# 大数据时代个人信息保护研究

Protecting
Information Privacy in
the Era of Big Data

李媛 著

华中·元照 中青年法律科学文库

华中科技大学出版社
http://www.hustp.com

华中·元照 中青年法律科学文库

# 大数据时代个人信息保护研究

Protecting
Information Privacy in
the Era of Big Data

李媛 著

华中科技大学出版社
http://www.hustp.com
中国·武汉

图书在版编目（CIP）数据

大数据时代个人信息保护研究 / 李媛著. -- 武汉：华中科技大学出版社，2019.1
（华中元照中青年法律科学文库）

ISBN 978-7-5680-4473-8

Ⅰ. ①大… Ⅱ. ①李… Ⅲ. ①个人信息—法律保护—研究 Ⅳ. ①D913.04

中国版本图书馆CIP数据核字（2018）第285197号

大数据时代个人信息保护研究 李 媛 著
Dashuju Shidai Geren Xinxi Baohu Yanjiu

策划编辑：王京图
责任编辑：李 娜
封面设计：傅瑞学
责任校对：梁大钧
责任监印：徐 露
出版发行：华中科技大学出版社（中国·武汉） 电话：（027）81321913
武汉市东湖新技术开发区华工科技园 邮编：430223
录 排：北京欣怡文化有限公司
印 刷：北京富泰印刷有限责任公司
开 本：710mm × 1000mm 1/16
印 张：17.75
字 数：263千字
版 次：2019年1月第1版 2019年1月第1次印刷
定 价：76.00元

本书若有印装质量问题，请向出版社营销中心调换
全国免费服务热线：400-6679-118，竭诚为您服务
华中出版 版权所有 侵权必究

重庆市教委人文社科项目："云中信息安全的隐忧与法律规制策略研究"；项目编号：17SKG007。

# 前　言

我们正在步入的世界，信息的价值受到人们的广泛重视。一个关于信息的海量生成、分享、挖掘、分析、应用的数据时代正在开启，我们所处的时代是一个正在发生变革的时代，是一个传统与前沿交替的时代。很多国家把大数据提升到国家战略的高度，美国政府更是把大数据看作“未来的石油”，大数据成为一种新的经济资产，如同货币或黄金。

《中国大数据发展调查报告（2017年）》指出：2016年中国大数据市场规模达168亿元，预计2017—2020年仍将保持30%以上的增长。近六成受访企业已成立数据分析相关部门，超过1/3的受访企业已应用大数据。大数据应用为企业带来的最明显效果是实现精准营销、智能决策、提升运营效率。而个人信息中潜藏的价值，已使得政府、企业走在了对个人信息如饥似渴地追逐的道路上，这刺激了各路竞争者对数据的挖掘。信息的交流、传递和获取，在解放我们的同时也在束缚着我们。

互联网与大数据的时代，似乎是一个人人为保护隐私而恐慌的时代，其实却是一个绝大部分人乐在其中不觉奇怪的时代。我们向互联网企业提供个人信息换取便利，我们向国家敞开胸怀以获得保护，我们把数据信息存储在云端，我们将个人的生活点滴在社交网络上进行分享，而与我们息息相关的信息碎片，广泛存在于虚拟与现实的世界里。当大数据技术成为一种革命力量席卷全球时，不论是在商业还是公共领域，大数据的运用给我们带来了诸多挑战：对互联网用户网上行为的定位与跟踪，对消费者个人敏感信息的记录与采集，数据预测分析带来的不公与歧视，云中信息的



此为试读，需要完整PDF请访问：www.ertongbook.com

安全顾虑与公权力机构的窥视，个人信息的频繁泄露与黑市买卖，网络侵权与电信诈骗以及信息跨境转移的法律冲突与数据壁垒……都使个人信息的收集方式、使用目的及影响后果日趋失控，个人信息的保护被前所未有地放置于洪荒之地，面临严峻威胁。面对这样的挑战，传统框架下用以保护个人信息与隐私的核心技术与法律手段呈现出有效性不足的特征。信息保护的最小采集原则、目的限制原则停留在纸面，没能在信息采集、传递、使用的过程中真正起到约束作用。而知情同意、匿名化、模糊化等传统隐私保护策略已不能有效保护个人的信息与隐私。

