

关于“大数据”及出版业转型升级的一点心得

兼谈“数字版权保护技术研发工程”相关情况

中国新闻出版研究院 张立

2016.9

目 录

- 一、 国家层面关于“大数据”的相关文件精神
- 二、“大数据”特点及在宏观层面和企业层面的不同含义
- 三、 传统出版业都有哪些数据？特点是什么？
- 四、 转型升级与出版业“大数据”应用
- 五、“数字版权保护技术研发工程”落地应用情况介绍

国家层面关于“大数据”的相关文件精神：

序号	发布时间	发布机构	文件名称	主要内容
1	3月5日2014年	国务院总理 李克强	2014年政府工作报告	设立新兴产业创业创新平台，在新一代移动通信、集成电路、 大数据 、先进制造、新能源、新材料等方面赶超先进，引领未来产业发展。
2	6月24日2015年	国务院办公厅	关于运用 大数据 加强对市场主体服务和监管的若干意见	以社会信用体系建设和政府信息公开、数据开放为抓手，充分运用 大数据 、云计算等现代信息技术，提高政府服务水平，加强事中事后监管，维护市场秩序，促进市场公平竞争，释放市场主体活力，进一步优化发展环境。提高 大数据 运用能力，增强政府服务和监管的有效性。高效采集、有效整合、充分运用政府数据和社会数据，健全政府运用 大数据 的工作机制，将运用 大数据 作为提高政府治理能力的重要手段，不断提高政府服务和监管的针对性、有效性。
3	7月1日2015年	国务院	关于积极推进“互联网+”行动的指导意见	鼓励制造企业利用物联网、云计算、 大数据 等技术，整合产品全生命周期数据，形成面向生产组织全过程的决策服务信息，为产品优化升级提供数据支撑。重点促进以移动互联网、云计算、 大数据 、物联网为代表的新一代信息技术与制造、能源、服务、农业等领域的融合创新，发展壮大新兴业态，打造新的产业增长点。
4	8月31日2015年	国务院	促进 大数据 发展行动纲要	按照党中央、国务院决策部署，发挥市场在资源配置中的决定性作用，加强顶层设计和统筹协调，大力推动政府信息系统和公共数据互联开放共享，加快政府信息平台整合，消除信息孤岛，推进数据资源向社会开放，增强政府公信力，引导社会发展，服务公众企业；以企业为主体，营造宽松公平环境，加大大 大数据 关键技术研发、产业发展和人才培养力度，着力推进数据汇集和发掘，深化 大数据 在各行业创新应用，促进 大数据 产业健康发展；完善法规制度和标准体系，科学规范利用 大数据 ，切实保障数据安全。通过促进 大数据 发展，加快建设数据强国，释放技术红利、制度红利和创新红利，提升政府治理能力，推动经济转型升级。加快政府数据开放共享，推动资源整合，提升治理能力。推动产业创新发展，培育新兴业态，助力经济转型。强化安全保障，提高管理水平，促进健康发展。
5	1月7日2016年	发改委	国家发展改革委办公厅关于组织实施促进 大数据 发展重大工程的通知	重点支持 大数据 示范应用、重点支持 大数据 共享开放、重点支持基础设施统筹发展、重点支持数据要素流通。在健康医疗、社保就业、教育文化、交通旅游等领域，推动传统公共服务数据与互联网、移动互联网、移动穿戴设备数据的汇聚整合，鼓励社会机构开展应用研究，开发便民服务应用，优化公共资源配置，提升公共服务水平。
6	3月18日2016年	国务院	中华人民共和国国民经济和社会发展第十三个五年规划纲要	实施国家 大数据 战略。把 大数据 作为基础性战略资源，全面实施促进 大数据 发展行动，加快推动数据资源共享开放和开发应用，助力产业转型升级和社会治理创新。深化 大数据 在各行业的创新应用，探索与传统产业协同发展新业态新模式，加快完善 大数据 产业链。加快海量数据采集、存储、清洗、分析发掘、可视化、安全与隐私保护等领域关键技术攻关。促进 大数据 软硬件产品发展。完善 大数据 产业公共服务支撑体系和生态体系，加强标准体系和质量技术基础建设。
7	4月13日2016年	促进 大数据 发展部际联席会议	促进 大数据 发展三年工作方案(2016-2018) 促进 大数据 发展2016年工作要点 政务信息资源共享管理暂行办法 政务信息资源目录编制指南	(四份文件尚未公开，仅有此次会议的相关报道) 会议审议通过了四份文件，部际联席会议召集人、国家发改委主任徐绍史主持会议并讲话，国家发改委副主任林念修、工业和信息化部副部长怀进鹏、中央网信办副主任庄荣文、联席会议各成员单位相关负责人出席会议。会议强调，要深刻领会实施国家 大数据 战略的重大意义。一是紧抓“数据大国”向“数据强国”转变的新机遇，发挥数据资源的战略作用，建设数据强国，积极构建国家竞争新优势。二是把握工业经济和信息经济交汇发展的关键期，深入发掘数据要素潜力，激活数据要素，有效催生经济发展新红利。三是顺应互联网时代形势，将 大数据 引入政府治理，实施数据决策，努力打造政府治理新手段。会议明确，全面落实《 大数据 纲要》，应找准着力点和突破口，围绕三个关键环节精准发力：一要加快数据共享开放，开展政府治理 大数据 示范应用，推进“互联网+政务服务”，深化数据创新应用；二要推动产业创新发展，做好 大数据 产业发展的规划，推动好工业 大数据 、互联网与制造业的融合发展；三要科学规范利用数据，建立完善 大数据 管理机制，加快相关法律法规和标准体系建设，强化数据安全保障。会议认为，2016年应加快推进“三大建设”：一是加快制度体系建设，尽快出台《政务信息资源共享管理暂行办法》；二是加快综合试验区建设，开展制度创新探索，推动数据创新应用，破解 大数据 发展难题；三是加快重大工程建设，坚持需求导向、统筹整合，充分发挥市场的决定性作用，助力经济转型发展。

