their thoughts, their emotions and their sensations. They conferred, as against the Government, the right to be let alone—the most comprehensive of rights and the right most valued by civilized men.”).

[24] For example, e.g. the 1790 Census counted white men “over 16” and “under 16” separately to determine military eligibility. United States Census Bureau, “History,”https://www.census.gov/history/www/through_the_decades/index_of_questions/1790_1.html; Margo Ander-son, The American Census: A Social History, (Yale University Press, 1988).

的私密性权利[25]。在 2014 年，奥巴马总统宣布了与私营部门合作发展的网络加密安全构架以加强国家基础设施的安全性[26]。

本章记述了这些倡议措施的交集。这些措施在保障公众与消费者权利的同时，正对以公共利益为目标的数据运用施加着积极影响。

奥巴马政府关于公开数据的举措

公开数据计划

根据政府的公开数据，我们凭借口袋里的智能手机就能知道我们所处的位置。几十年前，联邦政府将气象数据与全球定位系统免费对外开放，企业家们得以发明大量的新工具，提供新型服务，天气预报 APP、汽车导航系统等新发明因而不断涌现。

在过去，政府收集数据的方式主要是由政府机构自己进行收集，而奥巴马政府的一系列公开数据的倡议与决策，使得过去在健康、能源、气候、教育、经济、公共安全与全球发展等领域内难以收集的数据变得易于收集，开启了一个新的富有价值的数据宝库。奥巴马在 2013 年 5 月 9 日签署的第 13642 号总统行政令为联邦数据管理工作提出了新的准则：在保护好隐私安全性与机密性的同时，将数据公开化以及可读写化纳入政府的义务范围[27]。扩大公开数据的影响也同样是总统第二期管理工作规划的核心部分，例如管理和预算办公室(Office of Management and Budget, OMB)就已经建议其下属机构公开更多他们决策所依据的政府信息，因此，相信信息公开将可以惠及更多的人[28]。

公众在 Data.gov 网站上可以找到有关联邦消费者金融保护局(Consumer Financial Protection Bureau, CFPB)受到的所有抗议的信息，这些抗议主要针对于阿肯色州学生贷款的 911 个服务领域。这表明每个人可以利用 Data.gov 网站获得他们所需要的公开信息，而不需要对政府机构和这些机构所推动的工作项目有特别多的了解。感兴趣的软件开发者运用一些简单的工具，就能够自动获得这些数据包的信息。

联邦机构在某种程度上应根据公众的要求优先公布它们的数据以扩大数据的影响面，每一个机构都被要求需通过诸如邮件系统或是在线平台等数据反馈机制来征求它们

[25] President Barack Obama, *Consumer Data Privacy In A Networked World: A Framework For Protecting Privacy And Promoting Innovation In The Global Digital Economy*, The White House, February 2012, http://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

[26] National Institute of Standards & Technology, *Framework for Improving Critical Infrastructure Cybersecurity*, February 12, 2014, http://www.nist.gov/cyberframework/upload/cybersecurity-framework-021214-final.pdf.

[27] President Barack Obama, Making Open and Machine Readable the New Default for Government Information, Executive Order 13642, May 2013, http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government.

[28] Office of Management and Budget, Guidance for Providing and Using Administrative Data for Statistical Purposes, (OMB M-144-06), February 14, 2014, http://www.whitehouse.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf.

应当收集哪些数据。这样一来，任何倡议者、企业家、研究者就能第一时间联系联邦政府，建议哪些数据应该被公开。为了更进一步的形成反馈并促进政府公开信息的有效使用，政府官员一直在积极召开并参与编码马拉松(code-a-thons)、头脑风暴工作坊(Data Jams)、数据开放运动(Datapaloozas)与其他的一些以数据开放为主题的会议[29]。

根据 2013 年 5 月的总统行政令，管理与预算办公室以及科技政策办公室(Office of Science and Technology Policy, OSTP)发布了一个工作框架方案，为各机构管理运用即时更新的信息资源这一新形式财产提供指导，包括了对保护个人隐私、信息可信度的一系列要求[30]。政府机构根据开放程度已将信息资产划分为三个种类：开放性、半开放性、非开放性，并且只能出版发行开放性密级的信息。为了增进透明度，一些机构需将一些还没公布到网上的技术性公开数据纳入他们的外部数据财产清单。

“我的大数据”计划

使政府信息更公开透明、更易被机器读写仅仅是政府信息政策的一个组成部分。1974 年颁布的《隐私权法案》授予了公民可接触一些与其有关的个人信息的权利，公民行使这一权利应该变得更加安全高效，从 2010 年开始，奥巴马政府着手采取了一系列主题为“我的大数据”的倡议与措施，使得美国人可以更安全地获取他们的个人数据，用来更好地处理他们私人领域的申请活动和服务。

“我的大数据”计划具体包括以下部分：

“蓝纽扣”计划：“蓝纽扣”允许消费者安全地获取他们的健康信息，使得他们可以更好地管理他们的健康与经济状况，并与信息提供者交换相关信息。在 2010 年，美国退伍军人事务部(U.S. Department of Veterans Affairs, DVA)开始了“蓝纽扣”计划，退伍老兵可以通过该计划下载他们的健康记录。从那时起，540 万退伍军人利用“蓝纽扣”获取他们的健康信息，超过五百家私人公司允诺向“蓝纽扣”计划的参与者提供更多他们所掌握的健康数据，今天，超过 1.5 亿的美国人能够从健康服务提供商、医药实验室、零售药房供应商与州免疫信息数据库获得他们所需的个人健康数据。

“创建副本”计划：在 2014 年，美国国税局通过一个名为“Get transcript”的工具将纳税人的信息数据加以共享，纳税人可以通过它获得他们自己最近三年的纳税记录。个人纳税者可以借此下载过去的纳税申报单，这使得居民进行抵押、学生贷款、商务贷款等活动与填写纳税表更加便捷。

---

[29] These events have helped federal agencies showcase government data resources being made freely available; collaborate with innovators about how open government data can be used to fuel new products, services, and companies; launch new challenges and incentive prizes designed to spur innovative use of data; and highlight how new uses of open government data are making a tangible impact in American lives and advancing the national interest

[30] Specifically, the Open Data Policy (OMB M-13-13) requires agencies to collect or create information in a way that supports downstream information processing and dissemination; to maintain internal and external data asset inventories; and to clarify information management responsibilities. Agencies must also use ma-chine-readable and open formats, data standards, and common core and extensible metadata.

“绿纽扣”计划：在 2012 年，美国政府与电力行业合作推出了“绿纽扣”计划，这为家庭与企业提供了便捷的途径来获得他们的能源使用信息，并且有利于营造良好的消费者环境与电子化模式。今天，为 5900 万家庭与企业提供服务的 48 家电力供应商通过参与“绿纽扣”计划，帮助他们的消费者节约资源。凭借自身掌握的能源数据，消费者可以选择享受何种私人服务，以更好地管理他们的能源消耗状况来达到理财的目的[31]。

“我的学生数据”计划：教育部将助学金免费申请表与联邦助学情况的一些信息共享，这些信息囊括了借贷、补助金、注册与超额偿付等方面的具体事项，这使得学生与资助人能够上网下载所需信息资源。在这些计划中，信息都是通过“注重使用者体验”、“机器可读写”、“文本信息平面化”的方式实现共享的。

除了为人们提供安全、高效的个人信息，“我的大数据”计划帮助建立了一个有效的个人数据获得性模型，政府也希望将其推广到更多的私人与公众领域。获取个人信息的能力在未来将会变得越来越重要，生活的各个方面都将会逐步卷入到个人、公司与公共组织的信息交换之中。

大数据计划：“数据-知识-行动”

在未来，“大数据”将会成为这个信息交换过程的核心，使得数据转化为知识，并进而转化为行动的过程更加快捷。在 2012 年 3 月 29 日，六个联邦机构加入到“大数据研究和发展计划”(“Big data Research and Development Initiative”)中来，超过两亿的科研经费被用于工具与技术开发以推进对海量数据进行获取、组织与整理并发现有效信息的相关技术发展。

自从“数据-知识-行动”(“Data to Knowledge to Action”)计划实施以来，在 1 亿美金的“XDATA”项目支持下，美国国防部先进项目研究局(Defense Advanced Research Projects Agency, DARPA)创建了一个关于研究出版物与公开化资源软件的“开放目录”，努力发展能够处理分析存在缺陷的、不完整的海量数据的技术[32]。国家卫生研究院(National Institutes of Health, NIH)也拿出 5000 万美金支持开展生物领域的“数据-知识-行动”计划。国家科学基金会(National Science Foundation, NSF)赞助的大数据研究计划，为人类基因组研究节省了 40%的经费。能源部也宣布向“可扩展数据的管理分析及其可视化协会”(“Scalable Data Management, Analysis, and Visualization Institute”)提供一项 2500 万美元的赞助，这家机构所处理的气候数据信息使得季节性台风预报的准确性提高了 25%

[31] Aneesh Chopra, “Green Button: Providing Consumers with Access to Their Energy Data,” *Office of Science and Technology Policy Blog*, January 2012, http://www.whitehouse.gov/blog/2012/01/18/green-button-providing-consumers-access-their-energy-data.

[32] In November 2013, the White House organized a “Data to Knowledge to Action” event that featured dozens of announcements of new public, private, academic and non-profit initiatives. From transforming how research universities prepare students to become data scientists to allowing more citizens and entrepreneurs to access and analyze the huge amounts of space-based data that NASA collects about the Earth, the commitments promise to spur tremendous progress. The Administration is also working to increase the number of data scientists who are actively engaged in solving hard problems in education, health care, sustainability, informed decision-making, and non-profit effectiveness.

以上。还有许多针对大数据的研究支持计划，比如奥巴马总统 2013 年 4 月发布的创新神经技术脑(BRAIN)计划。作为政府大数据计划的组成部分，国家科学基金会为大数据中出现的社会、道德与公共政策问题的相关研究也提供了特别的资金支持。

美国隐私法案和国际隐私法框架

美国《隐私法》的发展

工业革命带来的技术革新浪潮使得社会发生巨大变迁，《隐私法》正是在这一社会背景上发展起来的。隐私权最初由美国学者沃伦(Samuel Warren)和布兰蒂斯(Louis Brandeis)在 1890 年由两人合著的著名法学论文《隐私权》一文中提出，初代可便携照相机的出现直接促成了他们观点的提出，在论文中，他们指出“最近的发明与商业应用将人们的目光吸引到个人隐私权的保护上来，…这项权利应不受侵犯…很多技术发明威胁到了隐私权，‘窃窃私语被公之于众’的预言可能被实现。”[33]提出建立普遍性的隐私保护法的倡议出现在 20 世纪，这一倡议富有预见性，建立了从政府到个人的涵盖各个方面的公民隐私权。[34]

案例法历经了上个世纪的发展，其中关于宪法第四修正案的解释条目随着时间与技术的发展也在不断进行调整。[35]在 1928 年，联邦最高法院受理了“欧姆斯戴德诉美国联邦政府”(*Olmstead v. United States*)一案并宣判在诉讼人屋外设置电话窃听装置并没有违反宪法第四修正案，即使政府以此获得了屋内谈话的内容。[36]但是，欧姆斯戴德案的裁定因为贾斯蒂斯·布兰蒂斯(Justice Brandeis)的抗辩而传播得更广，他写道：“国父们曾经授予公民其隐私不可侵犯的权利以限制政府的行为。”[37]

欧姆斯戴德案的法庭决议一直沿用，直到 1967 年“卡茨诉联邦政府”(*Katz v. United States*)一案才被推翻。法庭认为，联邦调查局(Federal Bureau of Investigation, FBI)在没有调查授权的情况下在公用电话亭外安装监听记录装置，侵害了个人使用公用电话时应有的同时也是符合个人期待的隐私权，即使这个装置没有置于电话亭内部，或是身体以及财物上。此后，主观期待的隐私权得到保护，社会也开始将这视为理所当然。[38]

民事法庭并没有立即将隐私权认定为一个公民向他者提起诉讼的正当理由——也就是律师们常说的“诉因”(“cause of action”)。直到 1934 年的《侵权行为法》中，无正

---

[33] Samuel Warren and Louis Brandeis, “The Right to Privacy,” 4 *Harvard Law Review* 193, 195 (1890).

[34] See William Prosser, “Privacy,” 48 *California Law Review* 383 (1960).

[35] Wayne Lafave, “Search and Seizure: A Treatise On The Fourth Amendment,” §§ 1.1–1.2 (West Publishing, 5th ed. 2011).

[36] *Olmstead v. United States*, 277 U.S. 438 (1928).

[37] Ibid at 478.

[38] *Katz v. United States*, 389 U.S. 347, 361 (1967) (Harlan, J., concurring); see also LaFave, supra note 35 § 2.1(b) (“[L]ower courts attempting to interpret and apply Katz quickly came to rely upon the Harlan elaboration, as ultimately did a majority of the Supreme Court.”).

当理由地严重侵犯个人隐私才被正式确定为可作为起诉的基本出发点。[39]大多数州的法院这才开始将隐私权认定为诉因，这份规定并不是对民事侵权行为的单独一款规定，而是由 4 款复杂的规定组成的：[40]

1. 行为侵犯个人私人空间或私人事务
2. 公开散播个人隐私
3. 为丑化某人信息而将信息公开
4. 为了非个人本人目的而挪用了个人肖像[41]

现在许多批评认为这四款并没有很好地解决隐私问题，市场经济下因商业目的而大范围收集、使用、散播个人信息的现象仍很严重。同时一些人声称，自动化的程序应该能减轻隐私问题给公众带来的忧虑，因为它是使用电脑来进行操作并完成一系列任务，而不是像过去一样由人来操作完成。[42]

信息公平实践原则

随着计算技术的发展与它在政府和私人间的应用更加普及，全球的政策制定者们开始重新审视它与隐私的关系。1973 年，美国卫生、教育与福利部发布了一份题为“录音、计算机与公民权利”(“Records, Computers, and the Rights of Citizens”)的报告。[43]报告分析了“自动化个人数据系统可能导致的不良后果”并建议建立信息使用的保障措施。这些措施，也就是如今广为人知的“公平信息实务法则”(FIPPs)，成为了当今数据保护制度的奠基石。

尽管这些法则在法律与国际公约中都有不同的表现形式，但本质上，“公平信息实务法则”清楚地表达了处理个人信息时的基本保护措施。它规定个人有权知道他人收集了那些关于他的信息，以及这些信息是如何被使用的。进一步说，个人有权拒绝某些信息使用并更正不准确的信息。信息收集组织有义务保证信息的可靠性并保护信息安全。这些法则成为了 1974 年《隐私法》的基础，这一法案规范了联邦政府在个人信息的维护、收集、使用与传播等方面的行为。[44]

19 世纪 70 年代后期，几个其他国家也相继通过了隐私法。[45]1980 年，经济合作及发展组织(OECD)发布了其《关于隐私保护和个人信息跨界流动管理的指导》(“Guidelines Governing the Protection of Privacy and Transborder Flow of Personal Data”)。[46]基于“公平信息实务法则”的经济合作及发展组织指导并提供了关于过去三十年里国家隐私法，特

[39] Restatement (First) Torts §867 (1939).
[40] Prosser, supra note 34 at 389 (1960).
[41] Ibid. See also Restatement (Second) Torts § 652A (1977) (Prosser’s privacy torts incorporated into the Restatement).
[42] Ibid.
[43] See, e.g., K.A. Taipale, “Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data,” V *The Columbia Science and Technology Review*, (2003), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=546782.
[44] Pub. L. 93-579 (codified at 5 U.S.C. §552a).
[45] Organization for Economic Cooperation and Development, *Thirty Years After The OECD Privacy Guidelines*, 2011, p. 17, http://www.oecd.org/sti/ieconomy/49710223.pdf.
[46] Ibid at 27.

