

# 大数据：出版产业的机遇与实践

文 / 石佳靓

出版产业大数据时代的来临是数字出版实践和大数据国家战略深化发展的结果。大数据仿佛一头庞然大物，拥有足够的能力践踏和重组已有的产业格局。大数据带来了新的数据分析技术，通过对各种业务数据的整合分析和挖掘，将极大改进企业的运转效率，或拓展企业的产品业务线，提升产品附加值，或改进终端用户体验，提升客户满意度。一方面，大数据分析是企业发展的助推剂，帮助企业精确解决运转过程中的问题，实现发展速度和质量的提升；另一方面，数据越来越成为企业的核心资产，而数据处理和分析能力将越来越成为企业核心竞争力的构成要素，对数据资产的管理和运营也能成为企业潜在的增值空间。

权威咨询公司高德纳（Gartner）2013年9月发布的大数据分析报告表明，媒体和通讯是最先接受并实施大数据项目的行业。<sup>[1]</sup>对出版企业而言，大数据的影响和发展前景更是不可估量。中国新闻出版研究院2013年发布的《2012～2013中国数字出版产业年度报告》<sup>[2]</sup>直接将“大数据分析挖掘将走进数字出版”写入未来出版行业的发展趋势。出版企业不仅能通过产业大数据的发展提升出版集团内部的业务整合能力，整合市场资源，拓展产品服务，而且也能通过海量的数字产品和出版资源构建自身的数据资产，提升企业盈利能力，增强企业竞争力。因此，在出版转型与变革时期，探索大数据环境下出版产业面临的新形势和新机遇，结合国内外相关的实践案例，对于国内出版产业更好地应对大数据时代的各种挑战，提前制定大数据产业战略，具有重要的产业价值和研究价值。

## 一、大数据环境下传统产业的新机遇

历史上每一次生产技术的革新都带来生产力的大幅提升和产业结构的重大调整，只有具备敏锐的嗅觉才能在变革的时代中不断推陈出新，传统产业尤其如此。在大数据时代下，传统产业凸现出三大显著的发展趋势，相应的6类商业模式也愈发清晰。

### 1. 产业发展的三大趋势

对于大数据催生的产业变革，宏源证券研究所高级分析师赵国栋等<sup>[3]</sup>认为，信息产业的发展将会呈现出三大趋势：数据成为资产、行业垂直整合以及泛互联网化。这三大趋势，是观察包括出版产业在内的企业在大数据环境下的产业变革和企业变革的重要视角，也是分析研究相关企业成长路径的顶层逻辑之一。

数据成为资产。数据资产将成为和土地、资本、人力同等重要的、独立的生产要素。尤其对信息产业来说，决定产业兴旺繁荣与否的关键因素，已不是传统的生产要素，而是一度被束之高阁的数据资产。从某种意义上来说，数据将成为工业化转向信息化的关键要素，成为促进产业变革重组的战略资源，成为推动传统企业转型的动力根源。

行业垂直整合。传统的逻辑是，新兴产业通过垂直整合来开辟市场，在产品市场成熟以后，产业链上专业分工才逐步细化，成本也逐渐降低，优势逐渐转向水平分工格局。而在大数据时代，行业垂直整合趋势越发明显，这种趋势仍然首先体现在信息产业，然

后向相邻产业蔓延,这是大数据改变产业竞争格局的一个缩影。因此,在大数据时代,越是靠近终端客户的公司在产业链上越是具备竞争优势。如甲骨文公司首先在数据库软件市场站稳脚跟,随即向应用软件市场进军,然后利用庞大的客户群优势,向产业链上游进军,一举收购了太阳(Sun)公司。

泛互联网化。泛互联网化是不论大型公司、小型公司、新兴企业还是传统企业理解大数据变革、实践大数据战略的一个重要的思想基础。在泛互联网范式中,强调终端、平台、应用的泛互联网化。智能手机、平板电脑、电子阅读器等移动终端的市场份额将逐渐占据多数,电子商务、在线教育等网络平台将更为深刻地影响生产生活,利用最新技术满足用户个性化的、碎片需求的大数据应用将物尽其用,这些都可以成为盈利的主要来源。

## 2. 产业发展的6种商业模式

基于以上在大数据环境下产业变革的三大趋势,有6种围绕数据资产的商业模式可以为出版企业在大数据环境中的战略思考和战略决策提供参考。这6种商业模式即数据租售模式、信息租售模式、数字媒体模式、数据使能模式、数据空间模式和大数据技术提供商。下面对这6种模式做一简单介绍。