孙寿山：以转型升级促进传统媒体与新兴媒体融合发展（《出版发行研究》2014年第6期，在第十届深圳文博会数字出版高端论坛上的讲话）

非常高兴我们再次相聚在深圳文博会数字出版高端论坛。当前，我国新闻出版业保持着欣欣向荣的良好局面，展现出无限的生机和活力。本次论坛以“转型促进发展，服务拓展市场”为主题，在展示转型成果的同时，对如何以转型升级促进传统媒体与新兴媒体融合发展进行了深入讨论。

转型升级是改革所依，发展所需，是行业潮流，大势所趋，这已是业界的普遍共识。在总局的持续推动和大家的不懈努力下，近年来传统新闻出版企业在思想认识、基础设施建设、研发应用、产品形态、商业模式、产业协作等方面均取得了突破性进展。然而，我们必须清醒地认识到，问题仍然不少，差距依然巨大，道路充满曲折，任务十分繁重。

近年来，传统新闻出版单位数字化转型升级虽然取得了令人欣喜的成绩，但也存在不少突出问题，距离实现融合发展这一目标，尚有着不小的差距。一是思想认识上的片面性，带来缺乏探索创新精神，存在较强的辅业意识、惯性思维以及畏难情绪；二是技术选择上的盲目性，带来对前沿技术、关键技术的认识不到位，对新技术应用和创新性思维的缺失；三是工作机制上的本位性，带来不善于借用外部力量；四是管理体制上的落后性，带来管理机制、运营机制滞后，缺乏顶层设计。

之所以存在这些问题，是因为许多新闻出版单位在实施转型工作中，思想还需要进一步解放，思路还有待进一步拓宽。要想在新的产业发展形势下谋求发展，革新图存是唯一选择，要重点在思维、技术、模式等三个方面，进行全方位变革，加快新闻出版业进入新的发展阶段。

