

理想国

imaginist

The Big Data Revolution

除了上帝，任何人都必须用数据来说话。

涂子沛
— 著

大数据

[3.0 升级版]

正在到来的数据革命，
以及它如何改变政府、
商业与我们的生活



GUANGXI NORMAL UNIVERSITY PRESS
广西师范大学出版社

涂
于沛
— 著

The
Big Data
Revolution

大数据

正在到来的数据革命，
以及它如何改变政府、
商业与我们的生活

[3.0 升级版]

广西师范大学出版社

· 桂林 ·

图书在版编目 (CIP) 数据

大数据：正在到来的数据革命，以及它如何改变政府、商业与我们的生活 [3.0 升级版]
涂子沛著. —桂林：广西师范大学出版社，2015.4

ISBN 978-7-5495-6410-1

I. ①大… II. ①涂… III. ①信息产业—产业经济—通俗读物
IV. ①F49-49

中国版本图书馆 CIP 数据核字 (2015) 第 038933 号

广西师范大学出版社出版发行

桂林市中华路22号 邮政编码：541001
网址：www.bbtpress.com

出 版 人：何林夏

全国新华书店经销

发行热线：010-64284815

山东临沂新华印刷物流集团有限责任公司

开本：710mm×1000mm 1/16

印张：25 字数：260千字 图表：145幅

2015年4月第3版 2015年4月第1次印刷

定价：69.00元（精装）

如发现印装质量问题，影响阅读，请与印刷厂联系调换。

一个真正的信息社会，首先是一个公民社会。

——作者题记

薛华成 [中国管理信息系统的主要开创者、澳门科技大学行政与管理学院名誉院长]

通过历史故事的讲述，用近似新闻评论的笔法，《大数据》这本书以美国为例介绍了由经验管理到数据管理这场正在发生的革命性变革，通俗形象地讲清楚了科学的道理，实属难得。这对学者们扩大视野，对管理干部深化科学发展观的认识，都有十分重要的价值。

施伯乐 [复旦大学计算机研究所所长、上海（国际）数据库研究中心主任]

《大数据》这本书揭示了数据在未来社会中的革命性作用。大数据将是下一个社会发展阶段的“石油”和“金矿”。无论是个人、企业还是国家，谁能更好地抓住数据、理解数据、分析数据，谁就能在下一波的社会竞争中脱颖而出。关于数据的知识，将成为个人知识结构中的必备要素和基础。

赵嘉敏 [东西网总裁、译言网创始人之一]

子沛兄的《大数据》这本书，帮助我们形象地了解信息自由、技术创新与社会进步的关系。我最感兴趣的是美国信息自由制度建立的过程。它让我看到，一个社会、一种文化，是如何通过自省式的努力，认清比自身存在更本源的客观规律。这种制度一旦建立，就如同打开了控制某个基因群组的开关，后面所呈现的演化，虽然眼花缭乱，反而显得平淡无奇、顺理成章了。

董洁林 [苏州大学企业创新和发展研究中心主任、教授]

18世纪启动的工业革命为人类创造了巨大的财富，也极大地推动了社会变革。20世纪中启动的信息革命对人类社会也将产生巨大的冲击，而这个进程刚刚开始。涂子沛先生所著的《大数据》一书，用有趣的故事带我们回顾信息时代的里程碑，使我们透过未来的变幻不清的迷雾，展望和畅想信息革命的大趋势。

汪小帆 [上海交大电子信息与电气工程学院教授]

《大数据》值得仔细阅读与思考。要让数据更好地说话，既依赖于数据科学与技术的发展，更需要政府、机构和公民的数据意识的不断提高。

张晖 [卡内基梅隆大学计算机系终身教授]

数据科学正在成为一个新兴的学科和领域。涂子沛先生的《大数据》这本书，为我们揭示了大数据浪潮发生的前因后果，以及这种技术进步对政治、经济、文化的种种影响。希望读者从本书中获益，以积极的态势应对中国的大数据挑战。

Ramayya Krishnan [卡内基梅隆大学海因兹学院及信息管理学院院长、终身教授]

Big Data has the potential to fundamentally transform Society. However, unlocking this potential will require careful attention to data governance and insightful application of data analytics combined with an environment that spurs managerial innovation. Zipei's book provides a roadmap for the Chinese professionals using the American experience as context. As the dean of the Heinz College and its School of Information Systems at Carnegie Mellon University, I am very proud of what Zipei has accomplished. His book combines government, business and technology concepts in a holistic manner, embodying the unique strengths of programs and faculty of the Heinz College. I believe this book will stimulate new ideas in China to meet the big data challenge.