别行业隐私法及其实践的信息。1981 年，欧洲委员会通过了《个人信息自动处理中的个人保护公约》("Automatic Processing of Personal Data", Convention 108)，这一公约采用"公平信息实务法则"的手段来凸显欧洲对于隐私权的保护。

尽管有一些关键的不同，但是美国和欧盟国家关于隐私权保护的框架都是基于"公平信息实务法则"。基于隐私权是基本人权这一认识，欧洲国家的保护措施通常包括自上而下的严格法制与对于个人信息的使用的全面限制或是要求信息主体的明确同意。相对的，美国则通常采用在例如医疗保障与信用体系等特别领域实施特别规定来管控特定的风险。这使得美国很少有对于信息使用的全领域普适规则，从而为产品与服务的创新留下空间。但是，但这也为潜在的数据跨领域使用留下了空间

"公平信息实务法则"形成了诸多部门法与国际公约的共同思路。他们被编入 2004 年《亚洲太平洋经济合作组织隐私权法则》("Asia Pacific Economic Cooperation Privacy Principles")，这一文件由亚洲太平洋经济合作组织（简称亚太经合组织或 APEC）成员国签署通过，并构成美国-欧盟与美国-瑞士的安全港框架基础，这一框架将以对于"公平信息实务法则"的一致观点作为沟通美欧法律的基础。[47]

美国特定行业的隐私法

上世纪七八十年代的美国，特别制定的行业法律开始出现并作为以侵权行为为基础的习惯法的补充。这些法律只对特定的数据提供保护。除了少数例外，大多数州与联邦政府都通过了相应法律。[48]

《公平信用报告法案》("Fair Credit Reporting Act", FCRA)最初颁布于 1970 年，这一法案旨在促进消费报告机构所收集的信息的准确性与公平性的同时，推进相关隐私保护。这些信息被用于信用与保险报告、雇员背景调查与租户筛查。这一法案赋予了个人访问与修正个人信息的权利，从而保护了消费者的权利。它要求那些提供消费者报告的公司确保信息的准确与完整；它限制这些信息的使用；它要求这些机构在依据报告进行不利于当事人的措施（例如拒绝贷款）时需尽到告知的义务。

1996 年的《健康保险携带与责任法案》("Health Insurance Portability and Accountability Act" ,HIPAA)规定个人健康信息只能被特定的、法案中明确的主体使用并

[47] The APEC Privacy Principles are associated with the 2004 APEC Privacy Framework and APEC Cross Border Privacy Rules system approved in 2011. See Asia-Pacific Economic Cooperation, "APEC Privacy Principles," 2005, p. 3, http://www.apec.org/Groups/Committee-on-Trade-and-Investment/~/media/Files/Groups/ECSG/05_ecsg_privacyframewk.ashx; *Consumer Data Privacy In A Net-worked World*, p 49-52; export.gov/safeharbor for information on the U.S.-EU and U.S.-Swiss Safe Harbor Frameworks. These enforceable self-certification programs are administered by the U.S. Department of Commerce and were developed in consultation with the European Commission and the Federal Data Pro-tection and Information Commissioner of Switzerland, respectively, to provide a streamlined means for U.S. organizations to comply with EU and Swiss data protection laws.

[48] California, for example, has a right to privacy in the state Constitution. Cal. Const. art. 1 §1.

披露，法案中也包括了用于帮助个人了解并控制其健康信息使用的标准。[49]《健康保险携带与责任法案》(HIPAA)的核心原则是“最小化必须”(“minimum necessary”)原则。[50]国会与卫生部会周期性地升级健康数据的保护。1998 年《儿童在线隐私保护法案》(“Children’s Online Privacy Protection Act”, COPPA)和联邦贸易委员会(Federal Trade Commission, FTC)的法令要求用于 13 岁以下儿童的在线服务或要收集儿童个人信息的在线服务需要获得父母的同意才能进行。在金融领域，《金融服务现代化法案》(“Gramm-Leach-Bliley Act”, GLBA)要求金融机构尊重客户隐私并保护客户非公共信息的安全与机密。在诸如教育，通信，录像带租借与基因信息等其他领域，也有相应法案保障隐私。[51]

消费者隐私权法案

2012年2月， 白宫发布了一篇名为消费者数据隐私权的报告：在全球数字化经济环境下保护隐私权与促进创新的新体系框架（“Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy”）。[52]这种“隐私”蓝图包含四个关键要素：基于信息公平实践原则的消费者隐私权法案；呼吁政府的多方利益相关者在特定的商业环境应用这些原则；对隐私权有效执行与对制定消费者隐私权立法基准的支持；对支持数据跨国流动的国际隐私权制度的承诺。

隐私权蓝图的核心是消费者隐私权利法案，它对消费者保护标准进行明确规定。这些权利是：

个人控制：消费者可以对企业从自己这里收集什么信息，以及如何使用这些信息进行控制

透明：消费者有权简单易懂地获取有关隐私权与安全实践的信息。

相关环境：消费者有权得知企业如何在消费者提供信息的相关环境方面进行收集、使用与披露用户数据

安全：消费者的个人数据必须得到安全与负责任地处理

可修改和准确性：因个人数据的敏感性，以及不准确的数据会对消费者有产生不良后果的风险，消费者有权查阅并更正个人资料

聚焦收集：企业在合理的限度内收集与保存用户数据

---

[49] See U.S. Department of Health and Human Services, Health Information Privacy, “Summary of the HIPAA Privacy Rule,” http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/index.html.

[50] This principle ensures that covered entities make reasonable efforts to use, disclose, and request only the minimum amount of protected health information needed to accomplish the intended purpose of the use, disclosure, or request. See U.S. Department of Health & Human Services, Health Information Privacy, “Minimum Necessary Requirement,” http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/minimumnecessary.html.

[51] They include: The Fair Credit Reporting Act of 1970, the Family Educational Rights and Privacy Act of 1974, the Electronic Communications Privacy Act of 1986, the Computer Fraud and Abuse Act of 1986, the Cable Communications Policy Act of 1984, the Video Privacy Protection Act of 1998, and the Genetic Information Nondiscrimination Act of 2008.

[52] See *Consumer Data Privacy In A Networked World*, p 25.

问责：拥有个人数据的公司有义务采取适当措施，以确保它们符合《消费者隐私权法案》("Consumer Privacy Bill of Rights", CPBR)

《消费者隐私权法案》更加关注消费者而非仅仅是以往用法律术语表达的隐私结构。比如，它根据“易接受性和准确性”("access and accuracy")的原则对权利进行描述，与以往对于“数据的质量和完整性”的公式化表达相比，更易为用户理解。同样的，它确保了公司将会尊重从消费者收集与使用数据的背景环境，从而取代“目的说明”("purpose specification")。

《消费者隐私权法案》还借鉴了公平信息实践的原则以更好地适应我们所生活的网络环境。

与要求企业遵循一系列专一、严格的条令不同，《消费者隐私权法案》建立了一般原则并提供给企业自由决定如何实施这些条令的权力。《消费者隐私权法案》的相关环境原则与其他六大原则相互间产生作用，确保消费者的数据将以符合他们的期望收集并使用。与此同时，相关环境原则允许了企业在信息的使用与“企业-用户”间的关系以及围绕如何收集数据的环境保持一致时，可以开展新的能够使用个人信息的服务。

互联网的复杂性、全球性与持续的发展需要及时的、可发展的创新扶持政策。为了应对这个挑战，《隐私法》的蓝图呼吁所有利益相关者聚集到一起，制定自愿性的，强制性的行为准则，明确规范如何将《消费者隐私权法案》应用到具体的商业环境中。《消费者隐私权法案》是基于广泛的基准原则与具体的行动守则的结合，能够在支持创新的同时保护好消费者。

提升全球互操作性

在其他国家与国际组织开始复核他们的隐私保护框架时，奥巴马政府发布了《消费者隐私权法案》。在 2013 年，经济合作与发展组织升级了自己的隐私权指导方针，这在机制上补充了公平信息实践原则，帮助落实并加强了隐私保护。在 2013 年发布的亚太经合组织跨境隐私规则系统，也在很大程度上效法了经济合作与发展组织的指导方针。[53]欧委会正在审核第 108 号协定（《个人信息自动处理中的个人保护公约》, "Automatic Processing of Personal Data"）。在这些不同的隐私保护框架之间建立桥梁，对确保国际贸易的强劲增长是至关重要的。

欧盟也正在推进其数据保护规则的改革进程。[54]现有的欧盟数据保护指令仅允许欧盟公民数据享有"充分的"隐私保护法案，或向拥有有效的数据安全保护机制的国家（如美国-欧盟安全港协议）流出。在 2014 年 1 月，美国与欧盟开始协商，如何加强安全港协议框架以确保它能继续提供有力的数据保护，并且能使提高其透明度，得到有效执行

[53] Organization for Economic Cooperation and Development, "OECD Work on Privacy," http://www.oecd.org/sti/ieconomy/privacy.htm.

[54] European Commission, "Commission Proposes a Comprehensive Reform of the Data Protection Rules," January 25, 2012, http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm.

与法律上确定性三者成为可能。这些谈判都还在继续，即使像欧洲、美国，也都在磋商这些隐私保护框架将如何适应大数据技术的同时，能够增加计算与存储能力。[55]

在 2014 年 3 月，联邦贸易委员会与欧盟机构的官员连同亚太经合组织一同宣布，欧盟与亚太经合组织将发布共同的计划文件，满足双方在隐私保护框架方面的共同需求。[56]这项筹划工作将帮助那些在欧盟与亚太经合组织地区同时进行贸易的公司解决在两方隐私保护中的认证问题，避免因双方框架不一致或重叠所带来的困难。[57]这样的努力澄清了公司的义务，帮助在全球隐私框架之间建立起相互间的操作性。

结论

目前最普遍的隐私风险依然是涉及“小数据”——定向妥协的内容，例如，以个人银行信息为目的的金融诈骗。这些风险并不涉及到大量的、急速的数据，或是繁多的信息种类，也没有隐含有与大数据有关的复杂化信息。对于“小数据”的隐私保护在美国已通过公平信息实践原则，借由特定的部门法律，强有力的执法部门与全球隐私保护机制得到有效的解决。

隐私权方面的学者，政策制定者与技术专家现在正转向大数据的问题，即如何在“公平信息实务法则”的基础框架下对大数据技术进行有效的管理。这份调查报告的剩余部分就将探索大数据在公共与私营领域的应用，然后将考虑大数据的整体应用对现有隐私保护框架的可能影响。

[55] See Joined Cases C-293/12 and C-594/12, Digital Rights Ireland Ltd. v. Minister for Communications, Marine and Natural Resources, et al. (Apr. 8, 2014) in which the European Court of Justice invalidated the data retention requirements applied to electronic communications on the basis that the scope of the requirements interfered in a “particularly serious manner with the fundamental rights to respect for private life and to the protection of personal data.”

[56] European Commission, Article 29 Data Protection Working Party, Press Release: “Promoting Cooperation on Data Transfer Systems Between Europe and the Asia-Pacific,” March 26, 2013, http://ec.europa.eu/justice/data-protection/article-29/press-material/press-release/art29_press_material/20130326_pr_apec_en.pdf.

[57] Article 29 Data Protection Working Party, Opinion 02/2014 on a referential for requirements for Binding Corporate Rules, February 27, 2014, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp212_en.pdf.

3、公共部门的数据管理

国家维护着和平，并同时保障食物的安全，确保空气与水源的干净。为此，它颁布法律法规来规范经济与政治行为，而大数据技术则有望使这些政府所提供的服务得到全面的提升。

本章将探讨大数据是如何帮助政府更好地履行它在医疗、教育、国土安全以及法律执行方面的职责，并指出大数据带来的挑战。自建国起，关于政府应该做什么、不应该做什么的讨论以及如何在科技日益发展的同时保护公民权利的疑问就不曾间断过。当合众国的奠基者们为这个年轻的国家制定法律与规范时，他们就为如何避免私人空间受到政府不恰当的干预而苦苦思索。而今天，大数据带来的改变或许会让他们大吃一惊：摩尔定律和泽字节正与宪法和权力法案一样，在国会的争论中起到举足轻重的作用。

从核心层面上讲，政府机构对于大数据的使用会加剧我们对政府与个人间权力平衡的担忧。公民信息一旦为了某个特定的目的而被编辑整理，它们就很有可能被用于其他目的，这在国家处于紧急状态时将显得尤为突出。政府在二战期间滥用其数据便是其中最为可耻的例子之一：本是在严格的保密条件下搜集的普查数据被用来确定日裔美国人的居住地并以此来将其扣留至集中营直到战争结束。

因为政府在为公众利益行使权利与权威的同时承担着特殊的责任，所以我们必须慎重考虑大数据在公共部门的使用方式以及对政府部门的数据使用的控制与限制方法。如果任其发展，大数据可能成为政府权力凌驾于公民权力之上的工具。而与此同时，大数据也能被用于进一步加强责任制，并设计一套从本质上更尊重个人隐私与公民权利的政治系统。

大数据与医疗保健服务

数据一直是医疗保健服务中的一部分。在过去的几年中，议会出台了相关法案来鼓励医疗保健服务供应商过渡至使用电子病历，这极大地提高了可供临床医生、研究者与病人使用的数据量。随着《患者保护与平价医疗法案》(“Affordable Care Act”, ACA)的制定，医疗保险的偿付机制正开始从相互分隔、具有潜在不协调性的“按服务收费”(“fee-for-service”)模式转变至基于更佳健康状况的付费模式。总而言之，这些趋势正在帮助形成一个“学习型”医疗保健系统，在此系统内，临床数据将迅速反馈给患者并指导治疗有效进行。

大数据可以确定饮食、运动、预防护理和其他生活方式因素对健康的影响，使得人们不必向医生寻求医疗保健意见。大数据分析能够帮助确定临床治疗、处方药剂以及公共卫生干预对于特定或广泛群体的效果，并对传统研究方式提供参考。从支付角度来看，大数据能够保证给患者提供治疗的医生有优秀的临床记录，同时，治疗的费用根据患者的康复效果而非治疗本身的次数确定。

预测医学的新起是大数据在健康领域的终极运用。这项强大的技术可以同时深入解析一个人的健康状况与遗传信息，使医生更好地预测特定疾病在特定个体上是否可能发生，并预测患者对于特定治疗方式的反应。与此同时，预测医学提出了许多复杂的问题。传统意义上，健康数据的隐私政策都力求在临床信息被分享与分析的同时保护相关患者的个人身份信息。而逐渐地，基于特定群体或人群的数据将在临床症状出现前或出现后不久被用于确定疾病的类型。

但是，预测医学挖掘出的信息所带来的风险将超出单一个体，一旦出现差错，不仅遗传信息提供者本人，他的孩子以及未来的后代等拥有与他相似遗传信息的人都将会受到牵连。因此，将基因组数据与医疗保健数据相连接的生物数据库便成为了个人隐私在医学研究与治疗领域中的无法回避的前沿话话题。

目前的隐私框架在不久前才包括了正在使用的健康信息，这一框架或许不能很好地解决上述发展带来的问题并推动相关研究的进行。运用大数据来改善健康状况需要先进的分析模型来摄取包括生活方式、基因组、医疗与财务数据在内的多种信息。生活方式与健康状况之间的紧密关系意味着个人数据与医疗保健数据之间的界限已经开始模糊。而这些类型的数据却收到不同的、有时甚至是相互冲突的联邦和各州政府的监管，其中包括《健康保险便利和责任法案》(“Health Insurance Portability and Accountability Act”, HIPAA)、《金融服务法现代化法案》(“Gramm-Leach-Bliley Act”, GLBA)、《公平信贷报告法案》(“Fair Credit Reporting Act”, FCRA)与《联邦贸易委员会法案》(“Federal Trade Commission Act”, FTCA)。当数据的来源多种多样时，同时遵守多个法律带来的复杂性随之增加，与此同时，医疗机构还会与不受上述法律约束的许多组织相互勾结，[59] 形成一整套利益链条，各种个人健康信息被一系列企业共享，甚至于州政府会违背消费者对个人医疗数据隐私保护的意愿而出售其相关数据。在此情况下，针对医疗保健领域的大数据部门的设立也就成为了迫切之需，此举同时有望进一步降低行业成本并激发发展潜力。

尽管医学技术不断变化，但健康数据仍然是我们生活中非常私密的部分。在大数据使得较之以往任何时候都更为强大的发现成为可能的同时，重新审视相关信息被所有医疗保健机构共享后的隐私保密方式也显得相当重要。医疗保健行业的领导者已经呼吁构建一个更为广泛的信用框架，使得不同来源、不同隐私保密程度的健康数据得以汇聚。这一框架需要附加《健康保险便利和责任法案》与《反基因歧视法》(“Genetic Information Non-Discrimination Act”, GINDA)中的隐私保护条款，并同时设计标准化数据结构以提高其跨平台适应性。在研究了健康信息技术后，总统科技顾问委员会得出以下结论：国家需要建立统一的数据标准与结构使不同类型的数据记录可以在受到控制的条件下方便访问。[60]

在医疗数据保密框架逐步跟进技术发展的过程中，需要全美医疗保健与保险的供应商之间细致协商，而这份努力，将为未来的国民经济与公民健康的福祉奠定基础。

[59] Latanya Sweeney, a Professor of Government and Technology in Residence at Harvard University, has studied information flows in the health care industry. A graphical map of data flows that depicts information flows outside entities regulated by HIPAA can be found at www.thedatamap.org.