1. 思维革新：技术、模式、产品上的升级发展，需要建立在思维和意识变革的基础之上。对产业发展形势的认知程度，决定了转型升级的深入程度。未来新闻出版业是基于互联网环境下的发展，互联网的发展趋势，就是新闻出版业的转型升级方向。这种转型升级，不是对传统业务进行简单的拓展与延伸，也不是对原有框架的修修补补，而是需要用全新的互联网思维、视角和发展规律来谋划和推进各项工作，实现从模式改变到角色转变的全面革新。
2. 技术更新：正如奇葆同志所言，技术与内容互为支撑，互相融合，是一体之两翼，驱动之双轮，二者共同构成新闻出版业的核心竞争力。因此，新闻出版业的发展，既离不开注重内容建设，也离不开技术的有力推动。传统媒体在内容信息积累、采集核实、分析解读等方面，有着当下新兴媒体无法比拟的优势，但如果不能将优势与先进技术进行有效的结合，则内容的传播力和影响力都难以实现突破性进展，上述优势也难以得到巩固。

近年来，传统新闻出版单位普遍加大了对新技术、重点技术的关注，但对新技术的适应能力及应用能力普遍偏低，可以说技术依然是我们的短板。因此，传统新闻单位必须瞄准发展趋势，紧盯技术前沿，并基于发展趋势，把当今可用的技术尽可能囊括到我们的视野中来，从中找到业务的结合点，将先进技术纳入到项目设计和产品研发中，不断以新技术、新应用推进新闻出版业发展。如可以利用大数据和云计算技术，实现资源的充分整合和业务流程的优化再造，提升数据处理能力，为内容的多元生产和个性化传播提供强大支撑；如可以利用4G、WiFi等移动互联网技术，通过智能手机、平板电脑等智能终端，借助客户端，加强互动化、分众化传播，优化内容信息，丰富表现形式，增强在移动互联网的覆盖面和影响力。

3. 模式创新

类型、资源、规模等方面存在的差异,要求新闻出版单位在转型过程中需要依据自身优势特点,准确把握整体形势,探索适合自身发展的数字出版商业模式,依据模式展开业务布局,比如搭建基于属地特色、集团优势或行业主流的特色化资源数据库平台,向用户提供内容和服务;比如向互联网企业借鉴成熟的商业模式,通过收、并购技术公司、游戏公司等,增强自身研发应用实力,实现产业链延伸;比如从数字教育、手机应用、O2O等行业热点模式中寻找与自身业务发展的结合点;比如通过提供个性化内容和服务,聚拢用户和流量,通过广告植入、内容链接“电商”等模式创新走向内容增值,跳出版权售卖的思维定式,等等。只有商业模式上多元创新,才能带动数字产品形态的丰富与多元,才能实现新闻出版业内容资源的充分整合与盘活,不断激发产业的活力与创造力。

