

开放数据

如何从无处不在的
免费数据中发掘创意和商机

乔尔·古林◎著 张尚轩◎译

OPEN DATA NOW

The Secret to Hot Startups,
Smart Investing,
Savvy Marketing, and Fast Innovation

继互联网之后，
又一轮改变全球商业模式的
信息化浪潮



中信出版社·CHINACITICPRESS

OPEN

DATA NOW

The Secret to Hot Startups, Smart Investing,
Savvy Marketing, and Fast Innovation

开放数据

如何从无处不在的
免费数据中发掘创意和商机

乔尔·古林◎著 张尚轩◎译

图书在版编目 (CIP) 数据

开放数据: 如何从无处不在的免费数据中发掘创意和商机 / (美) 古林著; 张尚轩译. —北京: 中信出版社, 2015.1

书名原文: Open Data Now: The Secret to Hot Startups, Smart Investing, Savvy Marketing, and Fast Innovation

ISBN 978-7-5086-4814-9

I. ①开… II. ①古… ②张… III. ①商业信息学 IV. ①F713.51

中国版本图书馆CIP数据核字 (2014) 第 218737 号

Open Data Now

978-0-07-182977-9

Copyright © 2015 by McGraw-Hill Education.

All Rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including without limitation photocopying, recording, taping, or any database, information or retrieval system, without the prior written permission of the publisher.

This authorized Chinese translation edition is jointly published by McGraw-Hill Education and China CITIC Press. This edition is authorized for sale in the People's Republic of China, excluding Hong Kong, Macao SAR and Taiwan.

Copyright © 2015 by McGraw-Hill Education and China CITIC Press.

版权所有。未经出版人事先书面许可, 本书出版物的任何部分不得以任何方式或途径复制或传播, 包括但不限于复印、录制、录音, 或通过任何数据库、信息或可检索的系统。

本授权中文简体字翻译版由麦格劳·希尔 (亚洲) 教育出版公司和中信出版社合作出版。此版本经授权仅限在中华人民共和国境内 (不包括香港特别行政区、澳门特别行政区和台湾) 销售。

版权 ©2015 由麦格劳·希尔 (亚洲) 教育出版公司与中信出版社所有。

本书封面贴有 McGraw-Hill Education 公司防伪标签, 无标签者不得销售。

开放数据——如何从无处不在的免费数据中发掘创意和商机

著 者: [美] 乔尔·古林

译 者: 张尚轩

策划推广: 中信出版社 (China CITIC Press)

出版发行: 中信出版集团股份有限公司

(北京市朝阳区惠新东街甲 4 号富盛大厦 2 座 邮编 100029)

(CITIC Publishing Group)

承印者: 山东临沂新华印刷物流集团有限责任公司

开 本: 787mm×1092mm 1/16

印 张: 17 字 数: 191 千字

版 次: 2015 年 1 月第 1 版

印 次: 2015 年 1 月第 1 次印刷

京权图字: 01-2014-3551

广告经营许可证: 京朝工商广字第 8087 号

书 号: ISBN 978-7-5086-4814-9 / F · 3268

定 价: 55.00 元

版权所有·侵权必究

凡购本社图书, 如有缺页、倒页、脱页, 由发行公司负责退换。

服务热线: 010-84849555 服务传真: 010-84849000

投稿邮箱: author@citicpub.com

献给我的父母——帕特·古林和杰里·古林，
是他们致力于利用数据改变社会的热忱激励了我。

同时献给我的妻子卡萝尔，我爱你，我很珍惜和你在一起的日子。

我的数据之旅

我很幸运，有一个充满变数却又让人满意的职业生涯。我最开始是一名科学记者；后来成为书籍作家和杂志编辑；曾是《消费者报告》杂志（*Consumer Reports*）的编辑主任，后来又做了很多年的执行副总裁；在华盛顿担任过一家联邦政府机构的局长；在几个非营利性出版组织担任过领导职位。但是这些都没有我在过去三年里有幸参与的开放数据工作那样，令人着迷，鼓舞人心。

