

大数据 大文化

BIG DATA BIG CULTURE

崔成泉 王晓芳 编著



云南大学出版社
Yunnan University Press

大数据 大文化

BIG DATA BIG CULTURE

崔成泉 王晓芳 编著



 云南大学出版社
Yunnan University Press

图书在版编目 (CIP) 数据

大数据 大文化 / 崔成泉, 王晓芳编著. — 昆明 : 云南大学出版社, 2014

ISBN 978-7-5482-2092-3

I. ①大… II. ①崔… ②王… III. ①信息技术—应用—文化产业—研究—中国 IV. ①G124-39

中国版本图书馆CIP数据核字 (2014) 第202806号

策划编辑: 陈 曦

责任编辑: 柴 伟

装帧设计: 庄海萌

大数据 大文化

BIG DATA BIG CULTURE

崔成泉 王晓芳 编著

出版发行: 云南大学出版社

印 装: 云南省昆明市卓林包装印刷有限公司

开 本: 787mm×1092mm 1/16

印 张: 8.75

字 数: 156千

版 次: 2014年10月第1版

印 次: 2014年10月第1次印刷

书 号: ISBN 978-7-5482-2092-3

定 价: 28.00元

社 址: 云南省昆明市翠湖北路2号云南大学英华园内

邮 编: 650091

电 话: 0871-65031071 65033244

网 址: <http://www.ynup.com>

E-mail: market@ynup.com

序

崔成泉

随着《大数据时代》^①在中国的广泛传播，“大数据”瞬时成为社会热点话题，2013年也被称作“大数据元年”。

随意打开一家网站的读书频道，输入“大数据”一词，相关书籍已经数不胜数。国内各大出版社使出浑身解数，挤着登上“大数据”的这趟潮流列车。但是翻看这些图书目录，大部分内容都聚集在大数据的技术本身，以及大数据在商业中的应用。而《大数据 大文化》这本论文集，想表达的是直击文化本身，用大数据思维来解读和思考文化面临的问题与挑战，以重构我们的思维方式、思维路径。

大数据、大文化汇聚一起，绝非偶然。信息技术迅猛发展，深刻地影响着文化的变革与发展，催生着新型文化业态的生成与发展。什么是大数据？大数据有何技术特征，大数据与大文化之间如何融合，这两者之间的关系如何解读？文化产业和公共文化应以什么样的发展路径去应对大数据时代？从国家治理的层面上来讲，大数据时代文化资源的整合，需要体制改革提供什么样的支撑和顶层设计？等等，一系列问题有待于我们的回答和进一步探讨。

不管你是否认可，大数据时代已然到来；不管你是否认同，大数据技术进入文化领域已是必然。“走进来”的大数据与“走出去”的大文化，均是时代运演的结果。

《大数据时代》一书前瞻性地指出，大数据带来的信息风暴正在变革我们的生活、工作和思维，开启了一次重大的代际转型，发动了一次时代的思维变革、商业变革和管理变革，并提出，放弃对因果关系的渴求，取而代之的是关注相关性关系。大数据时代的新变化，颠覆了千百年来人类的思维习惯，对人类认知和交流方

^① [英] 维克托·迈尔-舍恩伯格、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。

式提出了全新的挑战。

中共十八届三中全会报告指出：“全面深化改革的总目标是完善和发展中国特色社会主义制度，推进国家治理体系和治理能力现代化。”“建设社会主义文化强国，增强国家文化软实力，必须坚持社会主义先进文化前进方向，坚持中国特色社会主义文化发展道路，坚持以人民为中心的工作导向，进一步深化文化体制改革。要完善文化管理体制，建立健全现代文化市场体系，构建现代公共文化服务体系，提高文化开放水平。”

这一切告诉我们，在大数据时代，必须让大数据与大文化进行深入对接，必须对文化科技发展潮流有充分的认知和把握，从而为中华文化复兴与强国之梦插上大数据的翅膀，这是我们的期待，也是我们的呼唤。有鉴于此，2013年10月25日，首届“大数据——大文化”高峰论坛在上海宝山举行，受到业界的多方关注。当然，正是因为这次论坛，才催生了这本论文集。

