

大数据与隐私

利益博弈者、监管者和利益相关者



O'REILLY®

出版社

University Press

[美] 特伦斯·克雷格 玛丽·E. 卢德洛芙 著

赵亮 武青 译

该译著由东北大学哲学学科双一流学科平台建设项目、“辽宁省科技项目教育厅一般项目”（项目批号 W2015147）、“辽宁省科技项目社科项目省社科规划项目”（项目批号 L15BZX006）、“陈昌曙技术哲学发展基金”资助出版

大数据与隐私

——利益博弈者、监管者和利益相关者

[美] 特伦斯·克雷格 [美] 玛丽·E. 卢德洛芙 著

赵 亮 武 青 译



东北大学出版社

·沈 阳·

© 赵 亮 武 青 2016

图书在版编目 (CIP) 数据

大数据与隐私: 利益博弈者、监管者和利益相关者 /
(美) 特伦斯·克雷格, (美) 玛丽·E. 卢德洛芙著; 赵
亮, 武青译. —沈阳: 东北大学出版社, 2016.12
ISBN 978-7-5517-1510-2

I. ①大… II. ①特… ②玛… ③赵… ④武… III.
①计算机网络—数据处理—关系—隐私权—研究 IV.
①TP393.083②B82-057

中国版本图书馆 CIP 数据核字 (2016) 第 324707 号
辽宁省版权局著作权合同登记号 06-2016-233

出 版 者: 东北大学出版社

地址: 沈阳市和平区文化路三号巷 11 号

邮编: 110819

电话: 024-83680267 (社务部) 83687331 (市场部)

传真: 024-83687332 (总编室) 83680178 (出版部)

网址: <http://www.neupress.com>

E-mail: neuph@neupress.com

印 刷 者: 沈阳航空发动机研究所印刷厂

发 行 者: 东北大学出版社

幅面尺寸: 178mm×233mm

印 张: 6

字 数: 118 千字

出版时间: 2016 年 12 月第 1 版

印刷时间: 2016 年 12 月第 1 次印刷

组稿编辑: 刘振军

责任编辑: 孙德海

责任校对: 潘佳宁

封面设计: 张 健

责任出版: 唐敏志

ISBN 978-7-5517-1510-2

定 价: 25.00 元

O'Reilly Media, Inc.介绍

O'Reilly Media通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自1978年开始，O'Reilly一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了Make杂志，从而成为DIY革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版，在线服务或者面授课程，每一项O'Reilly的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar博客有口皆碑。”

——Wired

“O'Reilly凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference是聚集关键思想领袖的绝对典范。”

——CRN

“一本O'Reilly的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照Yogi Berra的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去Tim似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

前言

本书中的惯用格式

本书中使用如下格式约定：

斜体字——表示新出现的术语、URL、电子邮件地址、文件名与文件扩展名。

如何联系我们

美国：

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472

中国：

北京市西城区西直门南大街2号成铭大厦C座807室（100035）
奥莱利技术咨询（北京）有限公司

针对本书我们有一个专门网页，用于列出勘误表、示例及其他补充信息。您可以通过下列网址访问该网页：

<http://www.oreilly.com/catalog/9781449305000>

如希望评论或询问与本书相关的技术问题，请发送电子邮件到以下邮箱：

bookquestions@oreilly.com

想了解更多关于 O'Reilly 的图书、课程、会议及新闻等更多信息，请访问我们的网站：

<http://www.oreilly.com>

致 谢

没有来自很多人的帮助，我们就无法完成这本书的写作。

我们想要感谢我们的伴侣，他们的付出超过了期望，忍受我们平和的（如果真是这样的话）相互嘶吼，帮助校对并交流想法。无需多言，公司长期艰难运营，再加上我们周末的写作安排，使我们没有太多时间去做任何其他的事情。我们的伴侣因此受到的冲击最大，我们要永远感恩，我们做出了如此完美的对另一半的选择！

我们还想要感谢 Mike Loukides, Meghan Blanchette，以及奥莱利方面全体工作人员所给予的机会和支持。我们要特别感谢在我们的书稿出现延滞时，他们给予的温和敦促，这促使我们最终坚持到底。

我们要感谢《隐私身份创新》节目（PII）的执行制作人兼联合创始人 Natalie Fonseca。她举办的出色的研讨会教给我们太多从前所不知道的隐私方面的知识，我们要由衷感谢她对本书的大力支持。