自2010年秋天以来，我一直沉浸在开放数据的世界里。在成为联邦通信委员会（Federal Communications Commission，以下简称FCC）消费者和政府事务局局长之前，我已经在联邦政府工作一年。我被招募进FCC，是因为我有扎实的消费者研究背景，并且当时FCC的新主席尤利西斯·格纳肖夫斯基（Julius Genachowski）想把消费者问题设为工作重点。我一加入这个机构，就开始尝试

解决一个很棘手的问题：怎么能帮助消费者从一系列让人目不暇接的手机业务中选择出最适合他们的一个？

为了找到问题的答案，我开始和一些同事交流，他们当时正在攻克金融服务、医疗和其他领域相似的消费者问题。后来我在科技政策办公室遇到了丹尼·戈罗夫（**Danny Goroff**），他建议应该有人成立一个白宫工作小组来解决这个问题。结果是：我成为 **Smart Disclosure**（智能信息披露）白宫工作小组的组长，副组长是美国财政部的索菲·拉塞曼（**Sophie Raseman**）。我们成立这个小组是为了研究消费者如何能够在市场中，利用政府对产品和服务的开放数据来做出更合理的选择。

通过这个工作小组我认识了贝丝·诺韦克（**Beth Noveck**），她曾经为奥巴马政府领导过开放政府计划（**the Open Government Initiative**）。和她合作过几个项目之后，我作为高级顾问加入了她在纽约大学的政务研究室（**Governance Lab**）团队。这个被我们称为“**GovLab**”的研究室，尝试通过技术、数据和合作来发展通往政府的新途径，其核心工作就是开放数据。

工作过程中我有幸认识了不少的企业家、企业领导者、研究者、学者、记者和其他一些正在开创认识和使用开放数据新途径的人。我也遇到了一群新生代的科技传播者，他们的战斗口号是“数据解放”。此外，我也发现了一个正在成长的、充满生机的开放数据资源新世界——一个为商业、政府和社会转型变革提供素材的数据网络。

我很荣幸能够分别在私营企业、消费者权益机构和政府工作过，这本书很全面地反映了我从这三个角度看开放数据的观点。我希望这本书

能够对它的主要受众——有意促进企业发展、以更高效率开拓消费者市场、寻求新的投资机会和创建新公司的商业读者们有所启发。但是我同时希望它能对和这些公司合作的政府机构、强调公司和政府诚信的倡导者，以及希望通过开放数据提高生活质量的消费者有所帮助。

开放数据将会推进项目的开发、商业的发展……并将有助于企业家推出人们之前无法想象的产品与服务。

——美国总统奥巴马，2013年5月9日

全球最棒的免费资源

设想一下，如果突然间你拥有了一大批商业资源。一大批涉及诸多领域的新公司将会因此诞生：它们将能提供更好的医疗服务，深度分析投资工具，更有效率地供给能源，提高交通运输能力，通过手机应用软件和互联网为消费者带来更广领域的服务。已建立的公司可以更好地了解客户的真正需求，现货交易市场的早期趋势，以及如何选择风险最小的最佳合作伙伴。同时科学研究的速度也将极大提高。虽然创建这样的机会需要花费资金和精力，但是这些重要资源本身却是免费的。

这样的资源就是开放数据。开放数据的概念很简单：那些已经被政府或者其他组织发布，任何人都能获得并能用于任何商业或者个人目的的数据。只要你曾经上网预订过机票，用过手机的GPS（全球定位系统）定位功能，或者浏览过天气预报，你就使用过开放数据。

经过不断提升的数据分析方法，这些免费的开放数据将可以应用于确立新的商业风投，解决一些决策难题，提供新类型的商业情报以及更多其他方面。

今天，开放数据的革命将我们快速引入一个全新的领域。功能更强大的计算机，更经济的存储方式，以及呈指数增长的数字信息，庞大的数据库第一次变得触手可及。开放数据由此被提升到一个新的高度，并在逐渐改变我们的世界。