本论文集共收录了13篇论文，从三个层次来探讨大数据与大文化之间的关系，层层递进。先从宏观视角探讨大数据本身，最后到大数据与大文化之间的关系，从大数据与文化发展、文化产业、文化事业、文化管理、文化消费、公共文化服务等多方面进行深入研究，再到案例分析，每一篇文章都极具原创性和创新性。而文集最大亮点在于其作者身份，除了擅长学术研究的专家学者，还有来自行政机构的政府官员，有文化、科技企业的企业家代表，还有处在大数据前沿的一线数据挖掘专家，他们都从各自的身份背景和职业入手，切身感受大数据带给他们的冲击或是改变，深入浅出地描述了大数据与大文化之间的关系。

文化部文化科技司司长于平在《大数据时代的艺术学对策研究》一文中强调了大数据时代的思维转变。他认为大数据由于数量巨大，且来源庞杂、非结构性强，它通常用“概率”说话而并不给出“精确”的判断。大数据将改变我们理解社会的方法，无论是解决时代的“问题”还是应对“问题”的时代，我们都必须关注“大数据时代”的思维转变。就对策而言，我们要日益增强数据的分析能力，有效实现对未来的预测能力。

对于大数据与中华文化的传承与发展，中宣部改革办副主任兼财政部文资办副主任高书生则强调加快建设“中华文化素材库”。所谓“中华文化素材库”就是文化资源数字化催生的一座文化“金矿”，对于传承和传播中华文化、变革文化生产方式、发展文化生产力具有划时代的意义。正如他在论文中所述：“文化消费的数

字化‘倒逼’处于上游的文化创作、生产和传播的数字化，从而引发文化生产方式的根本性变革。”

“无论我们选择与不选择，大数据技术进入文化领域是必然的。”中国艺术科技研究所文化标准研究中心主任闫贤良在论文中这样描述：“大数据打开了文化的窗口，便有两种新型文化业态快速成长。传统意义的文化将失去往日的宁静，不再自在自为。”

不过，面对大数据时代，更需要一种理性的认知，正如江苏省文化产业集团董事长、党委书记，南京艺术学院文化产业学院院长李向民论文的标题所言：“相信大数据，但不迷信大数据。”确实，如果不能准确把握和应用数据，纵使有云计算，也是徒劳无功的。文化产业从业者应当敏锐地发现大数据对文化产品制作和营销的影响，如果能够很好地运用，对于文化企业的发展将有非常大的促进作用，但是过于迷信也可能会变成谬误。

四川省文化厅宣传信息中心主任、研究员赵红川在论文中阐述道，大数据重新定义了文化的创造方式，扩展了文化资源内容，丰富了创意生成的手段，推动了新的文化生态的形成，甚至改变了文化的传承方式，因此，大数据具有非常重要的文化意义。

“大数据并不是一个充斥着运算法则和机器的冰冷世界，其中仍需要人类扮演重要角色。人类独有的弱点、错觉、错误都是十分必要的，因为这些特性的另一头是人类的创造力、直觉和天赋。”这是《大数据时代》这本书中，我最喜欢的几句话之一。对于这本论文集的出版，初衷很简单，就是想告诉大家，大数据并不遥远，它就在你我的身边。

（作者系中国文化传媒集团国家文化产业发展促进中心主任、中国文化报社编委）

目 录

◎第一章 宏观概述：什么是大数据？

■主题 1：大数据技术体系

04 | 大数据时代文化资源产业化应用与思考

李洪波

■主题 2：大数据与大文化

18 | 选择与被选择——大数据背景下的大文化发展潮流

闫贤良

◎第二章 中观视角：大数据与大文化

■主题 3：大数据与学科建设

28 | 大数据时代的艺术学对策研究

于 平

■主题 4：大数据与国家文化

37 | 加快建设“中华文化素材库”