互联网技术在短短二十年的商业化浪潮中，以前所未有的速度谱写着改变世界的产业传奇，创造着全新的产业生态和经济模式。然而，信息技术的迅猛发展早已把法律远远地抛在了身后。大数据时代的商业模式缺乏保护个人信息的动因，个人信息保护现状不仅存在制度性缺陷又缺乏道德伦理的指引，再加上公权力机构基于国家安全、侦破犯罪、社会治理等需要，对个人信息保护的轻视，都使个人信息缺乏有效保护。强大的数据处理能力与普遍存在的对个人信息的侵害，使保护个人信息的呼声日益高涨。

随着信息科学技术的发展，信息的交流、传递和获取比以往任何时代都更加自由和迅速。在互联网带来的社会变革中，我们能便捷地获得商品与服务，能以前所未有的方式获得新知，交流与分享观点。这些发展革新了我们的自我表现，强化了我们的自由，为处于动态发展中的个人信息提供保护，为主体构建一个能自由呼吸的空间，它为人们的自我属性与社会性属性进行着边界管理。在伴随着互联网成长的过程中，在波及全球的保卫个人信息的战场上，中国身处复杂情境，在改革开放的进程中，无论是个人信息保护的理念还是相应的法律制度，都缺乏系统建构。伴随着都市化生活中个人信息保护意识的觉醒，个人信息保护的新时代已豁然降临，面对这汹涌而来的时代浪潮，背负传统前行的中国理应思考该何去何从。

作者

2018 年 12 月

# 目 录

# 第一章　大数据时代个人信息保护面临的典型风险

研究大数据的先驱麦肯锡给出这样的定义，大数据是一种规模在获取、存储、管理、分析等方面都大大超出了常规数据库软件工具能力范围的数据集合。具有海量的数据规模、快速的数据流转、多样的数据类型与价值密度低的四大特征。[1]学者维克托·迈尔－舍恩伯格则认为，大数据关注的是所有数据而不是随机样本，它并不注重精确性，而是注重多样与混杂性，它放弃对因果关系的追求，取而代之的是对相关关系的关注。[2]

我们所处的时代是一个人人有终端、物物可传感、处处可上网、时时在链接的时代。从科学研究到电子商务、从医疗卫生到社交娱乐，数据信息都呈爆发式增长。互联网公司已被海量数据包围，谷歌公司每天处理的数据是美国国家图书馆所有纸质藏书所含数据量的上千倍；脸书上人们在网站上点击“喜欢”按钮或写评论的次数，每天都超过30亿次；截至2012年，Twitter上的信息发布量，每天都会超过4亿条。[3]

人们自己生成并分享了与自己相关的大量信息；同时，也有大量的信息在信息主体没有参与甚至不知情的情况下，通过他人及相关机构生成。

---

〔1〕 *Big Data: The Next Frontier for Innovation, Competition, and Productivity,* McKinSey Global Institute Report, (2011-05) [2015-10-11], http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

〔2〕 参见[英]维克托·迈尔－舍恩伯格、[英]肯尼思·库克耶：《大数据时代》，盛杨燕、周涛译，浙江人民出版社2013年版，第27-96页。

〔3〕 参见[英]维克托·迈尔－舍恩伯格、[英]肯尼思·库克耶：《大数据时代》，盛杨燕、周涛译，浙江人民出版社2013年版，第11页。



此为试读，需要完整PDF请访问：www.ertongbook.com

我们的搜索历史、位置信息、网页浏览习惯、阅读习惯，休闲爱好、信用信息[1]等，被采集与使用的程度都远远超过我们的想象。互联网公司以前所未有的方式储存、分析用户的个人数据，而它们采集、分析数据的能力也与日俱增。[2]

大数据给我们的生活、工作与思维带来了急剧变革。移动互联的发展、手持设备的兴起、智能终端的普及、新型传感的应用都快速渗透到了地球上的每个角落。随着信息技术的发展，新的创新形式——大数据，呈现的是一种信息处理的能力。它能在极短的时间里筛选、分类、访问大量的数据信息。过程涉及数据挖掘、提取、储存、预测分析与对分析结果的应用等。

数据信息中潜藏的价值刺激了各路竞争者对数据的挖掘。大数据成为一种新的经济资产，如同货币或黄金。当今的技术可以进行前所未有的数据匹配，反匿名化操作，数据挖掘，这些都归因于广泛存在的数字档案与信息技术的发展。我们向互联网企业提交个人信息换取便利，我们向国家敞开胸怀以获得保护，我们把数据信息储存在云端，我们将个人的点滴生活在社交网络上分享。而个人信息的收集方式、使用目的及影响后果日趋失控，个人信息的保护被前所未有地放置于洪荒之地，面临严峻威胁。当大数据的技术成为一种革命力量席卷全球时，我们需清醒地认识到个人信息保护面临的严峻威胁。