开放数据正在全球范围内成为明智的商业精英制胜的法宝。投资者利用它去衡量不同公司的风险与收益，以期获得最佳的投资机会。公司老板利用它去理解自身品牌声誉的微妙内涵，并根据数据制定出市场策略。企业家凭借天气、房地产、交通运输等行业的数据，为公众提供更好的服务，造福社会。这些开发项目汇集了几千万的项目资金并能创造出几倍于此的收益。

2013 年 11 月，麦肯锡公司（McKinsey & Company）发布了一份关于全球开放数据的最新报告。报告显示，每年大约有 3 万亿美元的经济潜力是由教育、能源以及医疗保健领域的数据开发带来的。这是迄今为止对开放数据产出价值的最高评估，也许有所夸大，例如该价值涵盖了通过运用开放数据所带来的教育质量提升的价值。尽管如此，这份最新的报告也显示出了对开放数据不断增长的商业兴趣以及它所能带来的潜在价值的惊喜。

与此同时，对于私人领域之外的商业环境，开放数据也具有深远的影响。政府部门通过公布更多数据来增强公众对政府的信心，并吸引更多

多商业投资。开放数据也能帮助医药行业的研究人员更快地找到治疗疾病的方法。此外，媒体也能够借助开放数据来揭露社会的各种弊病，比如洗钱和医保欺诈，并且能够公开更多合作活动。

在过去的三年中，我曾与许多新兴开放数据运动的领导者会面，并先后担任FCC消费者和政府事务局局长，以及Smart Disclosure白宫工作小组组长。现在我就职于纽约大学。我曾经见过很多满怀激情的企业家，他们都在利用开放数据去构建新的商业机会，并推动诸如医疗、金融、教育以及能源领域项目的前进。（我曾接触过一些美国政府高官、英国政府官员、政策制定者、学者以及非政府组织，他们都致力于公开更多用于商业及其他领域的数据。）这本书反映了他们的观点、见解、认识，以及建议。

鉴于开放数据领域的快速发展，没有哪本书能够做到完全跟进潮流。因此，我建立了一个网站OpenDataNow.com，来密切关注最新的发展、争议以及趋势。作为以开放数据为中心的组群，该网站也提供日志、应用和最新消息的链接。我鼓励读者朋友们登录该网站，提出建议并帮助网站的发展。我在本书中也提到了一些专家的文章、访谈录以及播客。

本书提供了一些能够帮助读者进一步了解开放数据的资源。大部分章节对如何实际运用开放数据给出了建议。

无论你是企业家、决策者、投资家、小公司所有者，还是公共行业或者非营利组织的工作人员，开放数据都将为你工作的领域提供新的机遇。同时，本书也能引导你去理解这一新生事物，直面新的挑战，驾驭新的力量。

前 言 我的数据之旅 // VII

引 言 全球最棒的免费资源 // XI

第一部分

开放数据的力量

第一章 与互联网同样巨大的机遇 // 003

第二章 热门新兴公司：从政府数据中获利 // 019

第三章 消费者网站：Smart Disclosure 的决策引擎 // 041

第四章 管理数据泛滥的新兴公司 // 067

第五章 数据导向投资：商业分析的新工具 // 081

第六章 绿色投资：为可持续性数据一搏 // 095

第七章 精明的市场营销：如何用数据定义你的品牌声誉 // 113

第八章 情感分析的市场营销学 // 131

第九章 利用群体智慧进行快速创新 // 143

第十章 开放的实验室：通过公开合作进行革新 // 157

第二部分

商业环境：开放数据的新趋势

第十一章 个人数据的隐私、安全 and 价值 // 181

第十二章 在透明的社会环境下做生意 // 201

第十三章 政府和数据：为开放的世界制定规则 // 217

第十四章 开放数据的未来 // 241

致 谢 // 253

OPEN DATA NOW

第一部分 开放数据的力量

The Secret to Hot Startups,
Smart Investing,
Savvy Marketing,
and Fast Innovation

