



▲谭跃

主讲人简介：

谭跃，全国政协委员，中国出版集团有限公司原董事长、党组书记。曾获国家新闻出版总署“新中国60年百名优秀出版人物”称号、“2009中国书业年度评选”（《出版人》杂志主办）最高奖年度出版人奖、《中华读书报》2009年“年度出版人”奖等多个奖项；当选2009CCTV中国经济年度人物，是新闻出版界首次获此殊荣的人。

传统文化、大数据与出版

□主讲人：谭跃

大数据与中国传统文化

我曾读到这样一个观点，数据化将使世界一切皆可量化。一切，当然也包含出版。这个观点指出，过去年代的重点是在技术；而现在，人们聚焦在信息，也就是数据(Data)。Data的拉丁文本意是已知，是存在过的现实。

那么，数据化与数字化是什么样关系呢？学术上有很多表述。我认为，在二进制的数字条件下，数据又做了进一步开拓，让它更加实用，更加能够运用到社会领域的方方面面。美国很多专家认为，数据无处不在。

美国专家认为，人类现在正进入一个新的时代，这个时代的“数据”事实证明，世界的本质不是原子，而是信息。也就是说，世界是由万物组成的，过去我们对世界的认知是基于万物由原子构成，而今专家们认为万物的基础不是原子而是信息。这让我想到了老子在《道德经》中所讲：“大道汜兮，其可左右。万物恃之以生而不辞，功成而不有。衣养万物而不为主，常无欲，可名于小。”不管是原子，还是数据，都是从不同角度回答什么是道的问题。其实，我

们的祖先早在很多年前就已经阐释了它。道，就是原子形式的物质，数据形式的信息，不同领域的道表现不同，所以它无处不在。

任何事物都不是从天而降的，都有一个历史演化过程，都能找到它的根。在中国，这个根就是中国古老的思想文化。我们知道，佛教上有一粒公案，讲的是释迦牟尼在一次佛大会上，拈起了一枝花，高高举起，什么都不说，注视着全场。大家都不明白他的意思，面面相觑，只有他的大弟子迦叶破颜一笑。正是这一拈一笑，彼此心领神会，后来释迦牟尼就把衣钵传给了迦叶。

这也是“拈花一笑”的出处。对此阐释的学者很多，最有名的是南怀瑾先生。他与孔子的“一以贯之”联系起来。孔子跟曾子说：“曾参啊，我的道是什么呢？吾道一以贯之。”过去人们的理解是，讲道的人做任何事情都要一以贯之。而南先生的理解是，道是一，因而要一以贯之。老子《道德经》讲：“道生一，一生二，二生三，三生万物。”这与我们前面讲的“数据”是有联系的，只是我们的先贤在当时的条件下已经以自身独到的悟性作出了阐释。随后，孔子讲完，曾子就说：“是，我知道了。”等到孔子一走，曾子的同门就问，“一以贯之”到底是什么意思呢？曾子说，“夫子之道，忠恕而已矣！”老师的道，只有忠恕而已矣。

孔子讲的是“一以贯之”，而不是“二以贯之”，怎么用“忠恕”二字解呢？南怀瑾先生认为，这就是阴阳，就是一，二，二就是三，三就是万物。这就是中国古老哲学中的大数据。一是一切的一，一是一的一切，还有什么大数据比“一”还大吗？这就是古代哲学观与现在大数据的联系。

不只如此。《易经》的三大原则——变易、简易、不易，三大法则——象、数、理，也跟大数据的问题相关。象是现象，八卦就是8个现象挂在墙上，然后又演化成16卦、32卦和64卦，这些都是自然和社会的现象。大数据所描述的东西就是各个领域的现象。数，是数据。理，也叫辞，是对象和数的理

