

【按语】大数据产业是目前讨论热烈的话题之一。数据产业应当说归属信息技术领域,但数据应用,尤其是对已有数据二次或多次开发涉及的就不仅仅是技术领域,还有社会及人文领域,广而言之,涉及更为广泛的文化领域。从网络及信息应用发展趋势看,美国对数据的应用,尤其是数据在文化领域的应用是领先的,很多适用于现代社会文化消费方式的新理念值得我们借鉴。

法国阿维尼翁文化论坛是欧洲最具影响力的大型文化论坛,从2008年起,每年11月举办,现已举办6期。法国政府对这一文化论坛高度重视,2011年,时任法国总统萨科齐出席论坛发表主旨演讲并参加分会讨论。每届论坛法国文化与交流部长都参加。因其巨大的影响力,阿维尼翁文化论坛被誉为“文化达沃斯”。《大数据大文化:数据不断增长的力量及对文化经济的展望》(文中简称《大数据大文化》)是2013年阿维尼翁文化论坛最有分量且最具产业指导意义的研究报告,对于文化经济的发展具有重大导向作用。因文章篇幅较长,本文特对该报告的中心内容及其对中国的启示做一述评。

## 大数据应用与文化发展趋势

### ——《大数据大文化》研究报告述评

■(加拿大)林青

大数据就是通过改变数据应用规模和形态而形成的不断增长的数据及其能力。由于数据得到越来越多的分享和互联,因此在一定程度上它可以揭示出产业活动所透露的价值和意义,而这种有意义的数据只要被发现,经过加工处理,就可形成一定的应用,这正是大数据产业的使命。大规模数据的应用已经对文化经济产生巨大的推动作用。当今,人们对此越来越重视,因为大数据是一种不断扩张的数据能力,它在人类决策活动中日益成为不可或缺的依据。大数据应用于文化及关联产业相比电商和金融等行业来说相对滞后,但影响不可小视。从商业角度讲,大数据改变了文化产业运营模式和形态;而从社会角度看,它改变着人们的生活方式及社会结构。面对这一迅速膨胀的产业,如何把握时机,乘势促进文化产业,是我们需要思考的问题。

[关键词]大数据;大文化;文化经济;文化产业

[中图分类号]G114 [文献标识码]A [文章编号]1004-518X(2014)03-0247-09

林青,中国数字图书馆有限责任公司副总裁,蒙特利尔大学博士。

#### 一、数据产业的发展与现状

数据一词并不新鲜,从有计算机开始就有数据。但大数据则是近几年随着云计算、卫星定位及多媒体应用而产生的现象。简单地讲,大数据就是通过改变数据应用规模和形态而形成的不断增长的数据及其能力。就数据本身而言,它包括“原面貌的”数据资料,诸如数字化

的一个网页、一张照片、一段视频、一首歌曲,以及“经过加工提炼的”数据资料。由于数据得到越来越多的分享和互联,因此在一定程度上它可以揭示出产业活动所透露的价值和意义,而这种有意义的数据只要被发现,经过加工处理,就可形成一定的应用,这正是大数据产业的使命。

数据产业发展经历了两个阶段。第一阶段在20世纪90年代,互联网的信息改变了数据形态和应用规模,特别是与文化相关的、面向公共大众的信息。这一阶段的核心是利用信息系统分析信息、更有效地服务客户,它的主要逻辑是通过专家支持,检索出客户所需内容的网站。第二阶段发生在20世纪90年代末和21世纪初,90年代后期网站从百万数量级发展到2013年两亿个网站的规模,计算机要分析的数据规模已不仅仅是几百万个网站,而是数以亿计的网页或资料。鉴于全球网站数据迅速增长的状况,以雅虎为代表的那种专家支持形成的网站分级分类(年鉴)目录式的检索方式已不适用,逐渐被谷歌式的数据处理方式所取代,即将指令自动化处理,并更注重关联内容,而不是网站或机构本身。最有价值的是查询出来的内容,而网站或个人不过是所查内容衍生出来的附带结果。

在经历指令自动化处理和内容驱动查询这两个形态转变后,大数据处理又面临新的挑战。第三个转变要解决的问题是:如何利用数学方法分析数以十亿计的超本文链接结构,从而自动评价出网站内容的价值。

正是经历了上述几个阶段的演变,以及社交网络、移动终端、电商和计算机应用软件的普及。大数据的应用时代才真正到来,截至2012年底,谷歌的搜索量已经达到20 000亿次/年。根据资料显示,截止《大数据大文化》<sup>①</sup>报告发表时2013年的统计数据,谷歌和FACEBOOK各有近12亿用户/月;微软有9亿多用户/月;雅虎和腾讯各有8亿多用户/月;而苹果、百度各有约6亿和近6亿的用户/月,亚马逊的用户量也达到6亿/月<sup>②</sup>。这种大规模数据的应用已经对文化经济产生巨大的推动作用。当今,人们对此越来越重视,因为大数据是一种不断扩张的数据能力,它在人类决策活动中日益成为不可或缺的依据。但需要说明的是,这种数据呈指数形态的增长实际上是不可控的,它不断地发生规模和形态上的改变,而且几十年来不停地自我供给,不断激发新发明,不断完善由信息发掘而生的基本服务,不管这些新发明是必需的,还是随机而成的。