高书生

■主题 5：大数据与文化产业

44 | 大数据与我国文化产业发展战略

张晓明 宋革新

■主题 6：大数据与文化管理

52 | 大数据与政府文化职能转变

赵红川

■主题 7：大数据与文化消费

61 | 相信大数据，但不迷信大数据

李向民

■主题 8：大数据与文化企业

70 | 大数据时代枢纽型企业的历史机遇

陆小华

■主题 9：大数据与公共文化服务

82 | 大数据时代的公共文化服务模式转型

孙一钢

◎第三章 微观分析：大数据技术在文化领域的应用

■案例 1：智慧博物馆

96 | 智慧博物馆大数据的发展与应用

马维理

■案例 2：互联网公司的实操

104 | 大数据之锤与大文化之钉

李维刚

■案例 3：数字内容的个性化实现

113 | 大数据时代数字内容价值发现与内容整合

屈辰晨

■案例 4：信用模型的构建与应用

122 | 基于大数据的信用模型构建和应用

徐卫华

131 | 后 记



第一章

宏观概述：什么是大数据？



LI HONG BO

李洪波

中国人民大学工商管理学硕士。现任北京中新视界数字科技有限公司总经理、中科院自动化所信息可视化联合实验室联合主任。

长期从事信息可视化、文化创意产业、版权产业、企业信息化等领域研究。曾负责创意产业测评指标体系研究工作，组织进行了《基于价值链的创意产品商业价值评估与交易公共服务平台》项目的开发和研究工作。负责国家新闻出版广电总局（国家版权局）立项的《版权作品价值评估体系和模型研究》课题研究工作。参与起草《著作权资产评估准则》《中央企业信息化绩效考核体系》等工作。

大数据时代文化资源产业化应用与思考

李洪波

摘要：信息技术的迅猛发展，对文化的变革与发展有着深远的影响。2013 年被称为“大数据元年”，随之而来的数据仓库、数据安全、数据分析、数据挖掘等围绕大数据的商业价值的利用逐渐成为行业人士争相追捧的利润焦点，同时，大数据技术也深刻地影响了我国传统文化和文化产业的转型升级。

文化产业对于社会的进化与平衡有着巨大的责任，而科学技术能够有效地改变生产方式、提高生产效率，也可以更具体准确地进行数据统计和分析。如何将大数据技术作为一个概念性的工具，以科技的方式应用在文化领域，促进文化产业与科技的对接、转变生产和增长方式，以及推动文化产业高质量、高效率的发展，加快文化科技融合的应用模式和创新手段，是我们需要思考的问题。

本文在简要概括了大数据的概念与特点及发展趋势之后，结合典型案例，从大数据技术对文化产业价值链系统的重整以及文化产业大数据建设与应用路径方面进行了分析。最后，对基于全新价值导向的对策进行了思考和梳理。

关键词：大数据 文化产业 客户体验价值

党的十八大明确做出推动文化大发展、大繁荣的决定，提出“必须坚持以文化创新为动力，进一步推动科技和文化融合，加快形成构建现代文化产业体系”。在加强文化产业装备科技化、文化生产要素数字化的同时，对文化消费驱动的客户体验价值急需运用云计算、信息可视化和网络化等科技创新构建大数据体系，建立从文化资源到文化消费的闭环产业价值系统，推动文化产业真正实现跨越式发展。

2013 年作为全球“大数据元年”，美国已经将大数据作为国家最高战略制订了实施计划，奥巴马将“大数据”比作“未来新能源”。谷歌通过对搜索数据的大数

据分析准确预测了电影票房收入，Netflix^①以大数据分析为基础，用巨资打造的《纸牌屋》在全球43个国家热播，大数据对文化消费和文化创新的巨大作用与前景日益彰显。

费孝通先生在《社会进化到社会平衡》中说明了从社会学的角度来考虑“社会的进化”和“社会的平衡”这样的概念工具如何运用，土地和文化如何支持人口的增长和需求的增长，从而能够使得社会进化趋势动态地形成一个社会平衡。这带来三个方面的启示：首先，文化的作用和发展空间是巨大的。由于相关的资源越来越稀缺，而人口的增长和人的需求弹性也随之增长，对于如何满足人的欲望，如何形成新的平衡，文化产业有着巨大的责任。其次是技术。技术有效地改变了生产管理，使各个方面都有了很大的提高，也使我们能够更好地量化人的行为特征、了解人的内心需求，同时也支持创新发展的新型驱动模式来支持战略的转化。最后，大数据是一个“概念工具”。这个工具在不同的学科有不同的理解，技术学科中专门指云计算等方面的技术，对于产业来讲是企业的数据库资产，在社会学里是一种现象，影响人们的思维观念。