[60] President's Council of Advisors on Science & Technology, *Realizing the Full Potential Of Health Information Technology to Improve Health Care for Americans: The Path Forward*, The White House, December 2010, http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf.

对学习的研究：大数据与教育

如今，上到大学，下至幼儿园，众多科技帮助并提升了学生在课内外的学习过程。获取学习资料、观看授课视频、评价教学活动、进行团队合作、完成家庭作业、参加课程考试，这一切都可以在互联网上完成。

这些基于科技进步的工具与平台给予了学生与教师更多的可能性。仅需数代的革新，这些工具就能提供实时的评估来使学习资料能够按照学生的接受速度来进行演示。不仅如此，教育技术还能扩大受教育人数、增进学生间的互动并使教学内容的持续性反馈成为可能。[61]

除了个性化的教育，新的数据类型的运用使得研究者对于学习行为的研究能力有了质的飞跃。从大规模开放在线课堂等基于科技的学习平台上获取的数据可以被精确跟踪，借助这些数据，我们能够进行对远超传统教育方式的探索，对学生学习轨迹的移动进行更为准确与广泛的研究。具体包括：深入了解学生在学习活动中的接收效果，根据不同的学习目标，选择合适的学习资料，并进一步地运用这些数据帮助那些处于相似状况的学生。目前，教育部正在研究如何运用这些科技，并已开始整合国家教育技术计划下在线教学平台所产生的数据，并计划成立虚拟学习实验室，为进一步的研究提供方法论上的指导。[62]

教育领域的大数据革命同时也带来了一些亟待解决的问题：随着科技日益深入课堂教学，我们如何最好地确保学生的隐私不受侵犯。一方面，各州与本地社区历来都是教育的主要提供者；另一方面，大量的在线学习工具与课程都是由盈利性企业提供。这就导致了在谁有权获得线上教育平台产生的数据及这些数据应当如何被使用的问题上备受争议。对于这类教育记录，《家庭教育权和隐私法案》("Family Educational Rights and Privacy Act", FERPA)、《保护学生权利修正案》("Protection of Pupil Rights Amendment", PPRA)和《儿童在线隐私保护法》("Children's Online Privacy Protection Act", COPPA)中的相关条文在使用过程中都会遇到相应的挑战。

在大数据时代保护儿童的隐私

今天的孩子们是从识字前就接触数字设备的第一代人。在美国，青少年是移动应用与社交平台上的活跃用户。当他们使用这些科技时，关于他们的精确数据，其中一些甚至包含敏感信息，就在网络上被存储与处理。这类数据既包含能够大幅度提升孩子的学习效果并为其开启全新机遇的可能性，但同时，也可能在他们成人时形成一份入侵型的消费者个人信息，或通过其他方式对他们之后的生活产生影响。虽然年轻人一般与成年人一样乃至更加清醒地意识到数据会被商业机构与政府部门使用，但他们的数据还是会经常地受到父母、老师、大学招生人员、军队征兵人员与社会工作者的审查。他们

---

[61] President's Council of Advisors on Science & Technology, *Harnessing Technology for Higher Education*, The White House, December 2013, http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_edit_dec-2013.pdf.

[62] Department of Education, *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*, October 2012, http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf. For information about the National Education technology plan, see www.tech.ed.gov/netp.

中的弱势群体，包括寄养儿童与无家可归的年轻人，他们通常没有得到成年人的指导因而特别容易遭受数据滥用与身份盗窃。在强有力的监视之下，年轻人苦苦寻找保护他们隐私的方法，即使他们无法限制别人对于分享内容本身的获取，许多年轻人仍然尝试着用多种方式将所分享内容的含义变得模糊、晦涩，使得只有特定的对象才能理解其中的意思。[63]

因为年轻人是那么的年轻，他们需要适当的自由来探索与尝试而不至于因一时的疏忽在日后受到挥之不去的侵扰。儿童在线隐私保密法要求网站运营商与移动应用开发者在收集低于13周岁的儿童的个人信息时必须征得其父母或监护人的同意。而现在，我们对于儿童正在遭受什么“伤害”以及怎样的政策框架才能确保他们伴随技术成长是一种促成而不是阻碍都还没能得出一个确定的结论。

与医疗保健一样，青少年在与数字教育平台的交互中表现出的部分数据是极其私密的个人信息，这些数据包括对于特定学习方式的偏好和他本人相对于其他学生的表现。它甚至能够分辨出有学习障碍或注意力无法长时间集中的学生。根据学生在一天内的上线与在线时间，他个人的生活习惯甚至都可以被获知。教育机构应当如何使用这类数据来改善学生的学习机会？对于使用这些平台的，特别是处于基础教育阶段的学生，他们如何能够保证自己的数据是安全的？

为了回答关于这些数据的所有权与恰当使用方式的复杂问题，教育部于2014年2月公布了针对在线教育服务指南。[64]指南明确指出，只有满足《家庭教育权和隐私权法案》与《保护学生权利修正案》中规定的具体要求，学校或学区才可以才能够与第三方机构签订涉及学生数据的协议。随着越来越多的线上学习工具和服务可以为孩子们所使用，州与地区政府也正密切地关注着这些问题。[65]学校与学区以未来合法的教育效益为目的共享受到保护的学生信息，并且在分享的过程中必须对这些信息保持“直接控制”。即使在这新的指导之下，如何在大数据世界中最好地保护学生隐私仍必须是一个持续的议题。

当局正致力于解决这些问题，并通过教育部加以实施，来使得所有的学生在享受大数据在教育与学习上带来的创新效益的同时免于受到其潜在威胁所带来的伤害。[66] 正如

[63] danah boyd, *It's Complicated: The Social Lives of Networked Teens*, (Yale University Press, 2014), www.danah.org/books/ItsComplicated.pdf.

[64] Department of Education, *Protecting Student Privacy While Using Online Educational Services: Requirement and Best Practices*, February 2014, http://ptac.ed.gov/sites/default/files/Student%20Privacy%20and%20Online%20Educational%20Services%20%28February%202014%29.pdf.

[65] For example, California recently passed a law prohibiting online services from gathering information about a minor's activities for marketing purposes, or from displaying certain online advertising to minors. The law further requires online services to delete information that the minor posted on the website or service, a right for which the statute has now been dubbed "the Eraser Law."

[66] The Department of Education is exploring data innovation and use in a wide variety of contexts, including making more educational data available through application programming interfaces. See David Soo, "How can the Department of Education Increase Innovation, Transparency and Access to Data?," *Department of Education Blog*, http://www.ed.gov/blog/2014/04/how-can-the-department-of-education-increase-innovation-transparency-and-access-to-data/.

教育部长阿恩·邓肯(Arne Duncan)所说："学生数据必须是安全且珍贵的，无论它存储在何处，它都不是一种商品。"[67] 这意味着必须确保学生的个人信息与在线活动不受到不恰当的使用，尤其当这些信息是在教育环境下被收集的。

大数据在国土安全部

每一天，有两百万人次乘坐飞机穿梭于美国上空，超过一百万人通过陆路进入国内。验证每一个人的身份并确定他或她是否会构成威胁的工作落到了国土安全部（"Department of Homeland Security", DHS）的头上，后者必须在数秒内处理大量的数据来完成这项职责。这项工作不仅仅是从一个"草垛"中寻找一根"针头"，保护我们居住的家园往往需要从许许多多的"草垛"中找出那根最为关键的"针头"——一个典型的大数据问题。

确保国土安全部有效而合法地使用它搜集的信息是项艰巨的任务。在"911"恐怖袭击之后，国土安全部已经分离出22个独立的政府机构。如今国土安全部中的许多数据库仍分散各地，运行着陈旧的操作系统，而无法整合不同安全级别的信息。除此以外，国土安全部的职责组合方式多样，而这些职责本身却分别由在法律上独立的部门执行。在任何时候，这些信息必须在保护本国公民和进入或定居于本国的外籍公民的隐私与人身自由的条件下被用于授权过的特定目的，而这确保信息被正确使用的任务，就由国土安全部总部的六个办事处执行。

自 2012 年起，来自首席信息官、政策部门和情报部门的代表与个人隐私、公民自由和法律监督方面的官员一同开始运行第一个跨部门大数据应用试点项目——"海王星"（"Neptune"）与"地狱犬"("Cerberus")。[68]"海王星"项目计划将不同来源的未经分类的信息汇聚成一个"数据湖"，并在其内部设置多项安全保障措施，[69] 包括添加多条数据标签的权限与精确到"哪些用户可以基于哪些目的使用哪些数据"的访问规则。所有这些数据都依据一套精细的方案贴上标签。在政府使用的过程中，重点关注是否存在经授权的访问目的，访问任务和必要事项以及使用者本身在访问信息时是否具有合适的工作证明和明细。在这种方式下，通过对数据标签、用户属性与访问前后信息的三重定位就能确定哪条信息在何处被谁访问。

[67] Department of Education, Technology in Education: Privacy and Progress, Remarks of U.S. Secretary of Education Arne Duncan at the Common Sense Media Privacy Zone Conference, February 24, 2014, https://www.ed.gov/news/speeches/technology-education-privacy-and-progress.

[68] Department of Homeland Security, *Privacy Impact Assessment for the Neptune Pilot*, September 2013, http://www.dhs.gov/sites/default/files/publications/privacy-pia-dhs-wide-neptune-09252013.pdf; "Privacy Impact Assessment for the Cerberus Pilot," November 22, 2013,

[69] In the first phase, three databases, from different parts of the agency, are fed into Neptune, where the data is then tagged and sorted. From there, the Department of Homeland Security feeds this tagged data into Cerberus, which operates at the classified level. Here, DHS can compare its unclassified and classified information.

数据管理的一个案例

为了在它的大数据试点项目中确定数据标签的统一标准，国土安全部还将数据系统的所有者与来自个人隐私、公民自由与法律监督办公室的代表聚集到一起。对每一领域的数据，他们都共同制定其数据属性并针对不同的用户群体设置了对应的访问权限。在制定出一整套标签来对信息进行编码后，他们又针对特定的使用限制或一些法律法规下的特殊情况设计附加的规则与保护措施。通过这种添加标签的方式，不仅可以完成高精度的数据访问控制，同时也保留了源数据与其原始搜集目的之间的联系，最终形成了一套对数据从哪里来、到哪里去得到进行全程监控的分类规则。

每个数据库中的字段分为三类：核心身份信息:例如姓名，出生日期和公民身份；扩展身份信息：包括地址、电话号码和电子邮箱；具体的随机数据：衍生于国土安全部中国的电子信息与真人信息的匹配过程。随机数据是最为敏感的数据类型，它可能包括执法人员对被访者的观察记录以及对被访者提出的威胁国土安全的指控。此时详细的规则就能借助数据标签来确定哪些人可以以何种目的访问这些信息。在这两个试点项目中，大多数访问权限的规则设计都需要国土安全部内不同部门的数据使用者间的持续协商才完成。例如，许多数据使用者需要核心身份信息访问权限来获得完成相应的任务所需的特定数据，但由于特定的使用限制，一些规则要求这些数据者提供与所确定的标准更为匹配的信息。

“海王星”与“地狱犬”试点项目同时包含对数据使用者能够采用的搜索方式的重要限制。一个基础检查点可能只需要对一个特定的个人进行数据搜索，因为这个检查点仅需核实基本的身份信息。但是，移民局和海关在侦查案件时，就需要对个人基础身份信息和特征信息进行搜索。而国土安全部的情报分析员就可能需要综合身份、特征与行动趋势信息来分析国家安全的潜在威胁。同时，系统管理员也没有系统内部数据的访问权限，因此数据库的框架设计要允许管理员在不访问任何个人记录的同时也能维持整体系统的正常运作。

在这两个试点项目中，数据库以完全不同于国土安全部自2002年沿袭至今的方式进行了重新组织。在这些大数据的相应举措开始之前，跨越不同部门的数据库搜索较为不便，而将这些数据进行汇总更是难上加难。在过去，数据的使用者与系统管理员一旦登陆成功便很有可能获得所有的访问权限，这些访问有时甚至不会受到跟踪、监测。如今，国土安全局有能力精确设计根据任务需求而定的访问权限。最重要的是，在这些先进的数据库中，通过人为地制定标签和数据的重新组织，国土安全局能够在强有力的法律监督下实施新型的事件预测与异常分析。

国家安全局如此细致地设计数据处理系统并不是偶然的结果。国家安全局内部专门设有独立的隐私办公室与公民权利与自由办公室，每一个办公室都配有专业人员来帮助研究处理这一复杂领域的相关事务。[70] 每一个试点项目在实施前都会向社会公众公布详细的隐私影响评估报告。国家安全局同时向公众提供各项目的介绍并接受大众对于项目具体措施的问询。经过这一系列的工作，隐私与公民自由办公室的官员不仅批准了这两个试点项目的实施，还同时通过了服务于未来功能扩展的配置建设。所有这一切都将有

[70] For more information, see the Department of Homeland Security's Privacy Office website, http://www.dhs.gov/privacy, and Office for Civil Rights and Civil Liberties, http://www.dhs.gov/office-civilrights-and-civil-liberties.

助于推动国土安全部的计划能在确保隐私和公民自由自始至终得到密切关注的同时得到进一步的发展。

在执法过程中贯彻隐私价值观

在法律执行方面，大数据是一个强有力的工具。近日，美国国防部高级研究计划局(“Defense Advanced Research Projects Agency”, DARPA)的“记忆延伸”(“Memex”)计划下开发的高级网络工具已帮助联邦执法部门在查明美国人口贩运网络的工作上取得实质性进展。这些工具不仅整理众所周知的“表层网络”(“surface web”)信息，还同时收集“深度网络”(“deep web”)下那些不被常用搜索引擎索引的公开信息。通过对网络站点的大范围搜索，这一工具能够发现原本难以获得或需要大量时间才能发现的信息。执法部门现有的数据能够锁定可能的人口贩卖团伙，进而协助干警确认性贩卖与其他犯罪活动的联系。目前，该工具已帮助侦察出一个起源于亚洲并蔓延至美国多个城市的人口贩卖网络。这是大数据能用以帮助世界上部分最脆弱人群的有力佐证。

大数据技术为执法部门等安全保障机构提供了有效的工具，但大数据技术的合理运用也是个难题。整合多种数据源能够让我们更全面地了解嫌疑人在作案期间的相关活动，但同时，在仅有极少甚至没有任何人工干预的情况下，行为模式分析可以揭示犯罪组织的组成或用以预测未来可能发生的犯罪行为。广泛收集数据能够帮助抓捕罪犯，但同时也可能会读取到非调查对象的详细个人信息。具体到法律执行过程中，我们必须谨慎行事，确保大数据技术在用于保护社会安全、公正执法的同时，兼顾对公民自由与公民的合法隐私权益的保障。

大数据将自然而然地以各种形式合理应用于国家安全层面。一套汇集全球数据的强大智能系统将用以侦查恐怖组织网络，提供攻击预警，以及阻止大规模杀伤性武器的扩散，而与此同时，它将运作于各种法律机构的授权和监督之下，较之协助调配警力至高危犯罪区域的执法系统，将提供更多的隐私保护。即使应用于不同领域，在整个执法和情报背景下，维护公民的隐私和权利始终是一致的。隐私保护和执法官员必须确保在系统运行的任何情况下，减少信息持有的最小化原则和控制访问的信息标签技术均能够得以保证实施。

新的工具与新的挑战

新技术的使用已导致过重要的宪法判决，在执法领域更是如此。[71] 正如大法官阿托利在 2013 年最高法院关于警方在未收到法院命令的情况下擅自在嫌犯的汽车上安装全球定位跟踪器一案的判决中所指出的：“虽然几乎无法想象，但相似的情况在 18 世纪后期也曾发生过。”“你能够想象一位警官为了掌握马车夫的行程而将自己藏于马车内

[71] Most jurisprudence to date does not consider in their entirety big data technologies by the definition used in this report, but rather many of the advanced technologies, such as GPS trackers, that now play a crucial role in big data applications.