不少朋友和同事都参与了本书书稿的审稿工作。我们要感谢他们所有人贡献的见解和评论。毫无疑问，他们所有人的努力使得这本书更加完善。

无需再多言，这本书是一道汤，而现在是时候在甲板上喝点鸡尾酒了！

特伦斯的个人致谢

致我的母亲、父亲和我美丽的妻子：没有你们，就没有我。致我的俄罗斯籍雇员，要特别向 Slavik 和 Sasha 致意；致谢 B. Sujata 博士和 Elan，每次当我听到成功如何毁灭了一个人，我就会想到你们是如何突破这个定律，谢谢你们的不断支持；向我得克萨斯和北卡罗来纳的家致谢（你们知道你们是谁，我稍微有些离题了）；致谢我的全体员工，包括过去的和现在的，感谢你们帮助我创立了 PatternBuilders 公司，最后，当然并非谢意的最末，致我的合著者和亲爱的朋友——玛丽，感谢这份在我职业生涯中最有意义的合作。

玛丽的个人致谢

感谢我的丈夫和姐妹们，他们帮我收拾了所有的烂摊子，并使我能始终情绪饱满地面对这次“针对隐私的劳作”。（等等，我的意思是爱！）致我最亲爱的表妹，感谢你不断提醒我为什么我应该定期去上语法进修课，感谢你有如此敏锐的洞察力，（是的，我应该交给你更多的时间！）致我所有的朋友和家人，感谢你们忍受我为了获得对不同观点更全面的掌握，而围绕与隐私有关的一切无休止的发问。最后，致我的合著者和同样（见上文）亲爱的朋友特伦斯：我无法再找出一个更好的人，来共同努力写作完成一本主题如此微妙而复杂的书了。尽管我们有过展开争论的时刻，但我们从未失去大局观，我相信，这一切都是为了造就出一本更好的书！

目 录

第一章 完美的风暴	1
穿过魔镜	2
欢迎来到大数据时代	4
从散落的拼图到完整的画面：幻想提前来到	5
大坏狼般的广告	5
老大哥和遍布世界的大数据	7
踟蹰不前：隐私与安全的矛盾张力	9
原书参考文献	10
第二章 数字时代的隐私权	14
数字时代隐私意味着什么	15
美国的隐私权：不被侵扰的权利	17
欧洲的隐私权：名誉与尊严	19
总处于有色眼镜之下的隐私	21
隐私无边界	21
价值观的冲突	22
网络隐私权：个体小“我”面对集体大“我”	23
原书参考文献	24
第三章 监管者	27
“数字”隐私监管简史	29
隐私监管模式——互补还是矛盾	29
美国的监管模式——自下而上的路径	31

联邦贸易委员会 (The Federal Trade Commission)	31
联邦通讯委员会 (The Federal Communications Commission)	32
商务部 (The Department of Commerce)	33
能源部 (The Department of Energy)	33
卫生与公众服务部 (The Department of Health and Human Services)	33
消费者金融保护局 (The Consumer Financial Protection Bureau)	33
对于美国模式的结语	34
欧盟的监管模式——自上而下的路径	34
安全港效应	36
关于欧盟模式的结语	37
其他国家隐私法律概览	37
隐私 VS 安全	37
永不消逝的数据	38
进步与否，我们都在一条船上	39
原书参考文献	39

第四章 利益博弈者45

认识利益博弈者	47
在线广告活动简史	48
知识产权、可信计算和隐私	50
挑战隐私的极限，直到极致	54
前所未有的可及性进一步削弱隐私期望值	57
把妖怪放出魔瓶	59
以隐私之名实施保护和提供服务的利益博弈者	62
正在崛起的隐私经济	63
利益博弈者各显神通，消费者隐私继续受侵	65
原书参考文献	66

第五章 透析大数据与隐私71

问题的核心：商品抑或权利	72
我们都相互连通	73
我们愿意为了安全付出何种代价	75
数据的真相：一旦暴露，难以控制	76

走完的闭环	77
原书参考文献	77

后 记79

特伦斯的感想	79
玛丽的感想	81

关于作者84

完美的风暴

如果像我们一样，你在过去的20年左右时间里都工作在高技术行业，那么你就会对当今关于数据隐私的讨论有一个全面的看法。无论你处于隐私问题象限的哪一个区间——从以一种合情合理的方式方法采集和使用你的数据，到对你的数据信息可能被层出不穷的途径和手段加以盗用所持的更具功利性的，有时也许是更具争议性的，甚至极端可疑的看待——可以肯定地说，风险已达到前所未有的高度。