中文译文：大数据具有催生社会变革的能量。但释放这种能量，需要严谨的数据治理、富有洞见的数据分析，以及一个激发管理创新的环境。涂子沛的《大数据》这本书总结了美国的经验和实践，并为中国提出了一个路线图。我为他的成就感到骄傲。这本书完整地融合了技术、政治、商业三个角度，体现了我们学院师资教学的独特优势。我相信本书能够在中国激发新的思想和讨论，帮助中国迎接大数据的挑战。

献给

母亲魏木兰和妻子张炜婷

感恩无私的爱

1. 母亲，您是我生命中最爱的人，是您给了我生命，是您教会了我如何做人。

2. 母亲，您是我心中最温暖的地方，是您给了我力量，是您教会了我如何坚强。

3. 母亲，您是我人生中最美丽的风景，是您给了我希望，是您教会了我如何梦想。

4. 母亲，您是我生命中最珍贵的宝藏，是您给了我勇气，是您教会了我如何勇敢。

5. 母亲，您是我心中最柔软的地方，是您给了我安慰，是您教会了我如何温柔。

6. 母亲，您是我生命中最无私的奉献者，是您给了我无私的爱，是您教会了我如何无私。

7. 母亲，您是我生命中最伟大的创造者，是您给了我生命，是您教会了我如何创造。

8. 母亲，您是我心中最美好的回忆，是您给了我快乐，是您教会了我如何快乐。

9. 母亲，您是我生命中最坚强的后盾，是您给了我支持，是您教会了我如何坚强。

10. 母亲，您是我生命中最无私的奉献者，是您给了我无私的爱，是您教会了我如何无私。

11. 母亲，您是我生命中最伟大的创造者，是您给了我生命，是您教会了我如何创造。

12. 母亲，您是我心中最美好的回忆，是您给了我快乐，是您教会了我如何快乐。

数据可以治国，还可以强国

新版自序

今年春节期间，出版社告诉我，作为中国大数据领域的开山之作，《大数据》很受欢迎，他们希望满足市场的需求，推出第三个版本，嘱咐我写点东西。

之所以受到欢迎，可能是因为独树一帜。一提到大数据，很多人的第一反应，是数据在商业领域中的作用，所津津乐道的，也是精准营销，即通过数据分析增强对用户的洞察、扩大营销收入。但本书所强调的，却是“数据不仅可以治国，还可以强国”。

数据治国，是指要凭借对数据的有效收集、处理和分析来治理国家，决定国家的大政方针和具体政策。

从定义看，数据是对客观世界进行量化和记录的结果。量化，是人类进行科学研究最根本的手段，就此而言，数据治国也可以理解为“用定量分析、实证研究的科学方法来治理国家”。此外，数据表示的是过去，关注、表达的却是未来。数据之中蕴藏着社会发展的规律，“数据治国”的提法，也暗含着“人类可以更好地把握未来”之意。

回顾人类的历史，任何一个国家的治理，都需要参照数据，但由于信息技

术的进步、记录手段的普及，物理世界的状态、个人的社会行为得到了前所未有的记录，这种记录的粒度越来越细、维度越来越多、频度越来越密，形成了“大数据”。今天的数据治国，含义远远超出了“参照数据”。

例如，当前国人关心的头号问题，莫过于空气质量。要解决好这个问题，从环境的监测、传感器的安装，到空气标准的设立、污染原因的分析，再到高污染、高能耗、低产值传统产业的转型，和智能电网、智能电表的使用和普及，整个治理链条都贯穿了数据。又例如交通拥堵，如今很多城市，立交桥高达三四层、地铁线也在快速扩张，却赶不上车辆增加的速度，在资源增长、空间有限的情况下，唯一的出路，是用好大数据，对城市的状态进行实时的分析和预测，不断优化现有资源的使用情况。还有当下智慧城市的建设，大数据的作用可谓重中之重。