数据租售模式:即出租或售卖经过加工后的数据资产。这是最典型的数据资产模式。按照销售对象的不同,又分为为用户提供增值服务和为第三方有偿提供数据这两种类型。

信息租售模式:即围绕某一行业或领域,从多种渠道搜集、整合相关信息进行租售。其竞争优势主要在于专有的信息通道和强大的数据整合能力。这里信息指经过加工处理、承载一定行业特征的数据集合。

数字媒体模式:即利用其掌握的海量数据进行深度分析,提供精准营销和信息聚合服务。目前,传统媒体的广告市场正在急剧向数字媒体的广告市场转移,因此,这类模式也具有很好的前景。

数据使能模式:这类模式也是基于大量的数据和有效的数据分析技术。不同的是其业务范围较数字媒体模式大大拓展。如在金融行业,可以通过统计分析和预测解决业务流程中的一些关键性问题,大大降低风险,提高生产效率。

数据空间运营模式:这类模式主要提供数据存储空间。如Dropbox(一款网络文件同步工具)、微盘等,

发展潜力巨大,盈利模式相对多元。

大数据技术提供商:这类公司以技术见长,提供专业的大数据技术解决方案,在语音数据处理、视频数据处理、语义识别、图像处理等领域都大有可为。其他非技术企业可以通过产业分工实现与这些企业在大数据转型中的合作。

## 二、出版产业大数据发展的可能误区

尽管大数据浪潮来势凶猛,但也并非空穴来风、毫无征兆,有关数字出版在出版业内的理论探讨和实践摸索由来已久。数字出版可以看作是传统出版界对大数据浪潮的预先感知和提前反应,也可以看作是传统出版界数字化转型的先期阶段。即便如此,传统出版人在面对这个崭新领域时,难免会出现种种理解上的偏差和片面的,甚至错误的认识。《中国出版传媒商报》的李鲆对此作出了深刻的总结。<sup>[4]</sup>同样,在认识大数据在出版产业的应用和发展时,也有可能存在一些误区,这些误区不但可能混淆出版人对大数据的理解,而且会贻误出版产业转型变革的重要时机。正视并纠正这些可能的认知偏差,对于理解大数据乃至抓住大数据时代出版产业的战略机遇至关重要。

误区一:大数据时代出版产业的转型升级就是技术和结构升级。大数据时代出版产业转型升级的根本不是表面的技术和结构升级,而是价值功能转型升级,即从低层次传播向高层次服务转变。传统出版企业往往缺乏专业的技术人才,这种人力资源的缺失造成一种认知惯性和偏差,即认为只要补充技术人才从而带动出版企业技术转型,就能带来企业战略层面的转型。在同样的逻辑下,认为企业组织结构转型升级也就能实现传统型企业向数字型或者大数据型企业的转变。这种逻辑误区对出版企业来说是十分有害的。因为,能否认识到大数据时代出版产业的转型升级的根本在于价值功能的转型升级,决定了出版企业在未来能否应对越来越多从其他行业转型或扩展而来的信息服务商的蚕食,决定了企业的生存地图。<sup>[5]</sup>

误区二:大数据时代的出版仍然必须依靠传统出版提供内容。大数据时代的出版的确可以利用传统出版的资源,比如将传统出版内容数字化,从而实现新的盈收。但是,未必一定需要依托传统的出版资源,技术和产业环境的变化使得不同的主体可以创造出

自己独有的方式进行出版活动,而传统的出版内容资源只成为海量出版内容的一部分。外文图书影视作品以众包形式的翻译就是其中一例。近一两年来国内涌现出的百度、阿里和腾讯等互联网巨头整合并购网络文学企业的浪潮,也充分说明数字出版并非仅仅是传统出版企业的天下,传统出版企业也并非占据不可替代的竞争优势。面对读者行为影响内容走向、预测读者喜好实现精准推送、专属内容智能定制、数字阅读升级为知识解决方案等大数据出版的趋势,<sup>[6]</sup>出版企业应该及时转变观念,从传统出版向数据内容服务转向。

误区三:出版企业应用大数据重点在于建立数据库存储企业的数据。标准化的数据采集和存储基础设施仅仅涉及大数据的“藏”,更为重要的是源源不断地利用历史和现时不断产生的数据实施分析、描述、预测和决策,使数据的价值得以变现。也是由于传统的原因,出版企业或多或少都储存有一定数量的、分散的、类型多样的数据,这些数据已经形成了企业的数据资产。例如,在数字出版中的出版系统、发行管理系统和作者管理系统等企业资源计划(ERP)系统中存储的生产记录、财务记录、资源消耗统计等数据;在企业官网上有用户的注册信息、行为记录、需求信息、评价记录等数据;如果出版企业和一些社会化媒体或者电商平台开展合作,还能得到特定用户群体信息、市场信息、品牌口碑、社会评价等数据。大数据在出版企业的应用重点是对这些数据进行规范、整合和利用,即大数据的“用”。