性认识。象靠着，靠眼睛观察。数靠算，也即是运算。理讲判断，也就是理性思维。象、数、理，最终是判断。因此，有人提出的理论不存在的观点是不成立的，只是理论建立在什么基础上很重要；如果建立在象、数的基础上，理还是有必要、有道理的。所以，在《易经》64卦的基础上，孔子才写了《十翼》，作了10篇论文，在“理”的层面把《周易》讲清楚了。

了解这一文化内容，我们心中便有了数。俗话说“掐指一算”，“算”的背后是数，心中有数，算到心中有数的层面就定了。就像现在的出版业，如果对大数据背后的道理不理解就会产生畏惧，数字化还没明白，又来大数据了？

没有明白，什么事情发生了，大家会说“早有定数”，这是普通老百姓都会讲的。但这些思想是哪里来的呢？比如，先天八卦、后天八卦，这些是数字。《易经》在汉朝时的研究被称为“京房十六卦变”，离不开数字。黄道十二宫，还是数字。当然，“十二宫”里面的数字就复杂了。六十花甲，60年一个花甲，也离不开数字。十二生肖，还是离不开数字。

前一段，我又翻出《洛书》来看，之前只注意它的图案，这次带着问题看，发现图案的上面全是数字，还有一句话：“戴九履一、左三右七、二四为肩、六八为足。”《易经》中有“六爻之动”，讲的是八卦的每一卦里有六个爻，“六爻之动，三极之道也”，它告诉我们的第一层意思是，天地之间虽然数字很多，但是真正有用的不会超过“六”。中国文化博大精深，中国先贤也是极具智慧。对此后来的科学家做过研究，除了极少的现象，万事万物的物理现象，基本上都是六个阶段。它讲的第二层意思是，“六爻”都在动，动的本质是什么呢？它告诉我们是“三极之道”，“三极”指的是天、地、人，也就是说，“六爻”是三极变化的法则。

这些都是古代人们的一种表达方式，现代人也许会认为这是非理性的，甚至是一种迷信说法。其实，这就是古人科学研究的一种数据表达，只不过现在我们不使用、不研究了。

因此，从中国古代文化中可以看出，中国先人有很强的数字概念，而如今所讲的数字、数据、大数据等，其实古已有之。

正确看待大数据

对大数据的理解，我们要正确去把握。数据本来就是有的，古人早就意识到了，并做了高度抽象的概括。只不过到了新的条件下，特别是到了大数据时代、云计算的条件下，它的作用和潜能被释放了出来，只是和古代数据观已不可同日而语了。这是第一个观念。

第二个观念是，哲学上讲量变到质变，大数据其实也一样。当数据大到一定程度时，事物的形态就发生了变化了。大数据的意义就在于此，它已经不是原来的样子了，或者说它不是了。

举个例子，法国拉斯科洞窟壁画中有很多是马的造型。毕加索参观后开玩笑说，人类自这以后就没有再创造了，画来画去，都是马。然而，现在的科学家又从量变到质变的角度重新解释，虽然一幅马的图案大同小异，变到分子量级的程度，物质形态、本质就会发生变化。比如铜，铜是可以导电的，分子级别的铜就不导电了。还有陶瓷，分子级别的陶瓷，就成了柔软、带有弹性的物质。再说金属，金属是硬的，分子级别的金属是软的，可以任意摆弄它。

这些例子都在说明“大数据”的概念早已存在，只是到了现代技术条件下，较为集中地体现了出来，但体现并揭示的仍然是哲学上的定律，量变到质变，这也就是大数据的意义所在。

第三个观念，作为企业，过去最为强调的是有形资产，当然现在还作为一个重要的判断标准，也就是企业规模。我们知道，与有形资产同样重要的，还有无形资产。对出版业来说，那就是版权、商标、商号等。到了大数据时代，它又告诉人们，比这些更重要的是数据资源。一个企业做得好，数据也是投入，也是潜在的竞争力。比如美国苹果手机公司，如果用有形资产去评估它，那它就没法了了不起，如果用大数据的思维去评价它，那就不得了。因为像这样的公司，他们拥有巨量的数据，并且每天还在增加。