数据几乎无处不在,数据来源有公开数据、私人数据和半私人数据(社交网络),而且数据之间相互交叉、

相互激发,它们可以创造出多种机遇和灵感,应用更是五花八门。美国领先的大数据公司帕朗泰尔(Palantir)的“放大镜头技术”(Zoomer)就可以在任何时间,提供世界任何一个城市的一个社区的准确信息,监测到居民和企业的工作、消费和生活情况。这些技术的应用导致的问题是:数据采集发掘的技术可能性远远超出我们管理数据的能力。大数据产业刚刚诞生,发展超前,使人们既窥见到机会,也感觉到潜在风险。而就文化产业而言,《大数据大文化》报告指出:文化生态毕竟是以文化为核心的系统,不能只让数据技术人员来定义未来文化经济的游戏规则,对数据凶猛发展的态势应当迅速警觉并找出解决数据发展失控的方法。

## 二、快速增长的大数据市场形成龙头企业统治地位

大数据市场是个新兴市场,近年来发展强劲,逐渐成为个人资料价值化的产业。2012年全球大数据产业市场规模估计在63亿美元。虽然这是一个非常年轻的产业,但增长极为强劲,近年来每年的数据产业收入平均增长率为40%以上。展望2018年,这一产业的市场规模可以达到500亿美元。未来五年受大数据影响的最期待的产业是录制音乐产业。2012年录制音乐的收入为165亿美元,其中34%为数字化关联产品。

但对数据的控制和使用却集中在大型企业手中。根据2012年的统计,2%~5%的大中型企业开始体验并使用大数据;而对规模企业而言,17%~30%的企业使用大数据;而美国前500强大企业中有90%使用大数据。但在2012年,真正上规模的大数据项目不过数十个而已。实际上,拥有大数据基础设施对运营是至关重要的,但在世界范围内这些设施却集中在为数不多的几家大公司手中。在经历了创新浪潮,建设大规模信息处理设施后,大数据产业转向了数据增值应用,企业调整方向,投资更多倾注到不同行业、不同体量的应用解决方案之上。2013年,公共领域、电信运营商和娱乐/媒体产业成为大数据最主要的用户。在今后几年中,参与大数据项目的企业会越来越多,而且各有所求。一方面,吸引传统的互联网用户在移动互联网上使用数据;另一方面,大规模建设移动互联网和云计算设施也是为了缓解数据

龙头企业搜索业务的运营压力。

首先,大数据处理能力高度集中在少数龙头企业,势必形成行业统治地位。这些龙头企业在文化经济中取得的成就和影响力代表着一种稳重和持续发展的趋势。大规模处理数字化信息的能力使这些企业获得了得天独厚的优势,能够制定一般经济领域,特别是文化经济领域的游戏规则。比如,苹果就利用其优势地位在世界范围内强化一种应用标准,制定出广告、市场推广、分销、传播、销售以及在线音乐消费的新的游戏规则。同样,谷歌和FACEBOOK在各自强项领域内也独占鳌头。在信息搜索和基于互联网的社区服务中,这两个巨头分别占据了互联网50%以上的用户。它们各自用户量均超过10亿。这样的强势地位不仅使它们可以定义广告的游戏规则,而且还可以随意改变这些规则(Ad libitum)。例如,2012年推出的Knowledge Graph就试图直接回答互联网用户的问题,特别是有关文化的问题(地点、人物、艺术作品等)。这一新形态彻底打破了搜索与网页内容提供者相依共存的传统模式。谷歌不再像以前那样系统地将用户转到网址,而是通过直接回答客户问题的方式,让客户在谷歌网页上找到这些网址。谷歌的做法从某种程度上讲使得现有的互联网纯玩家的经济模式变得越来越不稳定,去中介化的倾向极为明显。由这种模式建立的语义学网络为这项新服务积累了大量的素材。目前已经有5.7亿数字化的物品以及经过加工的数字化信息,这些素材和信息完全可以进行“自我描述”,也就是说可以直接将这些作为答案回复给一些基本的、由事实构成的问题。

再有,所谓大数据的基础设施(大量的存储及网络运营设备,以及相关的软件)多为龙头企业所有,因此它们有能力通过采集用户的“投票”,评估其使用偏好,并依次加诸并实施新的游戏规则。而就所谓“投票”原则来说<sup>③</sup>,占统治地位的龙头企业表现出愿意相互接受这类游戏规则倾向,因为从本质上讲,这些游戏规则符合他们倡导的消费者偏好为先的逻辑。比如,Facebook近来推出一个新的广告服务,在这项服务中,Facebook可以将用户的姓名和照片置入广告,除非用户事先声明不允许这样做。在看到这项服务被广大用户接受后,谷歌也采