吗？”[7]阿托利进一步指出，“类似的事或许在 1791 年就已经发生过，只不过现在的‘马车’更大而‘警察’更小了。”[73]

这个“小型警察”(“tiny constable”)却有着巨大的影响。无论是全球定位系统的跟踪、闭路电视的监察还是肉眼无法识别的传感器，这些监视无处不在，这也使得对隐私的合理预期以及对执法技术的约束和合理运用的诉求越来越普遍。

近几十年来，监控器材的体积越来越小，监控成本也越来越低。得益于此，全美超过 70 座城市都配置了音频传感器来确定枪声发出的位置以便快速派遣警员到可能的案发现场。[74]不仅如此，随着数据访问速度的加快与存储成本的降低，各地警方也能够在全市范围内实时获取车牌与其他车辆信息，并加以存储以备后期使用。[75]

这些技术带来的便利是巨大的。从寻找失踪人员到开展复杂的搜捕行动，先进的监察技术使得联邦政府及各州、各地区政府能够对犯罪行为进行更加迅速与有效的反馈。同时，这也使得正义在网络犯罪的侦察中得到更好地贯彻：执法部门能够通过这些技术及时获取电子证据来将这些高科技罪犯绳之以法。

除了监控，大数据带来的预测技术为执法部门提升了更好地准备、干预或彻底阻止某些犯罪行为的潜力。以洛杉矶与孟菲斯警方所使用的程序为例，分析软件能够实施预测进而形成一个地区的“犯罪热点图”(“hotspots”)。[76]许多城市通过加强相应“热点”地区的巡逻警力，有效降低了辖区内财产犯罪的数量。

富有争议的是，预测分析技术如今已能被用于对某一独立个体的犯罪倾向分析。针对一场帮派火拼，芝加哥警方尝试将犯罪预测的侧重点由地理信息转移至身份信息。通过将警方数据与其他数据进行整合，同时加以社会网络分析，芝加哥警方根据与暴力犯罪的相关因素形成了一份涉及约400人的名单。据此，警方能够在已有的指控与犯罪记录证据之外，对一些特定个人提高防范。[78]

预测分析技术也被刑事司法领域的其他方面。在费城，警方正运用软件预测哪些假释犯在出狱后再次犯罪的可能性较大进而需要加强监督。该软件使用二十几类变量，包括年龄、犯罪史及地理位置等。

与此同时，这些新技术应当如何及何时应用的问题引发了巨大的争议。[80]一方面，这些技术能够帮助执法等其他公共资源得到更加精确地分配并同时减少犯罪的发生；另一方面，《宪法》与《权利法案》所赋予我们的相应权利必须得到捍卫。

---

[72] *United States v. Jones*, 132 S.Ct. 945, 958 (2012) (Alito, J., concurring).

[73] Ibid at n.3.

[74] Over 70 cities in the U.S. use gunshot detection technology developed and provided by SST Solutions called ShotSpotter. For more information, please visit www.shotspotter.com.

[75] International Association of Chiefs of Police, *Privacy Impact Assessment Report for the Utilization of License Plate Readers*, September 2009, http://www.theiacp.org/Portals/0/pdfs/LPR_Privacy_Impact_Assessment.pdf.

[76] The National Institute of Justice, the Department of Justice's research, development, and evaluation agency, provides detailed information on the use of predictive policing at law enforcement agencies. For more information, visit www.nij.gov/topics/law-enforcement/strategies/predictive-policing.

[78] The application of this particular predictive policing technology emerged out of a series of grants issued by the National Institute of Justice the Chicago Police Department, most recently involving Miles Wernick as technical investigator. For more information, see http://www.nij.gov/topics/lawenforcement/strategies/predictive-policing/Pages/research.aspx.

[80] Controversial aspects of the Chicago pilot's methodology are captured by in Jay Stanley, "Chicago Police 'Heat List' Renews Old Fears About Government Flagging and

警察部门通过运用一系列数据与算法来预测犯罪几率并在犯罪活动开始之前调配警力，这有着重大意义。它要求我们对宪法在监察方面定义的行为——“特别怀疑”(“individualized suspicion”)的含义进行仔细审视。长期以来，我们都信奉着“一个人的行为、运用与个人背景都受到执法部门的监控的局面，会对自由言论与结社的权利产生‘寒蝉效应’”的合理观点。下一节中我们将探讨大数据技术对法律中的哪些特定领域带来了改变。

大数据技术对隐私法的启示

第三方数据的访问权限

至今为止，个人文件与记录经历了由以纸为载体存放在家中，到以电子文档的形式存储于家用电脑的硬盘内，再到以多种文件格式同时储存在本地与可供多台终端访问的云端的三个阶段。随着远程控制技术与云端储存技术在个人电脑与文件管理领域普及，我们必须采取相应的措施以保证法律跟上技术发展的脚步。

什么是值得保护的，我们对于这一问题的思考大部分是建立在这一个体是否期望将这一行为归为自身隐私范畴的基础上的。正如大法官波特·斯图尔特(Potter Stewart)在1967年的卡茨案中所指出的：“第四修正案所保护的是人，而不是地方。当一个人的行为是其自身故意暴露给公众的，即使这一行为是发生在他自己的家里或办公室内，该行为都不适用于宪法第四修正案……但若是他将某一行为视为自己的隐私，即使处于公开场合，这一行为也可能会受到宪法保护。”[82]

两年后，最高法院的判决进一步阐述了宪法第四修正案中对于分享给第三方机构的信息的规定。在 1976 年美国诉米勒案(*United States v. Miller*)中，法院裁定宪法第四修正案并未禁止政府获取“那些提供给第三方机构并由其转交给政府的信息，即使这些信息是在其本身仅被用于特定目的的，同时要求第三方机构不得将这些数据透露给他人的条件下提供给该机构的”。[83]三年后，在史密斯诉马里兰州案(*Smith v. Maryland*)中，史密斯因其自愿向电话公司提供的拨号信息没有得到相应的与基于个人隐私的合理预期相符的保护而进行申述。最高法院重申：“它依然…认为一个人对其自愿转交与第三方机构的信息没有基于隐私的合法预期”。

米勒与史密斯案(*Miller and Smith*)是经常被引用来说明最高法院所具有的根本性的“第三方主义”(“third-party doctrine”)的案例。几十年来，这一学说始终认定，当个人自愿向诸如电话公司、银行甚至其他个人等第三方提供信息时，政府能够在不触及宪法第四修正案给予的个人权利的前提下，无需个人认可地从这些第三方机构中获取信息。执法部门依然根据“第三方主义”来获取在刑事案件侦破与国家安全调查中发挥

Tagging,” *American Civil Liberties Union*, February 2014, https://www.aclu.org/blog/technology-and-liberty/chicago-police-heat-list-renews-old-fears-about-government-flagging-and; Whet Moser, The Small Social Networks at the Heart of Chicago Violence,” *Chicago Magazine*, December 9, 2013, http://www.chicagomag.com/city-life/December-2013/The-Small-Social-Networks-at-the-Heart-of-Chicago-Violence.

[82] *Katz v. United States*, 389 U.S. 347, 351-52 (1967).

[83] *United States v. Miller*, 425 U.S. 435, 443 (1976).

重要作用的信息来保证全国居民的安全；而联邦法院则在之后的判决中将该主义广泛运用在实体信息与电子信息之中。

在此背景之下，国会与各州议会颁布法规，为某些类型的信息提供附加的保障。1974年颁布了用以保护联邦政府所持有的个人信息的《隐私法》(“Privacy Act”)；1986年通过了用以保护电子通讯记录（对象之一）的《电子通信隐私法》(“Electronic Communications Privacy Act”, ECPA)和保护拨号信息（对象之一）的《禁止监视记录器与/或追踪设置法案》(“Pen/Trap Act”)。这些法案填补了宪法第四修正案在保护第三方机构所持有的信息的不足，为相关隐私信息提供了法律保护。

随着技术的进步，特别是人际交往过程中电子记录的成倍增长，一些评论家呼吁对“第三方主义”进行复审。[85]2 2010年，在美国沃夏诉案(*United States v.Warshak*)的六审判决中，法院判决电子邮件“类似于信件与电话”并属于基于个人隐私的合理预期的范畴，政府不能要求商业网络服务供应商在不事先通知用户预期结果并获得许可的情况下提供用户的电子邮件信息。在近期的最高法院的判决中，大法官索托马约尔(Sotomayor)则表示，“在这个人们将大量的个人信息存储在第三方机构来完成日常工作的电子时代”，当前对于第三方机构的信息流出的监管做法是“不合适”的。[87]

虽然我们未曾获知自沃夏诉美国案后是否有任何法院曾通过下述判决：除非得到用户的同意，其个人的电子通讯信息不得在未经授权的情况下被访问，但是现在“第三方主义”正继续适用于此类通信的元数据，并经调整后应用于基站地址信息与无线网络信号信息上。[88]

这份关于大数据与个人隐私的调查报告还对个人隐私、市场信心与在政府强行迫使第三方机构透露电子数据中涉及的相关法律等方面进行了深入研究。为了跟上科技发展的步伐，我们需要不断检验自身的法律与政策，并考虑如何在隐私保护方面将存储在诸如云端等远程存储器中的内容数据与存储在家庭或办公室的硬盘上的内容数据

[85] Fred Cate and C. Ben Dutton, “Comments to the 60-Day Cybersecurity Review,” *Center for Applied Cybersecurity Research*, March 2009, http://www.whitehouse.gov/files/documents/cyber/Center%20for%20Applied%20Cybersecurity%20Research%20-%20Cybersecurity%20Comments.Cate.pdf; Randy Reitman, “Deep Dive: Updating the Electronic Communications Privacy Act,” *Electronic Frontier Foundation*, December 2012, https://www.eff.org/deeplinks/2012/12/deep-dive-updating-electronic-communications-privacy-act.

[87] This assertion was not part of the Supreme Court’s holding, but emphasizes the emerging discussion of third-party doctrine. *United States v. Jones*, 132 S.Ct. 945, 957 (2012) (Sotomayor, J., concurring).

[88] The doctrine has been adapted and applied to cell-site location information multiple times, most recently by the Fifth Circuit in *In re Application of the United States for Historical Cell Site Data*, 724 F.3d 600 (5th Cir. 2013) (finding cell site data may be obtained without a probable cause warrant); *United States v. Norris*, No. 2:11-CR-00188-KJM, 2013 WL 4737197 (E.D. Cal. Sept. 3, 2013) (finding defendant who hacked a private wireless network had no reasonable expectation of privacy in his transmissions over that network). Moreover, leading commentators have argued for the continuing vitality of the third-party doctrine in the modern era, including Professor Orin Kerr in Orin S. Kerr, “The Case for the Third-Party Doctrine,” 107 *Michigan Law Review* 561 (2009), and Orin S. Kerr, “Defending the Third-Party Doctrine: A Response to Epstein and Murphy,” 24 *Berkeley Technology Law Journal* 1229 (2009). See also *United States v. Perrine*, 518 F.3d 1196, 1204 (10th Cir. 2008); *United States v. Forrester*, 512 F.3d 500, 510 (9th Cir. 2008).

相联系。在过去的30年里，短信、电子邮件与其他交流平台已经成为了私人通信的重要手段，而这些信息大多储存在远程存储器中。

数据与元数据

从购买商品到上传数码照片，普通的美国人一天内会与企业进行数次形式各异的交易。这些交易会产生记录，而其中像药店票据一类记录还会涉及个人的敏感信息。在日常的行为中，用户也进行着大量的“数字化排放”(“digital exhaust”)或产生许多跟踪信息，留下更多碎片化的信息，手机传输时的地理坐标与服务器日志中的互联网通讯协议地址就是两个很好的例子。借助更加强大的分析工具，部分细小且彼此间毫无关系的数据碎片也能得到识别，进一步加强了第三方机构所搜集与持有的数据被加以合并与分析来揭示更多个人信息的可能性。如何使这些材料与从中挖掘出的信息得到应有的保护是当下一个紧迫的难题。

除此以外，对于关于某些类型的数据——特别是“元数据”(“metadata”)或是较通讯及文档内容之外的其自身的传输记录——他们是否应该收到较现在更为周密的隐私保护也是一个同样重要的问题。“元数据”是用来描述数据自身特征的。其中的一个经典例子来自电信行业。过去，电话拨出与挂断信号，作为一种元数据，较通话内容本身，被认为透露了较少的信息，而被给予了不同的隐私保护等级。如今，随着大数据的到来，无论是服务商的合约，还是机构出台的政策都不会对各种类型的数据进行如此简单的划分。

虽然专家们在元数据的隐私保护问题上存在着分歧，但是当下元数据的敏感性远超昔日的观点已得到充分的认可，并进而推动了对有关政策的重新审视。在情报信息方面，总统已经指示他的情报顾问委员会考虑这个问题，并针对时下关于元数据与隐私问题的设想做出长期可行性规划。这篇调查报告建议政府应该将该问题的范围扩大至情报信息以外，根据数据与信息对个体身份与行为的揭露程度制定相应的法律并实施其他保护。

商业数据服务的政府使用

私营部门强大的分析与数据挖掘技术不仅仅适用于商业领域。从土地管理到行政优化，各州、地方与联邦机构购买了大量私人数据库的访问权限以用于合法的公共服务。这些服务的数据来源有时是不对外披露甚至是作为商业专利而受到保护的。一些法律学者与隐私保护倡导者已经对包括执法与情报机构在内的政府部门使用商业数据服务产品的现象表示担忧。[89]

---

[89] See Robert Gellman and Pam Dixon, “Data Brokers and the Federal Government: A New Front in the Battle for Privacy Opens,” *World Privacy Forum Report*, Oct. 30, 2013; Chris Hoofnagle, “Big Brother's Little Helpers: How Choicepoint and Other Commercial Data Brokers Collect, Process, and Package Your Data for Law Enforcement,” 29 *North Carolina Journal of International Law and Commercial Regulation* 595 (2003); Jon Michaels, “All the President's Spies: Private-Public Intelligence Partnerships in the War on Terror,” 96 *California Law Review* 901 (2008).

而财政部一直致力于实施一项计划，具体包括减少支付中出现的对象错误、金额错误与缺少相应书面材料等现象，期望通过这些举措防止联邦支出过程中出现铺张浪费与徇私舞弊的行为。为了向联邦机构提供包括检查多个数据库、确定不合格的收款人与防止欺诈或错误等功能在内的一站式服务，财政部开发了“不付款”门户网站（“Do Not Pay”portal）。尽管目前其所提供的数据库均为政府数据库，财政部预期未来商用数据库也可通过该网站获得访问。

为了协助财政部，国家行政管理与预算局(Office of Management and Budget, OMB)已发行主要指南以确保个人隐私在该项目中得到充分保护。[90]该指南指出商业数据源“也可能带来或增加新的个人隐私风险，诸如数据库提供不准确或过期信息”。该指南进一步要求所有进入“不付款”门户网站的数据库在进行审查与批准前需进行为期30天的公示以征求大众意见。同时，所有数据库都必须是该项目中的不可或缺的相关资料，并且要足够精确以确保数据库涉及的每个人都受到公平对待，同时还严禁涉及任何描述宪法第一修正案保护下的权利行使方式的信息，除非此类数据的使用是得到有关法规的明确授权的。

鉴于通过商业信息源可以获取的个人敏感信息的范围越来越大，这项指南是确保政府决策中使用的私营部门数据的隐私得到保护的重要一步。类似的指南需要普及到更广泛的机构与项目中，使得不论信息的来源如何，民众都能从政府处得到应有的保护。

内部威胁与持续性评估

2013 年华盛顿海军工厂内部发生一起枪击案，尽管一系列的逮捕行动使当局对与处于特殊公职岗位的员工评定更为频繁，作为一名内部合约工人的嫌犯还是通过了秘密的安全调查。[91]这是包括切尔西·曼宁(Chelsea Manning)维基解密泄密事件、纳达尔·马里克·哈桑(Nidal Hasan)的福特胡德军事基地枪击事件与美国情报史上最严重的泄密事件——爱德华·斯诺登(Edward Snowden)泄露国家安全局(National Security Agency, NSA)内部文件事件在内的一系列国家安全检查的执行者的叛逃或暴力行为中的最新一例。

联邦政府的雇员与承包商都需根据其风险等级、职位敏感程度与访问敏感设施或系统的权限接受相应的不同级别的调查。目前，“绝密”(“top secret”)等级的雇员与承包商每五年需要重新接受调查，而“秘密”(“secret”)等级的调查周期则为十年。相关机构在此之外无法及时获取雇员新的或值得注意的信息。

试点方案的实施结果已经证明，综合适当的官方、商业数据库与社交媒体的自动审核机制来确定对象暴力或违规几率是行之有效的，这些“负面信息”(“derogatory information”)可能会导致相关部门对一位在职人员继续从事敏感职位的能力产生质疑。以国防部为例，近日其进行了一次名为“自动连续评估系统”(“Automated Continuous

[90] Office of Management and Budget memorandum M-13-20, *Protecting Privacy while Reducing Improper Payments with the Do Not Pay Initiative* (Aug. 13, 2013), http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-20.pdf.