误区四:大数据对出版企业来说完美无瑕。大数据背景下,信息的战略价值愈加凸显,出版业将获得更为广阔的发展前景。然而,大数据对出版企业来说并非完美无瑕,我们也应辩证看待。在战略层面,出版企业应该认识到,大数据不能替代文化风格与思想深度。过度地强调以用户为中心,过度地迎合读者的阅读兴趣,毫无疑问会降低作品的思想深度,进而可能对整个出版文化生态造成负面影响。在操作层面,随着大数据战略的广泛实施,数据所有权、版权保护、隐私保护等新问题将会日益凸显,进而也有可能产生新的“数字鸿沟”“信息孤岛”问题。<sup>[7]</sup>因此,出版企业也不能仅对大数据怀有美好的想象,忽视其硬币的另一面。对这些可能伴随大数据而来的问题给予前瞻性的关注和预判同样重要。

### 三、出版企业中大数据应用的实践案例

大数据概念发起于国外,在出版行业早有应用,与此同时,国内新闻出版行业有关大数据管理和应用的实践亦有成功典型。下文将以实际案例分析大数据在国内外出版行业的典型应用,为我国出版行业提供借鉴和参照。

1. 电子书领域的奈飞公司(Netflix):思科伯德(Scribd)和沃易思特(Oyster)

Nexflix 利用海量用户数据挖掘和分析,打造出一炮而红的原创剧集《纸牌屋》。而在电子书领域,Scribd 和 Oyster 两家在线图书馆公司通过手机内容网站订阅者的行为数据,为电子书或者个人出版物提供数据分析服务,有志于成为电子书领域的 Netflix。<sup>[8]</sup>它们以按月付订阅费的模式提供无限量的电子书租阅服务,以此跟踪读者的阅读行为,希望借此就读者接下来想选什么书作出精准的推荐。Oyster 与 Scribd 会主动邀请读者给书评分,也会跟踪某本书被读者读完的比例和点击率,或者看到选择某本书的人中有多少人会点击它去了解更多信息,借助这类信息再结合其他因素来推荐图书。

Scribd 和 Oyster 通过对读者行为数据的分析发现<sup>[9]</sup>:

人们如果在凌晨 5 点还在看书的话,极有可能会选择一本言情小说,而得克萨斯和佐治亚州的读者更是如此;在清晨时段,惊悚小说对人们更有吸引力;如果一位读者喜欢菲利普·K·迪克(Philip K. Dick,一位作家),那么他/她很可能也喜欢关于啤酒而非葡萄酒或烈酒的书籍;励志类的书籍非常热门,但是选择此类书的读者只有 20% 会把它读完;悬疑小说越长,读者就越快跳到最后一张查看“凶手”或者“大老板”是谁;如果书籍分成短章节,读者读完的几率会增加 25%,等等。

Scribd 还发现,与模仿实体书店的排行榜、好书排行或主题制定的榜单相比,个性化的推荐带来的浏览流量是前者的 10 倍。因此,Scribd 计划将算法与排行榜结合在一起,制定一个根据读者以往阅读习惯判断可能会让他们感兴趣的畅销书排行榜。除此之外,Scribd 还会利用其掌握的不同时段内读者阅读喜好的变化情况,在一天当中的不同时段推荐不同的书籍或题材。



## 2. 从传统印刷企业到新型文化企业：雅昌集团

雅昌集团成立于1993年，业务经营范围以传统印刷为主，曾多次获得美国印刷行业最权威、最具影响力的大奖——“班尼奖”。2000年，雅昌集团创立了雅昌艺术网，迈出了从传统印刷业向新型文化公司转型的第一步。尽管创立之初设备简陋，但雅昌艺术网通过年复一年、日复一日的漫长积累，形成了人类历史上空前的“艺术品数据库”。凭借这些数据资产，雅昌集团的业务范围也从书籍印刷扩展到数字出版，建成了雅昌艺术网、艺术品数据库、流动美术馆、艺术阅读体验中心等。可以说，雅昌集团的发展历程将大数据环境下产业变革的三大趋势体现得淋漓尽致。

