

基于政务云数据中心的政府数据开放共享服务集成平台研究*

唐长乐 王春迎 (武汉大学信息管理学院 湖北 430072)

摘要 文章分析了政务云数据中心、政府数据交换共享平台、政府数据开放平台和政府数据整合服务平台,并在这一大中心三大平台的基础上,从政务数据的储存、共享、开放和服务的集成管理出发,提出构建一个基于政务云数据中心的政府数据开放共享服务集成平台的设想,从而实现政务数据生命周期的全过程管理。

关键词 政务大数据 政务云数据中心 政府数据开放 数据共享 数据服务 集成平台

Research on Integrated Platform of Government Data Open and Sharing Service Based on Government Cloud Data Center

Tang Changle Wang Chunying (School of Information Management, Wuhan University, Hubei, 430072)

Abstract This paper analyzes the government cloud data center, the government data exchange and sharing platform, the government data open platform and the government data integration service platform. Based on one center and three platforms and from the perspective of government data storage, sharing, opening and service integration management, this paper puts forward the idea of building an integrated platform for the opening, shared and service of government data which based on the government cloud data center, so as to realize the whole process management of government data lifecycle.

Keywords government big data, government cloud data center, open government data, data sharing, data service, integrated platform

1 建设背景:政务大数据

大数据时代的来临引发了社会各个领域的数据革命,拥有海量的数据并对其进行分析挖掘和应用已然成了各业的潮流。大数据催生了政务大数据这一概念,与传统的政府数据不同:在数据类型上,政务大数据不仅包括结构化的数据,更包括了海量的、实时的非结构化数据;在数据来源上,政务大数据不仅包括政府在部门业务过程中采集和产生的数据,还包括统计调查、网络舆情治理等得到的关于企业、个人的社情民意

数据,以及通过物理设备采集到的气象、环境、交通等环境数据。同时,政务大数据也与其他行业领域的的数据不同,政务大数据涉及经济、社会、教育、文化、医疗、环境、工农业、旅游等各个领域,其复杂程度远大于特定的领域或行业,而且许多数据只有政府部门才能采集和拥有,一般企业和个人无法取得。

政务大数据作为一种丰富独特的大数据资源,对其进行存储是基础,但盘活它以发挥其价值,才能变得有意义。世界知名的管理咨询公司麦肯锡调查了各行业大数据价值发挥的难易程度以及发展潜力,如

* 本文系国家社会科学基金重大项目“面向国家大数据战略的政府数据开放共享对策研究”(编号:15ZDC025)的研究成果之一。

图1所示^[1],结果表明:政府数据价值的潜力要明显大于其他领域或行业,而且其数据开放的难易程度也相对较低。可见,政务领域是大数据价值发挥和技术应用的重点领域,政务大数据具有极大的开发价值和开发的可行性。大数据背景下的政府数据管理将极大地改变政府的管理模式,节约政府投资,提高政府决策能力,提升公共服务和社会管理能力;麦肯锡的这份报告还估计,欧洲发达经济体政府利用大数据可以节省超过1000亿欧元的政府开支。为此,欧盟正在力推《数据价值链战略计划》,用大数据改造传统治理模式,试图大幅降低公共部门成本,并促进经济增长和就业增长;2016年9月份我国出台的《政府信息资源共享管理暂行办法》指出,要加快推动政务信息系统互联和公共数据共享,充分发挥政务信息资源共享在深化改革、转变职能、创新管理中的重要作用,增强政府公信力,提高行政效率,提升服务水平^[2]。然而目前我国政府部门的政务大数据的开发深度和应用广度却十分有限,一方面政府部门较为重视事务性活动,技术活动相对较少,同企业相比,技术更新和对数据开发利用的重视程度不高;另一方面,政府内部各个部门之间相对独立,且缺乏协调机制,大量数据独立储存在不同业务系统中,数据割据影响数据的共享与应用服务,难以发挥政务大数据的真正价值。

因此,笔者结合大数据背景从政务数据的存储、共享、开放和服务出发,分析了与之对应的一个中心和三大平台,并以此为基础,分析了当前时代背景下发挥政务大数据的真正价值,需要结合大数据分析、云存储、