今天的政府，无论是东方、西方，都有人在不断尝试，试图把“凭借数据来治理社会”的模式推到极致。本书第七章“全国隐私风波”重点阐述了十几年前美国各方力量围绕“万维信息触角计划”(TIA)产生的纠结和较量。2002年，在国防部长拉姆斯菲尔德、海军中将波因德克斯特的主导下，美国国防部高级项目研究所(DARPA)试图建立一个超级数据库，把全社会每一个人的基本资料、信用卡交易信息、医疗保险记录、出入境记录、航空和酒店信息、电子邮件、电话记录、网络搜索记录等等数据都整合到一起，然后通过数据挖掘和监控，在大量的信息当中发现“信号”，实现有效的社会治理和管控。这个计划，因为可能侵犯公民的隐私，在美国国会不断受阻，但美国国防部却屡次变换名目、暗中推动项目进展。这段历史，堪称2013年“斯诺登事件”的前传，因为后来大面积曝光，最终震惊世界。

西方不亮东方亮。这期间，新加坡的国防部长 Peter Ho 在美国访问，他获知万维信息触角计划之后，大为赞赏。不久后，该计划在美国触礁流产，波因德克斯特被迫辞职，新加坡却重金引进了这个项目组的一批人员，并聘波因德克斯特为国防部顾问，视为上宾。在波因德克斯特的指导下，新加坡快速建立了一套万维信息触角，命名为“风险评估和全景扫描”(RAHS)。

RAHS 最初的目的，是通过全社会的数据联通和挖掘，在错综复杂的社会

现象中发现恐怖袭击、流行病疫情等突发事件的“信号”，达到未雨绸缪、有效治理的目的。之后，由于行之有效，新加坡政府将系统不断扩大，推广到经济、文化、社会管理的种种领域。房屋管理局用它来了解人们对于住房体系的批评和期待，人口部门通过它掌握大众的生育态度变化，旅游部门用它来预测各地游客数量的周期，食品部门通过它决策是否应该减少对进口食品的依赖等等。从这个角度说，新加坡是数据治国的一个实验室，它不仅将大数据用于情报和反恐，还用于构建和谐社会。

在中国，特别要强调的是，数据治国不是“数字治国”。一些官员一谈起辖区人口面积、GDP 增长幅度、项目投资多少、惠及领域若干、利税增加几何都有“数”在胸，还特别喜欢用“三项原则”、“五个重点”、“八项规定”等等数字范式对工作进行总结，各项工作的考核也都与数据挂钩，大有唯“数据”盛行之势。这不是不好，但这些都是孤立的、静态的数字，并不是“数据治国”的真义。大数据时代，数据治国更关注的是动态的数据、系统化的数据，以不间断“流”的形式存在的、成片的、活的数据，它们应该成为公共决策的资源为政府和社会所广泛使用。归根结底，公共决策最重要的依据将是系统的、成片的、动态的数据流，而不是个人经验或长官意志，过去深入群众、实地考察的工作方法虽仍然有效，但对社会治理而言，系统采集的数据、科学分析的结果更为重要。

我认为，数据不仅可以治国，还可以强国。其中的原因，是因为数据正在成为各行各业最重要的创新资源。

以农业为例，来自天气、土壤和农作物的数据，可以实现自动化的灌溉、防治虫灾、决定如何种植和收割、节省水利资源、提高单位产量；在教育领域，以大数据为基础的在线智能学习平台（MOOC），可以突破教室的限制，让成千上万的学生同时得到个性化的教学和辅导；再以医疗领域为例，IBM 公司的 Watson 系统可以快速地检索几百万名患者的病例、最新的医疗研究报告和成果，做出更快、更准确的诊断；工业领域更不待言，汽车、飞机正在被大数据改造，成为无人驾驶汽车、无人驾驶飞机；还有金融领域，利用电商平台的交易数据，“阿里小贷”可以在几分钟之内判断企业的信用，为近百万小微企业

发放贷款。今年年初，阿里巴巴又推出了基于个体消费者的“芝麻信用”，用的创新资源还是数据。一百年前的美国，用了几十年的时间才建立全民信用体系。因为拥有无处不在的充沛数据，今天中国的全民信用体系可能在更短的时间内、以更低的成本就可以建立起来。

毫不夸张地说，基于数据的创新将带动人类社会的各个领域都实现巨大的飞跃，这种飞跃是前人难以想象的。今天的年轻人面临的创新机会要远远多于前几代人，因为他们拥有人类有史以来最伟大的创新资源：数据。