一、一个依靠数据分析进行产品优化的真实应用

玩家行为分析

进入	广告			
留存				
跳出				
付费				

- 1.设计：如何设计能真正提高产品营收的采集要素？
- 2.工具：选择什么软件进行分析？
- 3.分析：具体分析过程

优化方案

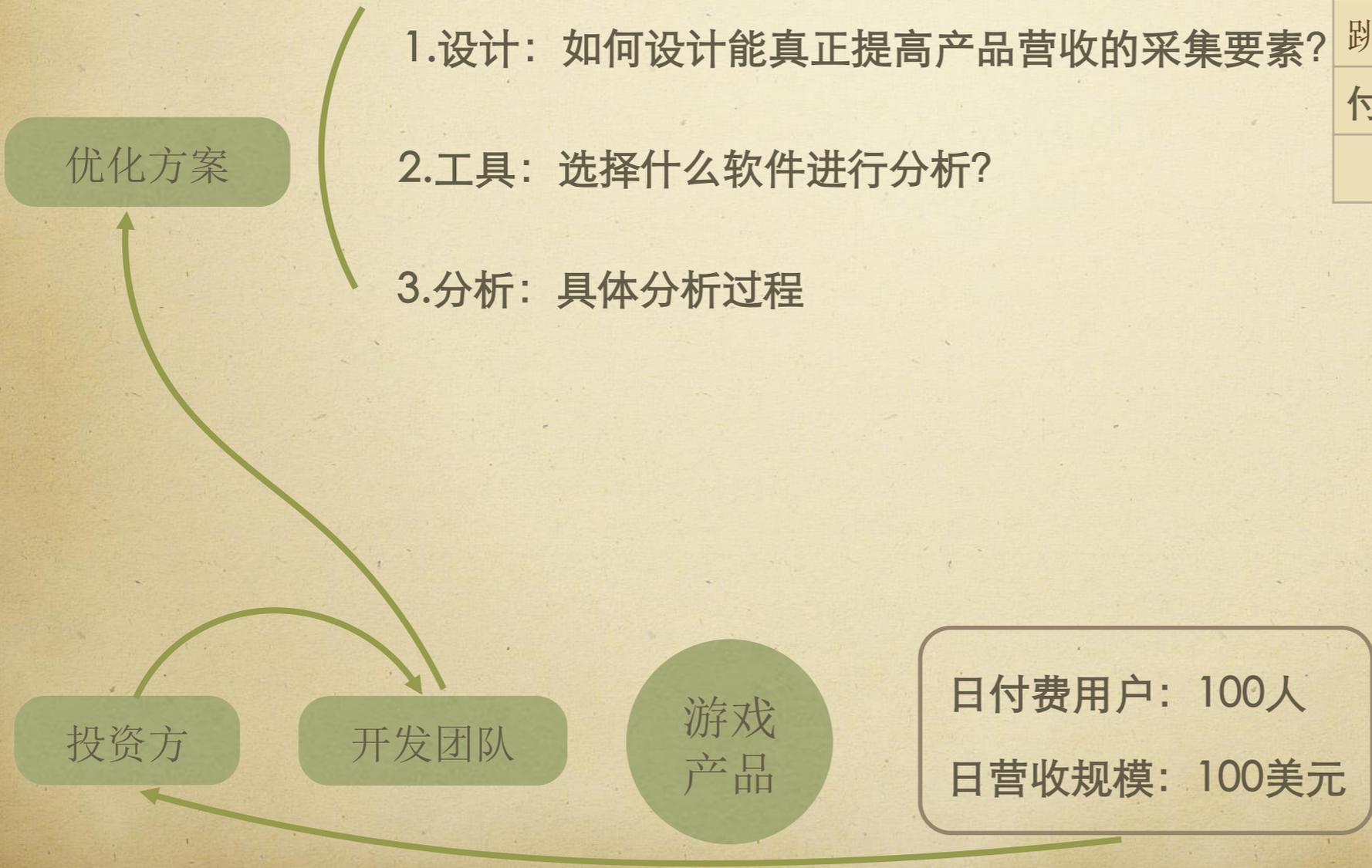
投资方

开发团队

游戏
产品

日付费用户：100人

日营收规模：100美元



广电总局关于新闻出版业“转型升级”的相关文件精神：

序号	发布时间	发布机构	文件/讲话名称	主要内容
1	2014年4月24日	国家新闻出版广电总局 财政部	关于推动新闻出版业数字化转型的指导意见	通过政府引导、以企业为主体，加速新闻出版与科技融合，推动传统新闻出版业 转型升级 ，提高新闻出版业在数字时代的生产力、传播力和影响力，为人民群众的知识学习、信息消费提供服务，为国民经济其他领域的产业发展提供知识支撑，更好更多地提供生活性服务与生产性服务。实现行业信息数据共享，构建数字出版产业链，初步建立起一整套数字化内容生产、传播、服务的标准体系和规范；促进新闻出版业建立全新的服务模式，实现经营模式和服务方式的有效转变。
2	2014年5月15日	孙寿山在深圳文博会上上的讲话	以转型升级促进传统媒体与新兴媒体融合发展	此前我们讲 转型升级 ，现在我们谈融合发展，这二者之间，有什么联系与区别呢？从实施路径上来讲，两者是基本一致的。转型升级，是传统新闻出版业实现融合发展的必备前提；融合发展，是传统新闻出版业转型升级的阶段性目标。传统新闻出版单位只有通过实施数字化转型升级，才能在巩固自身优势的基础上，充分借助新兴媒体的技术优势和渠道优势，借鉴新兴媒体的表述方式和信息内容加工方式，实现在全媒体背景下的深度融合，深入发展。二者的区别则在于切入点的不同。转型升级还是以传统新闻出版单位为主体的视角，融合发展则超越传统新闻出版的既有思维模式，以平行化、全局化的视角看待产业发展。
3	2015年3月31日	国家新闻出版广电总局 财政部	关于推动传统出版和新兴出版融合发展的指导意见	运用 大数据 、云计算、移动互联网、物联网等技术，加强出版内容、产品、用户数据库建设，提高数据采集、存储、管理、分析和运用能力。积极通过多种方式吸收借鉴、善加利用先进的传播技术和渠道，借力推动出版融合发展。充分利用新一代网络的技术优势，加快发展移动阅读、在线教育、知识服务、按需印刷、电子商务等新业态。 加强出版大数据分析 、结构化加工制作、资源知识化管理、数字版权保护、数字印刷、发布服务以及产品优化工具、跨终端呈现工具等关键性技术的研发和应用实践，着力解决出版融合发展面临的技术短板。建立和完善用户需求、生产需求、技术需求有机衔接的生产技术体系，不断以新技术引领出版融合发展， 驱动转型升级 。
4	2015年7月16日	孙寿山在新浪接受微访谈	孙寿山谈数字出版：融合引领创新 创新驱动发展	网友：随着云时代的到来，大数据越来越成为人们关注的热点。数字出版也正步入一个大规模生产、分享和应用数据的时代。如何利用大数据来发展数字出版产业？ 孙寿山：科技进步推动新闻出版业走过了“铅与火”“光与电”，迎来了“数与网”时代，智能化、万物互联、大数据将成为未来媒介发展趋势。在新常态下发展数字出版，必须坚持以内容为本、技术为用，内容为体、技术为翼，两者共同构成数字出版的核心竞争力。