从2011年开始，雅昌集团通过与惠普公司的合作，逐渐将精力聚焦于两个链条：一是核心业务链条，即艺术创作、传播和教育，以及艺术交易、收藏的业务链；二是产业价值链条，即如何通过业务链条创造更多的经济价值和社会价值。<sup>[10]</sup>从目前的业务经营范围和商业业态来看，雅昌集团现在的经营属性更多的是数据资产。如今的雅昌集团实现了从包括创作、传播、宣传、交易、收藏以及再流通等环节的前一个链条，到包含数据采集、存储、管理和应用等环节的后一个链条的提升。它在推动解决艺术教育、艺术学术、艺术市场健康和艺术产业发展这4个艺术产业课题的过程中，逐步形成了其大数据应用的4个重要方面，<sup>[11]</sup>从而也形成了自身独具的行业竞争优势。

在转型升级的过程中，出版产业必然能通过大数据战略充分挖掘和利用行业用户数据、产品内容数据、市场信息数据等数据资源，提高生产效率，优化营销模式，拓展产业链条，提升内容价值，改进用户体验。洞察并顺应传统产业在大数据环境下“数据成为资产”“行业垂直整合”以及“泛互联网化”这三大趋势，理解产业变革的逻辑和路径，是抓住新机遇、迎接新挑战的重要前提。数据产业的六大商业模式也是产业转型中的出版产业可资参考的。

新事物的诞生总会伴随着疑惑和误解，大数据在出版产业中的发展应用也不能避免。出版人应该认识到：第一，大数据时代出版产业转型升级的根本不是表面的技术和结构升级，而是价值功能转型升级；第二，大数据时代的出版并非一定需要依靠传统出版提供内容，出版企业应从传统出版向数据内容服务转变；第三，出版企业大数据应用的重点不在于“藏”，而在于“用”；第四，大数据也具有一枚硬币的两面，在战略层面和操作层面对其辩证理解和看待同样重要。

国外的Scribd和Oyster以及国内的雅昌集团是出版企业实施大数据应用的典型案例。前者通过对读者阅读行为数据的采集和分析，得出系列传统方法无法获得的规则和发现，这对于公司预测读者阅读兴趣、改进用户体验、优化营销策略等方面具有很大的启发意义。后者是本土企业通过大数据战略，从传统印刷企业成功转型为以数据为主要资产的新型文化企业的案例。这两个案例，尤其是雅昌集团的实践，充分体现了数据资产、行业垂直整合和泛互联网化的产业趋势，对我国传统出版企业的战略转型具有很好的借鉴意义。

（作者单位：南京大学信息管理学院）

## 四、结语

大数据浪潮带来新一波产业变革的机遇和挑战，作为信息和数据密集产业的出版产业自然首当其冲。

### 参考文献：

- [1] Gartner. Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype[EB/OL]. [2014-04-30]. <https://www.gartner.com/doc/2589121?ref=SiteSearch&stkw=big%20data&fml=search>
- [2] 中国新闻出版研究院. 2012～2013中国数字出版产业年度报告[R]. 中国书籍出版社：中国新闻出版研究院，2013
- [3] 赵国栋，易欢欢，糜万军等. 大数据时代的历史机遇：产业变革与数据科学[M]. 北京：清华大学出版社，2013：35-38
- [4] 李鲜. 关于数字出版的一切，你可能全错了[EB/OL]. [2014-04-30]. [http://www.cnepaper.com/zgtssb/html/2014-02/11/content\\_8\\_1.htm](http://www.cnepaper.com/zgtssb/html/2014-02/11/content_8_1.htm)
- [5] 王亚非. 大技术与大数据时代坐标下的出版产业[N]. 中国出版传媒商报，2013-08-27
- [6] 尹琨，任晓宁. 大数据开启数字出版瑰丽想象[N]. 中国新闻出版报，2013-11-28
- [7] 吴赞. 产业重构时代的出版与阅读——大数据背景下出版业应深度思考的五个关键命题[J]. 出版广角，2013(23)
- [8] Los Angeles Times. Oyster, the Netflix of e-books, arrives on the iPad[EB/OL]. [2014-04-30]. <http://www.latimes.com/business/technology/la-fi-tn-oyster-netflix-ebooks-ipad-20131016,0,206906.story#axzz30v1P5hNC>
- [9] 中国图书出版网. 电子书也可打造自己的“纸牌屋”[EB/OL]. [2014-04-30]. <http://www.bkpon.com/Web/ArticleShow.aspx?artid=118093&cateid=A1804>
- [10] 郭涛. 雅昌集团：大数据揭示艺术密码[EB/OL]. [2014-04-30]. [http://news.xinhuanet.com/info/2013-04/15/c\\_132310503.htm](http://news.xinhuanet.com/info/2013-04/15/c_132310503.htm)
- [11] 潘剑平. 雅昌 & 惠普大数据应用最佳实践媒体分享会[EB/OL]. [2014-04-30]. <http://soft.chinabyte.com/359/12759859.shtml>