和其他的创新资源相比，数据之所以伟大，是因为它不会被它所激发的思想和创新所消耗，它可以重复使用，可以同时被无数人使用，此数据和彼数据整合，还可以产生新的价值和效用。在空间的拓展中和时间的延伸中，数据的能量将在人类社会层层放大，数据的不断积累是资源和知识的持续增加。

但要成就数据的“伟大”，还有个前提，这就是数据的开放。正是出于这个原因，本书用了相当的篇幅来讨论“开放数据”。

对于“开放数据”，中国社会目前还存在不少误解。一是将开放等同于公开，其实开放和公开是两个完全不同的概念。公开是信息层面的，是一条一条的；开放是数据库资源层面的，是一片一片的。开放数据，指的是将原始的数据及其相关的元数据以可以下载的电子格式让第三方自由使用。开放也不一定代表免费，因为开放是有成本的，企业可以以收费的形式开放数据。开放也可以有层次、有范围，针对某个特定的组织和群体开放，不一定面对全体社会大众。

开放数据之于当下中国社会的意义，在于推动知识经济和网络经济的发展，在于促进中国经济由粗放向精细、从“制造”向“创造”的转型升级，在于释放社会生产力、催生创新。通过开放更多的数据，让创新的资源自由地流动，李克强总理提出的“大众创业，万众创新”才能更快、更好地在中国大地形成燎原之势。

大众创新、创业的趋势，也可以在美国社会的发展过程中观察到。根据美国科学基金会（NSF）的数据，1981年，大公司（超过25000人）的研发投入占全美研发投入的70%，可谓主导了美国的创新；到2007年，大公司的研发投入仍然在上升，但格局却发生了改变，尽管绝对数量增长了4倍，其占全美

研发投入的比例却下降到 35%。同期, 1000 人以下小公司的研发投入增长了 50 倍, 其占全美的比例从 1981 年的 4% 上升到 2007 年的 24%。这些数据表明, 研发和创新呈现分散化, 在向小公司和个人转移, 其中最根本的原因, 就是知识、信息特别是数据资源的深度开放, 不再受大公司、大组织的垄断, 不再是精英阶层的专利。

文化靠沉淀, 文明靠创新。创新, 才是一个国家持续发展、不断迈向强大的真正动力。今天的中国, 创新已经成为共识。就此而言, 数据强国我们正在路上。

记得《大数据》首版上市, 是 2012 年 7 月。当时我在美国, 半年多之后, 美国才出版了第一本有影响力的专著《大数据时代》, 比中国晚了整整半年。抚今追昔, 我感叹不已, 新的思想和观念在全球自由地流动, 在奔涌的浪潮之中, 我也从美国硅谷回到了中国的杭州。

除了增补自序、重要的专栏文章以及我在第一届世界互联网大会上的演讲等, 本版还增加了索引。一本好书不能没有索引, 这是我一直的观点。几经曲折, 《大数据》第三版终于加上了索引, 除了出版社的支持, 还要感谢冯启娜、赵瀚林、陈竞芬等志愿者, 因为他们的协助, 这个心愿得以达成。

涂子沛

2015 年 3 月 22 日于杭州

序言一 大数据：为华文世界提出一个重要话题

许倬云

涂子沛先生的新著《大数据》，已经完成，是一部300多页的大作。最近他将这本书的打样稿送来给我看，并且要求我作序文。可惜的是，我将要远行，而他索序甚急，我只能在浏览一遍之后，发抒自己一些感想。至于细细咀嚼和消化这本有趣作品的内容，必须要在一两个月以后，也就赶不上涂先生大作的出版时间。我事先声明，这些只是我浏览他大作以后的一些感想，非常立即的直觉，也一时谈不上深刻的见解。

这本书的内容，强调今天是一个大量数据公开于大众的时代。自从资讯革命以来，资讯工具、硬件、软件，平行发展、与时俱进。数据在我们生活之中，日常接触，已是处处可见的现象。收集数据，当然是靠电脑的快速分类和记录，然而更重要的是搜索引擎的进步，与网络之间网际的交流。到今天，一个一个网，不但可以串连在一起，互相沟通，而且“云端”（云计算）的设计，可以将每一个地区个别数据，储成一个大的数据库，有助于我们更迅速广泛地搜索。