▲各种书籍的出版

第四个观念，是主因。我们分析得知，大数据古已有之，只是现在集中体现了出来，那么主因是什么？我认为，是处理数据的能力在迅速提升。

大数据首先来源于哪里呢？一是来源于天文学，因为天文学的信息量巨大。二是来自生物学中对基因的研究。一个是宏观，一个是微观，这两极的数据都是巨量的，以至于大到现在的计算能力都无法实现。因此，到了大数据时代，这就倒逼着科学家们处理数据的能力迅速提升。

再举几个例子。有资料显示，2000年在美国新墨西哥州启动的斯隆数字巡天项目启动，用望远镜几周之内收集的数据，比自人类有史以来收集的所有数据还要多。但到了2010年，智利利用巡天望远镜5天的时间

就完成了它所有的运算，打破了新墨西哥州的记录。2013年，中国的“天河二号”超级计算机，已经把计算速度提高了几倍，成为当时全球最快超级计算机。2020年，我国的神威·太湖之光超级计算机已经实现千万核心并行第一性原理计算模拟，排名于世界前列。

10多年前，全球科学家联手用10年的时间完成了31.6亿个基因碱基对的排序。而今，这样的工作量只需十几分钟完成。谷歌公司每天要处理超过24拍字节的数据，这意味着其每天的数据处理量是美国国家图书馆所有纸质出版物所含数据量的上千倍。Facebook每天更新的照片量超过1000万张，每天的点击量或者写评论大约有30亿次，点击一次，就有大量的数据保留下来。当然，这些数据还在不断增长。

那么，问题就来了：第一，大数据的核心是什么？第二，大数据的特点是什么？第三，关键在哪里？通过不断学习，思路就慢慢清晰了。

首先，大数据的核心在于预测。对于出版业来说，早在几年前，亚马逊公司就已经开始使用大数据预测，预测市场需求，市场反应。这样的预测，可以明确某一个人的阅读领域是什么、最近的阅读热点在哪里，等等。不管是手机端、电脑端，只要使用一次就能留下数据，公司后台通过数学模式计算运转，结果就出来了。

其次，有人提出，大数据的特点有三个：一是数据更多，二是数据更杂，三是数据更好。这有一定道理。

数据更多，不是随机样本，而是全体数据。就是对一件事用全数据——即

数据更杂，不是精确性，而是混杂性。说的是越复杂、越纷繁的数据，预测将更有效、更接近真实。这种预测不是在追求每个数据的精确性，当一件事物的数据达到临界点的时候，它的形态性质就可能就发生了变化。这与抽样又不一样，抽样强调的是随机性，不能带主观意愿，牺牲的是数据的量。而大数据追求的是量，放弃的是精确性。

数据更好，不是因果关系，而是相关关系。比如沃尔玛的历史数据表明，一旦有飓风的时候，蛋挞的销量大增，不问为什么，不问因果关系，只问相关关系，就是一旦A出现了，B必然出现。所以沃尔玛就把蛋挞和雨伞、手电筒等飓风用品放在一起，销量大增。这就是相关关系，也就是让数据说话，只说现象，不问原因。

最后，关键在哪里？我归纳有这样几条：一是整体性，就是关于某个事物数据的整体性，而不是随机、抽样、代表。二是既然有了大数据，就必然呼唤云计算处理能力。三是容错，如谷歌公司为了做翻译平台，它建立了上万亿的语料库。这些数据都是互联网上已经发生过的数据，其中大多是废旧数据，所以会出错，那么容错才能接近真理。四是相关性。它的核心是量化两个数据值之间的数理关系。简单说，就是此长彼长或此消彼长的关系，A情况出现时B情况必然出现的关系，就是相关关系。