一个完美的风暴正在逐渐形成。这个由创新活动所激发的风暴已经改变了我们相互沟通和交流的方式。在20年前，谁又能够预料到，互联网络将对我们的生活产生如此翻天覆地的影响？除了睡觉之外，我们无时无刻不连接在网上，用我们的手提电脑、移动电话或平板电脑查看电子邮件，阅读喜好的博文，搜索就餐地点和工作机会，浏览好友的脸谱墙，买书，转账，导航，分享我们的位置信息，在世界的任何一个地方抗议独裁。欢迎来到数字时代！

数字技术已经创造和培植出了一套新的世界秩序，很多之前不可想象的事情现在都有可能实现。也许我们还没有供个人使用的喷气式发动机或会飞的小汽车，但我们已经有了可视电话和无人机。也许我们并没有在乔治·奥威尔在其反乌托邦小说《1984》所预言的世界里生活过——在那个世界里不存在隐私权，政府利用监视和误导手段控制其民众——然而，有些国家的政府确实已经出于自身的利益在利用人们的私人信息，甚至它对人们的透彻了解要远远超出奥威尔的想象。

我们的世界已经发生改变，可能有人认为变得更好，而其他人则认为更糟。今天，我们泄露了更多关于自身的数据信息，我们比人类历史上任何族群都拥有更多用以聚集和聚合的关于我们自身的数据。其中大多数时候，我们只是出于单纯的方便及利用“免费”或几乎免费的服务的考虑，而泄露了自身的数据信息。其中有些数据则被悄悄地加以收集，或通

过主动的官方行为获取，例如2009年800万人质问美国司法部通过其在移动通讯设备里安装的全球定位系统快速大量锁定用户位置的行为。

我们的线下生活现在搬到了线上。我们用私人信息换取网上的各种便利，如电子商务、即时通讯、与数以百计的友人或商业伙伴保持联络，以及在网络社区里就我们所关心的问题进行讨论，甚至寻求浪漫邂逅等。而作为交换，我们成为行销的目标。我们的数据以多种多样的方式被聚合和切分：按照年龄、性别、收入、所在的州或城市或村镇、偏好、行踪等，被依照我们的行为进行分组，而这些分组数据则被租借或售卖给想要把各种商品卖给我们的行销者。

大约如多数论者所言，大多数隐私方面的论辩都围绕着“行为定向行销”展开。加州大学伯克利分校和宾夕法尼亚大学最近联合开展的一项研究中，66%的受访者表示他们并不希望经销商根据他们的喜好发送精准定向广告。当受访者被告知他们的各种活动是如何被追踪的时候，反感率变得更高，达到73%~86%。在最近的一项由欧朋软件公司（opera software）开展的调查中，美国人表示相对于恐怖袭击、个人破产或家庭入侵，他们更害怕网上的隐私侵害。

精准定向广告这一概念已不再使人陌生。确实，在今天要通过数据踪迹跟踪任何东西，并对踪迹进行归纳整理，继而做出我们将会购买什么的合理推测，易如反掌。但是这种更具侵入性的广告行销是我们所真正害怕的事情吗？不妨想一想，同样的过程可以被用在对一系列更为广泛的活动进行有效的推测上。安全机构可以使用这种办法来识别潜在的恐怖分子，国税局可以使用这种办法来判断可能存在的虚假纳税申报，执法部门可以使用这种办法来监控可能发生的犯罪活动，信贷公司可以使用这种办法来确定信用风险的高低。数据本身可能是中性的，而数据被加以利用的方式却能够使原本无害的东西变得极度有害，即有些人所谓真正意义上的邪恶。

面对数据隐私问题，要讨论的并不仅仅是我们如何被投放广告。要讨论的应该是以何种商业和政治立场出发去收集和使用我们的私人信息。在我们为了获取产品和服务的便利而泄露私人信息的同时，我们也已经对政府机构以国家、州和地方安全的名义所进行的更具侵入性的监视活动敞开了大门。如果说我们步入此种境地是技术创新和企业家精神联合作用的结果，那么接下来要走向何方则取决于我们自己。

穿过魔镜^①

一切都要从1969年阿帕网（ARPANET，高级研究计划署网络）的诞生说起，这是一个由

^① 魔镜（the looking glass）这一典故来源于爱丽丝穿过魔镜到达虚幻世界的故事。（译者注）

分布于不同地理位置的电脑组成的网络，设计用来保护军事设施之间的信息流。它为后来的互联网奠定了工作基础，互联网则是由无数网络所联结成的一个大网络，如今成为将千万个人、公众、政府、商业和学术网络都连接在一起的蜂巢，传送着巨量的数字信息。