这些现象，20 世纪最后四分之一以来，已经呈现加速度发展的新事物，在 21 世纪进展速度之快，更是铺天盖地、无所不在。涂先生在这本书里强调：不仅数据经过管理而大量地存在，而且，在现代的国家，尤其是民主社会，开放的社会与政府之间，经由数据，彼此一目了然、无所隐瞒。一个自由的社会，掌握公权力的政府，跟任何其他政府一样，有压倒社会的庞大力量，因为他们手上掌握了人生需要的许多数据。不过，这些数据，在自由的社会，公民也可以一样取得，使政府所作所为，可以摊开在天地之间，让我们检验。过去封建专制和集权各种政体，其执政者能压迫老百姓，而老百姓没有办法回制公权力的压迫。涂先生特别标榜，美国奥巴马接任以后，尽力将数据开放于大众，固然奥巴马是一个有开放心胸的政治人物，如此将政府掌握的数据，大量地开放于群众，也是拜时代之赐，有如此的机缘，才能将数据公开。

涂先生引用胡适之先生与黄仁宇先生的话。胡先生说中国人习惯于“差不多先生”，凡事马马虎虎、不求精确。黄仁宇先生认为，中国不懂得用数字来管理国家。涂先生引用这两位先生的名言，当然是要彰显传统中国和今天美国之间的巨大差异。不过我必须有所说明：胡先生和黄先生的话语，都是“爱之深而责之切”的心态，他们身经当时中国的混乱，激愤而出此感言。

从历史上看，不论中国和西方，任何国家发展到可以有一个复杂文官系统管理以后，没有不依照数据来治国的。人口、资源、土地、财产种种的统计数字，在中国历史上，自从战国时代形成列国的国家体制以后，没有一个朝代不具有一定的数据库；只是以今天的标准来讲，粗糙和细密之间，古今有很多的差别而已。以汉代为例，汉简所显示的家户统计，每一户中的人口，男女老小，以及拥有的资产数目字，都详细统计，而且不论是居延边塞，或是荆州内郡，格式一致。汉简各种家户统计，与唐代西域州府的记录对比，其内容格式也是相当一致。这种基本的数据，在列朝的会典中，都见到其大概。当然，各个朝代的数据，有做得好的，也有做得差的。大致讲起来，外族侵犯中国建立的朝代，以武装力量强制建立政权，也往往依靠暴力的掠夺，取得他们所需要的资源。一个上轨道的朝代，其数据还是相当完整。

再看西方历史。希腊时代，我们了解的资料不够。罗马帝国时代，全帝国包含各种不同的政治单位，并没有一个大一统的文官政府；因此，全国性的资料库似乎不存在。等到中古黑暗时期，国不成国，地方不过是大小封建领主占有领土而已，他们并没有建立详细的资料库。近代以来，列国各自组成完整的主权国家，这些数据也纷纷出现了。

这是以历史上政权掌握数字而言。一家大型的企业，例如，中国清代的票号，如果他们手上没有复杂的数据库，就不能进行汇兑、放款、存款等等活动。英国的东印度公司，手上握有丰富的资源，他们也不能不具有一个相当完整的数据库，否则无以经营这么复杂的开拓业务。

今天的数据时代，我已经在前面提过，不仅公司单位都有搜集数据的能力，而且更重要的，有搜寻引擎可以将资料迅速检索，从其中归纳出条理，有助于了解情况。举一个例说，最近我们才看到，数据资料显示，美国百分之一的人口，拥有全国财富百分之四十以上，百分之九十九的人口，拥有全国财富才过半而已。对于许多长期习惯于美国是开放社会的一般人民，这一组数据显示的现象，几乎可说是理想的破灭，使大家必须检讨：美国真是如此开放吗？还是相对地在逐渐关闭？是不是财富与权力，已经逐渐集中到社会顶端一小撮的人手中？他们以财富作为魔法师的指挥棒，安排了我们的生活、决定了我们的未来。这种现象能够暴露于众，当然就因为在美国究竟资讯是公开的。

相对而言，在极权的国家，他们手上拥有足够的资讯，足以利用这资讯，掌握每一个人的日常生活；“老大哥”的影子，可以无所不在。资讯时代，对于极权的掌权者，他们拥有无可比拟的强大工具，甚至于比坦克车和催泪弹更为有效。资讯管理、资讯控制，是无影无踪，又是无所不在。说到这里，我们不能不更多警惕。

作为一个史学工作者，看惯了世间的灾难和创伤，不能不提出警告：这个中性工具，也只有了解到资讯工具阴暗面——双刃剑的特性，在权势独占这一工具时，可能出现的危险。有此认识，我们才能善于利用这了不起的工具，开创更好的未来，也防治不虞的灾害。