用了这样的逻辑并在其Google+社交网络上应用。

而在软件功能的游戏规则制定上,谷歌公司也占据着绝对的统治地位,2012年谷歌的安卓系统在已售出的智能手机市场占据80%的份额。目前,这两大巨头一个在搜索和手机软件方面占有绝对优势,一个在社交网络上独领风骚。两个巨头公司就这样用新的游戏规则强化各自的地位,使所有的业内其他企业望尘莫及。

掌握在巨头公司手中的信息基础设施本身就反映出大量的资金投入和对其他相关企业的收购,这是大部分其他公司可望而不可即的。对于文化企业而言,这些巨头获得的领先地位是它们在中短期无法赶超的,而且就目前而言,差距越来越大。报告预见:在未来几年中,文化企业只要能挖掘利用数据,更好地服务于文化活动,并从这些变化中获利,就有可能成为这些巨头的客户,有些企业还可能成为它们的追随者。

### 三、从数据作为市场工具到数据作为管理依据

对文化产业影响最大的,也是改变其经营形态的数据使用有两种:一种是利用数据作为市场推广的依据,即对客户消费偏好的数据加以分析,有的放矢地进行市场推广,既激发消费,又可引导数字化产品的生产。这方面在美国做得最好的是以视频文化产品租赁服务为主的Netflix公司和以个性化自动音乐组合服务为主的Pandora公司。另一种是利用数据参与公司本身的管理。数据不仅是文化产品消费和分析的结果,具有一定的导向性,而且数据本身也参与公司的运营管理之中。这方面比较有代表性的公司是美国的游戏公司Synga和社区游戏公司EA。

#### (一)Netflix<sup>④</sup>的发展与特色

1997年Netflix创建于硅谷,业务是通过网站租赁DVD,并通过美国邮局做租赁物流,它的服务简单,选择丰富,比当时美国Block Buster的价格要便宜。2007年起,Netflix抓住机会,在网上开展了视频流点播业务。客户可以上网租看电影和连续剧,消费成本大大降低,公司的竞争力得到提高。2013年Netflix在网上实现了通过各种上网终端(电视、游戏机、智能手机、平板电脑等)点播电影和电视连续剧的服务。Netflix从2006年600多万用

户发展到2013年4000多万注册用户,其中3100万为美国用户,业务已覆盖近40个国家。

Netflix运营的最大特点是,它的订单引擎不仅有一般的搜索功能,还能够根据客户的趣好和意见为客户推荐新的选择。这一系统的核心就是知识应用(Savoir-faire)。从2000年起,Netflix推出了Cinematch的个性化服务。这项推荐服务不仅在回答客户搜索中体现,而且配以相应的广告。还有部分推荐是通过邮件传递给客户的。2006年Cinematch每天已经能够产生10亿个推荐。这样巨大的数据处理有时难免产生与客户需求不相符合的推荐,为此,公司利用从48万个客户那里获得的上亿条涉及17770部电影的匿名资料对推荐进行了改进。2009年,Netflix开始了视频流(Streaming video)注册用户服务,客户选片错误可以瞬间改正。这项服务的成功使推荐价值(是否采用)更少地依赖于对电影的注释,而更多地依赖于消费者使用和行为的偏好。由此Netflix将改善推荐的准确度转变成对客户消费信息的研究,诸如客户使用什么设备观看、观看的时间表(每周的哪些天,什么时间)、他们的地理位置等信息。需要指出的是,直到2013年,Netflix推荐机制仍然没有完全自动化,公司有超过40个人每天以手工方式对每一部电影或电视连续剧做种类、时长、演员名称等注释。Netflix目前75%被观看的内容都是通过推荐引擎所实现的。实际上,每天由客户产生的电影注释多达400万,注册用户每天在线消费视频的总时数超过10亿小时(晚上流媒体占美国互联网流量的33%,为垄断的份额)。因此,Netflix仍在努力并拿出每个季度10亿多美元营收的一部分投入到运营及推荐数据的自动分析系统。

## (二) Pandora 广播电台节目创作及编程的自动化

2005年,提姆·外司特哥林和他的同事利用“音乐类型计划”(The Music Genome Project, MGP)创办了基于网络的“Pandora”<sup>⑥</sup>音乐电台,这个电台的独特之处就在于它通过计算机的算法将音乐题目和风格自动分门别类。Pandora运营的基本思路与Netflix一致。Pandora在运营中也遵循着以听众趣好为导向的逻辑,听众可以在网上建立自己的一个或几个电台,这种创建电台的过程完全是自动化的,是由程序根据听众的趣好而形成的。

2013年,Pandora已经拥有超过100万个标注识别的曲目,分成500个不同的音乐类别。每个月有超过7200万人使用这个音乐平台,而仅2013年第二季度,Pandora就为听众播放了40多亿小时的音乐。

