

BIG DATA

大数据真相

谁动了我的数据？

杨绪宾 刘 洋 编著

BIG DATA



华南理工大学出版社
SOUTH CHINA UNIVERSITY OF TECHNOLOGY PRESS

BIG DATA

大数据真相

谁动了我的数据？

杨绪宾 刘 洋 编著



华南理工大学出版社
SOUTH CHINA UNIVERSITY OF TECHNOLOGY PRESS

· 广州 ·

图书在版编目 (CIP) 数据

大数据真相：谁动了我的数据？/杨绪宾，刘洋编著. —广州：华南理工大学出版社，2018.8

ISBN 978 - 7 - 5623 - 5726 - 1

I. ①大… II. ①杨… ②刘… III. ①数据处理 - 研究 IV. ①TP274

中国版本图书馆 CIP 数据核字 (2018) 第 178541 号

Dashuju Zhenxiang: Shui Dongle Wo De Shuju?

大数据真相：谁动了我的数据？

杨绪宾 刘 洋 编著

出 版 人：卢家明

出版发行：华南理工大学出版社

(广州五山华南理工大学 17 号楼，邮编 510640)

<http://www.scutpress.com.cn> E-mail: scute13@scut.edu.cn

营销部电话：020 - 87113487 87111048 (传真)

策划编辑：袁 泽 刘 锋

责任编辑：袁 泽 王荷英 刘 锋

印 刷 者：广州市新怡印务有限公司

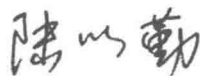
开 本：787mm × 960mm 1/16 印张：14.5 字数：266 千

版 次：2018 年 8 月第 1 版 2018 年 8 月第 1 次印刷

定 价：68.00 元

版权所有 盗版必究 印装差错 负责调换

大数据的发展是技术驱动的，与需求驱动的技术不同，当数据一旦成为核心资源被不断“挖掘”其价值的时候，数据所有者对如何保护自己的利益是非常模糊的。大部分组织和个人，一方面对自己的数据被非法利用或者隐私被泄露的现象深恶痛绝；另一方面对自己拥有多少数据资产，拥有什么数据资产，谁在利用自己的数据资产，如何保护自己的数据资产而不得其解。《大数据真相》针对这些社会热点问题，用生动活泼、接地气的写法系统而生动地揭示了“大数据真相”，相信本书会带给人们更多更新的思考。



广东省计算机信息网络安全协会会长、华南理工大学博士生导师、教授

《大数据真相》将与每个人息息相关的个人大数据抽丝剥茧，分析了大数据的经济价值，对大数据如何成为资产、如何转化为财富进行了深入浅出的阐述，揭示未来财富增长的“秘密”，并预言数据可以让个体在虚拟世界中实现“永生”的神话！未来已来，数字化生活就在眼前！让作者带我们进入数据时代的变革吧！



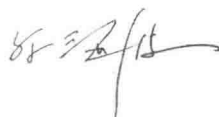
蓝盾股份有限公司副董事长兼总经理

《大数据真相》从保护个人数据安全角度出发，告诉我们个人数据的重要性及防范个人数据泄露的主要方法。学习和思考本书的内容，有助于我们懂得如何使用好数据、发挥数据的价值，促进大数据朝更好的方向发展。



启明星辰集团高级副总裁

大数据的应用越来越广泛，不论数据还是其他事物都具有多面性。《大数据真相》从数据的发展等多方位视角进行说明，告诉我们保护个人数据的同时也要合理利用数据，让数据从正能量的角度带给我们更加美好的生活。



浪潮集团高级副总裁、广东区总经理

大数据已渗透到各行各业中，是高新科技时代的产物，它正在协助众多企业不断发展新业务、创新运营模式、促进企业精准营销，让各方面的资源得到最大化利用。《大数据真相》从人们生活中所涉及的数据入手，深入阐述大数据与企业资本收益的关系、大数据与发展趋势的关系。整本书着眼于时代动向，意义深刻。大数据，让未来可感可知！



广州市宏方网络科技有限公司创始人

人类的文明历史是一个不断被数据化的过程，形成了对应于物理世界的数字世界。在数据信息大爆炸的今天，人类已走到全新文明的前夜，对数据世界和数据权利的探研是整个人类层面的重大课题。《大数据真相》对数据本质、数据权利、数据使用、数据未来做出了新颖独到且系统深刻的前瞻探析，意义重大，价值非凡，是透悉数据世界的上乘佳作！