数字出版讲求“用户至上，需求为先”，这就要求我们的出版企业要不断提升科技应用水平， 充分运用大数据技术 ，准确把握用户需求，提升资源整合能力，实现内容生产的分众化、特色化、精准化。充分运用先进信息技术和多媒体技术，丰富和优化内容呈现，满足多元用户需求，不断完善用户体验；充分运用社交类应用和技术，借助微博微信，促进社交平台与传播平台的有效对接，拓展信息传播渠道，提升传播效果，实现传播的即时化、互动化。 网友提问：今年4月，总局出台了《关于推动传统出版和新兴出版融合发展的指导意见》，本届数博会的主题也跟“融合”有关，“融合”将是未来出版业发展的主基调。现在都在 讨论传统出版转型升级 ，如何才是转型，又如何才能与数字出版实现融合发展？ 孙寿山：以互联网为代表的信息技术的快速发展，重新定义了人们获取信息包括获取知识、获取能力的方式，对传统出版带来全方位、深层次影响。融合发展是我国乃至全球新闻出版业的发展趋势，是传统出版与新兴出版需要共同面对的课题。融合发展不是局部的融合、个体的融合，也不是传统出版与新兴出版的相互替代，而是双方在内容、技术、渠道、经营和管理等方面多样化、多层次、全方位的互融互补。无论是传统出版企业还是新兴出版企业，首要的是思维观念上的融合。传统出版单位应不断强化和灵活运用互联网思维，逐渐摆脱在新媒体传播环境下的“水土不服”，向互联网企业学习灵活的经营管理机制，借助互联网的先进技术与新兴媒介渠道，将内容优势延伸至网络空间。新兴出版企业虽然与互联网共生共长，但同样需要具有融合发展意识，不断优化内容生产，完善服务水平。思想革新引领手段创新，企业应充分利用各项政策利好，积极推进各种业态创新、模式创新。努力实现“一个内容多种创意、一个创意多次开发、一次开发多种产品、一种产品多个形态、一次销售多条渠道、一次投入多次产出、一次产出多次增值”的新闻出版业融合发展“七个一”目标。
5	2016年4月20日	国家新闻出版广电总局	2016年新闻出版广播影视科技工作要点	加快推进广电融合媒体服务云平台和 大数据 系统建设，着力推动在技术支撑上融为一体、合而为一，尽快从相“加”迈向相“融”。继续推进各级电台、电视台数字化网络化全面升级，实现台内全流程数字化、一体化，推动省级以上和有条件的电台、电视台依据融合媒体平台建设的技术白皮书，建设传统媒体与新兴媒体融合的制播云平台和基于用户互动的制播 大数据 系统。推动实体书店利用互联网、 大数据 、数字印刷等技术，加快数字化改造，促进线上线下融合发展。
6	2016年8月20日	冯宏声在2016年上海书展“CNONIX国家标准应用研讨会”上主题演讲	冯宏声：促进新闻出版业大数据体系建设与大数据应用的总思路	促进 新闻出版业大数据体系建设与大数据应用 ，应统一认识，共同探索实现路径。要从顶层设计、政策发布、制度建设、机构建设起步，加强标准研制应用，加速技术研发与系统建设，搭建数据开放、共享、交换、运营平台，积极探索大数据应用模式，提高数据应用价值，推动行业大数据建设，带动新闻出版大数据在行业内外的产业化应用，构建完整的新闻出版业大数据体系；应立足解决行业共性问题，加强资源整合与信息共享，形成政府指导、行业机构组织、产学研联合的工作格局，鼓励、支持企业牵头参与，跨地区协作、跨领域合作，实现“平台整合共建、系统分布部署”，循序渐进地推进大数据建设工作。