Pandora的成功取决于根据听众的趣好和音乐家人工分析所形成的把握听众脉搏的智慧,并通过程序提供自动化的服务。这种听众与程序之间的互动,构成了一个持续性的“触点”,再加上人工分析干预,使整个系统可以不断捕获新的音乐趋势、音乐类别和听众趣好。这个机制实际上就是一种客户的投票,所选曲目可以帮助系统不断完善,并更贴近客户的需求。2013年8月,Pandora就已收录了近300亿条选曲形成的“投票”。

## (三) 以 Zynga 为代表的社交游戏使数据介入管理和盈利模式

游戏市场是文化经济的支柱之一,近十年来,全球基于互联网的数字游戏玩家从2亿发展到10亿。个人电脑、智能手机以及平板电脑都成了游戏的载体。商家更注重对游戏平台衍生出的资料进行开发,使这些数据成为游戏管理和盈利模式的基础。在这方面成就最大的是以社交游戏为主的2007年在旧金山创立的Zynga公司。这家公司最独特的开发思路是:游戏创新、挖掘和升级由始至终要以数据为导向,游戏要通过社交网络获得客户和玩家的忠诚度。最成功的一款游戏是2012年9月上线运营的“农庄城2”(Farmville 2)。从当年11月起,每个月玩家都超过6900万,游戏覆盖180个国家。在最活跃的时候,玩家平均每天生产出3.5亿个收获,每秒钟平均种下约8000块田地。2013年第一季度,Zynga游戏每天平均产生近700亿个事件(鼠标按动一次即为一个事件),这些活动是由每天5200万玩家完成的。这样大规模的数据分析和开掘可以更准确地根据玩家的趣好改善产品,引导消费,促使他们购买更多的游戏虚拟物品,同时,还可用游戏点或游戏币招募其他玩家。时间货币化是游戏盈利的重要手段之一,为了更长时间地留住玩家,玩家资料数据就成为开发者和设计者创造游戏,创造游戏环节,延长游戏时间的最好依据。资料数据还可以用来引导盈利模式,诸如根据游戏中的交互性回馈机制,在游戏中加入虚拟物品买卖(比如购买农庄城的拖拉机)、产

品植入和广告穿插。

但是利用数据也不是成功的保证,游戏市场是有周期的,最好的游戏玩家最终也会失去兴趣,因此必须持续性地发掘一些新的游戏感觉并将其产品化。掌控数据也不是万能的,在数字游戏竞争的环境中需要不断地创造新的开发背景及其标准。

#### (四) EA 游戏的情景创新与数据参与管理

EA(Electronic Arts)最早是做电子游戏的,2001年公司转型投资在线游戏平台和在在线游戏。2012年公司已经成为世界上第三大数字游戏供应商,年营业额超过40亿美元,其中有10多亿来自数字化游戏。EA 独具创新的游戏是将活生生的场景植入到游戏当中,而玩家正是这种游戏最重要的角色。为了将不同场景呈现给玩家,设计者将玩家每天的不同时段和游戏体验进行分析,做出适合各个时段和玩家设备的游戏场景。比如:早餐时可以在电脑上玩一局短游戏;上班路上重新看看游戏状况并在论坛上聊上几句;中午吃饭时可以按部就班地推进游戏;下班的路上可以玩几把在智能手机上设定好的游戏,晚上回到家可以玩一局时间长的游戏。

除了这种贴身适时的游戏设计外,EA 也和 Zynga 一样特别重视增值服务,诸如:注册用户收费、虚拟物品买卖、产品植入、广告元素设定等。而这些增值服务模式是根据客户趣好和游戏规则组合而成的,并且随时可变。EA 的商业策略完全像编码一样写入了游戏规则之中,其目的就是要刺激玩家兴趣,刺激他们花更多的钱,刺激他们更长时间的滞留,从而间接带来广告和商品植入的效果。

利用数据进行管理和分析是为了更好地引导玩家消费。比如,在一款游戏中,玩家资料就可以显示出玩家开的饭店(Burgers)的数量,以及他们为此花费的虚拟币(Donus)。这种持续的数据统计有利于改进游戏,并可以证实公司的决策和新的应用是否受到玩家的赞赏,是否有利于游戏的回报。而最值得称道的是 EA 利用数据资料制成了玩家活动地图,玩家可以一目了然地看到哪些游戏空间用得最多。这些地图标识可以起到建议作用,让玩家有一个更好的环境,游戏更适合他们的趣味,甚至可以建议有同样趣味的玩家组合在一起玩游戏。再

有,EA 利用玩家数据可以分析不同时段的收入和玩家数量之间的关系。同时,在线游戏还可以改造成适于其他媒体的电子游戏,对此公司根据不同载体的特点和玩家使用的频率做出相应的市场推广。

#### 四、大数据可否成为文化经济的催化剂?

数据挖掘和应用能否成为文化产业发展的催化剂?

