

地方综合年鉴构建大数据应用的设想

余贤林^{*}

摘 要 进入 21 世纪,信息化技术在各领域飞速发展。对于集中反映年度资料集合体的地方综合年鉴来说,在信息化的推广运用方面面临着更大的机遇和挑战。年鉴的性质决定了其在大数据运用中的巨大优势,因此,年鉴工作理应树立引进和运用大数据的理念,并在编纂过程中运用信息化的处理技术,只有这样,才能有效利用地方综合年鉴形成的庞大海量信息资源,为年鉴功能的拓展发挥作用,更好服务于地方经济社会发展和高效的行政管理,成为实现年鉴转型升级的有效途径。因此,无论是大数据平台建设还是针对大数据的研究,都为大数据的应用打下基础,也是年鉴创新的必然要求。

关键词 大数据 年鉴功能 拓展创新

地方综合年鉴(以下简称年鉴),是由地方政府主办的,系统记述本行政区域自然、政治、经济、文化、社会等方面情况的年度资料性文献。近年来,各地对年鉴工作进行了有益的探索和创新,从各地的情况来看,年鉴的改革和创新大多仍集中于年鉴本身的框架、栏目、内容等方面,而年鉴的传播范围窄、利用率不高、信息化程度低等问题依然存在。

大数据是一个新兴事物。关于大数据的概念、基本特征和一些典型分类,由于各方面的专家论述和研究较多,具体的定义和阐述并不统一,本文也不一一赘述。总的来说,大数据是以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合,正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析,从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态。^①

年鉴的主要功用在于记录和存史,但是修鉴为用,在大数据时代的背景下,年鉴工作不能仅仅局限于单纯的适应大数据时代的要求,更要充分有效利用大数据的技术手段,结

^{*} 余贤林,男,四川省梓潼县人,梓潼县地方志办公室主任,主要研究方向为地方史志发展、志鉴应用研究。

^① 国务院:《关于印发促进大数据发展行动纲要的通知》,2015 年 8 月 31 日。

合年鉴编纂所形成的海量信息资源,建立一个基于年鉴信息和数据体系的、服务于政府和公众的大数据平台,建立“用数据说话、用数据决策、用数据管理、用数据创新”的管理机制,实现基于数据的科学决策,将推动政府管理理念和社会治理模式进步,加快建设与社会主义市场经济体制和中国特色社会主义事业发展相适应的法治政府、创新政府、廉洁政府和服务型政府,逐步实现政府治理能力现代化。^① 推动政府数据全方位的开放共享,这既是年鉴改革创新的有效途径,也必将有力地提高年鉴的利用率和信息化程度,促进年鉴为当代政治、经济、社会、文化、生态建设服务。

一、年鉴在大数据运用中的优势

年鉴由于“一年一鉴”的基础特性,决定了年鉴的资料信息的滞后性,与大数据的速度性质并不相融。但是在实践中不难看出,年鉴的编纂与出版也有其本身的不与大数据技术相排斥的特点和特性,其海量的信息资源,本身就是一部大数据,加之“述而不论”的科学编纂方法,这也是年鉴与其他平台比较的巨大优势所在。年鉴从其表现形式和内容来看,大部分为文本、图片、表格,部分年鉴创新发展中还加入了视频、音频等其他多媒体,要有效利用这些海量的数据资源,采用大数据的技术手段来实现也是势所必然。

(一)权威性

从年鉴的资料收集过程来看,年鉴的资料内容除基础性资料外,既是各级各部门的资料汇总,也是各级各部门各项数据的汇总,其来源可靠,是正式的、公开的,能够反映其真实的状况;同时,在年鉴中也收录了各级统计部门的统计公报。从年鉴的出版来看,各地基本上已经实现了“一年一鉴、公开出版”的目标,已经形成了作为地方政府对外的权威发布的载体。

(二)系统性

大数据的难点在于数据资源的整合和共享,目前,各地由于历史的原因,条块分割的状况仍然普遍存在,各级各部门所形成的资料、数据,作为行业内来说,有一定的意义,但是他们只能从一个侧面来反映当地的状况,而不能全面的、可比的、更高的层面、更广的范围来描述和反映整体的情况。而年鉴对地方来说,是实现了跨部门、跨平台的一个载体,能够全面反映当地的政治、经济、社会等状况,是一部政府的“百科全书”。