“文化体制改革”的相关文件精神：

序号	发布时间	发布机构	文件/讲话名称	主要内容
1	2002年11月8日		十六大报告	继续 深化文化体制改革 。根据社会主义精神文明建设的特点和规律，适应社会主义市场经济发展的要求，推进文化体制改革。抓紧制定文化体制改革的总体方案。把深化改革同调整结构和促进发展结合起来，理顺政府和文化企事业单位的关系，加强文化法制建设，加强宏观管理，深化文化企事业单位内部改革，逐步建立有利于调动文化工作者积极性，推动文化创新，多出精品、多出人才的文化管理体制和运行机制。按照一手抓繁荣、一手抓管理的方针，健全文化市场体系，完善文化市场管理机制，为繁荣社会主义文化创造良好的社会环境。
2	2003年7月	中宣部、文化部、广电总局、新闻出版总署	关于文化体制改革试点工作的意见	正式确定北京、上海、广东、浙江、重庆和深圳、沈阳、西安、丽江九地为文化体制改革综合性 试点地区 ，山东大众报业集团、国家图书馆、中国电影集团公司等35家单位为改革试点单位，试点工作正式启动。
3	2004年9月19日	十六届四中全会通过	中共中央关于加强党的执政能力建设的决定	深化文化体制改革 ，解放和发展文化生产力。根据社会主义精神文明建设的特点和规律，适应社会主义市场经济的要求，进一步破除制约文化发展的体制性障碍。坚持把社会效益放在首位，实现社会效益和经济效益的统一，把文化发展的着力点放在满足人民群众精神文化需求和促进人的全面发展上。以体制机制创新为重点，增强微观活力，健全文化市场体系，依法加强管理，促进文化事业全面繁荣和文化产业快速发展，增强我国文化的总体实力。推动中华文化更好地走向世界，提高国际影响力。抓好队伍建设，引导文化工作者深入实际、深入生活、深入群众，为人民奉献更多无愧于时代的精神文化产品。加强文化发展战略研究，抓紧制定文化发展纲要和文化体制改革总体方案。
4	2006年1月12日	中共中央办公厅、国务院办公厅	关于深化文化体制改革的若干意见	积极稳妥地推进 文化体制改革 。文化体制改革涉及面广，是一项社会系统工程，政治性、政策性很强。既要大胆探索、勇于创新，又要细致稳妥、有序推进。要把深化改革与加快发展结合起来，把加强宏观管理与增强微观活力结合起来，把加强思想政治工作与解决实际问题结合起来，充分调动广大文化工作者和各方面的积极性、主动性和创造性，支持和参与改革。要完善配套政策，使文化体制改革与劳动、人事、分配、社会保障、行政管理等各方面的改革相衔接。认真总结文化体制改革试点工作的成功经验，及时加以推广。重大问题及时请示报告，重大改革措施和政策调整按规定报批。
5	2011年10月18日	第十七届中央委员会	中共中央关于深化文化体制改革推动社会主义文化大发展大繁荣若干重大问题的决定	<p>(一) 深化国有文化单位改革。以建立现代企业制度为重点，加快推进经营性文化单位改革，培育合格市场主体。科学界定文化单位性质和功能，区别对待、分类指导，循序渐进、逐步推开，推进一般国有文艺院团、非时政类报刊社、新闻网站转企改制，拓展出版、发行、影视企业改革成果，加快公司制股份制改造，完善法人治理结构，形成符合现代企业制度要求、体现文化企业特点的资产组织形式和经营管理模式。创新投融资体制，支持国有文化企业面向资本市场融资，支持其吸引社会资本进行股份制改造。