这恐怕还要经历几年的发展才可做出判断。但利用数据资料开发应用,并且促进文化经济发展已经是不争的事实。文化经济越来越倚重于由数据作为导向的管理。文化企业多为中小企业,往往不具备技术能力,因此需要专业的数据运营及服务商帮助它们实现文化领域的的数据服务,比如云计算服务商、移动运营商、新兴的数据服务企业和 IT 巨型公司。同时还需要指出,数据产业也要与文化相关的行业紧密结合,形成所谓文化生态系统,这个系统的外延远远超出相对局限的文化产业,比如旅游产业。同时还要注意的是:利用大数据发展文化产业不能被大型数据服务商所捆绑,未来数字化文化经济的游戏规则不能单独地凭由它们制定。

##### (一) 大数据可为旅游行业提供全方位的服务

每年赴法国的外国游客高达7700万人次,这样大的群体完全值得数据产业很好地挖掘。世界各国都非常重视旅游产业的发展,金砖四国这几年来旅游业的增长一直是两位数。而旅游业在法国带来的直接就业人口为100万,关联的就业人口也达100万之多。2012年,旅游消费产值达1380亿欧元,占法国国内生产总值的7%。这样一个与文化息息相关的产业正是各个国家高度重视发展的产业。所谓旅游景点大多是由不同的文化遗迹、文化设施、文化场所或其他文化元素构成的,因此广义地说,旅游业就是文化经济的一部分。因此大数据在这一领域的开发应用理所当然就是对文化的促进。

2011年法国旅游发展局与 MFG 公司合作发起了一个项目,该项目旨在通过社交网络分享法国旅游及文化照片。由此可以统计出哪些是外国人最喜欢、去的最多的景点。这项实验的结果可以针对外国游客促销法国旅游并为管理提供依据。2012年,法国蓝色海岸旅游区域委员会(CRT)与法国著名的移动通讯运营商 Orange 合

作,发起了一个导航体验,一方面要将游客的存在与流动进行量化与模型化试验。诸如:检测游客最喜欢参观的景点、最喜欢的住宿服务、每个景点平均过夜数、从到达蓝色海岸旅游区到离开的天数等。另一方面从移动电话收集的信息中提取潜藏的意义,以为当地旅游管理部门提供便捷化、产业化的决策依据。这些数据制成统计表格和形象的图标,解释并导出相关的旅游服务决策:最佳的酒店配置、欢迎外国游客的方式、不同类型不同国籍的游客对景点和最流行的旅游活动的认同度、餐馆/酒吧/演艺夜间的光顾频率等。比如,在2012年旅游旺季的夏天,Marineland地区附近的酒店入住不足,这种反差直接导出管理建议:改善当地酒店居住条件,加大景点晚间的吸引力以增加游客流量和黏度。

而从游客在社交网络上发布的旅游照片做数据分析,也可以试图理解游客是怎样感知一个国家的。“走遍法国”是法国旅游发展局的网站,这个网站就利用社交网络用户的信息资料(照片)推导出不同国籍游客所选出的最喜欢的旅游景点。而这些信息反过来又可以指导“走遍法国”网站的景点推荐和根据游客人数制定的市场宣传策略。

这些在社交网站发布的照片数据经过认证和分类,然后与“走遍法国”抽样调查的结果进行比较,就可以得出多种不同的结果。在几百万张照片基础上,根据照片源头信息和位置信息,制出一张由游客参与而成的带有标识的参考地图。这样的地图建立在这样一种假设上:一个景点越多人拍照,也就越有名。实际上一张照片就是一个摄影师的选票,对法国文化遗产和爱国情绪有影响。

“走遍法国”对分享在社交媒体上的资料挖掘是非常有效果的,被认作旅游情报学。它与过去在机场采访外国游客,给人多种选择的问卷调查大不一样。社交媒体上的资料虽然弱,但数量巨大,可以从中提炼出很有价值的信息。当数据自动地从社交媒体上提取出来并被分析,其成本相比传统的问卷和其他类型的调查要低廉得多。2013年,“走遍法国”还更有针对性地利用数据进行旅游促销,主要针对亚洲客户(中国人和日本人)和俄罗斯人,他们在亚洲最知名的社交网络投放了一些照片

使游客能更好地理解法国,精准地做一些市场促销。2013年9月,“走遍法国”还和西班牙的世界票务预订专业公司 Forward Data SL 达成战略合作关系,通过国际旅客预测旅游目的地的客流趋势和频率。