在此过程中，几个重要拐点的出现，最终使得互联网占据了我们的工作与个人生活的中心。

- 互联网成为一个日常词汇。1990年，蒂姆·伯纳斯·李爵士为万维网（world wide web）设计了最初的方案和协议。1993年，美国数字设备公司（digital equipment corporation）正式“开设”了其首个商业网站。到20世纪90年代中期，以网络浏览器的出现为先导，以个人计算机的普及使用为基础，三分之二的上班族和三分之一的家庭都接入了网络。
- 网上购物。易贝（eBay）和亚马逊（Amazon）都创立于1995年，以全新的商业模式直接单纯面向网上消费者开展业务。这促使传统的商业实体及实体经济拉开了其在网络世界中的重塑序幕，与此同时也涌现出了一批像Zappos和Netflix这样新兴的只在网上开展经营活动的企业。
- 网络搜索成为主流并催生了一种强大的新兴广告模式。1998年，继搜索引擎界先锋雅虎（Yahoo）和莱科思（Lycos）等之后，谷歌开发出了一套更好的搜索算法，以及一种更高级的精准定向广告策略。这不仅改变了人们搜索信息的方式，也完善了基于内容和基于付费查询的广告模式，使谷歌仅在2010年的第四季度就实现了高达84.4亿美元的收益。同时，这也产生出了历史上最大体量的个人线上行为数据集。
- 社交媒体网站起飞。2003年，继社交网络先锋Friendster（现在是一个社交游戏站点）的蓬勃壮大之后，Myspace创立并迅速成长为最受欢迎的社交网站，直到后来被脸谱网（Facebook）超越。2004年，社交媒体这个词被创造出来（最先被克里斯·夏普使用），脸谱网紧接着发布。2005年YouTube上线，紧随其后的是2006年出现的Twitter。所有这些网站（包括其他更多）产生了巨量的关于个人行为、人际关系（个人社交网站的理念所在），以及其位置信息（来自诸如Fourquare这样的网站所提供的服务）的数字数据。
- 个人设备的兴起。1996年，诺基亚9000手机成为首部接入互联网络的移动通讯设备。2001年，黑莓发布首款支持电子邮件使用的手机系统。2007年，苹果推出了iPhone，其为大量移动互联网应用和商务搭建了舞台。到2008年，相对个人电脑，出现了更多提供网络接入的手机。2010年，由iPad所引领的平板电脑设备引发了市场风潮，它具有更多的应用来催生更多的数据。这里，有史以来第一次，用户的地理位置信息成为了一个设备自身的必备属性。现在不用别人告诉你，你就可以知道任何时候其所处的位置。

- 通讯具有了即时性。美国在线（AOL）的即时通讯软件（instant messenger，即IM）于1996年推出了实时通讯功能，它继Skype和微软的MSN通讯软件（MSN messenger）推出之后得到个人用户和商务用户更加广泛的关注。1984年开发的短信服务（short message service，即SMS）协议使得移动设备发送文本信息成为可能，这也是现在青少年和年轻人沟通的主要方式。据估计，截至2014年将有超过35亿即时通讯（IM）用户。和社交媒体网站类似，即时通讯也产生了巨量的信息，这其中不仅包括个人用户的信息，还包括反映他们与他人和各种组织之间关系的深度和效度的信息——这些社交图谱都是最有价值的信息。

今天，我们生活在一个总是处于线上状态的数字世界中：我们在线工作、社交，我们在线关注新闻、收看自己喜爱的节目，我们在线交税和处理银行事务，甚至可以追求各种感官刺激。而我们所做的一切都会留下数字踪迹，以致我们不得不给它命一个名字：大数据。

欢迎来到大数据时代

除非你在过去的几年里一直在沉睡，否则你很可能已经领教到数字化世界所产生的数据量之巨大。像“淹没在数据中”“数据的海洋”“指数级数据增长”这样的辞藻都被引用来试图描述大数据的体量之巨。为什么？因为它实在是太大了，以致难于把握，如互联网数据中心（IDC）研究结果显示。

- 2009年，数字化世界的疆界扩张了62%，达到800000PB（不妨想象一下，每1PB即约等于100万GB，如果将这些数据都写入DVD光盘中，那么将这些光盘连接起来的长度，是从地球到达月球距离的两倍）。
- 到2010年，数据化世界的疆界预计会扩张到120万PB（这还是不完全统计的结果）。
- 截至2020年，数据化世界预计将扩张到2009年的44倍（写入DVD的话叠起来将会有到火星距离的一半那么长）。