---

〔1〕2014年腾讯征信、芝麻信用等八家民营机构取得了征信业务从业牌照。其中，依托阿里云的技术力量，芝麻信用包含了对3亿多实名个人、3700多万户中小微企业数据的整合。而腾讯征信则拥有8亿的QQ账户，超过5亿的微信账户，超过3亿的支付用户，以及QQ空间、腾讯网、QQ邮箱等多重服务上聚集的庞大用户资源与数据。互联网公司通过海量数据挖掘和分析技术可在更广范围预测用户的风险表现和个人信用，并将社交数据纳入信贷应用与生活应用中。参见姜琳、吴雨：《从哪来？如何用？咋保护？——三问个人征信市场“开闸”》，新华网，(2012-01-15)[2015-01-15]，http://news.xinhuanet.com/politics/2015-01/15/c_1114003768.htm；《央行要求腾讯征信等八家机构要做好个人征信准备》，中国新闻网，(2015-01-06)[2015-12-20]，http://news.xinhuanet.com/politics/2015-01/06/c_127361814.htm。

〔2〕2014年7月，阿里云计算发布了大数据产品——ODPS。通过ODPS在线服务，小型公司只需花上几百元就能分析海量数据，而ODPS可在6小时内处理100PB数据，相当于1亿部高清电影。参见《阿里云发布大数据产品ODPS 6小时处理100PB数据》，新浪科技，(2014-07-08)[2015-12-20]，http://tech.sina.com.cn/it/2014-07-08/14289482287.shtml。

# 第一节　网上行为的定位跟踪

早在 1971 年，学者米勒（Authur Miller）就曾谈到："电子计算机将使得预测个体或群体行为的虚拟活动成为可能。"[1]他同时也忧心，在未来，相关的组织机构会利用计算机与信息技术去描绘与影响消费者的行为，操控消费者的行为选择。[2]这样的预言在当今成为现实。

## 一、行为营销与广告支撑的生态系统

过去，各大公司采用的营销模式都是大众营销。这种营销模式利用的是一般人口统计学信息，在特定的期刊与特定的电视节目中投放广告。随着信息时代的来临，当今的广告公司在投放广告时，非常具有针对性。它们跟踪互联网用户的在线行为，研究消费者的需求、分析消费者的喜好并向消费者投放他们感兴趣的广告。如今，网络广告早已实现使用技术对用户信息进行全方位的精准分析，并根据分析结果，有针对性地向消费者定向投放广告，实现广告的精确营销。[3]这种收集不同人的喜好、行为、特点，并据此投放不同广告的方式就是行为营销。它利用个人过去的行为数据信息与其他人在相似情况下会做出的反应，对目标消费者进行一对一产品、服务的广告推送。学者达文波特（Thomas Davenport）与哈里斯（Jeanne Harris）指出，行为营销在进行分析时，需要充分利用自己掌握的或从第三方得来的信息，进行统计学分析。各商业机构基于对分析结果的判断，再对营销模式进行预测，决定向消费者推送的广告类别。[4]

---

〔1〕 Miller A.R., The Assault on Privacy: Computers, Data Banks and Dossiers, Signet: 1972, p. 42.

〔2〕 ibid, pp. 42-43.

〔3〕 参见林子杉：《互联网精准营销，是在偷窥还是帮助用户》，载《人民法院报》2015-09-21，第 6 版。

〔4〕 Davenport T.H., Harris J.G., *Competing on Analytics*, Boston: Harvard Business School Press, 2007, p. 7.

商业网站建立之初，网络广告就成为互联网经济增长的一个强劲动力。普华永道在2015年6月发布的《2015—2019年全球娱乐及媒体行业展望》的报告中预计，全球互联网广告总收益将从2014年的1354.2亿元增长到2019年的2398.9亿元。未来五年，互联网广告占中国广告市场的份额会不断攀升，估计将达39%~48%。[1]另外，有研究也揭示出，在美国，由广告支持的互联网企业为美国提供了数百万个与就业相关的重要职位，其中，每年在互动销售领域就能为美国贡献数十亿美元的经济增长份额。[2]百度2014年的广告收入总计将近500亿元，2015年将突破700亿；腾讯公司前两年的年报也显示，其年广告收入亦增长89%。[3]这当中，行为营销贡献巨大。在行为营销基础上搭建起的网络广告市场，不容小觑，将会呈现不断上涨的态势。