## (二) 公共数据在不同领域之间的应用产生的合力

数据应用最典型的例子是迪士尼大型娱乐公园为游客准备的神奇手镯。迪士尼是世界级的媒体巨人,也是世界第一的大型娱乐公园。全世界所有迪士尼公园总游客量超过1.2亿人次/年。在移动互联网时代,游客越来越习惯于通过互联网得到服务。神奇手镯是近年来迪士尼研发的一个名为“我的神奇帮手”(My Magic Plus)的项目。这个项目就是要为游客在游览全程中提供便捷的人性化的服务:从预定娱乐景点门票,园中全程游览一直到客人返回住所。这个手镯(也叫“神奇手带”)有个数字存储器,可以将公园门票、景点和餐馆预订、客房钥匙和付款方式的数据存储下来,同时可以收集游客在公园的任何时刻的信息,包括他的购买记录、他的选择和爱好等。这种设计的最初动机是:尽可能地满足游客个性化的体验,让他们玩得简单舒适,而且需要时还可以随时改善他们不满意的服务,即所谓的“神奇体验”。第二个动机是:为迪士尼创造更多的收入,比如为游客提供个性化的建议,影响游客消费;或使公园的服务更符合游客消费习惯和偏好。

采集和分析游客数据的平台需要的投资预计为10亿美元,第一个试点是最大的佛罗里达迪士尼公园,已经有20万游客享受到这项服务,预计在接下来的几个月中(即2013年第四季度)迪士尼将会为所有游客佩戴“神奇手带”。如果试验成功,平台将会为全球所有迪士尼公园的1.26亿(2012年)的游客提供服务。

对公共数据的专业挖掘是数据行业细分的必然趋势,也是很多中小数据公司得以生存的缝隙。在这方面做得比较出色的是“旅行顾问媒体集团”(Tripadvisor Media Group)。该集团公司在五大洲27个国家拥有不同品牌的17家旅游网站。公司根据各个网站的用户交流推荐的旅游信息,衍生出超过一亿条有关酒店、餐馆、景点等服务的评审和意见,这些信息经过加工整理,分门别类地上传到各个网站,为全球3200万注册用户提供服务。

2013年9月“旅行顾问”推出了一项名为“高明的旅行顾问”的专业服务:对客户交流和建议的数据(文字/图片)进行统计,为旅游专业人员提供旅游目的地的发展趋势和建议。“旅行顾问”在中国的官网是“到到网”(www.daodao.com)。

另一个是Yelp公司,该公司利用公共数据提供特色鲜明的地图服务。2004年该公司成立于旧金山,2011年在纽交所上市(NYSE:YELP),在此之前,曾避免了谷歌和雅虎的收购。2013年7月,Yelp推出以城市旅游和文化景点地理信息为数据的地图服务,地图通过不同标识标出旅游“热点”,同时游客还可以通过关键词寻找文化场馆。

类似服务模式也应用到政府的公共服务中,2004年,英格兰艺术委员会就曾发起一个研究计划,名为“伦敦文化地图”。这个项目的目的是要让老百姓了解伦敦城的文化建设基金都用到了什么地方,提高政策执行的透明度。“伦敦文化地图”可以让观众和专业人士获得有关伦敦文化活动的信息。地图数据同时还可以显示出文化基础设施地图标识、每个居民和每个居民社区的文化消费支出情况、文化习惯等。同时这些信息又和交通等信息结合在一起,看上去非常直观、量化。比如,可以看到在哪些社区举办经典的、市民踊跃参加的文化活动。对区域文化基础设施密度及消费程度差别也有鲜明的标注。2011年这项计划又和英格兰艺术委员会支持的“伦敦文化普及”任务相结合,以更好地理解公众的需求,精准地为他们服务,增加服务黏度,让他们更愿意光顾文化场所。这一举措是伦敦市利用城市及市民文化活动数据对文化进行管理的第一阶段尝试,它同时牵涉到社会、技术、经济和政治等因素。

#### 五、智慧城市:大数据在文化方面应用的理想生态系统

智慧城市的本质是要使城市基础设施投资能够互相利用,使城市平衡发展,达到综合效益最大化,更好地回报参与建设的各方。智慧城市在社区、大学和政治领域,而不是技术层面,对城市发展的管理有帮助。智慧城市是大数据最有发展前景的计划,是发展文化经济最理

想的实验目标。

2010年前在西班牙Santander市启动的一项智慧城市计划被认为是智慧城市领域的典范。该项目布设了覆盖率极高的大数据基础设施,现在已利用数据开始在多个领域提供相关的服务:在健康领域,智慧城市系统检测采集的污染源(空气、水和噪音)并发布相关的信息;在能源服务方面,三年中安装了20000个数据采集设备,通过数据分析,制定出节能策略,安排好照明、供暖方面服务;在垃圾回收方面,数据可以帮助实现垃圾桶满装情况下的最优收集安排;在交通方面,客流数据可以用来确定公共汽车时间表,并给出可用的停车位数量。2013年起,智慧城市的数据应用开始向区域商业、旅游业和文化业扩展。特别是通过和相关企业、居民和游客合作,文化产业在充分利用这些数据的基础上,可以打造出城市旅游文化的特色。

其他的智慧城市计划(旧金山、温哥华、西雅图、纽约、伦敦、巴黎、柏林、东京、首尔、上海等)也表明:不同领域的数据分享有利于创造有价值的经济和社会新形式,形成新的城市生态系统,当然也包括文化产业。