第一章 与互联网同样巨大的机遇

2012 年 11 月，我在伦敦参加了由刚刚成立的开放数据协会（ODI）召开的会议，会场座无虚席。该组织由英国政府提供 1 000 万英镑作为项目资金建立，并由政府和个人共同监管，集合了所有在此领域拥有独特见解的精英。该组织位于伦敦的“硅谷”——位于老街（the Old Street）环形路口最靠近地铁被称为“硅环”的地方。那里类似于旧金山南部市场周边以及曼哈顿下城区的硅巷，周边地区并不繁荣，却恰为新兴公司的发展提供了充足的空间，促进了革新的不断涌现，并抬升了房地产价值。

作为这个组织的一员，我与 24 位来自企业、政府和非营利组织的同事一起探讨开放数据。在会议日程的第二次连续会议上，与会的各方代表合并了来自麦克阿瑟基金会的资金。代表们分别来自白宫，唐宁街

10 号，英国政府内阁及商业部，世界银行，英国零售商巨头，两个高科技咨询公司，前沿科技出版社，大学法律、计算机、人工智能、物理和认知神经学系，两个基金会，以及致力于提升公司透明度、公众参与度和绿色商业行为程度方面的非营利组织。

我们受到了该协会首席执行官加文·斯塔克斯（Gavin Starks）的热烈欢迎。加文从事网络开发工作已 20 余年，曾任职于多个公司、机构，如维珍集团、Google（谷歌）、英国政府部门，以及联合国儿童基金会（UNICEF）。他曾经大胆地提出，开放数据将会产生类似万维网那样巨大的影响。

今天的开放数据，正如他所说，“像极了 1994 年时的互联网，那时的我仍在试图劝说人们接受电邮，开发网页。尽管人们都对互联网的前景感到非常兴奋，但是没有人确切地知道它的未来究竟会怎样。20 年间，我们已然见证了大量的革新、创造，甚至是破坏。今天，我们同样无法得知开放数据的明天将会怎样，但是我们确信，它将会带来新的变革。商业运作的方式将会得到改进。对我们来说，如何控制这一切将成为新的挑战。正如网络出现早期我们所看到的潜在发展机会那样，我认为，开放数据的潜力同样不可小觑”。

这不是人们第一次将一项新技术的发展与互联网发展初期的情景相比较。但是开放数据运动的领导者们有理由做出这样的论断。开放数据协会的总裁和联合创始人蒂姆·伯纳斯-李（Tim Berners-Lee）爵士，在他的个人经历中这样简明扼要地指出，他在欧洲核子研究组织（CERN）的欧洲粒子物理实验室创立了万维网。开放数据协会主席奈杰

尔·沙伯特（Nigel Shadbolt）先生，是网络与人工智能领域的先驱，在英国政府开放数据政策的制定过程中发挥了重要作用。他们都是美国、英国或者其他国家诸多致力于发展开放数据的梦想家的代表。

对开放数据的最佳描述是：公众、公司和机构可以接触到，能用于确立新投资，寻找新的合作伙伴，发现新趋势，做出基于数据处理的决策，并能解决复杂的问题。它不同于大数据（Big Data），虽然它们的确有所重叠。开放数据有一个宗旨，就是它将提供免费、公开、透明的数据信息，并能适用于我们需要的任何领域，比如商业经营、政府运作，以及处理各项事务。正如参加那次会议的人员一样，开放数据的工作人员背景多样，来自企业、技术、政府部门、学术、非营利组织，以及医疗健康、教育和环境科学等领域。

开放数据运动最初的目的在于推进民主，让政府公开数据，以便纳税人了解他们所支付税金的去向。除了带来良好的社会效益外，开放数据还创造了巨大的商业机会，这也是本书的关注重点。别忘了，互联网创立之初也是一项由政府出资打造的项目（即ARPAnet），由美国高级研究计划署（ARPA）负责，经总统艾森豪威尔批准，以回应苏联政府发射的人造地球卫星Sputnik。该项目在当时成为驱动经济发展的因素之一。同样，由政府推动的开放数据建设也将为数据管理提供经济方面的资源与基础。

经济增长和增加就业的需求也推动了美国和英国政府联合制定开放数据政策。这在奥巴马政府于2013年5月推出的关于开放数据的新政策中体现了出来。这项政策使得空前庞大的联邦数据能为公众所用。值