这件事的几乎所有数据来预测，准确率将会大大提高。比如人口普查，与抽样是相反的理念，抽样最大的特点是用少量的数据反映更多更真实的信息，而大数据预测则与之相反。

数据更多，不是精确性，而是混杂性。说的是越复杂、越纷繁的数据，预测将更有效、更接近真实。这种预测不是在追求每个数据的精确性，当一件事物的数据达到临界点的时候，它的形态性质就可能就发生了变化。这与抽样又不一样，抽样强调的是随机性，不能带主观意愿，牺牲的是数据的量。而大数据追求的是量，放弃的是精确性。

数据更好，不是因果关系，而是相关关系。比如沃尔玛的历史数据表明，一旦有飓风的时候，蛋挞的销量大增，不问为什么，不问因果关系，只问相关关系，就是一旦A出现了，B必然出现。所以沃尔玛就把蛋挞和雨伞、手电筒等飓风用品放在一起，销量大增。这就是相关关系，也就是让数据说话，只说现象，不问原因。

最后，关键在哪里？我归纳有这样几条：一是整体性，就是关于某个事物数据的整体性，而不是随机、抽样、代表。二是既然有了大数据，就必然呼唤云计算处理能力。三是容错，如谷歌公司为了做翻译平台，它建立了上万亿的语料库。这些数据都是互联网上已经发生过的数据，其中大多是废旧数据，所以会出错，那么容错才能接近真理。四是相关性。它的核心是量化两个数据值之间的数理关系。简单说，就是此长彼长或此消彼长的关系，A情况出现时B情况必然出现的关系，就是相关关系。

编者的话：

今年是中国共产党成立100周年。100年来，在中国共产党的领导下，我们国家的面貌发生了前所未有的改变。随着科技飞速发展，以互联网为核心的新一轮科技和产业革命蓬勃兴起，不断给人们的生产和生活方式带来革命性变化。出版，凝结着人类的思想和智慧，集聚了人类的发明创造和社会实践活动的经验与成果，反映了社会生活的各个侧面。出版的历史就是人类文明的历史。正确看待大数据与传统出版的关系，积极改革创新，以推动出版业的繁荣发展。本期讲坛邀请谭跃委员从大数据的角度，讲述出版业在新时代的发展。

出版与大数据

联系实际，我有这样的体会：一方面，传统出版人要正视这一问题，有必要清楚大数据的由来、爆发的主因，及其核心、特点和关键所在，这是大势所趋，也是发展方向；但另一方面，要充分估算到互联网、数字化、大数据、云计算等对出版的影响，以及又是如何影响的，还要对此进行总结。

15世纪印刷机被发明出来以后，它带动了一次世界性的信息爆炸。印刷机面世以后，大约出版了1.3亿册图书。到2010年，也就是谷歌的数字化图书计划实行7年之后，大约有2000万图书被扫描成了数字图书，这几乎相当于人类所有书写文明的15%；也就是说500多年产生的信息量，它只用了7年就完成了。这还只是一家公司，还没有完全展开来做。

大家知道亚马逊，它的优势是Kindle（电子书阅读器）。在Kindle上阅读的重复率、标记次数以及画线次数会有数据留存，这都是读者的重要信息。可是亚马逊都把它藏在那儿，不愿意跟出版商共享，因为出版商也不愿意把版权跟它分享。前面我们也讲了数据无处不在，如今大数据已经运用到各个行业，从未来的眼光看，它一

定会继续深刻的、长期的影响社会生活的各个方面。

再回到出版业，我们需以辩证的眼光来看待。

第一，究竟改变的是什么呢？数字化已经并将持续改变我们图书生产和管理流程，并部分改变呈现方式、阅读方式和营销方式。我还是比较保守的，不太相信颠覆论。因为结果往往不是专家预测的，而是消费者来决定的。我的简单判断是，像我这样的人，正常情况阅读还有30年，我们的选择倾向主要还是纸质书。但是改变是必然的，现实已经改变了，并且还将更大地改变。

第二，凸显的问题是什么呢？是内容的海量，以至于泥沙俱下，难以选择。我们打开各种数字化的端口，只要打开就会感觉到什么都有，但困惑的是不知道如何选择。现在有了监管，效果是好了很多。