下面举两个智慧城市应用的例子。

2013年6月,Ubisoft公司做了一个“我们是数据”的推广活动。这个项目的目的是要建立一个网站(Jeu Watch Dogs),将伦敦、巴黎和柏林的社交网络(Tweet, Photo d'instagram, Flickr, Checks de Foursquare)<sup>⑥</sup>和其他热点(地铁、自行车道、互联网上网点、高速无线上网点、移动电话入网点、自动银行机及自动售票机、交通指示灯、公共卫生间、监控摄像头等)的数据状况制成实时的区位地图。这类地图还可以分区域得出一些统计数字,如居民收入状况、失业率、犯罪率、在线消费状况,以及来源于Facebook的持续活动状况。

卡耐基大学发起了一个名为Livehoods的项目,这个项目是要研究一个城市的居民在社交网络上的活动是如何帮助人们理解这个城市的社会结构的。卡耐基大学项目小组与大型社交网络(Twitter, Foursquare)合作收集公共信息,用来分析北美几个大城市的社会活跃程度及其构成,以及不同城市区域人口流动的频率和相关的形态,诸如纽约、旧金山、温哥华、西雅图等。这个获得几百

万美元资助的项目得到很多公共和私营机构的关注,特别是政府机构和不动产促销机构,因为它们认为这种方法可以用到公共卫生领域(在传染病流行时可以预防病毒或疾病的传播)。文化活动密度和场所的客流数据与城市公共健康数据的结合是非常有意义的,它可以指导文化消费场所的疾病预防和文化产业的发展,比如在推出新的文化产品和服务时,如何让老百姓意识到大型文化活动有可能造成有损公共健康的风险。

智慧城市利用大数据已不单单在一个领域,它涉及诸多的领域,只有社会、经济、政治等领域的共同介入才构成真正意义的大数据生态系统。实际上在不同领域数据的背后多少会有一定的关联,而这种关联反映出人类生存更深层次的自然关系。智慧城市绝不是一小部分技术专家想当然发明并施展的一种包含海量信息的技术构架,而是真正有需求的合作方采集并分享数据资料而形成的服务形态。这种数据间的耦合恰恰可以使智慧城市计划不断进出新的价值。数据耦合产生新的价值链,不仅在文化行业之间,而且在智慧城市发展中所涉及各个领域之间发生关系:文化、教育、旅游、交通、住宿/社区、城市化、能源、健康、餐饮/食品供应、农业、贸易,以及以数据为原材料的信息和通信产业。

## 六、对中国的启示

《大数据大文化》研究报告的目的是要解析大数据行业如何影响文化产业发展,探讨文化产业如何接受大数据带来的新的文化消费形态,从而描述大数据时代文化生态系统的状况。报告的亮点是较为全面系统地介绍了西方,尤其是美国互联网文化公司在数据与文化相结合方面的成功经验。有些商业应用在中国已经实现,如社交网络的一些应用;有些则是比较新颖的,如以客户需求为导向的视频和音频数据分析和软件算法。这些海外的创新对中国文化产业发展有借鉴意义。另外,需要指出的是,这篇报告对文化领域的大数据产业的宏观描述及相关数据并不太关注,而比较注重具有垄断地位的互联网公司的市场数据。一方面,互联网巨头公司已经形成垄断,它们的市场数据某种程度上代表了市场走向;另一方面,每个公司的应用都有各自特点,代表一个

领域,相互之间的数据可比性不大,而且难以归类。因此,宏观数据的采集不作为这个报告的重点。通过上面的介绍和评述,我们认为西方数据和文化产业紧密结合的发展思路对中国文化产业发展有一定的借鉴意义。

该研究报告的题目为“大数据大文化”,顾名思义正是有了大数据,才有了大文化的概念。大数据应用于文化及关联产业相比电商和金融等行业来说相对滞后,但影响不容小视。从商业角度讲,大数据改变了文化产业运营模式和形态;而从社会角度看,它改变着人们的生活方式及社会结构。面对这一迅速膨胀的产业,如何把握时机,乘势促进文化产业,是我们需要思考的问题。

第一,在发展文化领域大数据产业时,大公司基于对数据占有和处理能力的优势,制定有利于巨头公司运营的游戏规则,这种做法既改变了传统的产业链相互依存的模式,又改变了文化消费形态和行业游戏规则,衍生出更多的文化产品。从文化产业发展角度看,无疑有正面的作用,但在文化产品消费形态上有可能形成“一统天下”的趋势,如国内出现不久的微信产品。微信在短时期内聚集了数以亿计的用户,初期在微信平台上推出了几款产品,随后根据客户使用行为的数据,又不断完善丰富产品,从而聚集更多的用户。在拥有数量巨大的客户前提下,制定有利于商业公司的游戏规则是不可避免的,从某种角度讲也是正当的。这种互联网行业的一边倒的现象算不算垄断?又如何界定?现在下结论还为时过早。笔者个人认为,只要公平竞争的市场环境可以保障,只要所涉及的市场利益足够大,一边倒或暂时垄断的现象就有可能打破。以前常说,互联网行业只有老大,没有老二,但从近十多年的发展看,情况似乎并不是这样的,早几年门户网站相互之间打得不可开交,有你无我,可现在却形成了各大势力平分天下的局面;电商市场也如是,先是为数不多的几家,后来多家蜂拥而上,形成“诸侯割据,群雄争霸”的局面,有天猫、京东、苏宁易购、亚马逊、国美系(库巴、国美)、当当、易迅、一号店、凡客、唯品会等。模式也较以前更加丰富且各具特色。微信出来后,又有“来往”、“易信”冒出来竞争。也许更应当引起注意的是,基于大数据导致的文化消费形态的变化,迅速挤压传统的文化行业,改变着我们的生活方式,