以客户体验价值思维模式对传统文化产业进行颠覆性创新的时代已经来临，利用大数据技术可以实现对客户体验和消费行为的累积化进行相关性分析，从而在促进文化资源资产化的同时，也促进作品更好更快地商品化，同时反向支撑文化产业新的价值评估体系的确立。在更好地满足人民日益增长的文化需求的同时，推动文化产业资本化进程。

一、大数据技术及发展趋势

（一）大数据的定义与特点

1. 大数据的定义

从技术视角来看，目前全球均比较认可 IDC (Internet Data Center)，即国际数据公司，对“大数据”的定义，即为了更经济地从高频率获取的、大容量的、不同结构和类型的数据中获取价值，而设计的新一代架构和技术。此定义也可以概括为4个特点，即高容量 (Volume)、多样性 (Variety)、速度 (Velocity)，以及价值

^① 是一家美国公司，在美国、加拿大提供互联网随选流媒体播放，定额制 DVD，蓝光光碟在线出租业务，换言之，就是一家在线影片租赁提供商。

(Value)，简称为4个V，包括基础架构、数据管理、分析挖掘和决策支持4个层面。当然，也有其他不同的观点，IBM对于大数据的定义便是规模性（Volume）、多样性（Variety）、高速性（Velocity）和真实性（Veracity）的“4V理论”。

2. 大数据炒作期与投资误区

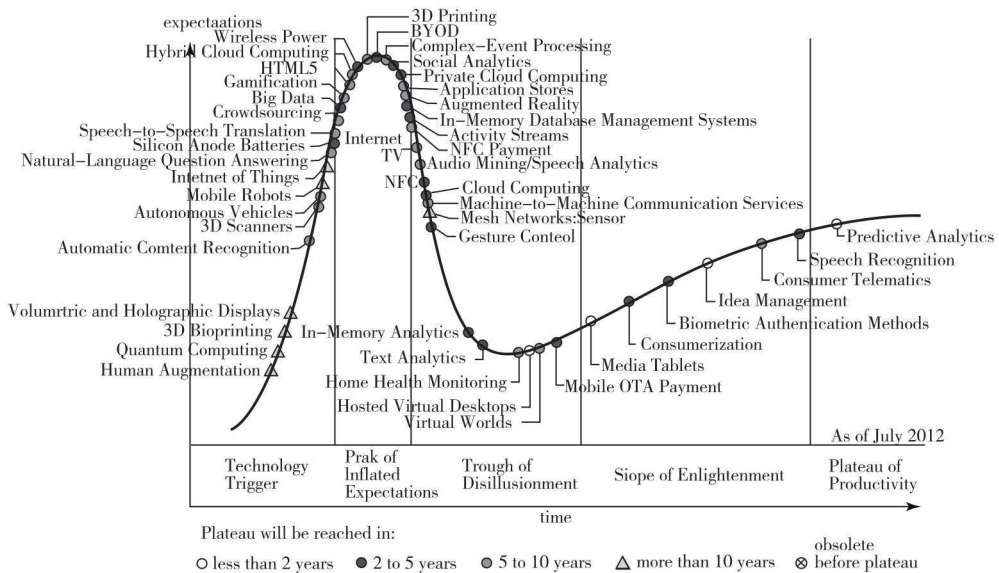


图1 大数据炒作期

来源：Gartner 2012 新兴技术炒作周期（Hype Cycle）报告

在Gartner^①的2012新兴技术炒作周期（Hype Cycle）报告和Readwrite^②的文章当中提到，大数据实际处在一个炒作期，相关技术存在着发展规律，2013年虽然被称作全球“大数据元年”，但仍需正视大数据能够带来哪些真正的价值，《大数据时代》的作者舍恩伯格在他的另一部著作《删除》里面着重论述了对大数据的取舍之道。同样，Readwrite在近期博客中提到了现在盲目投资应用先进的技术，如hadoop架构^③，没有进行很好的大数据治理设计，仅实现了对大量垃圾数据的收