大西洲科技公司总裁

大数据技术的飞速发展，使大数据应用悄然渗透到人们生活的方方面面。当人们热议着大数据给生活带来便利的时候、当 IT 大鳄享受着大数据饕餮盛宴的时候，作者以冷静的观察、缜密的思维、全面的视角，为你打开了了解大数据真相的另一扇窗……



广东省计算技术应用研究所咨询部部长

大数据时代，你还有隐私吗？

不知何时，你的手机暴响，一个暧昧的女声问你：大哥，你最近股票赚吗？你将手机调为振动。不久，它又开始肆意地振荡：先生，我们是××养生堂，中年男人普遍压力过大，需要功能调理，我们真诚地……有朋友都快崩溃了：自他去过男科医院，他的手机和座机便不得安生，一会儿介绍“威尔刚”，一会儿推荐“秃鸡散”，一会儿强调“秘鲁丹”——总而言之，所有的声音对他都是一种暗示性羞辱：阳痿、阳痿、阳痿……“我要告他们!!!”他高叫，他跺脚，但是他该告谁呢？他去过的男科医院或相关门诊并非一两家，所有的病历卡必须填上相关资讯，可以是手机，也可以是座机，但绝不容许空白，因此所有医院都有出卖他的可能。

理财推荐、房屋中介、移民指南——银行、保险、医院、电信、快递、网站……个个都拥有你的资讯。到底是谁把我们出卖了？是谁把这些鬼魅引来的？是谁让我们惹上了无穷无尽的询问、搭讪、骚扰、窥探？

作为普通人，你可能对自己的数据被利用到何种境地还缺乏感受。新闻里的“徐玉玉案”令公众不寒而栗——因为考生信息被泄露，家境困难的高中生徐玉玉，在接到大学录取通知书后，随即也接到一通电信诈骗电话，被以办理助学金为由骗走9900元学费，伤心欲绝之下，年轻的女孩心脏骤停而亡。

某朋友至今对某邮箱“疑似被黑”的事故印象深刻——过亿条用户名、密码、登录IP、生日等信息被窃。尽管该公司否认数据库遭到攻击，称是黑客获得了部分用户在其他平台相同的账号和密码，撞库

所得。有人养成好习惯，每次接到快递包裹，都会用黑色记号笔逐条划掉快递包裹上的所有信息，包括收货人姓名、联系方式、收货地址等等。因小区楼下收废品的人，可能会把纸箱上的信息单撕下来，然后转手倒卖出去。不久这笔数据交易会招致一通推销甚至诈骗电话找上门来。这些事可能发生在每个人身上，平日多做些预防工作总没错。但不为人知的是，无论你划掉多少张快递单，个人信息泄露防不胜防。有朋友说收废品捡快递单的方法效率太 low 了，现在都直接从淘宝店主那端买发货单，5 块钱一张。

世界从未变得如此数据驱动。大家都在谈论人工智能、个性化推荐、精准营销，世界从未如此渴望知道“你”是谁；从金融到医疗，从广告到电商，各行各业从未像现在这样对数据充满渴求。

大家都意识到了：数据，是新商业时代最重要的议题。

一位行业内人士透露，为了描述和分析一个用户画像，阿里巴巴构建了 741 个维度来收集数据。“弱数据甚至更多。所有的数据，你买过什么，购买频率和价格，你住在哪里，你有多少钱，他全知道。”

当前大公司之间为了争夺数据，爆发了前所未有的争斗。某物流公司举报对手盗取 6000 万条竞争对手货运数据，某微博诉对手过度攫取用户数据，腾讯控诉华为 Magic 手机侵犯用户隐私……过去则少有这样的情况。两年前，马云说阿里巴巴要做数据公司、未来最大的能源是数据时，大家可能根本没有意识到数据的价值。

大数据时代，商业和用户隐私之间发生激烈对撞。

大数据、人工智能改变世界，提升诊断的效率和准确度，这是一幅美好蓝图，唯独忽略了一点：在不知情的情况下，你的隐私，比如医疗数据正从医生电脑里流向商业公司。

尽管技术无罪，但商业对数据的贪婪缺乏克制。在利益驱使下，各种诸如网络爬虫、盗取手机 Root 权限的技术，让个人隐私数据前所未有的暴露在交易市场上。

隐私数据与商业的冲撞是如此激烈，引起了国家的高度关注。2017 年，公安部和国家网信办两次集中审查大数据企业，多家数据公

司被询问数据来源和运营模式。不仅如此，掌握大量个人信息的互联网公司，比如某招聘网站，也收到了执法部门的警示。政府在为我们不断打造安全数据应用环境，但同时也需要我们自己有很好的防范意识。