(三)全覆盖

年鉴的篇目设置各地虽然有所异同,但是均基本涵盖了当地的政治、经济、社会、文化、生活等社会组成的各个方面,既包括政府各级及其组成部门,又包括经济开发区、重点企业;既包括了经济所属的各个行业,又包括了社会建设、社会事业,更是包括了社会生活和民生的各个层面,整个资料和数据体系的完整性是其他出版物无法比拟的。

^① 国务院:《关于印发促进大数据发展行动纲要的通知》,2015年8月31日。

(四)连续性

一是时间的连续性,年鉴较为系统的编纂始于20世纪80年代末90年代初,在前期一二部为综合卷,后期大多实现了“一年一鉴、公开出版”,近30年连续出版,个别的行业年鉴时间更早。二是有明确的时间界限,年鉴编纂规定了其上限和下限,除个别特殊事件外,均为上年度的1月1日到12月31日。三是事物的发展变化的连续性,年鉴对一项事物的产生、发展、过程、结果基本都有记载。四是数据的标准基本统一。

(五)数据的多样性

各地地方综合年鉴的编纂方式、方法和传播的形式发生了很大变化,除传统的纸质媒体外,还有光盘版本、U盘版本、手机APP等多种形式,在原有大量的文字与数据之中,采用了大量的图片、音频、视频等素材,从年鉴的发展方向来说需要运用大数据的成果分析和指导,从而提高年鉴的利用率和运用效果。

二、年鉴大数据的运用设想

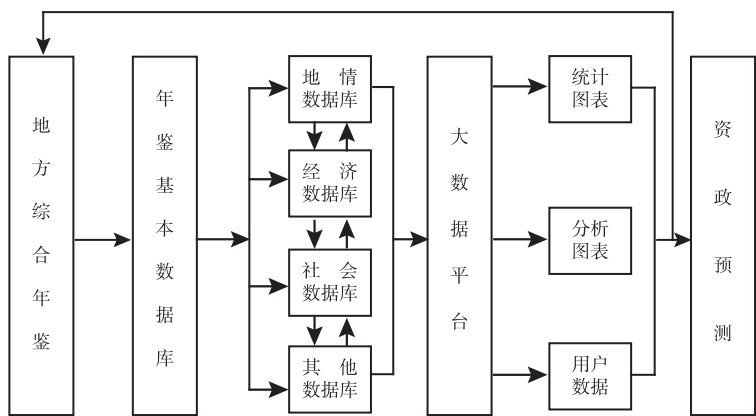
地方志的特点是庞大的历史文献和资料。洪民荣指出:“适应地方志智库作用的方法,典型的可以通过比较研究形成创新。一是“集成创新”,典型的是通过地方志成果的横向整合,得出新的结果和结论,提出新的建议。二是模仿创新,即借鉴历史已有成功经验或做法,提出改进后的意见和建议。”^①他还提出了地方史志的研究和创新的方向,但没有提出研究和创新的手段和方式方法。那么年鉴怎样发挥“资政”的作用,发挥“资政”的方式和实现的手段是什么?即以地方综合年鉴为基础所建立的大数据平台,以及所形成的全国的年鉴大数据的联网,为政府及各个部门的管理决策、改进服务等提供可视化的大数据支撑,为各地提供横向的、纵向的、跨地区的、跨部门的大数据支持。

目前,以大数据为技术手段运用在各个领域,尤其在商业应用领域比较广泛,如我们经常使用的淘宝(商品的定点推送)、支付宝(金融的风险控制)、百度地图(路径优化)等等;在政府层面,目前应用大多集中在行政审批、民生事项办理等方面,而全面反映政府经济社会运行,政府本身的运行、行政管理、政策的执行,为人民群众参与政府各项事务提供服务等方面缺乏大数据的支撑。目前为止,已经有省市走在了前列。2018年3月14日,贵阳市地方志办公室与同方知网签署了全国首个地方志云平台大数据项目协议,协议的签署将促进地方综合年鉴在大数据应用方面的研究,也必将带来地方志工作的重大变革。综上所述,年鉴的大数据运用平台基本架构设想如下。