[91] Department of Defense, *Security From Within: A Report of the Independent Review of the Washington Navy Yard Shooting*, November 2013, http://www.defense.gov/pubs/Independent-Review-of-the-WNY-Shooting-14-Nov-2013.pdf; Department of Defense, Under Secretary of Defense for Intelligence, *Internal Review of the Washington Navy Yard Shooting: A Report to the Secretary of Defense*, November, 2013, http://www.defense.gov/pubs/DoD-Internal-Review-of-the-WNY-Shooting-20-Nov-2013.pdf.

Evaluation System”)的试点项目。此次试点项目调查了包括陆军服役人员、文职雇员与承包商在内的3370名人员，结果其中21.7%的人员被认定在自上次调查以来产生了未曾报告过的负面信息。其中99名人员在此次试点项目中被发现陷入严重的经济危机、家庭暴力、吸毒或卖淫的指控并最终对他们进行了临时或永久性撤职。[92]

当局在近日公布的一份关于人员的岗位合适性与安全性的调查报告中，呼吁在联邦政府内部扩大持续适应性能力评估的使用范围。[93]尽管该项目中涉及的具体信息类型，尤其是来源于社交媒体的信息仍待确定，当局的调查报告中还是建议将以上措施在各部门及各安全级别中进行普及。

这些改革将会设计一套全新的流程来确保安全调查能够提高我们的安全程度。随着当局在联邦机构中大力推广跨部门持续性评估，雇员与承包商的隐私也需要得到细致的考虑。员工在此过程中必须拥有反驳或纠正调查结果中错误信息的能力来拒绝或撤销安全调查的结果。我们必须确保基于大数据技术的持续评价体系能够以保护社会大众安全与确保社会大众的服务者——联邦政府中的员工的公民自由与隐私权利的方式进行。

结论

当我们被大数据技术在公共部门引起的种种令人烦恼的问题纠缠不休时，我们很容易忽视这些技术在改善公共服务、促进经济增长和改善社区健康与安全方面带来的巨大机遇。这些实实在在的机会必须被置于大数据有关讨论的核心位置。

大数据技术拥有巨大的力量，它能使遍及整个政府行为框架的服务条款更加高效，它能够侦测徇私舞弊与铺张浪费的行为。不仅如此，大数据技术还能创造全新的价值形态。新型高精度气候模式数据源能够为气候变化带来有意义的科学发现；同时，了解能源与自然资源的使用方式有助于提高产出、降低能耗。数据的移动、存储与分析都变得越发高效与有力。以能源部为例，其内部正在开发新型计算机内存并设计超级计算机框架，此举有望产生全新的分析工具，使得大数据革命的突进更为迅猛。

几乎没有任何一个政府部门不是为了能更好地服务普通民众而设立的。大数据革命将不仅仅在已经涉及相关科技的部门与机构进行，它将席卷整个政府。那些以往没有大范围使用高级数据分析的部门与机构或许最有可能利用大数据技术为普通民众提供更好的服务。

大数据能量的释放将不仅仅停留于联邦政府，它将同时用于各州与自治市镇的机构革新。一些城镇已然成为一批最具创造力的大数据使用者来提供更优质的服务。相关联邦机构与计划为城镇、乡村提供财务与技术援助来完成市政技术革新以效仿纽约数据分析办公室与芝加哥智能数据的成功经验。

[92] Ibid.

[93] Performance Accountability Council, *Suitability and Security Processes Review, Report to the President*, February 2014, http://www.whitehouse.gov/sites/default/files/omb/reports/suitability-and-security-process-review-report.pdf.

让大数据技术为社会公众造福还需要供不应求的技术人才。一份近期关于公共与非盈利部门所具备的吸引与留住技术人才的能力的报告为我们敲响了警钟。[94] 尽管许多年轻技术人员深切关注公共服务并乐于在政府部门工作，但相比之下，私营部门给出的机会是如此地富有吸引力以至于这些技术人员倾向于将他们的大数据才能应用于消费市场而不是公共部门。这意味着作为科技方面的另一投资，联邦政府必须营造一个更富吸引力的工作文化的氛围并且移除将部分专家拒之门外的聘用屏障，正是这些专家的创造力与技术上的想象力，将充分激发大数据在政府部门的发展潜力。

[94] Ford and MacArthur Foundation, *A Future or Failure?: The Flow of Technology Talent into Government and Civil Society*, December 2013, http://www.fordfoundation.org/pdfs/news/afutureoffailure.pdf.

4、私营部门的数据管理

大数据是事关全球经济的重大技术革新。在接下来两年中，大数据科技与它的服务市场将会继续保持增长势头。[95] 本章将探讨大数据是如何让产品和服务更好地服务于消费者与企业，并提出一些由于一些消费者质疑他们的个人信息是如何被采集、分析与运用而带来的挑战。

奥巴马政府利用大数据来激发创新、生产力并保护个人隐私的传统价值观，以此来巩固美国的领导地位。然而，在最近的持续地采集、转移与重新设计大数据下的信息结构的同时也引发了关于个人对其私人数据的掌控问题，以及利用大数据确定易受伤害的民众时所产生的风险的重要问题。在大数据将成为经济增长与创新的有力驱动力的同时，也存在着一种令人不安的消费者与生产商之间的信息不对称的可能性。

大数据对消费者与企业的益处

大数据为消费者与企业都在创造着价值。无论是大型企业还是小型企业，大数据的访问以及处理数据的工具在都进一步普及，它带来的益处可以在各个领域都有所体会。在大企业，在投资大数据科技方面有几个驱动因素：分析运营与交易数据的能力；洞察客户线上消费行为，给市场带来新的极其复杂的产品；对组织中的机器与设备进行更加深入的了解。

科技公司利用大数据来分析上百万的声音样本，以提供更精确更可靠的语音接口；银行利用大数据技术来提升诈骗侦测能力；医疗提供者借助更精确的数据以改善对患者的治疗。大数据被生产商用来提升机器保修管理与设备监控，同时使物流最优化。零售商同时通过线上与线下的渠道与客户进行各种各样的互动，来为后者提供量身打造的建议与最优的价格。[96]

对消费者来说，大数据为影响人们日常生活的产品与服务的增加提供动力，这让网络安全专家得以保护这个体系并使之安全处理大量的网络与数据应用（从信用卡读卡机到数据应用），同时用它指明异常与威胁之处。[97] 它也使将近 29%的美国人，包括一些没有银行账户，或正在申请银行账户的人通过使用一些更广泛的非传统信息的方式建立信用资格并获得信用额度支持的资格，如租金、水电费、移动用户、保险、儿童保险与学费。[98]

---

[95] Dan Vesset and Henry Morris, *Unlocking the Business Value of Big Data: Infosys BigDataEdge*, IDC, 2013, http://www.infosys.com/bigdataedge/resources/Documents/unlocking-business-value.pdf

[96] Ibid.

[97] Centre for Information Policy Leadership, *Big Data and Analytics: Seeking Foundations for Effective Privacy Guidance*, February 2013, p. 3-4, http://www.hunton.com/files/Uploads/Documents/News_files/Big_Data_and_Analytics_February_2013.pdf.

[98] FDIC, *2011 FDIC National Survey of Unbanked and Underbanked Households*, 2012, http://www.fdic.gov/householdsurvey/2012_unbankedreport_execsumm.pdf.

这些新技术嵌入在网络中，高精度传感器等监测设备现在可以检测声音、速度、温度，甚至一氧化碳水平，并从停车场、学校与公共道路上提取数据，以此来提高能源效率与公共安全。车辆记录以及行驶范围与使用状况的相关报告将为先进的交通系统及其安全性的提升铺平道路。家电用品现在可以告诉我们什么时候可以在千里之外减弱我们屋里的灯光。政策必须在一定程度上适应改变，随着网络技术的到来，联邦贸易委员会(Federal Trade Commission, FTC)已经开始制定由物联网带来的政策框架的重构问题，续写他们长期以来保护消费者权益的历史。

本章接下来的部分将讨论网络广告与数据服务行业，它们都有过使用处于建立已久的监管框架之下的数据集的历史。

广告支撑的生态系统

从商业网站建立初期以来，网络广告一直是互联网增长的一个重要动力。一项研究估计，广告支持的互联网部门涉及美国数百万的重要职位，其中互动销售领域每年就为美国贡献数十亿美元的经济增长份额，[99] 它本身一个能让大数据扎根立足并蓬勃发展的行业。越来越精确的消费者数据包含了：他们在哪？用的是什么设备？他们上百种的兴趣是什么？再加上强大的分析，使得广告商更有效地了解客户需求。昂贵的电视节目空挡和整页的国家级杂志上的广告与精确分割、即时测量的网络广告相比显得粗糙不堪。一项研究表明，广告商愿意为针对性强的网络广告多支付 60%-200%的费用。[100]

消费者获得了稳定的数字生态系统，得到了一系列的免费内容、产品与服务。互联网还让国内与国际的广告商不仅仅与大公司接触，还会使其关注一些小型家庭企业的羽翼未丰的品牌。因此，消费者可以从更好、更实用的广告中获得更广泛的企业信息。这在市场上最终表现为更有竞争力、更具创新性。

在让这个生态系统发挥作用的过程中，很多不同的人起到了重要作用，包括消费者、直接参与进来的企业与一系列其它的提供分析或保险类服务或导出共享数据的经济实体。在网站的发布者与付钱在该网站的页面上显示自家广告的广告商之间，是一批令人眼花缭乱的公司。广告网络与广告交换有利于广告商和出版商之间的交易；广告内容与广告活动有相关机构、优化者与媒体来策划并加以投放。广告的效果由另一批专业公司来分析与测评的。[101]

总体来说，与消费者直接产生联系的公司从消费者处收集信息，它们被称为“第一方”(“first parties”)，具体包括新闻网站、社交媒体、在线或离线的公司零售商。但如上所述，作为不同业务之间总体关系的一部分，部分公司常以一种汇总或去身份化的形式代表“第一方”处理数据或访问数据，消费者的信息因而也可能被间接收集。这些第三方公司包括很多数字生态系统中的“中间玩家”、办理付款处理的金融交易公司、填写

[99] John Deighton and Leora Kornfeld, *Economic Value of the Advertising-Supported Internet Ecosystem*, Interactive Advertising Bureau, 2012, http://www.iab.net/economicvalue.

[100] J. Howard Beales and Jeffrey Eisenach, *An Empirical Analysis Of The Value Of Information Sharing in the Market for Online Content*, Navigant Economics, 2014, https://www.aboutads.info/resource/fullvalueinfostudy.pdf.

[101] LUMA Partners, “Display Lumascape,” http://www.lumapartners.com/lumascapes/display-ad-tech-lumascape.

订单的公司等其它公司。“第一方”既可以自己使用数据，也可转售他人以开发广告或用于其它用途。消费者往往无法理解其自身在这个市场中各个级别下被商品化程度。

消费者与透明度挑战

十多年来，网络广告业一直致力于为消费者提供自我监管框架下的选择自由并提高其透明度。在广告生态系统的边缘，消费者可以识别网站的管理者与广告的发布者，后者会将相关隐私政策或其他形式的通告送达给消费者告知他们的信息会被如何使用。在这种自我监管的制度下，当涉及到消费者行为模式与多站点广告投放时，公司同意遵循一系列原则以便收集消费者随着时间推移而产生的行为活动并从多站点的监测中推测消费者偏好。这些原则包括告知用户自己的数据收集方式；为用户提供退出某些跟踪形式的选项；限制敏感信息的使用，例如：孩子的信息或医疗、金融数据；要求删除相关数据或去身份化。

提高透明度与加强线上隐私的技术发展得很慢，并由于各种原因而没有被消费者广泛运用。例如在广告商与发布商采取的自我监管制度之下，许多线上的基于消费者行为的广告，都会包含一个标准化的图标来表示信息正在为了广告定点投放而被采集，同时也提供了供以消费者取消该信息采集的网页链接。[102]根据网络广告业的数据，这个图标已经出现在广告中上十亿次，但只有极小部分的用户使用到它的功能或了解它的意义。大型网络公司运营的广告网站也向用户提供了详细的仪表盘，用户可以在其中看到他们信息的基本使用状况，并且给予他们取消该服务的能力，这同样没有得到用户们的关注。[103]有很多相关理论来解释用户为什么不用这些隐私功能。一些人断言，隐私工具被隐藏起来了或者浏览起来太困难。[104]另一些人争论，接二连三的隐私条款与设置给消费者带来隐私疲劳，他们必须辛苦地亲自做完而不是接受服务。[105]也有可能是因为大部分人在享受可供选择的免费且功能强大的内容、产品与服务的同时，并不会被个性化的广告打扰。

当我们为跨平台信息收集的发展势头与广告投放日益提升的精确度而欢欣鼓舞时，对消费者透明度与有意义的选择的威胁也在日益加深。如今即使采用相对而言简单直接的技术，使消费者对其浏览器与基于浏览广告的目的而访问的网站间数据流有更大的控制能力，亦即“请勿追踪”(“Do Not Track”)的浏览器设置，也会遇到一些问题，因为防欺诈与网络安全的活动现在都依赖于这些相同的数据流进行追踪、阻止恶意活动。

---

[102] For information about the industry’s opt-out program, see http://www.youradchoices.com/.

[103] See Google Ads Settings at http://www.google.com/settings/ads; Yahoo! Ads Interest Manager at https://info.yahoo.com/privacy/us/yahoo/opt_out/targeting/; Microsoft at http://choice.microsoft.com/en-us/opt-out.

[104] See, e.g., Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor, “Why Johnny Can't Opt Out: a Usability Evaluation of Tools to Limit Online Behavioral Advertising,” Proceed-ings of the SIGCHI Conference on Human Factors in Computing Systems, 2012, http://dl.acm.org/citation.cfm?doid=2207676.2207759.

[105] See, e.g., Sarah Kidner, “Privacy Fatigue Hits Facebook: Have You Updated Your Settings?,” *Which? Conversation*, Oct. 18, 2011, http://conversation.which.co.uk/technology/facebook-privacy-settings-rivacy-fatigue/; Aleecia McDonald and Lorrie Cranor, “The Cost of Reading Privacy Policies,” *4 Information Society: A Journal of Law and Policy for the Information Society*, 543, 544, 564 (2008).