通过科技手段收集来的数据，像传统的问卷调查一样，也可淡化甚至完全抹去用户的个人信息。大数据时代，我们怎么确定这些企业搜集信息的过程中，不会精准定位到个人呢？政府不能保证，企业无法保证，广告商同样无法保证。

“斯诺登事件”大家肯定都知道，英国《卫报》和美国《华盛顿邮报》在2013年报道，美国国家安全局（NSA）和联邦调查局（FBI）于2007年启动了一个代号为“棱镜”的秘密监控项目，直接进入美国国际网络公司的中心服务器里挖掘数据、收集情报，包括微软、雅虎、谷歌、苹果等在内的9家国际网络巨头皆参与其中。

多名专家表示，在“斯诺登事件”之后，绝对隐私已经不存在。物联网的兴起，则更加剧这种矛盾。如今任何物品都可以通过传感器搜集和发送信息，智能家居让家具不仅会记录你的一言一行，还会把这个数据汇入终端；你在某购物网站搜索的商品，会出现在其他网站的广告推送中。这需要产业自律、政府监管，但最主要的还是提高自我防范意识。

如今，我们已离不开用数据来驱动社会的发展。根据相关统计，2017年微信用户已经超过了9亿，支付宝用户已经超过5亿，京东日均订单量已经达到29.5万单……新经济随着现代商业的发展已融入所有人的生活中。

而身处这样一个大数据时代，我们除了“裸奔”，似乎别无选择。

新时代“数据财富真相”

本书作者之一刘洋参与过很多城市的土地和房地产开发，认为土地开发权和数据开发权具有很多相似之处，是将土地开发权引用到数据开发权的第一畅想人。他认为土地开发权和数据开发权都依赖“大数据”，譬如对于规划师，一个楼盘只是一个设计项目，容积率对规划师只是一个数字，但对开发商却意味着金钱，是一栋栋“固定的银行”。数据开发权也同样如此，对于个人，它是一个个微不足道的数据，但对于网络科技寡头而言，是一个个“移动的数据银行”。

大数据在当今社会是一个巨大的财富池，你就是财富池的创造者。房产开发商和网络科技巨头都是拥有数量巨大的资源才雄踞富豪排行榜前列！特别是网络科技巨头们，他们财富的背后是对海量的个人数据进行的技术处理和应用。“技术虽好，只能养家；生意虽小，却能发家”。商业时代的财富逻辑就是现在的大数据，无论你拥有什么，只要数量巨大，那就是财富。“大妈月入三万，还会差你一个鸡蛋”的故事大家都知道，一句月入三万让多少“北上广”的白领们捂脸而去。一个鸡蛋虽然只能挣你五毛钱，但一个月轻轻松松地卖十万个，这背后的逻辑就是大数据。所以，我们不要轻看自己的数据价值，数据价值虽小，但是不容侵犯！况且，你的数据积少成多，也会成就你个人的大数据资产！

数据可谓是这个时代的“石油”。根本不用企业主动出击，网络大数据几乎可以搞定一切。我们听的每一首歌，叫的每一次外卖，搭的每一次车，其数据都可以被获取，成为海量数据中的一点一滴，被用于深度数据分析，转而用于商业领域，实现商业价值。数据的应用具

有多面性，政府会正面引导，我们也要有保护个人数据的意识。

土地被称作“财富之母”，类比土地，数据则是新一代的“财富之母”。数据与土地相比，人人都拥有，人人都创造。世界上大多数人拥有一块土地很难，但是拥有个人数据却很容易。个人数据如影随形，不离不弃。数据与土地相比，还有一个无与伦比的优势，那就是土地只能开发一次，而数据可以无限次地开发利用；它也不像机器、厂房一样会随着使用次数的增多而贬值，相反，数据很容易实现共享，而且使用的人越多，重复使用次数越多，其价值越大。土地需要开发商从政府手中购买使用权，然后再根据容积率等规范要求行使开发权。我们的数据是否也应该有规范来行使开发权？

关心你的数据，它与你在这个世界上的存在密切相关！

前言

你的数据，你的权利

我们已经建立了一个浩瀚程度超乎我们理解的网络世界。可以这样形容这个网络的大小：地球上每个人可以有 50 000 个网址，可以为 50 亿台设备提供服务，到 2020 年预计有 220 亿台设备互联。