① Gartner Group 公司成立于1979年，它是第一家信息技术研究和分析的公司。它为有需要的技术用户提供专门的服务。

② 一个关注技术的博客网站，创建于2003年。

③ 一个分布式系统基础架构，由Apache基金会开发。用户可以在不了解分布式底层细节的情况下，开发分布式程序。

集，投入 1 美元产生的回报只有几美分，因此，业界应该将大数据这个概念工具与自身实际需要相结合来进行理性投资。

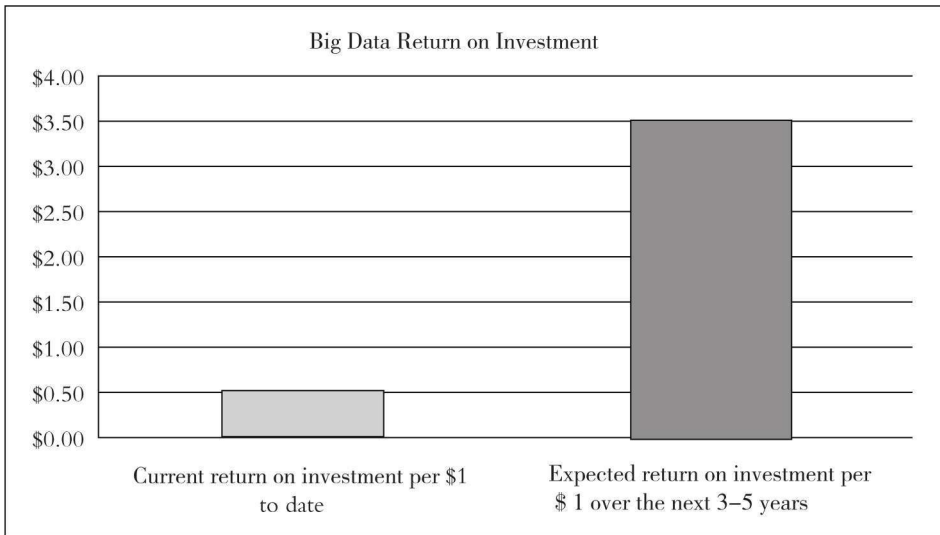


图 2 大数据投资回报对比

来源：Readwrite 博客文章

这里提到了一个隐含的概念，大数据如果从整个数据的特点来讲，更多的是在思维特点里面强调真实性，从技术的角度来讲对它的价值性则要提得更多一点。凯文·凯利^①在评价相关的技术成长时提到，技术已经变成了一个生命，它有自己的发展特征，因此，企业可以通过对有价值的小数据进行不断积累，来帮助企业更好地了解客户的价值。在了解客户价值的基础上获取数据，实际上就把用户的行为特征人为地形成了企业或公共事业单位的数据资产，相关的云计算、物联网、穿戴式设备使得自我量化能够形成可计算的内容，并在其中运用一些创新的方式，形成一些价值的转化开发，最后使反馈经济不断地进行循环。

（二）大数据发展趋势与应用前景

大数据技术的发展将带来大规模的交易数据、大规模的行为数据和大规模的融合数据，同时带来三大趋势：数据将成为资产、应用软件泛互联网化，以及行业应

^① 《连线》（Wired）杂志创始主编，被看作是“网络文化”（Cyberculture）的发言人和观察者。

用的垂直整合。数据资产是未来企业的竞争核心战略资源，而拥有数据的规模和活性以及对数据的解释和运用，将是下一阶段企业着重要构建和发展的能力。围绕数据，可以演绎出多种新的商业模式，如：租售数据模式、租售信息模式、数字媒体模式、数据使能模式、数据空间运营模式、大数据技术提供商，以及提供大数据的处理技术等。行业应用的垂直整合，使得越靠近终端用户公司，就会在产业链中拥有越大的发言权。

图灵奖^①获得者、著名数据库专家 Jim Gray 博士观察并总结人类自古以来在科学研究上，先后历经了实验、理论和计算三种范式。当数据量不断增长和累积到今天，传统的三种范式在科学研究中，特别是一些新的研究领域已经无法很好地发挥作用，需要有一种全新的第四种范式来指导新形势下的科学研究。基于这种考虑，Jim Gray 提出了一种新的数据探索型研究方式，被他自己称之为科学研究的“第四种范式”（The Fourth Paradigm）。