“请勿追踪”的挑战

“请勿追踪”的隐私设置的背后意义是提供给消费者一种简单易行的控制方式以限制对其进行的多网站行为追踪。一些浏览器在默认设置下就阻止第三方 Cookie 的追踪或使消费者自己能够选择这项功能。一些浏览器也让消费者发送不要跟踪自己的指示服务信号。虽然不跟踪技术相当简单，但在接受发出“请勿追踪”信号的用户访问的网站标准上，相关政策已被证明很难达成一致。一些网站自愿同意默认的访问者并同意“请勿追踪”的要求，但另一些则没有，或者依然进行局部追踪，敷衍着消费者，打击着隐私倡导者。

一个万维网联盟的工作组，包括了技术人员、开发人员、广告业代表与隐私倡导者，三年多来致力于创造一个“请勿追踪”的信号实施标准。近日，该工作组发布了最终候选的“请勿追踪”技术规范，并将向更大的社群征求审核意见。

在此期间，欧盟在 2009 年修订其电子隐私指令(E-Privacy Directive)，要求在使用用户的 cookie 与其它在线追踪设备时必须得到用户的允许，除非它们所请求的服务是绝对必要的，例如在线购物车。各地对于该指令的贯彻并不统一。虽然现在很多欧洲国家获取的 cookie 是得到一次性明确同意的，但这种行为被普遍认为是笨拙的并被批评在某些情况下并未在该指令设想的隐私保护方面给予用户有意义的选择权。

虽然不完美，但这些努力仍然表现了对于开发出一种技术手段以允许个人对商业实体获取并使用其信息进行控制这一领域日益增强的关注度。

数据服务业

除了主要专注于在线广告的公司之外，还有其他一系列企业从消费者、公用记录与其它数据集中提取信息。数据服务部门有时也被称为数据经纪人，它包括一些收集多个信息源数据的企业，它们将数据进行汇总分析，并共享这些信息以及由其派生出的信息。通常情况下，这些公司与它们所收集的信息的用户之间没有直接的关系。相反，他们为政府或其它企业提供服务，包括产品营销、验证个人信息、人肉搜索，或检测欺诈行为。其中一些公司也有“消费者报告机构”(“consumer reporting agencies”)的具体业务线路，例如为信贷申请、保险、就业医疗提供报告。

从监管的角度看，数据业务分为三大类：

1. 根据《公共信用报告法》(“Fair Credit Reporting Act”, FCRA)规范消费者报告，前者通常保存数据并将之收集在一个单独的系统对其进行分析，同时，对在一个分离的系统中出于上述目的的数据行为进行报告，并保证其同时遵循其它数据服务业务的具体规则
2. 风险减轻服务，例如身份验证、欺诈监测与人肉搜索或者查找服务
3. 包括确定潜在消费者、提升广告推送精度与其它相关服务在内的市场推广服务

第二章中所讨论的《公平信用报告法》(“Fair Credit Reporting Act”, FCRA)向消费者提供了肯定行的权利。提供报告以确定信贷资格、保险、就业的消费者报告机构(“Consumer reporting agencies”)，需按照《公平信用报告法》或者《平等信用机会法》

的相关规定施行。当有诸如接受被拒绝或是信用成本过高等情况发生时，该机构需要基于相关报告与法律需要告知消费者相关信息。消费者有权知道他们的档案与信用评分的状况，了解纠正与删除不正确信息的方式。[106]《公平信用报告法》授权信用报告机构在一定时间后删除负面信息，例如逾期付款与税收滞留的记录将在 7 年后从档案中删除，破产的记录则将在 10 年后被删除。某些类型的信息，如种族、性别与宗教，不得纳入作为确定资信的因素。

这些法定权利不以风险调控或市场推广为目的，事实上，数据服务公司可以提供查阅及改正机制来进行消费者的身份认证。在市场推广方面，一些公司允许消费者选择删除其在市场推广活动中使用的个人信息。

不受监管的数据代理服务

为了协助市场推广，数据经纪人可以提供一个人与某一品牌之间的互动、或是他通过各个渠道寻求帮助的从网页端到社交媒体账户到移动终端的信息数据。数据经纪人通过汇总一个人的购买模式、网站活动、在社交媒体上的活跃方式与他/她和网络广告间的互动，或者直接的客服记录信息，这些信息将借由公共记录信息或者其它通过商业可以取得的信息得到进一步的强化。依据这些信息，数据经纪人能够描摹出一名顾客的概貌，并进一步对其活动记录或约定进行监控，以帮助市场推广人员确定应该何时发送何种信息。

这些身份文件可以是非常详细的，包含最多上千条信息，一些大型企业数据对亿万消费者都有相应的身份文件。他们通过算法分析这些信息，对客户精确分类并辅以描述性的名称来帮助他们的企业客户识别人群，从而进行有针对性的广告投放，一些具体的客户分类如下："苦苦挣扎着的少数民族二等市民"("Ethnic Second-City Strugglers")、"一无所有的退休单身汉"("Retiring on Empty: Singles")"艰难的开始：年轻的单亲父母"("Tough Start: Young Single Parents")、"消耗殆尽的信用：一个城市家庭"("Credit Crunched: City Families")、"勉强度日的乡下汉"("Rural and Barely Making It")，[107] 这些身份文件既包括个人的事实性信息，还含有通过其他数据"模拟"得出的信息。数据经理人接下来可以出售符合特定标准的消费者"原始名单"("original lists")，同时他们也可能提供"附加数据"("data append")服务，公司可以通过这种方式买到更多特定消费者的数据，进而帮助他们形成更为完善的个人身份信息并据此保持它们的信息优势。[108]

什么是信用报告机构(Credit Reporting Agency)？

从 18 世纪 90 年代开始，信用报告公司（现在的信用报告机构），已经能收集并报告个人的信息，并用于决定信贷资格、保险与工作等领域。在一个典型的场景中，信用报告机构收集个人的信用记录，例如他们是否按时支付账单，他们所持有的银行账户的类别与时间，他们是否已经是贷款收回的对象，他们是否有显著的债务。之后该机构使用统计程序将这些数据进行对

[106] Federal Trade Commission, "A Summary of Your Rights Under the Fair Credit Reporting Act," http://www.consumer.ftc.gov/articles/pdf-0096-fair-credit-reporting-act.pdf.

[107] U.S. Senate Committee on Commerce, Science & Transportation, Majority Staff, "A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes," p. ii, December 18, 2013.

[108] Ibid at 22.

比，给予具有相似贷款记录的消费者相应的分数以反映其个人信用：它有多大的可能性按时还贷。这个分数代表着消费者买房买车的能力，抑或是代表着债权人是否可以或是在何种条款下可以向其发放贷款。

虽然这种消费者的精确分析可以带来许多好处，它同时也代表了私营部门有时会在未经当事人允许的情况下收集信息并利用算法来计算个人身份的强大能力。这项大数据技术如果使用不当，可能会对特定的个体产生显著的不利后果。在 2012 年的隐私报告(“Privacy Report”)中，美国联邦贸易委员(Federal Trade Commission, FTC)建议在《公共信用报告法》尚未覆盖的领域中，数据经纪人应该更加透明化；进一步的，根据数据的敏感程度与使用方式，授予消费者对其相关数据的合理的访问与选择权限。[109]

算法、替代计分和歧视

商业模式与大数据策略，特别是第三方数据服务公司，围绕着消费者数据的收集与使用，提出了如何提高透明度、实施问责制度的重要问题。强大的算法可以在释放企业可获得的信息的价值的同时，帮助普通消费者，但这样也会在自动化决策方面引起编码歧视。在数据访问途径的扩大与强大的分析功能的推动下，现在许多产品可以通过不同于法律规范下的传统信用评级机制对个体进行评分，[110] 这些产品试图地数字化地描述包括消费者的购买力、基于他/她在社交网络上的活跃状况所判断出的社交影响力（是他们影响社交圈，还是他们是社交圈的影响者）在内的任何事物。

这些评分也许是为市场目的而产生的，但是它们也可以在个人购买房产、预测职业安全与估计健康程度等方面发挥作用，这就像《公平信用报告法》与《平等信用机会法》监管下的信用指数一样。[111] 而什么类型的数据包含在评分指标之内、用什么样的算法对个人行为进行归因等细节都会受企业控制，而不为消费者所知。这意味这些评分无论对于消费者伤害的确认，还是在消费者本人在实际负责的决策链条内对于实际发言权的掌握中，都不会带来有意义的作用。

由于缺乏透明度与可信度，个人几乎没有能力来获取从他们身上直接收集或是经过分析后得到的信息。[112] 在网络公司自愿提供个人数据而《公平信用报告法》要求个人数据的规范化的今天，却迟迟没有出现一个全方位的门户网站为消费者与数据公司

---

[109] Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Business and Consumers*, 2012, http://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers.

[110] Frank Pasquale, *The Black Box Society: The Secret Algorithm Behind Money and Information*, (Harvard University Press, 2014).

[111] Pam Dixon and Robert Gellman, “The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future,” *World Privacy Forum*, April 2014, http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf.

[112] The Government Accounting Office conducted a gap analysis of privacy laws and regulation in its September 2013 report on Information Resellers. See GAO, *Information Resellers: Consumer Privacy Frame-work Needs to Reflect Changes in Technology and the Marketplace*, GAO-13-663, 2013, http://www.gao.gov/assets/660/658151.pdf.

的沟通牵线搭桥。这样的政策对于那些身份被盗用，出现了一时疏漏的人来说尤为不利，他们的得分会受到影响，这相应地会使其参与经济活动的能力会受到限制。

算法是什么？

简单来说，算法是处理数据的一系列的步骤与指示。算法生成类别并筛选信息，对数据进行操作，寻找数据间的模式与关系，或者帮助进行信息分析。算法的步骤由其作者的知识、动机、侧重点与预期产出决定。一个算法的输出可能无法显现出上述任何因素，也不会在它产生的判断中展现一个错误结果或是任意的选择的概率。人们常说的“学习算法”(“learning algorithms”)，它支撑了从搜索引擎的结果排序到数据库的内容过滤等各个方面，它们给每个变量分配不同的权重变量，并最终生成从预测行为到否定机遇等一切结果的决定，这种方式能够在消除偏见的同时保持科学的客观性。

考虑上述原因，民权社会所关心的是，这样的算法决策在数字经济中带来的“底线”问题——在中性算法的幌子下可能产生的对于社会的最弱势阶层的歧视。[113] 近日，一些线下零售商就被发现在同一款商品的销售商根据算法推算出的消费者居住地的不同附加了不同的折扣。尽管这些价差可能是由于特定居民区竞争对手的缺少，但事实上，高较之低收入住宅区的人们，收入住宅区的人们通常会享有更高的折扣。[114] 同一商品在不同地区以不同的价格出售有着有完全合法的理由。但是这种细分消费群体的方式对用户的需求进行如此紧密的划分以至于让消费者几乎无法察觉需要更好的服务，尤其当它涉及到差别定价与其它价格歧视的可能性时。因此，对于算法驱动的决策是如何扩大社会经济系统内部物价、服务乃至教育与劳动力配置方面的差距是值得进一步的检验的。

结论

广告支持的互联网通过提供有用的服务、新闻与娱乐节目，不考虑财务成本地为消费者创造了巨大的价值。更精确地广告投放能力对公司来说是具有巨大的价值的，它可以有效地提高观众购买他们的商品与服务的可能性。然而，大数据在私营部门的使用必须保护社会中的弱势群体使其免于不公正的对待。算法在相关资格认定的决策中的广泛使用必须得到谨慎地监管，否则即使没有歧视意图，也有可能产生对于弱势群体的歧视结果。美国联邦贸易委员会在相关产业与社会公众对这个复杂的话题的持续讨论上所给予的帮助是值得褒扬的，并应继续其重点关注数据经纪人这一新兴行业的计划。我们期待着他们将来在这一重要议题方面的精彩表现。为切实增加消费者关于其不规范评分的访问权限，尤其是其中更改并禁止其发布不准确信息的权限，相关的工作还需要进行开展。同样地，在衡量由于使用评分方式或算法而产生的不良后果

[113] The Leadership Conference on Civil and Human Rights, “Civil Rights Principles for the Era of Big Data,” http://www.civilrights.org/press/2014/civil-rights-principles-big-data.html.

[114] Jennifer Valentino-Devries and Jeremy Singer-Vine, “Websites Vary Prices, Deals Based on Users' Information,” *The Wall Street Journal*, December 24, 2012, http://online.wsj.com/news/articles/SB10001424127887323777204578189391813881534.

方面，我们需要开展更多的研究以进一步了解这些工具及将来它们在私营与公共部门的广泛使用所带来的影响。

5、为大数据构建的政策框架

高速发展的今天，信息时代从根本上重新决定了数据是如何影响人们的日常生活和宏观经济的。全球有超过 6000 多种数据，国际数据的流量具有持续性和多样性。为了达到更深于以往的层次，政府和企业家开始利用大数据来了解人们的生活，并改进他们的服务。

大数据在社会和经济的应用中创造了巨大的价值，这对国家来说是非常重要的战略。科技的创新会给美国的经济带来新的活力。随着时代的来临，大数据在工业和制造业上将会有更有意义的产业目标，让工业和经济上的信息更为快速地增长。

政府应该用政策来支持大数据技术，机构也必须继续对公众公开对数据的研究。联邦政府也应该投资大数据的技术发展，尤其在教育、医疗和能源方面。在上一个章节提到过的，调整现行政策，让大数据的某些应用符合公众利益，如医疗等。大数据的政策框架的制定需要公众和私营部门的合作，以进一步加快排除识别障碍并进行大数据创新，推动大数据的蓬勃发展。

像产品的其他改革因素一样，大数据产生的价值对于个人、组织和社会是不同的。虽然它的很多应用是明确有利的，但是它的使用会与个人的隐私以及公平、公正和独立这些核心价值产生冲突。

技术的启用使数据的收集更加普遍、有侵略性、有价值。使用新的高速缓存来收集和导出数据具有很大的潜在价值，但也难以保持均衡。一些私营和公共机构将有机会获得更多的数据和更多的资源来进行计算，这也可能增加机构和个人之间的不对等。

政府的职责是使技术的改革能够被公平地利用于可以产生公众利益的地方。以下为政策需要探索的四个主要方向：

1. 政府如何利用大数据为公众产生利益，而非做一些让公众无法接受的事情
2. 大数据通过影响核心价值观，改变消费者组成结构的程度
3. 如何保护公民不被可能用大数据技术的新的形式下歧视
4. 大数据怎样影响了从 1970 年就开始用的隐私的核心准则，通知和协议

大数据与公民

大数据将加强政府对公共服务的管理，并能够创造出全新类型的价值，但是这一工具无疑会使政府的权力愈发不受控制。地方警察如今拥有了比冷战时期更强大的监控设施。新的监控设施被阿里托的法官形象地比喻为管控着生活方方面面的“小警察”，这连同基于算法的居民调查方式，并据此在调配警力中采用的新型执法技术，引发了大数据技术对宪法第一修正案中所保护的言论自由与结社自由产生的影响的疑问。

当私人文件大量储存在家中时，国会通过了许多关于执法监管访问电子信息的法律。该储存通信法案是《电子通讯隐私法》（“Electroni Communications Privacy Act”，ECPA）中的一部分，并阐明了取得电子通讯信息的规则，包括电子邮件和云服务。《电子通讯隐私法》最初是在 1986 年通过的，它起到了保护个人通信存储私密性的作用。但随着

时间的推移，一些法律已经过时，不能用于评判今天的技术。我们在考虑如何更新法律时，也应该想到未来的很多障碍，包括公民的隐私利益，我们需要执法机构和民事执法机构来保护公民的安全，并执行刑法和民法。电子邮件、短信等其他私人数字通信已成为通讯的主要手段，云计算越来越多的被用在私人文件的储存，它们应当受到保护。

同样，许多给予了元数据的保护措施也随着电脑，网络，手机和云计算的发展而进行调整。没人能想到现在的我们能够将以前作为例行公事的数字痕迹还原成个人的隐私资料。如今，元数据的大多数用途仍然在于“小数据”世界的执法方式上，如确定和调用一个犯罪嫌疑人的电话号码。未来，元数据作为“大数据”世界的一部分，将会有越来越多的调查涉及到元数据。从而也引发了一个问题，即我们应该给予它怎样的保护。现在，书面或口头的通讯的内容受到了较多的保护，而对于元数据保护的则相对较少。

尽管政府使用的大数据技术引发了政府的权利如何被监督这一重要的问题，但在大数据背景下，我们可以增强问责制，保护公民的隐私等权利。这些措施包括，在采集或产生的数据时由当局通过复杂方式将数据标签化，其目的是限制用户的访问，跟踪用户并了解其访问目的，同时提醒上司数据可能被滥用。所有的这些方法都被部分联邦政府用在公民权利的保护和规范大数据技术中，以使其被正确的使用，并应用到更多的机构中。总体来说，如果大数据能够正确使用，它可以增加公民的实质自由和权利，推动公共服务的转型和改善。

大数据与顾客

大数据的采集和分析技术可以用在社会和经济中每个部分，并有很多被运用于商业领域。而其运用最广泛的地方是网络广告行业，即凭借人们浏览的网页或者手机记录的行程而推送的私人服务广告。另一方面，随着从现实世界获得的数据越来越多地被结合在网络活动中，信息的采集和用途变得十分广泛，且变化十分迅速。

最终的结果是数据中大量增加了个人私密信息。这些信息在各式商业中非常有价值。它们被买卖、交易、销售，整个行业中都存在着将这些信息所产生的结论商品化的现象。目前市场上销售的商品都包含着一些客户评分，它们描述了客户的基本情况、习性喜好、社会影响力、理财习惯、金融状况等，甚至是租户、工作保障和弱点。虽然有些数据被高度监管着，但其他的用途却没有。

将针对性的广告与消费者在网络和现实世界被跟踪和提供服务的活动相结合，将会产生了巨大的利益。广告和市场有效的补贴了网上的许多免费商品，带动了整个行业的消费应用软件的发展。正如有人在报告中直接指出：“我们不愿意把时间放在网络调查中。”

数据采集中十分重要的一点是在线身份的验证。数据服务和金融产业采用很多途径以确保顾客能够用移动设备和电脑进行安全的交易。相同的验证技术可以让顾客和私营部门交易，也可以让政府和公民在网上互动交流，轻松打开公共服务的新世界。

但是运用这种方式提供商业服务也会付出代价。组合大量客户的数据正间接地破坏着法律，联邦政府的急需新的政策来替换这个模糊的国家标准。在新标准中，若大量的参与者进入到数据的采集、储存、汇聚和销售中，这对消费者是没有好处的。因为一般的消费者不太可能知道数据被哪些范围内的人采集和持有，这就让他们很少有机会参与到其数据的准确性和范围确认中，这限制了他们了解这些信息是如何反馈到算法，进而决定其客户体验和市场准入的。