着眼于突出公益属性、强化服务功能、增强发展活力，全面推进文化事业单位人事、收入分配、社会保障制度改革，明确服务规范，加强绩效评估考核。创新公共文化服务设施运行机制，吸纳有代表性的社会人士、专业人士、基层群众参与管理。推动党报党刊、电台电视台进一步完善管理和运行机制。推动一般时政类报刊社、公益性出版社、代表民族特色和国家水准的文艺院团等事业单位实行企业化管理，增强面向市场、面向群众提供服务能力。</p> <p>(二) 健全现代文化市场体系。促进文化产品和要素在全国范围内合理流动，必须构建统一开放竞争有序的现代文化市场体系。要重点发展图书报刊、电子音像制品、演出娱乐、影视剧、动漫游戏等产品市场，进一步完善中国国际文化产业博览交易会等综合交易平台。发展连锁经营、物流配送、电子商务等现代流通组织和流通形式，加快建设大型文化流通企业和文化产品物流基地，构建以大城市为中心、中小城市相配套、贯通城乡的文化产品流通网络。加快培育版权、版权、技术、信息等要素市场，办好重点文化产权交易所，规范文化资产和艺术品交易。加强行业组织建设，健全中介机构。</p>
6	2012年2月1日	中共中央办公厅、国务院办公厅	国家“十二五”时期文化改革发展规划纲要	<p>五、加快文化体制机制改革创新</p> <p>(一) 培育文化市场主体。以建立现代企业制度为重点，加快推进经营性文化单位改革，培育合格市场主体。完成一般国有文艺院团、非时政类报刊社、新闻网站转企改制，拓展出版、发行、影视企业改革成果，加快公司制股份制改造，完善法人治理结构，形成符合现代企业制度要求、体现文化企业特点的资产组织形式和经营管理模式。推动国有文化企业积极参与市场竞争、自觉承担社会责任。把改革、改组、改造与创新管理结合起来，把深化改革与调整结构、整合资源结合起来，把建立现代企业制度与推进政企分开、转变政府职能结合起来，在政府引导下发挥市场机制的积极作用，充分发挥国有文化资本的控制力、影响力和带动力。</p> <p>(二) 深化文化事业单位改革。按照国家分类推进事业单位改革的总体要求，科学界定文化事业单位的性质和功能，突出公益属性、强化服务功能、增强发展活力，全面推进人事、收入分配和社会保障制度改革，明确服务规范，加强绩效评估考核。国家兴办的图书馆、博物馆、文化馆(站)、群众艺术馆、美术馆等公益性文化事业单位，要创新公共文化服务设施运行机制，探索建立事业单位法人治理结构，吸纳有代表性的社会人士、专业人士、基层群众参与管理。深入推进党报党刊发行体制改革和电台电视台制播分离改革，进一步完善管理和运行机制，不断扩大主流媒体的覆盖面和影响力。推动一般时政类报刊社、公益性出版社、代表民族特色和国家水准的文艺院团等事业单位实行企业化管理，增强面向市场、面向群众提供服务能力。</p>
7	2014年3月1日	十八届三中全会	深化文化体制改革实施方案	为今后一个时期的文化改革发展规划了路线图、确立了时间表、布置了任务书。总的思路和布局是，紧紧围绕一个核心目标，着力抓住两个关键环节，加快构建五个体系。“一个核心目标”，即培育和弘扬社会主义核心价值观、建设社会主义文化强国；“两个关键环节”，即完善文化管理体制、深化国有文化单位改革；“五个体系”，即构建现代公共文化服务体系、现代文化市场体系、优秀传统文化传承体系、对外文化传播和对外话语体系、文化政策法规体系。