在这个广告支撑的生态系统中，参与者众多。有互联网用户、网站、互联网广告商、企业及一系列提供分析或保险类服务或导出共享数据的经济实体。这是一批令人眼花缭乱的公司。总体来看，这些公司包括社交网站、新闻媒体、购物网站及其他在线或离线的产品、服务的销售商。它们从用户处采集的信息不仅供自己使用，也常常转售第三方。这些信息可能会用于有针对性的互联网广告开发也可能用作其他用途。事实上，互联网用户根本无法理解自身信息在市场各个层面中被交易的商品化程度。网络广告自身成为一个能让大数据立足扎根并蓬勃发展的行业。越来越精准的行为营销分析，包含了消费者的地理位置、兴趣爱好与即时需求，而对于针对性强的行为营销广告，有研究表明，广告商愿意为此多支付

〔1〕参见刘亚澜：《普华永道：互联网广告中国份额或达到48%》，(2015-06-03)[2015-12-08], http://tech.qq.com/a/20150603/029575.htm.

〔2〕Deighton J., Kornfeld L., "Economic Value of the Advertising-Supported Internet Ecosystem-2012", Interactive Advertising Bureau, (2012-09-30) [2015-12-30], http://www.iab.com/insights/economic-value-of-the-advertising-supported-internet-ecosystem/.

〔3〕参见林子杉：《互联网精准营销，是在偷窥还是帮助用户》，载《人民法院报》2015-09-21，第6版。

60%~200% 的费用。[1] 行为营销背后潜藏的巨大商业利益，使得在线广告商无法控制自身的行为，而能从行为营销中获利的网站也不断采用新技术对用户的网上行为进行跟踪。

## 二、新兴跟踪技术的迅速发展

互联网时代，为了追踪、分析与说服消费者，广告商已经开发出了很多便捷与成熟的营销跟踪技术，在线广告营销伴随着每一个上网浏览网页的用户。广告行业借助不同的技术，如 Cookies、Flash cookies、beacons、Html5 canvas fingerprinting，对用户行为进行追踪。而商业机构还在不断投入大量的资金与技术人员，对新型跟踪技术进行研发。

Cookies 是网站服务器在用户的内存或硬盘中保存的用来记录用户浏览的网页地址、网页停留时间、网页上键入的用户名、密码、用户浏览习惯等方面的小型浏览文件。它并非是由本机的浏览器生成，而是当我们浏览网页时，从所浏览的网站发送过来的，用来检测我们在做什么的小型数据包，只不过通过浏览器保存在了本机上。它不仅可以对用户行为进行跟踪，还可以为用户推荐曾经访问的网址，省去用户重新输入网址的麻烦，用户不必重新输入用户名和密码，就能实现登录。

Cookies 引发的最大问题是在用户完全不知情的背景下，对用户行为进行跟踪、记录，这往往会引发第三方（如行为广告商）的介入。广告商在采集到 Cookies 数据后，会有针对性地通过行为营销的方式向用户投放其可能感兴趣的广告，而在线广告商会向网站服务商支付报酬。行为营销的技术通常在用户不知情的情况下使用。因此，这种模式会涉及对用户个人信息的侵犯。[2]

---

〔1〕 Beales J.H., Eisenach J.A., "An Empirical Analysis of the Value of Information Sharing in the Market for Online Content", *Navigant Economics*, January,2014, http://www.aboutads.info/resource/fullvalueinfostudy.pdf.

〔2〕 参见石佳友：《网络环境下的个人信息保护立法》，载《苏州大学学报（哲学社会科学版）》2012 年第 6 期，第 89 页。

随着信息技术的发展，网站开发人员找到了一种更好的方法——Flash cookies，用以跟踪用户的网上行为。网站的开发人员认识到，传统的 HTTP 下的 cookies 并不稳定，用户可能会随时清除掉浏览器中的 HTTP Cookies，或者在浏览器选项中，手动将它设置为禁用模式。加利福尼亚大学伯克利分校的研究人员指出，被用户删除的 HTTP Cookies 可以利用 Flash cookies 中的信息进行重写，获得重生，这样 HTTP cookies 里原来保存的数据就会重新呈现在分析者面前。用户所采用的传统禁用或清除浏览器中 Cookies 的方法，无法抗衡网站对用户网上浏览历史的重写、跟踪与记录。[1]