获取知识的途径,以及文化产品的形态。《大数据大文化》只是在数据及基础设施占有导致商业游戏规则制定方面提出了问题,但并未涉及由此导致的社会及生活方式的改变,而在此方面产生的深层影响还需观察。但有一个趋势是比较明显的,为了保持相对垄断的地位,各个领域的巨头公司在不断扩张地盘,大肆兼并收购。中美皆是如此。

第二,在利用数据推动文化企业发展方面,西方一些公司的数据开发应用和创新经验值得借鉴,比如 Netflix 和 Pandora。数据本身蕴含着大量的商业机会,对裸数据的二次甚至多次开发实际上就是一种商业模式的创新。视频和音频产业在中国有很大的发展空间,关键是要在数据开发深度上下功夫,尤其是音频(音乐)市场,需要专业音乐人士和软件开发人员共同探索自动处理海量数据的方式。中国的在线音频市场目前发展情况与美国市场相比还有一定差距,虽然有众多的 mp3 形式的音乐网站,但在运营模式和商业模式上似乎不太成熟,没有形成像美国的 Pandora 和 Sirius XM(美国车载卫星广播公司,Nasdaq 上市公司)那样以注册客户为主的清晰盈利模式。这里既有如何利用数据开发出符合消费习惯的软件问题,又有知识产权保护困难的问题。而相比音频市场,中国的视频市场的发展要规范得多,大的门户网站都有视频服务,且版权保护做得比较到位,包括百度奇艺、搜狐视频、QQ 视频等,而且还有独立视频网站,如优酷、PPTV、乐视等。在盈利模式方面,中国和美国的公司侧重不同:中国公司注重贴片广告,而美国倚重注册收费模式。但只要坚持规范化运作,尊重知识产权,随时根据客户需求调整运营模式,中国公司的发展前景是可以期待的。

第三,要逐渐打破行业壁垒,使各个行业的数据得以相互分享整合,衍生新的产品和服务。从《大数据大文化》所列的行业看,旅游和智慧城市是综合利用数据,提供文化服务最合适的领域。中国旅游产业发展很快,智慧城市也被列入重点发展的领域,这两个领域中的每个

环节和局部都有很大的数据挖掘潜力,应当说是今后几年数据产业发展的理想空间。旅游和智慧城市都关系到民生,需要多行业、多领域的协作整合,因此,政府主导扶植的角色是不可或缺的。

第四,大数据的来源主要是个人信息,有游戏、娱乐、旅游、社交网络、位置、文化爱好等信息。互联网、移动电话、移动互联网和监控摄像都可以透露出个人的位置信息,甚至一些内容信息。在大力挖掘这类信息用于商业的同时,要有相应的规章制度约束。公开信息和私人信息的界限应当有明确的规定和使用限制,尽管这些数据对文化产业是必要的。

#### 注释:

①《大数据大文化:数据不断增长的力量及对文化经济的展望》(BIG DATA: BIG CULTURE Le Pouvoir grandissant de la data et ses perspectives pour l'économie de la culture),作者:Philippe Torres/Mathieu Soulé,法国巴黎银行研究所(L'Atelier BNP Paribas),参见 www.forum-avignon.org。

②《大数据大文化》,第24页、31页和9页,需要指出的是在报告中采用的数据因公布的时间不同略有偏差。

③这一原则即所谓的“测试并了解(test and learn)”,数字领域领军企业视其为“民主”的基本原则,同时也把它当作“不可抗拒”的市场工具,来刺激消费者大规模地接受这些游戏规则,而消费者并不能看到所涉及的问题。

④ Netflix 为美国 Nasdaq 上市公司。

⑤“音乐类型计划”利用150个技术指标或标准来定义 Pop/Rock(流行/摇滚)风格的音乐;350个技术指标来定义 Hip-Hop/Electronica(说唱/电声)风格的音乐;400个定义爵士乐;300和500个分别定义世界音乐和古典音乐。参见:《大数据大文化》,第44页。Pandora 为美国 NYSE 上市公司。

⑥这些社交网站大多都有视频或图片分享服务。

【责任编辑:王立霞】