涂先生这部书，清楚地叙述了资讯时代对我们生活的影响与社会的控御力。他讨论的范围方方面面、极为广泛。我盼望有了这本书作为起头，还有很多对资讯工具有研究、也有心得的人，参加讨论，让我们更清楚地了解，这个 21 世纪正在坐长的新的知识工具。为此，我们要对涂先生致敬与致谢，因为他为华文世界提出一个重要的话题。

2012 年 4 月 8 日 于匹兹堡

序言二 中国的雄心应该拓展到大数据领域

托马斯·H. 达文波特*



无论是对中国政府，还是就中国的商业组织而言，《大数据》都是一本重要的书。大数据及其分析，将会在未来 10 年改变几乎每一个行业的业务功能。任何一个组织，如果早一点着手大数据的工作，都可以获得明显的竞争优势，正如早期在“小数据”时代脱颖而出的竞争者一样，如第一资本金融公司、前进保险公司、万豪酒店等等。时光荏苒，现在到了抓住大数据机遇的时候了。

大数据之所以产生，是因为今天无处不在的传感器和微处理器。我们正在迈进普适计算的时代。其实，所有的机械或电子设备都可以留下数据痕迹，这些痕迹表明了它的性能、位置或状态。这些设备和使用它的人，通过互联网互相交流，又形成了另外一个庞大的数据源。当这些数据和来自其他媒体、无线或有线电话、有线电视、卫星等等来源的数据相结合的时候，更加显得庞大无比。

这些数据可以被使用，这意味着我们可以把所有的商业或组织活动都视为大

* 托马斯·H. 达文波特 (Thomas H. Davenport), 哈佛大学商学院访问教授、巴布森学院 (Babson College) 信息技术与管理学总统杰出奖教授, 2003 年, 他被《咨询》杂志评为全球“最优秀的 25 位咨询大师”之一, 2005 年被《优化》杂志评为世界商业与技术分析顶级三强之一。

数据的问题。如今的制造业，大多数机器上都已经安装有一个或多个微处理器，已经进入了大数据的状态。消费营销行业，无数顾客的交易触点和网上点击的流量，也成了大数据的问题。谷歌甚至认为其无人驾驶汽车也是一个大数据的问题。

世界各国的政府也开始认识到，他们坐拥海量数据，这些数据都有待分析。在亚洲国家的政府，也出现了大数据战略以及基于数据分析的方案和倡议。去年，新加坡成立了德勤数据分析研究所（DAI），这个新的机构是由新加坡政府经济发展委员会资助成立的。德勤数据分析研究所的目标，就是引领政府和企业对于数据的研究和应用。新加坡政府还资助了几所大学开展大数据和数据分析的研究活动。

任何一个组织，要抓住大数据的机遇，就必须做好几个方面的工作。从技术角度来看，首先要收集并且开发特定的工具，来管理大规模并行服务器产生的结构化和非结构化数据，这些数据，可能是自己专有的，也可能来源于“云”。其次，每一个组织都需要选定分析软件，用它来挖掘数据的意义。但可能最重要的是，任何组织都需要人才来管理和分析大数据。这些人被称为“数据科学家”，他们集黑客和定量分析员的优势和特长于一身，非常短缺。聪明的领导人，将想方设法留住这类人才。

不少公司都意识到了这种难得的机遇，现在已经采取了行动。例如，通用电气将投资 15 亿美元在旧金山湾区建立一个全球软件和分析中心，作为其全球研发机构的一部分。这个中心拟雇用至少 400 名数据科学家，现在已经有 180 名各就其位了。通用电气在全球拥有超过 1 万名工程师从事软件开发和数据分析工作，通过共同的分析平台、训练、领导力培训以及创新，他们的努力得以协调合作。通用电气对于大数据的研究活动，相当一部分集中在工业产品上，例如机车、涡轮机、喷气发动机以及大型能源发电设施。

对任何一个试图通过大数据获得成功的组织来说，通用电气的投资规模和雄心都是一个榜样。在很多领域，中国政府和中国的企业都有雄心勃勃的计划，这引起了全世界的关注，这些雄心和计划，现在应该拓展到大数据的领域。涂子沛先生的这本书，将在这个重要的领域，为中国政府和企业的努力提供引导和帮助。