三、传统出版业都有哪些数据？特点是什么？

2014 年数据（不含复本数）：

2014年书刊报合计新出版内容数据量为：987.45亿汉字（图书）+336.35亿汉字（期刊）+439.40亿汉字（报纸）=1763.2亿汉字=3526.4亿字节=0.32 TB。

截至2014年底，全国共有出版社583家（包括副牌社33家）；平均每个出版社新出图书数据量为：
 $987.45\text{亿汉字} \div 583\text{家} = 1.69\text{亿汉字} = 3.38\text{亿字节}$ 。

三大出版集团2014年新出图书（不含复本数）数据量为：

江苏凤凰出版传媒集团：29.4亿汉字 = 58.8亿字节 = 5.48GB。

湖南出版投资控股集团：16.87亿汉字 = 33.74亿字节 = 3.069GB。

中国教育出版传媒集团：41.64亿汉字 = 83.27亿字节 = 7.754GB。

全国出版业近10年（2005—2014）累计出版书刊报内容数据量（不算复本数）：

2005—2014十年全国书报刊累计内容数据量约为：图书6899.17亿汉字，期刊2878.57亿汉字，报纸4700.30亿汉字，十年合计约为：14478.04亿汉字，换算成字节约为28956.08亿字节，即约2.63 TB。

四、转型升级与出版业“大数据”应用

五、“数字版权保护技术研发工程”
落地应用情况介绍

一个依靠数据分析进行产品优化的真实应用

举个例子吧。一个只有10名员工的公司，要统计每个人中午吃什么，直接向每个人问一下，脑子就记住了；一个有100名员工的公司，要统计每个人中午吃什么，可能就得借助纸和笔这样的工具了；一个有1000名员工的公司，要统计每个人中午吃什么，说不定得拿EXCEL表汇总一下了；假如要想实时了解互联网上的用户中午用餐行为，EXCEL表恐怕也未必管用了。此时，“大数据”粉墨登场。

再举一例子。如果求一个正方形面积，长乘宽就够了，何必非要使用微积分呢？是说我们进入了一个“极限时代”吗？极限思想肯定是人类认识史上的巨大飞跃，其伟大之处是面对复杂问题有了特殊的解法。但如果问题简单到像求正方形面积一样，使用微积分就未必合适了。

上述两个例子是想说明，“大数据”既意味着一种数据的量级，也意味着数据的复杂程度，这正是“大数据”的两大主要特点。

EXCEL SPSS MATLAB