在考虑什么样的政策能够让大数据在消费者处蓬勃发展时，有一个尚待解决的重要问题，即如何使用采集到的信息。一方面，这意味着用大数据来划分顾客的营销目的，从而提供更有针对性的时机来使顾客购买商品和接受服务。另一方面，也意味着更为严肃地应用信息来计算消费者获得住房，医疗保健，信贷，就业，教育的资格。

大数据与歧视

除了创造巨大的社会价值，政府和私营部门在数据的使用中也可能会带来很多危害。这些危害可以分为有形的物质危害，如财产的损失，和无形的危害，如私人生活被侵犯和名誉受损。研究中最重要的结论就是，大数据造成的损害不仅仅是隐私上的，还有包括对个人和群体的歧视等。这种歧视是非有意的、由于大数据的结构和使用方法而产生的结果，但也是源于一些掠夺弱势阶层的意图。

一个来自波士顿的例子展现了怎样在大数据技术的使用中杜绝无意中产生的歧视。该市和波士顿市长办公室下辖的新型城市机械局(Mayor's Office of New Urban Mechanics)合作开发了一款实验性的应用，[115]为利用智能手机的加速计和 GPS 反馈出凹坑等道路情况，并将其报告至城市公共工程局的移动应用程序。这是一个城市利用大众来改善服务的范例。但这个程序也有一个潜在的问题，即穷人和老人可能没有智能手机，且不会下载这个程序。但它可以使用在在比较富裕的城市，以产生引导城市服务的效果。

值得称道的是，波士顿和坑洼街道（Street Bump）的开发者在推出程序之前就想到了这一点。他们首先平均部署了为整个城市各个区域服务的城市道路视察员，然后为公共提供了额外的数据。这具有防止不平等结果的先见之明，并且得出的数据表明这是值得的。该应用已记录了 36992 个“坑洼”，来帮助波士顿市民找出比较结实的没有坑洼的地面。

一些用户因为验证了他们的身份，而在和复杂的数据库信息交互的时候受到更多的歧视。拥有多个姓氏人或因结婚而改名字的女性通常遇到的错误最多。例如，在电子验证系统中，民权倡导者一直对国土安全部和社会保障局所共同运行的数据库表示担心。

电子验证系统有提供雇主确认新员工是否有资格在美国合法工作的能力的功能。当考虑到这个查询系统的进程与它所组合的数据来源是在不断地改变的时候，该系统的绝大多数结果都是能够及时和准确地提供雇主所雇佣的人是否有权在美国工作的的信息。定期评估能提高电子验证系统在表现出不同群体所占的比例方面的性能。该系统在 2009 年的一项评估发现，相比较 2.1%，有权工作但未确认的美国公民占 0.3%。几天后，这些工人的身份便被确认了。[116]

国土安全和社会安全部对这个问题相当重视。最近的一项评估程序发现，人们能够较快地且低错误率地验证他们的工作。五年后，美国公民的首次配比失误率下降了 60%，

[115] See New Urban Mechanics, http://www.newurbanmechanics.org/. All information about Street Bump comes from its former project manager James Solomon, who was interviewed by officials from the office of the White House Chief Technology Officer.

[116] Westat Corporation, *Findings of the E-Verify Program Evaluation*, December 2009, Report Submitted to Department of Homeland Security, http://www.uscis.gov/sites/default/files/USCIS/E-Verify/EVerify/Final%20E-Verify%20Report%2012-16-09_2.pdf.

非公民则下降了 30%。[117] 如果这个问题得不到解决，个人或团体的就业就会出现问题，所以必须纠正大数据系统来做到准确、透明。

这两个无意间产生歧视的例子说明了检测结果至关重要的原因。有时大数据技术并没有表现出歧视，并且应用过程并没有不公平，但是其整体对大数据的关注和利用则会造成了歧视。

在社会的特定领域，包括就业、信贷、保险、医疗、住房和教育，我们已经采取大量措施强制保持公平性。现行的立法和监管保障应管理个人资料在以上情况下的使用方式。尽管预测算法被允许在特定的情况下使用，里面所提供帮助决策的数据也要保持一定的透明度，且能够被改正。对于就业、信贷、保险等方面的重要决定，顾客有权知道为什么电脑的决定不同于自身决定，做这个决定用了哪些信息。同时，如果该信息有误，则需要其可以从根本上被改正。

由于美国历史上长期存在着歧视问题，所以这些保护措施有必要存在。自 20 世纪初，银行和贷款人使用位置数据来做出关于个人的假设。但直到 1975 年房屋抵押贷款披露法案签署成为法律之前，在考虑是否贷款时他们的考虑因素一直是其居住地，而非个人还贷能力，所以个人贷款并不普遍。银行毫不夸张地“画着”批注，并在周围划定不提供贷款的范围。这种现象存在了几十年，特别是对于非裔，拉美裔，亚裔美国人和犹太人，这更是产生歧视问题的工具之一。

社区可以作为种族或民族身份的代理，这会产生新的担心，即大数据技术可以划“数字红线”来区别非预期人群，无论是客户、雇员、租户或者有信用的收件人。这份报告中明确发现，大数据提供了歧视和掠夺的新形式。

然而，造成歧视的算法和数字挖掘技术，同时也可以帮助群体通过鉴别和经验上证实歧视现象及其危害。 民权团体可以使用大数据这个强大的新工具来要求受到平等的服务和对待。大数据给美国带来的是增强平等还是加剧不平等，这完全取决于它在这几年的应用方向，以及现阶段法律的保护和法律是如何执行的。

大数据与隐私

以物联网为工具的大数据打破了许多私人空间。家中的无线网络信号（WiFi）中可以显示出屋中的人数及其位置，也可通过采集功耗数据来显示出你在屋中的移动。[118] 当你走出房间时，在线面部识别技术也可以将你从图像中识别出来。始终开启的有音频和

[117] Westat Corporation. *Evaluation of the Accuracy of E-Verify Findings*, July 2012, Report Submitted to Department of Homeland Security, http://www.uscis.gov/sites/default/files/USCIS/Verification/E-Verify/E-Verify_Native_Documents/Everify%20Studies/Evaluation%20of%20the%20Accuracy%20of%20EVerify%20Findings.pdf.

[118] Stephen Wicker and Robert Thomas, “A Privacy-Aware Architecture for Demand Response Systems,” *44th Hawaii International Conference on System Sciences*, January 2011, http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5718673&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5718673; National Institute of Standards and Technology, *Guide-lines for Smart Grid Cyber Security: Vol. 2, Privacy and the Smart Grid*, 2010, http://csrc.nist.gov/publications/nistir/ir7628/nistir-7628_vol2.pdf.

视频接口的可穿戴设备以及整个物联网设备的出现只会产生越来越多的信息采集量。在合法使用的传感器的海洋中，限制信息采集是一个巨大的挑战，几乎是不可能的。

这种无处不在的信息采集是由大数据技术本身性质所决定的。[119] 无论是产生模拟信号还是数字信号，数据都被重复使用着，并且以前所未有方式结合，这便激励着更多的数据采集。数据的潜在价值推动着“土地战”，机构的重点也转向尽可能多的采集和利用数据。公司不断地发掘他们已有的数据，同时寻找他们需要的数据来提高其市场地位。当今世界，数据存储的成本已经大幅下降，同时仍具有尚无法预测的未来创新潜力，所以采集尽可能多的数据是至关重要的。

大数据的另一个现实就是，数据一旦被采集，就很难保持提供者的匿名性和隐私性。虽然有研究希望在大数据的采集中模糊个人识别信息，或重新标识“无名氏”的信息。融合数据技术集资要比隐私保护技术方便许多。

总之，这些趋势要求我们关注四十年中，告知与同意框架是如何为隐私保护提供支持的。在结构性过度采集的技术中，重新鉴定要比识别功能更强大，并将重点放在了信息的采集和保存上，个人的隐私就没有那么受关注了。总统委员会科学技术的顾问说：“告知与同意框架已经被大数据所带来的正面效益打败了，大数据所带来的是新的、并非显而易见但十分强大的使用价值。[120]

[119] President’s Council of Advisors on Science & Technology, *Big Data and Privacy: A Technological Per-spective*, The White House, May 1, 2014, whitehouse.gov/bigdata.

[120] Ibid at 36.

联邦政府关于增强隐私方面技术的研究				
增强隐私的技术的研究和开发一直是奥巴马政府的优先事项。网络和信息技术研究与发展（Networking and Information Technology Research and Development，NITRD）计划机构每年在隐私上的研究总花费超过 7000 万美元，[121] 且主要分为四个部分：对隐私安全扩展的支持，对企业如何遵守隐私法的研究，医疗保健的隐私，保护基础搜索的隐私。下表总结了 NITRD 的一些正在进行中研究项目。总统委员会的科技顾问赞成加强研究美国研究隐私相关的技术和围绕其使用的社会科学问题。				
研究领域	**对隐私安全扩展的支持**	**对企业如何遵守隐私法的研究**	**医疗保健的隐私**	**保护基础搜索的隐私**
机构	空军研究实验室、国防高级研究计划局、国家安全局、高级情报计划、研究活动处、海军研究办公室	能源部、国土安全部、国家标准与技术研究所	远程医学和先进技术研究中心、国家协调员卫生信息技术、美国国立卫生研究院的办公室	美国国家科学基金会
预算（共 $77M/年）	$34M/year	$10M/year	$8M/year	$25M/year
重点项目摘录	匿名化技术、 保密协作与沟通、 同态加密、 隐私保护数据汇总、 交通安全路线	自动化隐私协议、 位置隐私保护工具、 保护个人身份信息、 法律合规标准、 智能电网的自愿原则	收集和使用限制、 数据分割隐私、 患者知情同意和隐私、 病人的数据质量、 保持医疗保健数据匿名	隐私和工具算法基础、 隐私经济学、 隐私权在社会心理上的建构、 隐私政策分析、 云计算，数据整合，挖掘的隐私保护

[121] Networking and Information Technology Research and Development, *Report on Privacy Research within NITRD*, April 2014, http://www.nitrd.gov/Pubs/Report_on_Privacy_Research_within_NITRD.pdf.

预测大数据变革的下一篇章

对于现在绝大多数的普通交互来说，告知与同意框架充分保护了隐私。但是总统委员会的科技顾问表示，技术轨迹正在转向采集、使用和储存对消费者和个人并没有直接联系的数据上来[122]。假若该框架被违背，比如由我们的家庭设备采集的数据，我们则需要重新关注数据的使用，这一政策转向正在被专家、学者广泛讨论[123]。数据的使用情况是极为重要的，它对社会有利有弊，如“双刃剑”一般。

负责任地使用政策框架会带来许多潜在优势。将责任从个人转移到采集、保存和使用数据的实体，由于个人在目前市场中的位置，他们并不能很好地理解和抗争告知和同意框架。关注于使用责任制，也可以使数据的采集者和使用者对数据的管理及其可能产生的危害负责，而不是狭隘地将其责任定义为是否通过正常途径采集数据。

更多地关注责任并不意味着忽视收集的环境。对数据负责，一方面就是要尊重原始数据的采集。实际上，如同在消费者隐私权法案所阐述的尊重环境原则，这一规则并不令人惊讶。虽然数据的收集不能立即用在就业上，但技术的发展正在向这个方向转变。先进的数据标记技术可以已采集和用户授权使用的信息细节进行编码，从而使许可使用的信息可以一直跟随着数据。若是该技术得到良好发展和广泛使用，即使不能解决大数据中所有的问题，也可以用于应对一些关键挑战。

或许最为重要的是，为了更负责地使用大数据，我们应该将关注的重点放到如何平衡大数据所带来的效益和对隐私以及其它由于大数据采集信息的不可避免性而受到危害的价值。我们是否应该制定规则，不能在任何环境下使用没有得到使用授权的数据，即只使用得到使用授权的数据？对于医学研究中为了治愈癌症而使用的数据，和商业营销中对消费者的广告定位而使用的数据，我们应该如何区分和界定它们？

正如奥巴马总统在人权消费者隐私条例草案的发布会上所说，“尽管我们生活在一个能够比过去更自由地共享个人信息的世界，但我们必须坚决否认隐私价值已经过时。”隐私“从一开始就一直是我们的民主制度的心脏，而现在，我们比以往的任何时候更需要它。”这在利用大数据的时代更是如此。

[122] President’s Council of Advisors on Science & Technology, *Big Data and Privacy: A Technological Perspective*, The White House, May 1, 2014, p. 20, whitehouse.gov/bigdata.

[123] Craig Mundie, “Privacy Pragmatism: Focus on Data Use, Not Data Collection,” *Foreign Affairs*, March/April, 2014, http://www.foreignaffairs.com/articles/140741/craig-mundie/privacy-pragmatism.

6、结论与建议

奥巴马总统在 2014 年 1 月 17 日宣布的白宫对大数据及隐私的评估报告，设想研究大数据技术更广泛影响，他认识到，大数据革命广泛开展于公众与私营部门之间，其影响需要被一并考虑在当局对信号情报的审查当中。

90 天里，白宫大数据工作组着手于研究大数据技术是如何改变政府、商业和社会。我们试图了解大数据将会带来怎样的机遇、怎样的进步，我们希望领会已有科技和遇见不远的未来。与此同时，总统科学技术顾问委员会对相关技术进行了评估，他们的研究结果支撑了本报告中许多对科技的论断。

大数据工具为我们提供了惊人、强有力的契机，以解锁已存在的和新采集的数据、发现先前难以接近和洞察的现象。大数据可以推动医疗、教育、农业、能源使用的发展与突破，并对企业如何组织其供应链、监控其设备提供启示。大数据具有精简公共服务供给的潜力，从政府的各个层面提升税款的利用效率，并大大加强国家安全保障能力。大数据的承诺，需要政府数据被视为一种国家资源，并被负责任地提供给那些能够通过它创造出社会价值的人。同时，它为塑造下一代计算工具与技术提供机会，这也将反过来进一步地推动创新。

然而，大数据也引发了许多困惑。就其本质而言，许多部署在我们的手机、家中、办公室、城市路灯柱和屋顶的监测传感技术正在采集越来越多的信息。分析上的不断进步激励我们采集尽可能多的数据，这不仅是为了当下的使用，也是为了日后的应用潜力。从技术角度讲，这促使了数据具有功能性上的永恒性和普及性，使我们留下的数字痕迹被采集、分析、组合，揭示出关乎我们自身与生活的数量惊人的事物。这些发展挑战了人们长期以来的隐私观念，引发了“告知与许可”框架下，用户对被采集数据的初始权限的质疑。然而，这些趋势将不会阻止我们创造途径使人们参与到对其信息的处理和管理中去。

这份评估报告的重要发现之一为，在大数据被用来造福社会的同时，也有可能被用来制造社会危害。纵使在并非有意歧视的情况下，大数据的使用仍然可能导致有失公正的结果。微小的偏见具有潜在的积累性，可能给某些弱势群体带来大范围的结果影响。社会必须采取措施以防止这些潜在危害，无论是公民与政府、消费者和公司或是员工与企业之间，都应保持权力在个人与机构间的适当平衡。

大数据变革正处于其最初阶段。我们需要数年才能理解其完整的技术内涵、其对健康、教育、经济的强化作用，及更为关键的是，它对美国核心价值观的影响，包括隐私权、非歧视、自我决定权。

即使是在当下大数据变革的早期，本评估报告的作者仍认为重要的结论已然出现，即大数据可以从多个领域的层面告知当局者该如何迈出下一步。特别是以下五个方面，它们将引发美国关于在大数据世界如何最大化利益和最小化危害的全民讨论。