而大数据并不只是数据体量的问题，还涉及可用数据源的海量存在，其不同的数据格式，以及绝大部分数据是由用户所产生的这一实际情况：事实上数据化世界70%的内容都产生于我们所有人通过电子邮件、Facebook、Twitter、LinkedIn、Flickr、YouTube等进行的数据交互。例如：

- 谷歌每天接受1兆次不同的URL地址访问和20亿次搜索请求。
- 在YouTube上有7000万视频资源（它们每天至少被点播1亿次）。

- 网上有 1.33 亿篇博客。
- 网上有超过 290 亿条推文（并且以每天 300 万条的速度增加）。
- 脸谱网有超过 5 亿活跃用户，他们每个月在网站上总共驻留超过 7000 亿分钟。

除此之外，还有越来越多来自联邦、州和当地政府机构及学术与研究机构的供开放使用的数据源，以及地理空间数据、经济数据、人口统计数据等，类似情况不胜枚举。随着数字化数据的激增，借由政府或商业组织来保护个人隐私成了一件实现起来非常困难的事情。

从散落的拼图到完整的画面：幻想提前来到

随着关于我们自身的数据量不断增加，审视和分析这些数据的技术能力也相应提升。我们已经从过去将关于自身星星点点的零散数据信息七零八落地存放于线下及线上各处，过渡到现在都已拥有了关于自身的全景式数据。所有这些数据都是被数字化地加以采集和存储的。

以往，有两个困难一直阻碍着数据挖掘科学、预测建模及探索性分析领域的发展：过大数据体量的难于存储与过高处理成本的难以实现。今天，数据存储与处理的成本正在呈指数级下降，并且这种趋势大概会一直持续。与此同时，出现了一种前所未有的数据聚合方式，能够将关于每一个人的数据以数字化格式加以整合。这使得各种规模的组织及政府机构易于找到关于任何个体的信息，并使用分析模型来预测其未来的行为。

我们比以往任何时候都更加透彻地被了解，而被了解到的那些信息则可以被用于预测各种各样的行为，包括购买行为、政治行为或犯罪行为。同样的信息也时常被用来创建概要描述文档，以识别各种针对国家安全或国际安全的潜在威胁，对于那些过分专制的国家政权来说，这意味着一旦其公民中有符合预测模型得出的高危概要特征的，那么他们无论是否有问题，都有可能遭受严酷对待。

大坏狼^①般的广告

行为预测广告（behavioral advertising）面对我们的隐私时，真的只能充当大坏狼角色吗？当然，行为预测广告并不是什么新鲜概念，它只不过是通过对你的行为来推测你可能有兴趣购买什么产品或服务的一种方法而已。

① 大坏狼（big bad wolf）这一形象源自同名电影 *The Big Bad Wolf*。（译者注）

在前数字化时代，有一些专门从事购买行为分析的公司，诸如尼尔森市场研究公司（AC Nielsen），还有一些从事“出租”其按照收入水平、性别、婚姻状况、购买行为等属性加以分类的顾客列表的公司。你的邮筒里可能会像所有人一样，不时被各种各样的促销广告塞满，也可能每个小时都会接到关于某些商品买卖的推销电话。更有可能的情况是，那些促销广告正是你提供给银行、信贷公司、百货商店的信息或在订阅报刊杂志中提供的信息所导致的结果。

但就某种意义上说，这种信息本身是漫无目的的。好比一个人的姓名和住址信息被出租，通常只是作为一组信息中的一个字段，但承租者（为广告行销买单的企业或组织）还并没有直接获得信息，除非对他们有所回应。而一旦回应之后，就会出现在承租公司的邮件列表当中，承租公司便开始着手建立其关于这个人的个人信息档案。而即使这样，一家特定的公司或组织基于其行为，还会在多个主要数据库或客户数据库中拥有多重关于其信息的档案。

在互联网时代，如果我的网络浏览踪迹显示我喜爱夏威夷（定向行为），那么当我在网上冲浪时就将会看到去夏威夷旅游的广告，同样，喜爱阿拉斯加的某人看到的就将是去阿拉斯加旅游的广告。这仅仅是线上广告能实现的一项稍微更具个性化的功能而已。所接收到的广告都是根据行动去向和行为意向所发送的，因为行为随时在被追踪，而对你下一步的推测假设则是通过这些行为做出的。广告商青睐这种模式，因为他们能够由此将针对性更强的广告送达到更多对商品感兴趣的消费者终端，这也意味着他们能够卖掉更多的商品。