(一)年鉴大数据应用的基本模型设想

以大数据的技术,将年鉴所形成的数据经清理、整合、转换、关联、集成后,以一定标准实现格式化存储,建立起相应的地情数据库,如经济数据库、社会数据库等。

^① 洪民荣:《地方志:既是“知库”,也应成为“智库”》,《解放日报》2017年7月18日第9版。



年鉴大数据结构模型示意图

（二）数据库的建立

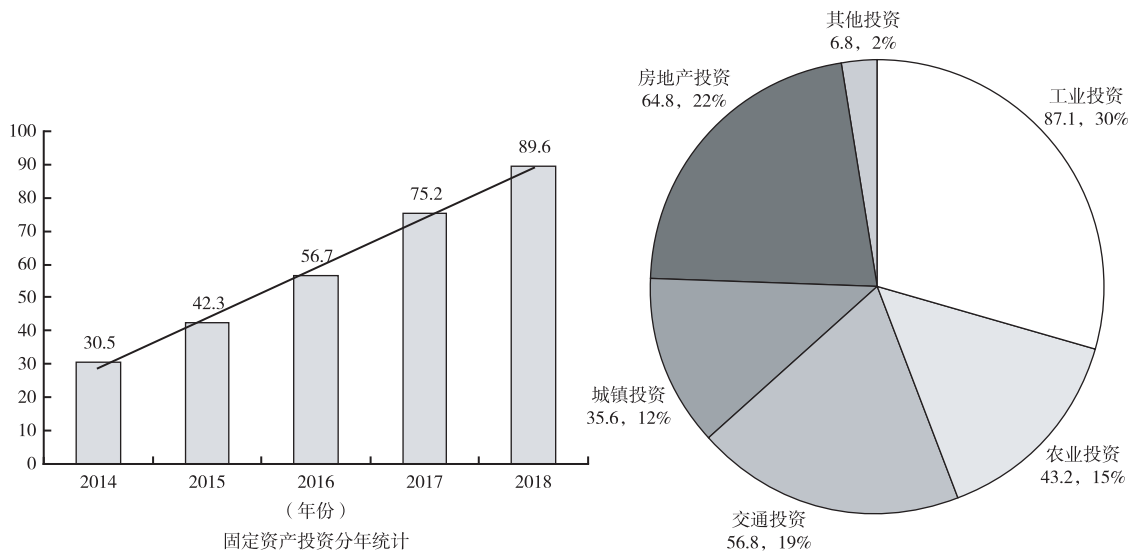
一是数据库的建立。即是将所形成的年鉴全部实现电子化,这里的电子化不是简单的将年鉴做成 PDF、WORD 等文档,而是要实现年鉴条目、全文、图表、统计等内容的分级检索,这方面已经比较成熟,在许多省市已经建立起了或者部分建立了年鉴的数据库。但是个别的数据库仍然只是一个简单的收录,还没有实现表格化存储以及篇目、条目、分目检索等服务功能。二是数据采集和清理。根据县域经济的特点,分类建立各种类型的专项数据库。如上所述的基础地情数据库,这里面可以包含土地、人口、水资源、森林资源、气象等基础信息;经济数据库,包含县域经济一、二、三产业经济,也可以分产业建立相应的数据库。社会数据库,包含行政事务、社会事务、民生事务等以及其他类的数据库。三是数据库的关联。各数据库相对独立,但是数据库之间通过特定算法实现相互关联,成为联机分析处理、数据挖掘和预测的基础。如需要查询固定资产投资状况,那么除经济数据库里的固定资产投资外,同时相关联的农业农村投资、交通建设投资、城镇投资、工业投资等都应有相应的关联。四是针对地方综合年鉴在大数据方面滞后的特性,在平台建立之后,可与县级其他部门的即时数据库进行关联,以解决即时数据更新问题。

（三）统计分析结果的可视化

检索结果为可视化界面,结果应按照一定规律排列,而不是简单罗列出来。可视化的结果能够直观的呈现所检索数据的特点,同时又能体现可视化结果的关联性,简单易用,即使未经专门培训也能很快接受和使用,就如同看图说话一样简单明了。如前文提及的固定资产投资统计分析结果的表达,就可以实现相应的可视化分析结果,并同时显示关联性数据的统计分析结果。如需查询其中一年数据,只需点击相应链接即可。

（四）资政和预测

修鉴为用,编纂年鉴最终目的还是要用于为当今时代服务,预测性分析更是大数据分



可视化统计分析结果示意图

析最终的应用领域之一。年鉴所形成的大数据经过采集、清洗、关联,通过统计和分析,展现出事务的表面特征以及相关因素之间的内在联系,从而在政府决策支持方面以及经济和社会事务活动方面发挥重要的作用。数据挖掘作为大数据的高级应用,以完整的数据基础,通过科学的计算手段,能够在一定的层面实现预测的效果。