物联网等新兴技术,构建一个统一的基于政务云数据中心的政府数据开放共享服务集成平台,以达到多方面的集成协调,实现政府数据生命周期的全过程管理。

2 基础奠定:一个中心三大平台

基于政务云数据中心的政府数据开放共享服务平台作为一个综合性的政务数据管理平台,它的建设离不开政务云数据中心、政府数据交换共享平台、政府数据开放平台和政府数据整合服务平台。传统模式下,一个中心三大平台独立存在,发挥各自的功能,但它们的建设经验为集成平台的建设提供了基础,所以有必要了解它们各自的含义和功能。

2.1 政务云数据中心

云存储系统采用大规模分布式并行文件系统,以大量的服务器和存储设备为基础,构建一个大规模存储集群,提供上百PB的存储容量,并能够在线进行容量的扩充,由此搭建的大容量存储系统整体成本远低于传统存储架构,并且具有良好的可扩充性和灵活性^[3]。

政务云数据中心的主要功能就是一个云存储系统,能利用云存储技术实现政务大数据的收集与存储。政务大数据来源广泛,既包括外部数据源,如互联网、移动互联网、物联网、社交媒体等产生的舆情评论、社交关系、实时位置、旅游度假信息、车位使用等数据,还包括内部数据源,如政府业务及相关系统产生的司法、公安、设计、财政、商务、民政、教育、环境、卫生、城建、交通运输等业务数据。海量数据对政府部门的存储能力提出了严峻的挑战,利用云存储技术可以提高资源利用率、减少硬件设备数量,构建分布式数据中心架构。将以政务外网为基础的横向外部数据源,以垂直部委纵向网为基础的纵向业务数据源,以及部分委办厅局牵头的行业数据源,一点式接入统一的政务云数据中心互联网络,可以实现对政府数据的集中存储和管理。将各政务部门的数据资源存储于容量可扩充、高传输率的政务云数据中心,既节省了部门系统建设初期及后续的存储软硬件设施的投入成本,也为后续其他平台功能的实现,

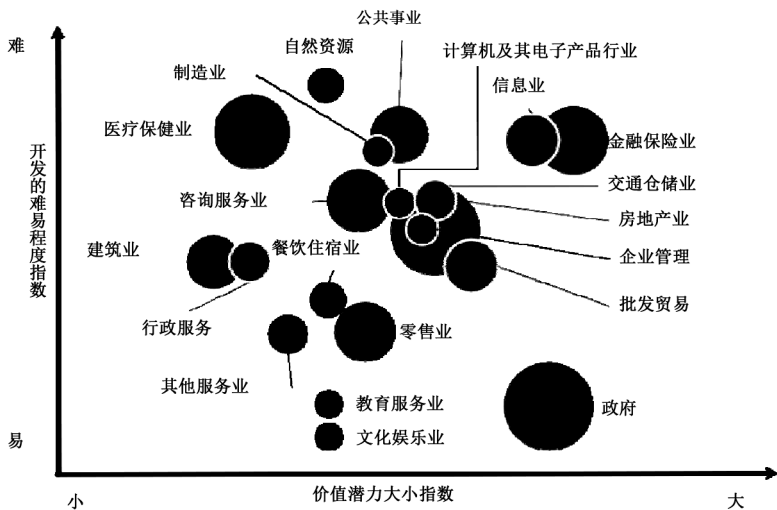


图1 各行业大数据价值潜力和开发难易度

如政务数据的交换共享、开放和服务提供了基础和接口。

2.2 政府数据交换共享平台

政府数据作为公共产品,其共享包括两个层面:一是在资源的积累和建设期间,有共同参与,共同建设的理念,以节约有限的资源,避免重复建设;二是在资源的使用上,提倡开放、分享,以最大程度上提高使用效率,发挥其最大效益,避免部门分割、单位所有的现象^[4]。而依托政府数据交换共享平台等先进技术手段,可以有效降低成本,使共享成为现实可行。