1.保护个人隐私的价值：在美国以及全球协作的隐私保护体系中，通过在市场上对个人信息的保护来维护个人隐私的价值

2.稳定/持续负责的教育：要认识到学校（尤其是 K-12）是使用大数据以提升学习机会的重要领域，同时也要对个人数据的使用进行保护，强化数位素养和技术

3.大数据与歧视：防止大数据使用过程中可能带来的新的歧视方式

4.执法和安全保障：在执法过程、公共安全、国家安全中，确保大数据的合理负责使用

5.数据公共资源化：将数据作为公共资源，用于提升公共服务，投资于能够推动大数据革命的科学研究

政策建议

本评估报告还指出了六条值得政府及时关注并制定相应政策的建议，分别为：

- 改进消费者隐私权力法案：商务部应当采取咨询手段，来征求利益相关者与公众对大数据发展及其是如何影响消费者隐私权法案的意见，然后制定立法文本草案供利益相关者审议，并向国会提交
- 通过关于国家数据外泄的立法：国会应当通过立法，沿袭当局 2011 年 5 月的网络安全立法建议，制定一套国家统一数据外泄标准
- 将隐私保护扩展至非美籍人士：管理和预算办公室应与各部门和机构协作，在可行领域将 1974 年隐私法应用于非美籍人士，或制定可替代的隐私政策，应用于各种国籍的人士，以对其个人信息进行合适、有意义的保护
- 确保对在校学生的数据采集只被用于教育目的：联邦政府必须通过法规确保学生的数据恰当共享或使用，尤其针对几种在同一个教育环境下的数据
- 发展技术以阻止歧视：联邦政府首席公民权利和消费者保护机构应当发展技术专长，识别通过大数据分析而对受保护阶层产生歧视性影响的做法和结果，并制定调查和解决方案
- 修正电子通信隐私法：国会应当修正电子通信隐私法，以确保对在线数字内容的保护标准与现实所提供的相一致，包括去除未读的或留存一定年限以上的邮件中因超时而产生的差别

1、保护个人隐私的价值

大数据技术正在推动巨大的创新，同时也产生了新的隐私问题，其影响远远超出了当下备受关注的线上广告问题。这些影响使我们在未来的隐私保护中急需一个更具广泛性的国家审核模式，包括 2012 年发布的政府的消费者隐私权利法案。相对于传统的告知与同意模式，即专注于采集数据之前获得的用户许可，对其进行重新审核是尤为重要的。虽然在许多情况下，告知与同意仍作为一种基本模式存在，但如今，我们需要做出判断，在大数据环境下，更侧重于数据的使用和重复使用的研究方式是否会成为使隐私权管理更为高效的基础。或许，建立一种使个体参与到其个人数据采集后的使用和分配问题的机制，将会是更好的授权方法，以使人们能够从其个人信息中获利。隐私保护的方式也必须不断发展，以适应大数据所带来的社会效益。

推进消费者隐私权利法案

正如 2012 年 2 月奥巴马总统所明确的那样，消费者权利法案和与消费者隐私相关的蓝图代表了“一种动态模型，使得在新的信息技术条件下，提供强大隐私保障、推动创新”消费者隐私权利法案为基于公平信息实物准则。一些隐私专家相信，这些原则在一些细微处有足够的灵活性以解决和支持数据的新兴用途，包括大数据。但其他人对此并不太确信，尤其是技术人员，因为不可否认的是，大数据确实挑战了当下一些支撑隐

私框架的关键假设，尤其是在采集和使用上。值得考虑的是，在告知和同意模式的背景下，大数据的发展该如何可行地保护隐私，以及存在哪些实际限制。

> **建议：**商务部应当立即征求公众意见，针对消费者隐私权利法案如何在支持大数据创新的同时，又对其风险作出反应，以及如何负责地使用框架，就像第五章所阐述的那样，包含在消费者隐私权利法案确立的框架之内。根据评价过程，商务部应当制定立法文本草案供利益相关者审议，并向国会提交。

提高数据服务行业（俗称为“数据经纪人”）的透明度

消费者有权更清楚地知晓，在与他们进行直接交易的企业以外的第三方数据采集者处，其数据是如何被共享的。这就意味着消费者应当清楚地知道数据采集与再利用的范围，即参与调解其用户体验或从用户多样化中采集信息的企业数量。数据服务行业应该跟随线上广告和信贷行业的脚步，建立专门的网站或门户网站，将公司名单及其数据使用方法列于其上，从而为消费者提供途径，以便他们更好地控制自身信息的采集和使用，同时也可使消费者对其数据的使用途径进行自主选择。

纵使我们越来越关注数据的使用，消费者仍希望借助于“不追踪”工具来掌控其数据被采集的时间和方式

由于现在出现了越来越多的记录个人行动、行为和位置数据的设备和服务，所以我们极其需要加强隐私保护工具的性能。民意调查显示，人们对这类工具有十分显著的需求，政府和私营部门必须继续发展隐私保护技术，改善对消费者的服务。

卫生保健服务方面，大数据使医学水平进一步提升、成本进一步降低，故政府应当建立协商程序，评估如何使健康保险流通与责任法案及其它相关的联邦法律法规能够更好地适用

在预测、检查、治疗疾病方面重大突破的实现，一方面需要最大限度的公众政策关注，另一方面，若想挖掘出其全部的潜能，必然少不了医药数据隐私体制的实质性的进步，这样才能使科研人员结合并分析各种生活方式和健康信息。任何改革还必须考虑，在监管和法律保护下，由健康保险流通和责任法案管辖范围外的企业组织散布的大量个人健康信息。

美国应当引领全球大数据对话，重申政府建立全球协作的隐私框架的承诺

大数据的优势有赖于全球信息自由流动。由于这将对不同国家的传统和法律框架产生影响，故美国应当集结国际合作伙伴参与对话，探讨大数据带来的好处与挑战。

具体来说，国务院和商务部应积极发展政府间的双边合作关系，包括欧盟、亚太经济合作组织（APEC）、经济合作与发展组织及其他利益相关者，对现有的及拟议的政策框架应对大数据的方式进行评估。

当局也应努力加强美国与欧盟安全港架构协议，鼓励更多的国家和公司加入亚太经合组织跨境隐私规则体系，通过努力，使欧洲绑定合作规则体系与亚太经合组织跨境隐私规则体系结盟，促进美国、欧洲和亚洲之间的数据流合作。

美国尊重全球化的个人隐私价值，并应将其体现在处理全体相关人员数据的方式上

因此，美国应扩大对非美籍人士的隐私保护。

建议：管理和预算办公室应与各部门和机构协作，在可行领域将 1974 年隐私法应用于非美籍人士，或制定可替代的隐私政策，应用于各种国籍人士，以对其个人信息进行合适、有意义的保护。

2、数字时代负责任的教育创新

大数据给孩子和青年提供了提升教育经历的重大机会。大数据与教育的交叉主要在两个方面。随着学生开始与教育机构共享数据，他们所期待的是发展知识与技能，而不是被用作建立个人优缺点的档案，从而对日后产生不良影响。教育机构也处于帮助孩子、青年及成人应对大数据世界的特殊位置。

在提升教育创新中应确保数据保护

随着网络设备的发展，个人学习变得越来越普遍，提升教育有赖于大数据的发展。接下来的五年，在总统的连接教育倡议下，科技设备将大量走进美国教室，具有加强教课与学习的极大潜力，特别是对于弱势群体。以网络为基础的教育工具和软件使教育技术和商业的重复与创新成为可能。无论是在教室内还是教室外，这些设备中都已被部署了强大的学生隐私安全保护系统。家庭教育权和隐私权法案以及儿童在线隐私权保护法案提供了联邦监管框架，来保护学生的隐私，但是前者制定于网络普及前，后者则制定于智能手机、平板电脑、应用软件、云计算和大数据产生之前。学生及其家属需要强劲的保护手段来应对当下出现的威胁，但同时，他们也应当获得途径来学习科技带来的益处，以保证学生能够充分发挥潜能。

建议：联邦政府应确保学校采集的数据是用于教育用途，并继续支持投资和创新，以提高整个学校的绩效水平。为了促进创新。学校应当探讨如何在现代化

背景下实施家庭教育权和隐私权法案以及儿童在线隐私权保护法案的联邦监管框架，以确保两个相辅相成的目标：(1)确保学生的数据适当共享或使用，尤其针对几种在同一个教育环境下的数据，(2)确保教育科技中的创新，包括新的方法和商业模型，有足够的机会蓬勃发展。

数位素养是 21 世纪的重要技能

为了确保各年龄层的学生、公民和消费者在数据使用中有权充分保护自己，以防数据滥用，对于他们来说，顺畅地理解数据被采集和共享的方式、算法被采用的方式和目的，以及他们可以使用什么样的工具和技术来保护自己尤为重要。即使这些技能将不会取代监管保护方式，增加数位素养也可以使人们更好地在一个充斥着大数据的世界中生活。数位素养，即理解个人数据是如何被手机共享和使用，应当被看做 K-12 教育中的关键技能，并融入标准课程中。

3、大数据与歧视

自动化决策技术是不透明的，基本无法被普通人应用。然而，他们正在承担越来越重要的作用，并在有关个人获得医疗、教育、就业、信贷、商品和服务环境中被使用。环境和技术的结合造成了许多困难，即如何确保发现、判断和纠正在自动化决策过程中有意或无意产生的歧视效应。我们必须开展关于大数据、歧视、公民自由的全国对话。

联邦政府必须关注大数据技术，避免与国家法律和价值观不符的歧视的产生潜力

建议：联邦政府的首席公民权利和消费者保护机构，包括司法部、联邦贸易委员会、消费者金融保护局和公平就业机会委员会，应当扩大技术专长，来识别对受保护阶级有歧视性影响的大数据分析所促进的做法和结果，并制定计划，调查和解决违反法律的此类事件，在评估和解决潜在问题时，这些机构可能会考虑数据的分类、采集背景，以及对某些特别值得关注的群体，如残疾人的基因组信息。

消费者期望于有权知道，他们所接受的商品和服务的价格是否与其它的有系统性的不同

令消费者难以置信的是，呈现给他们的数据和算法作为全部参数，塑造了他们的线上和线下生活。尽管如此，由于消费者的体验是基于其个人信息，所以一定的透明度也是适宜的，特别在不同公司给消费者提供不同的定价时，例如消费者利用网络搜索引擎

或大型零售商的网上商店比较机票价格。经济顾问委员会应评估线上和线下不同定价的演变方式，评估市场有效运作的意义，思考是否需要保证消费者公平的新途径。

数据分析可被用于保障公民自由

同样的大数据技术，在导致歧视的同时，也可以帮助群体行使其权力。相关应用和数据挖掘功能可以识别以及经验性地证实的歧视现象，并揭示其造成的危害。联邦政府民权办公室和民权社会，应采用新且有力的大数据手段，以确保最易受攻击的群体能够得到公正的对待。

为了树立公民意识，联邦政府消费者保护和技术机构应当组织公开研讨会、落实问题报告，其应当针对的问题有，这些新技术潜在的歧视性做法，差别定价的做法，和在信贷、就业、教育、住房和医疗保健上使用代理评价重复规范评分的做法。

4、执法与安全保护

大数据的合法使用可以使社区更安全，使国家的基础架构更具弹性，并加强国家安全。十分关键的是，国家安全、国土安全、执法和情报机构应当积极尝试和合法运用大数据技术，同时也坚持全面问责制，进行监督并保证隐私。

应当重新修订电子通信隐私法

建议：国会应当修正电子通信隐私法，以确保对在线数字内容的保护标准与现实所提供的相一致，包括去除未读的或留存一定年限以上的邮件中因过时而产生的差别。

执法中使用的预测分析应当继续受到仔细的政策审查

至关重要的是，在预测刑事调查的情况外，法律管理下的大数据分析应当被适当部署对个人隐私和公民自由的保护系统。无罪推定是美国刑事司法系统遵循的基本原则。针对言论和社交自由的宪法权利，为防止寒蝉效益，公众必须了解这些项目的存在、操作方式及功效。

联邦机构中隐私和数据方面的专家应为国家、地方和其他联邦执法机构提供技术援助，以探索并部署大数据技术

执法机关应继续研究如何培养大数据监控技术的联邦拨款项目，使其能够被负责任地使用，同时研究在国家和地方建立全国大数据试点项目登记处的潜在效用，以跟踪、

识别和推广最佳途径。针对未来一年有助于推进隐私社会的隐私保护技术的发展，联邦政府机构与技术领导者、专家也应当报告其进度。

评估政府对合法获取的商业数据的使用，以确保其与我们的价值观一致

在了解长期的基本商业记录对搜索犯罪嫌疑人的作用时，联邦政府应该承担对美国公民市售数据用途的审查，注重雇佣大数据技术服务的运用，并确保它们与适当的监督结合，以保护公民隐私和自由。

联邦政府应当实施最佳途径，建立制度协议与机制，以帮助确保数据的使用控制和安全存储

美国国土安全部、情报界和国防部领导者正在制定隐私保护和个人信息处理政策。其它公共部门机构应当评估这些做法，特别是通过数据标记来强制限制使用、控制访问政策和定格的审计，是否能整合到他们的数据库和数据处理中，向其中融入对隐私、公民权利和公民自由的保护策略。

利用大数据分析和信息共享来加强网络安全

保护驱动经济发展的网络、支持公共安全和保护国家安全，这已经成为关键的国土安全使命。联邦政府与私营部门合作伙伴计划中，在试点项目和研究中使用大数据，以保证网络安全，并保护关键基础设施，加强我们的应变能力和网络防御能力，特别是在越来越多网络威胁数据被共享的情况下。当局继续支持立法保护隐私，为公司特殊威胁数据共享提供目标责任保护，并在此基础上适当保护其网络。与此同时，政府将继续采取行政措施，增加奖励，减少数据共享和分析的屏障，帮助公众和私营部门预防和应对网络威胁。

5、数据公共资源化

政府数据是国家资源，并应被尽可能广泛地向公众提供数据，以提高政府效率，确保政府问责制，推动经济繁荣和社会良好，同时也要继续保护个人隐私、商业机密和国家安全。这意味着寻找新的机会使政府释放大量数据，并确保所有机构最大限度地使用Data.gov，即联邦数据的工具与资源存储库。大数据可以改善公共服务，带来对政策制定的新认知，从政府的各个层面提升税款的利用效率。

政府数据应当被准确、安全地存储，并最大程度地开放访问

政府数据，尤其是统计和人口普查，由于其高精确性、高可靠性和高保密性而区别于其他数据。类似的，如今的“我的数据”倡议使美国人可以通过有效的方式轻松安全地访问个人数据，其格式构成的模型使得个人数据拥有可获得性，这应当被尽可能广泛地应用于政府中。

所有部门和机构应其高级隐私官员和公民自由官员密切配合，评估他们是如何驾驭大数据，从而最好地执行任务

过去未大范围采用过高级数据分析工具的部门和机构应当最大程度地理解:大数据革命对他们及其服务的公民意味着什么。他们应该尝试开展试点项目，发展内部人才，扩大研究和开发。各机构应从最早阶段就开始与他们的隐私和公民自由官员协商建立这些项目。

特别的，大数据分析为美国人民在提供政府服务时增加价值和绩效带来了重要机会。大数据也有发现和报告浪费、欺诈和滥用的强大力量，从而能够节省税款、提高公信度。大数据也可以帮助进一步识别出政府高绩效的做法，从而这些做法可以重复应用于类似的机构和程序，并可能提供新的使公共部门管理有效化的方式。

我们应在隐私保护技术的研究和开发上大幅增加投资，鼓励计算机科学和数学、社会科学、通讯和法律等学科的跨领域研究

政府应致力于引导研究，以确定在哪些领域中，大数据分析可以给美国人民的生活水平带来最显著提升，同时鼓励数据学家进一步发展社会、伦理和政策知识。为此，科学与技术政策办公室应在与整个机构中专家的合作中，致力于明确可以带来显著公共利益的领域，例如城市信息学，并对可以使其受到适当关注、获取适当资源的方式进行评估。

关注有发展前景的基本领域，例如数据源、去身份和加密等，但同时，我们也应当鼓励关注那些可以迅速应用到消费者中的市场型新科技工具。由于我们需要越来越多的从事数据方向的干部和能将关键策略转变为技术基础设施的社会科学家，我们将资金投入到一些研究中，例如针对从社会伦理角度传授科学技术知识的科学技术研究，对数据科学家和工程师进行模块化教学，使他们了解这份事业所具有的更广阔的社会影响力。

译者信息与版权说明

翻译人员：

给总统的一封信	阮海博	（浙江大学计算机科学与技术 2012 级本科生）
第一章	陈新	（浙江大学历史学系教授）
	邱桐	（浙江大学工业工程 2013 级本科生）
第二章	李政毅	（浙江大学社会学 2012 级本科生）
第三章	鄢龙	（浙江大学金融学 2012 级本科生）
第四章	陈曼珂	（浙江大学会计学 2012 级本科生）
第五章	付文鑫	（浙江大学机械工程及其自动化 2012 级本科生）
第六章	王依琪	（浙江大学环境工程 2012 级本科生）

校对人：
陈曼珂、阮海博、王依琪、鄢龙

统稿人：
鄢龙

版权顾问：
赵越（华东政法大学知识产权专业 2010 级本科生）

联系人：
鄢龙，电话 18868111770，电子邮件：yourslongly@outlook.com；
王依琪，电话 15967171025，电子邮件：wyq0717@gmail.com。