三、大数据背景下的地方综合年鉴的创新

年鉴的创新是个永远的话题,近年来各地在年鉴创新上都在不断的积极探索,但从实际效果来看,还不尽如人意。个别地方对于年鉴创新还是停留在形式上的创新,装帧越来越豪华,印刷越来越精美,篇目结构不断调整,一些地方的年鉴在传统光盘年鉴基础上,做优盘年鉴、微信年鉴、年鉴口袋书等,这些创新实质上并没有实现年鉴应用的根本性改变。实现年鉴创新,要从根本性改变年鉴应用的瓶颈因素,利用大数据技术是实现年鉴创新更有效的途径。大数据在年鉴编纂和研究中的应用可从以下几方面着手

(一)加强年鉴创新应用基础理论研究

新时代的信息技术不论是从硬件上,还是在软件上均发生了大的改变,特别是在云存储、云计算、网络通信等技术的发展,给我们的办公方式、传播方式、应用方式带来了翻天覆地的变革。年鉴的编纂要适应这种改变,年鉴的基础理论研究也要适应这种改变,要从中央到地方,将大数据思维融入年鉴工作的始终,对新时代年鉴的功用发挥进行再认识。在年鉴学学科理论研究之外,要逐步建立跨学科、跨平台的年鉴基础理论研究和实践探索,加大对大数据研究的人、财、物的投入,为年鉴的深度开发利用注入活力。

（二）年鉴的大数据平台“四步走”发展战略

按照目前的具体情况,建立大数据平台可分四步走,一是实现年鉴的电子化,研究制定统一的年鉴电子化的规范和标准。年鉴的电子化是年鉴创新应用的基础,应尽快就年鉴电子化的标准、规范和时间要求进行明确,进而在全国实现年鉴电子化的全覆盖。二是建立年鉴全文数据库。以年鉴的规范为基础,全国统一标准,以县为单位,建立以年鉴为基础的县级年鉴分类数据库。三是选择有条件的县区进行试点,以县为单位建立年鉴的大数据平台,并逐步实现县到市、省、全国的联网。四是将全国各类专业年鉴纳入年鉴大数据平台,实现地方综合年鉴跨平台、跨领域对接。

（三）加强年鉴编纂与大数据应用的对接

为了保证大数据分析的真实和有价值,对数据的规范性提出了更高的要求。在年鉴的编纂中,总体框架、篇目设置、资料的采集均要适应大数据环境下的要求,实现无缝对接。一方面,我们的数据采集要有真实性、系统性、连续性、可比性、及时性,即我们的资料的收集要全面准确地反映当地的政治、经济、社会、文化、生态状况,做到应收尽收,尽量做到连续可比。另一方面,在年鉴的内容上创新,既要反映出年鉴的地方特色,同时也要反映出年鉴的年度特色,即便是一些工作不是连续可比的,但这些特色工作按照大数据对于数据的挖掘算法,都能在大数据中发挥其应有的价值。

（四）利用大数据平台的数据反馈,提升年鉴的质量

受众的体验是目前年鉴编纂过程中最难以掌握的,他们读后的感想、他们的兴趣点在什么地方等,我们都不清楚。就笔者在年鉴编纂实践中来看,我们在出版发行之前,一般都会请一至两名读者通读全书并提出意见,他们作为第一读者往往会发现我们自己在编纂过程中难以发现的错漏,但是这仅仅是一部年鉴反映出来的极小部分。年鉴的大数据应用,其中既有年鉴中的大数据统计分析及预测应用,也有用户在使用过程中所产生的交互数据,通过对这些数据的统计和分析,揭示受众的行为特征,定点推送读者更感兴趣的部分,既能不断提升受众与年鉴的良性互动,又能促进编纂思路 and 内容的调整,更加符合时代的要求与针对性,促进年鉴编纂质量的提升,更好地为社会服务。

四、结语

改革与创新是新时代的关键词,年鉴既要为国存史,更要积极主动参与社会实践,为当今时代服务。只有这样,才能不断地焕发生机与活力。年鉴与大数据技术相融合,是年鉴发展创新的必然趋势,也是年鉴实现科学化应用的有效途径。可以想象,以年鉴为基础的大数据应用平台实现全国联网,必将带来年鉴事业的飞速发展,也必将带来年鉴事业发展的广阔前景。