《政府信息资源共享管理暂行办法》规定,按照资源共享属性,政务信息资源可分为三类,无条件共享、有条件共享和不予共享三类,同时规定政府信息资源共享平台是管理国家政务信息资源目录、支撑各政务部门开展政务信息资源共享交换的国家关键信息基础设施^[5]。这里的政务信息资源既包括了政府部门加工处理后的信息,也涵盖了政府大量的基础数据。按照《政府信息资源共享管理暂行办法》的规定,政府数据同样也应当具有不同的共享属性,这就需要提前鉴别哪些数据可以共享,哪些属于有条件共享,哪些需要保密。对于可以共享以及具备部分共享的数据,则需要相应的技术平台支持不同部门间数据的共享,而政府数据交换共享平台就是利用目录技术、交换技术,实现不同政府间数据交换共享的技术平台,是国家信息基础设施的重要组成部分。早期的政府信息系统建设由于独立软件繁多,基层数据工作独立,数据共享性差,缺乏整体规划。早期为解决这一问题采用了多对多的网状共享模式,但其缺乏统筹管理,协调成本过高,周期长,而且已有投资难以复用,具体操作困难较大。而基于一对多星型结构的政府数据交换共享平台,具备多级管控的数据交换体系,通过以交换域为单位的管理模式,为各委办单位提供跨部门的数据交换和管控的能力,从而有效节约软硬件投入,实现统一的交换共享管理。

2.3 政府数据开放平台

公众数据使用意识和能力的提升,企业数据处理和开发的多样化,以及全球开放政府趋势的加快,使得建立一个透明、开放、民主的数据政府成为各国政府改革创新、提高国际竞争力的重要目标之一。通过政府数据开放平台,用于企业和个人下载和利用,利用外部技术优势,可以将政府数据的潜在价值转变为实际效

益,实现价值增值和创新应用。

从全球经验来看,政府数据开放平台往往采用专门数据网站的形式,以满足社会公众对数据访问和下载的需求^[6]。作为全球第一个政府数据开放平台——美国 data.gov 自 2009 年上线以来,在全球引发了政府数据开放平台建设的热潮,英国、加拿大、澳大利亚、新加坡、韩国等国家陆续建立了自己的政府数据开放平台。2013 年八国集团(C8)签署的《开放数据宪章》承诺“用可以重复使用的格式,及时主动地向社会开放高质量的信息,包括原始的数据”,该宪章成为世界各国推动开放政府数据的基本依据。相关的开放政府数据评估项目也紧随其后,如世界银行的“开放数据准备度”(Open Data Readiness Assessment)、万维网基金会的“开放数据晴雨表”(Open Data Barometer)、开放知识基金会的“全球开放数据指数”(Global Open Data Index)、经济合作与发展组织(OECD)的“开放政府数据指数”以及联合国的“开放政府数据调查”^[7]。我国自 2011 年起,在北京、上海等地区率先开启了政府数据开放的工作,有学者调查发现,目前我国已建设了 13 个地方性政府数据开放平台^[8]。2015 年 9 月 5 日,国务院颁发了《国务院关于印发促进大数据发展行动纲要的通知》(国发 2015[50]号),该通知明确提出,我国要在 2018 年底前建成国家政府数据统一开放平台^[9]。

2.4 政府数据整合服务平台

传统的政府服务数据缺乏有效的整合机制,不同业务部门甚至同一个业务部门内部由于服务数据的独立性,造成了业务服务功能的分散,致使业务服务步骤繁琐,如开展计生服务工作,需要分别从公安局获取人口出生数据和死亡数据,从民政局获取流动人口数据,从社保局获取医保数据,从教育局获取学生数据,这种状况降低了政府自身的服务效率,增加了协调成本,也给服务对象造成了时间和精力上的负担。