第三，最终稀缺的是什么呢？稀缺的是有效内容的搜索，以及重大思想成果的选择和获得。这一方面表现的不光是互联网，还是大数据，都在不断成长中，另一方面也说明传统出版的努力依然有着巨大空间。

第四，比较难办的是什么呢？难办的是内容主体的孤岛化。刚才说到亚马逊，各大书商和亚马逊之间，虽然各有资源优势，但之间却是屏蔽的。这虽然是几年前的数据，但主体的孤岛化以及内容数据化的商业模式还是有待探索与开发。

第五，数字化和数据化，改变着关于内容生产的市场预测、加工方式、管理过程、呈现样式、交易方法，但是没有改变内容本身的价值。没有改变传统出版人原来所拥有的立点，就是内容这个立点没有改变；也没有改变内容创新的主体地位，特别是不会改变精神产品生产的规律，即是说规律是不会改变的。

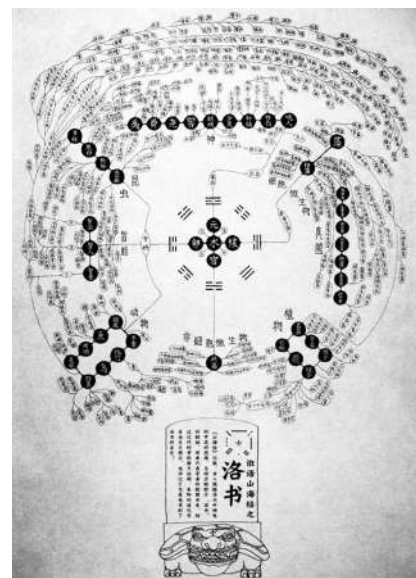
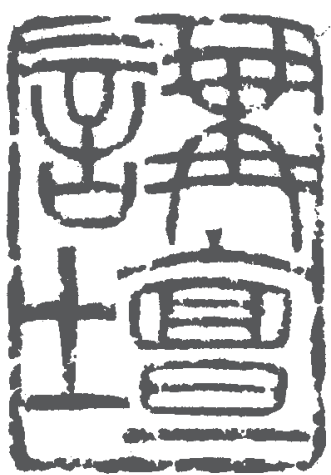
前些年，美国人在欢呼、崇拜大数据的同时，也发现了问题。比如，有人问乔布斯，苹果公司经营得这么好，市场是怎么调研的？乔布斯说没有调研，并接着说，消费者没有义务去了解自己想要什么，这是生产者的事。这是一个反例的例子，不用大数据预测却成功的例子。还有一个，500多年前哥伦布发现新大陆，没人做过预测，没人有数据概念，更没有

大数据概念，但是一个历史性的重大发现诞生了。还有，福特时代没有数据显示，汽车可以代替马车，但是汽车这个重大发明也诞生了。这些都说明，大数据是管用的，但是更大的数据，比大数据更大的数据是什么呢？是源于人本身，是人的创造力、直觉和天赋。我们的先贤哲人，那个时候都不具备大数据的运算能力，依靠的就是他们的智慧，给我们留下了丰厚的文化遗产，这种智慧源于直觉，靠的是人自身的悟性，对事物规律的把握。

因此，我们要看到，数字化、大数据对传统出版带来了知识的集成、需求的判断、趋势的把握、搜索的便捷以及跨界的链接。这不仅是挑战，更重要的是商机，是发展空间，是文化的传承。科学再怎么发展，它不会改变出版的本质——选择。比如，孔子年代占卜书不少，著名的、上了《左传》的和《周易》一样地位的就只有《周易》。没有这方面内容的选择，就不会有后来《周易》的地位。《诗经》也是一样，也是孔子选择的结果。看上去一个简单的、具体的编辑工作，带来的是长久的社会影响和伟大的思想成果。



扫码读讲坛



▲《洛书》