因此，大数据技术将广泛应用于基于标准化和元素化的文化资源产业化应用。文化资源是人们从事文化生活和生产所必需的前提准备，是人们从事一切与文化活动有关的生产和生活内容的总称，而文化资源数据化是指将不同时期、不同载体的文化资源内容进行碎片化和结构化的数字化处理，以文字、图片、音乐、故事、影像、声音等方式进行数字化记录及集成。

文化产业之丰富、巨大、多元化的信息和素材，对大数据式的解决方案及思路有着必然的渴求，凭借大数据的处理及价值创新方式，文化产业基于文化资源的开发利用，能够更好地预测客户需求或喜好，洞察市场走向，进而有的放矢地创新文化产品与服务。同时融合信息、金融、科技等社会服务体系，衍生出更丰富、更有创造力的文化产业形态。

二、大数据技术对文化产业价值链系统的重整与再造

（一）大数据重塑产业价值链为闭环价值系统

大数据对文化产业价值链的重塑和闭环价值系统的形成，为作为投入要素的文化资源活化利用提供了广阔的空间。中华文化源远流长、博大精深，不论从时间、

^① 图灵奖（A. M. Turing Award，又译“杜林奖”），由美国计算机协会（ACM）于1966年设立，又叫“A. M. 图灵奖”，专门奖励那些对计算机事业做出重要贡献的个人。

空间，还是从发展过程、多元化程度等维度来看，都有其无限广博、厚重的资源信息。在社会发展的进程中，分散的文化资源很容易被忽视、淡忘，甚至没有机会被发掘出来，以至于造成传统文化的断层、流失。文化是一个民族生生不息、代代相传的基础，文化资源的保护和有序梳理是一项复杂艰难而又必需的任务。

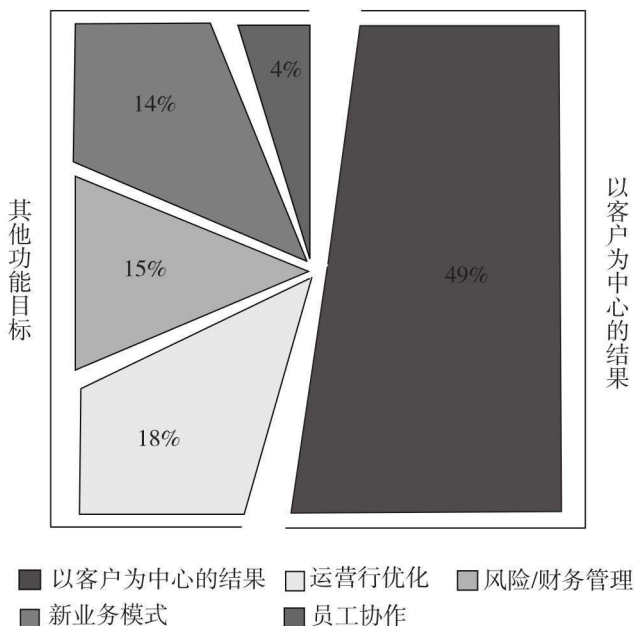


图3 大数据重塑产业价值链

来源：牛津大学赛德商学院与 IBM 合作研究

在大数据环境下，综合运用信息技术、互联网技术、可视化技术等科技手段，以市场为导向，从内容创作、产品化、市场营销、消费体验和 data 反馈将文化产业价值链重塑为闭环价值系统，从而遵循市场有效配置资源的规律，形成对文化资源的合理开发，惠及全民。

（二）大数据重塑产业价值链典型案例

1. 谷歌电影票房商业预测案例

2013年5月，谷歌发布《谷歌搜索预测电影票房白皮书》，公布了一项重要研究成果——电影票房预测模型。该模型能够提前一个月预测电影上映首周的票房收入，准确度高达94%。谷歌票房预测模型的基础是与电影相关的搜索量与票房收入的相关性分析。大数据分析在电影行业的应用前景是把模糊的行业经验变得更科