Becons 技术则可以让广告公司实时监控用户浏览网页的行为。监控的内容包括用户如何移动鼠标、打印了哪些信息、检索了哪些信息、在表格中填入了哪些信息。随着营销跟踪技术的不断发展，一些网络服务提供商已经开始使用“深层封包监控技术”监控公司用户在互联网上的行为。[2]

近两年，很多网站与跟踪软件都开始使用 Html5 Canvas Fingerprinting 这项新技术对互联网用户的网上行为进行跟踪。实际上，每一个浏览器都有自己的特征。网站可以检测用户的浏览器版本、操作系统类型、安装的浏览器插件、屏幕分辨率、所在时区、下载的字体及其他信息。[3] 而 Canvas 是 Html5 中动态绘图的标签，每一种浏览器会使用不同的图像处理引擎，不同的导出选项，不同的压缩等级，这会使得每一台电脑绘制出的图形有些不同，而这些图案能够用来给用户分配特定编号，这个编号被视为是“指纹”，能够用来识别不同的用户。[4] 学者埃克斯利（Peter

〔1〕 Soltani A., Cantly S., Mayo Q. et al., “Flash Cookies and Privacy”, AAAI Spring Symposium: “Intelligent Information Privacy Management”, Palo Alto, March 22-24, 2010; Ayenson M. D., Wambach D. J., Soltani A., et al., “Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning”, http://dx.doi.org/10.2139/ssrn.1898390.

〔2〕 “Deep Packet Inspection and Privacy”, Electronic privacy information center, http://epic.org/privacy/dpi.

〔3〕 参见《网站是如何跟踪监视你的》, (2013-12-12)[2015-03-22], http://www.bitscn.com/network/protect/201405/199320.html。

〔4〕 参见《基于 HTML5 的 Canvas 指纹跟踪技术》, (2014-08-15) [2015-07-22], http://bbs.csdn.net/topics/390861386?page=1.

Eckersley）进一步指出了与浏览器指纹相关的隐私风险，其中服务器端程序可以向浏览器查询相关配置的信息，将特定的计算机识别出来。[1]有学者指出，如果要避免指纹跟踪，人们需要禁用网站的一些关键功能，如JavaScript与Adobe's Flah技术。[2]而截至2010年，BlueCava，一家在线跟踪公司，声称拥有2亿个设备的指纹。[3]即使是电脑高手，面对Canvas指纹跟踪技术，也很难保护自己的隐私。Html5 Canvas在使用功能上，不仅可应用于图片的处理，它还能应用到监视用户的鼠标移动、键盘输入等。目前来看Canvas指纹跟踪很难被阻挡，只要用户使用浏览器上网，用户的网上行踪就如同处于公开状态一般。

大量的实证研究揭示出互联网广告商不断地使用新的、不为人们熟知的技术跟踪消费者行为，很多跟踪技术对于众多消费者而言，甚至都不曾听闻。各大网站普遍使用的行为跟踪技术，在隐私保护的政策中很少得到披露。即使是反跟踪工具与最强大的信息隐私保护设置，也不能有效且普遍地对它进行制衡。消费者在互联网上对跟踪技术进行防范与选择，往往也达不到效用。新兴跟踪技术的发展，已成为了网络信息隐私保护领域的新威胁。纯粹运用市场调节的手段，将损害到消费者践行个人自治的实践与能力，消费者手中的选择权，也将成为一项徒有虚名的权利。此领域需要引入法律干预。[4]

如何从法律规制的路径阻断Cookies、Flash cookies、beacons、HTML5 Canvas Fingerprinting等技术对我们的持续跟踪、骚扰，如何披露并告知用户谁是其背后的控制者，并为用户提供是否分享个人信息的选择权，成为

〔1〕 Eckersley P., "How Unique Is Your Web Browser?" 6205 Lecture Notes Computer Sci. 1 (2010).

〔2〕 Hoofnagle C. J., Soltani A., Good N., et al., "Behavioral Advertising: The Offer You Cannot Refuse", *Harvard law & Policy Review,* vol.6, issue.2, 2012, p. 285.

〔3〕 Angwin J., Valentino-DeVries J., "Race Is On to 'Fingerprint' Phones, PCs", Wall St. J., (2010-11-30) [2015-10-05], at A1.

〔4〕 Hoofnagle C. J., Soltani A., Good N., et al., "Behavioral Advertising: The Offer You Cannot Refuse", *Harvard law & Policy Review,* vol.6, issue.2, 2012, p. 285. ibid, pp. 273-274.



此为试读，需要完整PDF请访问：www.ertongbook.com