在这个网络上我们创造着无比浩瀚的数据。智能手机能对你的健康状况做出记录，可以方便搜集任何想要的信息，计算出你前往机场的最佳时间；当你快要到家时，家里的灯光会打开，空调也会自动调节到你选择的适宜温度；购物助手可以告诉你，你需要什么样的书、什么样的日用品，你喜欢什么样的电影。

毋庸置疑，科技的发展是为了让我们生活更方便。我们是否可以问一下这些我们交付了信任的企业，我们享受舒适的生活时是否就应该舍弃我们的数据权利？就像互联网大佬所说的“以隐私换方便”？我们必须回答这些有关生活方式的深层次疑问。谁拥有哪些信息，信息公开应该遵循哪些原则？能否进行有效数据管理，如果可以，该从哪里着手？政府、企业和网民应该发挥什么样的作用？

如今，我们增加一个新的电子设备，基本就“舍弃”自己的一些数据权利。我们对数据交给了谁一无所知，对这个拿到数据的组织的道德水平和价值观更是一无所知。表面上我们会被告知数据“尚待开发”，但是谁知道这些企业拿到我们的数据后会做什么？实际上我们看到的只是企业的营销策略表象，而背后“开发”我们数据的人姓甚名谁，道德水准如何？我们知之甚少——而他们却对我们了如指掌。长远来看，随着网络接入点的增加，我们也要问问自己愿意交出多少个人数据。

互联网刚出现的时候，潜意识中我们感觉网络会知道我们在哪里、我们看了什么。那时候，一个人的信息分布在不同的“知情者”手里：银行知道一些，医院知道一些，政府也知道一些。但现在互联网巨头可谓已经掌握了

我们几乎全部的资料——衣、食、住、行、游、购、娱，无所不包，并将其储存在一个触手可及的地方。这样做当然方便，不过如果存储者用某种我们并不那么首肯的方式来使用这些数据时，是否应该另当别论？这就不是“以隐私换方便”这么简单，而是涉及法律问题。现在个人数据已经成为资产，现行的网络商业模式该如何去补偿数据的创造者？尤其值得一提的是，广大网民要求互联网企业停止使用那些令人费解的服务协议，而对个人数据采集和销售却没有提出任何建议或要求。

曾经有互联网公司用新闻订阅信息来测试阅读新闻是否会对其情绪造成影响，这引起了公众的强烈抗议。事实上，我们并不记得有过这样一个授权选项，起码出发点绝不是如此。

医疗数据被视为个人隐私，但为了治病，病人会对医生如实透露信息。现在，医疗和科技产业之间的界限已经变得模糊，一些“可穿戴设备”和软件的制造商就在不停宣扬，让他们的产品不被列为医疗用品，这样各种信息保护监管就游离在医疗条例之外。

隐私仅仅是关于数据所有权、数据垄断、安全、竞争等群众热点问题的一部分。民众关心的热点问题还包括数据控制、大数据未来等问题，当然也包括数据使用权，即选择个人数据被使用的方式，自己的数据被什么人在使用等问题。

一些对此问题有认识的企业已经有所准备，会成立风险委员会，会有人专门处理有关数据道德守则的问题，还会监管数据如何收集和使用——尽管有时候也不是监管而是在试错。一些后来进入的小公司可能没有如此正式的协议也没有人力来处理这种问题，可能也没有独立的董事会成员来专门管理此事。如果真的发生严重的问题，很多消费者就不再使用他们的服务，无论他们的商业模式前景多么无限。

我们都喜欢尝试新的应用，因此会把我们的QQ或者微信账号作为其他应用的入口，这样等于不加思考地把个人信息从监管比较严格的大公司转移到部分管理没有那么严格的小公司。消费者都相信或期待会有专门人员来管理这些事情，但事实上，这个人是谁却无从知晓。我们应当如何解决这个问题？

近期来看，新兴企业不应该将个人数据仅仅作为营销手段，而应该作为一个核心战略来考虑。政府相关部门更应该进行自我教育，同时对公众进行

教育并实行更强有力的监管，就像许多国家几十年前监管汽车安全带的使用一样，应当发起公共安全教育行动，从司法解释和宣传推广等方面来开展工作。

这些信任问题不是单靠政府监管就能解决的。科技企业必须想方设法去补偿数据创造者，以及开展受客户欢迎的创新性业务来赢得用户的信任。事实上，任何创新型企业都必须这样做才能取得消费者的信任。

未来世界里所有的生活细节都会转化为数据（包括人际关系），所有的一切都会永存于网络世界。在这种情况下个人能够保护和控制好个人数据，并从个人数据的开发利用得到实实在在的益处？