现在跟以往的不同之处则在于，你在网上所做的一切都会被数字化地加以捕捉，然后被分析并对你采取相应反馈。谷歌利用存储在用户本地终端上的相关数据（cookie）来追踪用户的线上行为、群体特征与喜好品味。还有很多公司都在追踪你的线上行为——大多是通过你在知情或不知情的情况下容许安装在电脑桌面或其他个人设备上的各种 cookie 实现的。还有一大批公司，诸如 eXelate，在售卖个人信息。但就绝大多数情况而言，这些信息并不会明确地识别出具体个人。实际上，这些信息将你归入到一组具有相似群体特征和口味喜好的人当中，而之后这组群体信息会被“租让”给有（线上）广告行销需求的人。

然而，与其建立繁复的多重信息档案，当然不如将其整合起来，以获得关于个人身份和行为记录的更加确切详尽的掌握。例如，社会化网络聚合工具 Spokeo 能够从电话簿、社交网站、市场调查、房地产表单、商业网站及政府机构等渠道公开获取并聚合出关于你的有效信息。如果以你的名字作为对象进行网络搜索，可能会惊讶地发现你的详细居住信息（通过谷歌地图），购房消费及房产税（通过政府数据库），配偶的姓名（通过政府登记），家庭成员情况（通过人口统计数据），所有的电话联络方式（通过线上黄页查号簿），先前的住址及居住花销，以及（根据你选择的社交媒体数据信息呈现的公开程度）远大于你希望任何你的家庭小圈子和朋友小圈子之外的人所知道的信息范围。这些信息大多数在前网络

时代就能加以收集，但可能需要花费大量时间和精力去拜访各种机构，填写各种表格，并且通常需要付费。而今天，人们所要做的仅仅是将自己或其他任何人的名字输入到搜索栏中再点击提交。

如今已不仅仅是关于各种cookie的问题了。比如，一些可能会包括私人信息的公共数据能够被抓取（或称网络抓取）、收集并加以分析。又出现了一种相关的新概念——定位营销（location marketing），会基于人的地理定位有针对性地推送广告（这是借由手机里的GPS芯片实现）。因此，一旦GPS定位显示你正位于某个特定商店附近，就会收到特别针对这家商店的广告或优惠信息。

关于自身能被通过公开渠道或私密源头加以收集的一定量数据信息，或者可以被视为对个人侵扰，或者可以被直接理解成生存在一个数字世界中所要付出的代价。毕竟，你常要用到种种功能的各种网站——诸如Facebook、Twitter、LinkedIn、谷歌、Foursquare等——都需要一种商业模式来保证其持续生存。你跟你所访问的任何一个网站之间不言自明的游戏规则是：我从你那里得到一些好处，我也回馈给你一些好处。这个所谓好处就是私人信息，这些信息通常会被网站外租给各种广告商。而当关于私人信息足够充分，使广告商能够轻而易举、惠而不费地展开广告行销时，私人信息的价值便如金子般宝贵了。

但是现在问题来了：单从概念上讲，只要消费者对行为预测广告有足够的心理准备，那么在道德上行为预测广告就没有什么错处。如果私人数据单纯是被收集和使用于广告目的，其影响还尚属良性。关于隐私的论辩焦点并不在于行为预测广告，而是针对一切非良性的私人数据挖掘与利用方式。如果作为消费者只将数据隐私同广告行为联系在一起，便忽略了一个更大得多的问题：谁正在使用我们的数据？他们正在出于何种目的使用我们的数据？当我们连是谁正在监视我们也不得而知时，我们又能如何保护自己免受隐私侵害的威胁？

老大哥^①和遍布世界的大数据

各国政府正在不断加大资金投入来捕捉和分析数字踪迹，以打击犯罪和恐怖主义，夸张的话语表述被赋予刺激广大民众的功能，保证他们在出现隐私论辩时将安全考量置于隐私考量之上。毕竟，如果我们的人身安全都得不到保障，谁还去关心隐私问题呢？

英国采用数字视频技术来跟踪民众和游客。他们安装了超过185万部闭路电视监控摄像头，平均每32个人分配一部。每天任何经过伦敦的人都会上百次地被监控摄像头捕捉到。

① 老大哥（big brother）这一典故源自乔治·奥威尔的名著《1984》，后来成为专制统治的代名词。（译者注）