我们要重塑新的互联网秩序，使它惠及广大用户而不是变成吞没个人利益的“黑洞”。我们必须构建具有道德和价值导向的数字化生活方式，确保个人数据安全和公平利用是其核心，因为个人数据已经成为个人权利的重要基石。

在编写本书的过程中，引用了学术和业界有识之士的部分研讨观点，在此对以下单位和个人特别致谢（排名不分先后）：

广东省计算机信息网络安全协会会长、华南理工大学博士生导师陆以勤教授；

启明星辰集团常务副总裁张晓东、高级副总裁李春燕；

广东省计算技术应用研究所深圳分所负责人雷唯；

蓝盾股份有限公司董事长柯宗庆，副董事长、总经理柯宗贵；

浪潮集团副总裁、广东区总经理孙海波；

浙江大华技术股份有限公司中国区副总经理戴勇；

中国新电信集团董事会主席列海权；

广东省餐饮协会供应链大数据中心主任杨弘德（中国台湾著名供应链管理专家）；

大西洲科技公司总裁彭顺丰；

广州市宏方网络科技有限公司创始人董事长祝杰。

本书借鉴和引用了互联网部分文章、新闻报道，在此对原作者表示由衷的敬意和谢意！

一、从“暗偷”到“明抢”

个人大数据到底属于谁？ /1

数据利用乱象 /3

数据开发乱象 /15

华为与腾讯间的数据之争 /24

数据平台与用户的版权利益之争 /29

被遗忘的权利 /33

数据保护法刻不容缓！ /35

立法正逢其时 /36

二、从“IT”到“DT”

什么是个人大数据？ /39

万物皆数据 /41

大数据的内涵与外延 /42

“数据”与“信息” /43

数据化与数字化 /45

大数据时代 /46

大数据类型 /47

“量化”和“追踪” /50

人生即数据 /52

人性也是大数据 /59

延伸阅读：大数据时代人们的焦虑 /62

三、从“资源”到“资产”

如何理解数据资产和产权？ /69

数字经济 /71

数据资产 /73

大数据资产 /78

数据资产保护模式 /81

数据财产化 /87

数据权利体系与土地权利体系 /90

四、从“无视”到“重视”

如何掌控个人数据产权？ /103

个人数据流失真相 /105

个人数据泄露途径 /109

手机泄密途径 /111

黑客攻击 /117

AI 犯罪 /120

建立政府保护伞 /122

培育个人信息安全自我保护素养 /125

数字化生存 /128

五、从“积少”到“成多”

如何量化个人大数据产权？ /131

给数据估值 /133

数据资产增值 /136

区块链与数据价值衡量 /138

时间刻度 /146

“数说”即“数据说话” /148

个人数据价值量化 /151

颠覆微信！ /157

六、从“垄断”到“共享”

如何推动个人大数据产权交易？ /163

数据垄断的“三座大山” /165

国外的数据垄断 /168

数据资本家 /170

数据奴隶 /178

互联网金融 /179

数据流动 /185

数据重组 /186

开放公共数据 /187

完善数据资产市场体系 /189

七、从“消逝”到“永生”

如何步入未来的数据智能时代？ /193

数据为王！ /195

数据使用共享 /197

未来的竞争本质 /199

数据宗教 /200

人工智能的食物 /201

数字世界的牛顿定律 /205

万物互联世界 /206

数字化永生 /207

有人可成为神仙？ /210

未来人类命运 /212

信数据，得永生 /213

斯人已逝，数据永存 /215

一、从“暗偷”到“明抢” 个人大数据到底属于谁？

■数据可谓是这个时代最重要的资源，已成为各位“互联网大佬”的必争之物。但是由于数据产权法律关系的不甚明确，在社会实践中数据资源被各大巨头公司无偿“侵占”，并构成公司巨大估值的重要组成部分。各大公司为了经济利益对用户数据争来夺去，唯独忘记了产生数据的用户。离开了用户数据这个“1”，其后面估值再多的“0”都是没有意义的。

在“0”和“1”之间，国内由于立法等原因，现只能由网络安全类企业用技术手段解决恶意提取个人数据的问题。蓝盾股份有限公司副董事长、总经理柯宗贵先生曾提到：大数据技术的兴起，在对社会实现巨大贡献的同时，也因数据盗取、数据篡改、数据滥用等问题给社会以及个人带来了重大安全隐患，大数据安全防护迫在眉睫。