从数据整合层面来看,通过构建统一的政府数据整合服务平台,可以实现不同服务数据特征的整合以及不同服务数据标准的兼容,实现从分散异构数据到综合信息的转变。例如某人的特征信息分散在不同的系统中,同一特征信息具有不同的数据表示方法和标准,如性别可表示为:“男”、“M”、“1”等,通过数据整合,构建性别代码维度,实现数据间共享和互操作;还可利用数据画像技术刻画某人的不同特征信息,如教育情况、参保情况、生育情况、婚姻情况、教育情况等,从而

实现个人数据到个人综合信息的转变。

从业务整合层面来看,以服务数据的整合为基础,政府数据整合服务平台可以整合不同的服务事项,如技术服务、劳动保障、民政业务、房屋租赁、法律援助、城管业务等,提供统一服务,由“指定窗口”转变为“综合窗口”;实现需要同一业务数据的不同服务之间的数据共享,如民政服务、劳动业务、残联业务可以共享计生业务中的计生证明,无需多次开具证明,大大提高政务服务效率,降低了服务协调成本。

3 核心主体:政府数据集成平台

政务云数据中心实现了政府业务数据的集中存储;政府数据开放平台实现了外部对政府数据的增值开发利用;政府数据交换整合平台实现了政府内部各部门之间的数据共享;政府数据整合服务平台实现了政府综合服务数据以信息、产品的形式提供社会利用。政府数据集成平台则既整合了政务云数据中心、政府数据开放平台、政府数据共享平台和政府数据整合服务平台的核心功能,又具有四者所缺乏的功能和协调性,融入了物联网技术、云计算技术、大数据技术等先进技术,实现了整体功能大于部分之和,发挥统一整体的优势。

3.1 集成平台构建的必要性

构建基于政务云数据中心的政府数据开放共享服务集成平台既是政务数据存储、开放、共享和服务功能集成的要求,也是政务大数据创新应用、智慧政务和互联网+政务等大趋势的要求。

3.1.1 数据的储存、开放、共享和服务功能集成的要求

数据的存储是后续数据开放、共享和服务的基础,数据的开放、共享和服务则是数据价值发挥的三个方向。上升到平台层面,四者对应了不同数据平台的不同功能。以海量政府数据集中存储为基础,快速识别政府数据的用途和功能,以统一的技术体系、标准模式和接口方式,实现政务大数据的规范、实时、安全、高效的分类分级利用,已经成为大数据时代提高政府数据处理能力和业务服务效率的重要目标。通过构建以政务云存储为基础的政府数据开放共享服务集成平台,可以有效整合数据的存储、开放、共享和服务等功能,同时利用统一的云技术、大数据技术、网络技术,可实现数据分布式存储、数据预处理、数据实时调用、数据脱敏、数据分类分级、数据统一调度、数据深度挖掘、数

据可视化呈现等。

3.1.2 政务大数据创新应用、智慧政务和互联网+政务趋势的要求

大数据的创新应用必须解决政府数据开放、共享和服务过程中涉及的一系列问题,包括如何解决共享数据的冗余性,如何解决开放数据的安全隐私保护问题,如何解决数据的一致性、及时性和稳定性问题,以及如何实现服务数据的分析挖掘和再利用问题。利用统一的标准规范、技术手段和安全措施,政府数据开放共享服务集成平台无疑是集中解决这些问题的最佳方式。

智慧政务强调开放政府、互动沟通、高效协同和一站式服务,政府数据作为政府实现各项业务功能的支撑,是实现智慧政务理念的基础,因此有必要构建一个集中的平台来管理政府数据。政府数据开放共享服务集成平台能够有效的整合政府数据,实现政府数据的透明开放、增值利用,达到内部各部门的有序互联、有效共享以及各类跨部门的业务联动、整合服务,打造智慧政务的“数据航母”。

互联网+政务基于“开放、互动、参与、融合”互联网思维,要求打破政府部门之间的数据鸿沟,搭建“以人为本、市民参与、社会协同”的公共新型服务平台,通过互联网与政务数据的融合,实现政务数据资源的统一协调与共享。这一思维与构建政府数据开放共享服务集成平台的理念和宗旨一致,是实现互联网+政务的重要途径。

3.2 集成平台的构建

“平台”一词所表达的含义有三个方面:(1)它是一种具有基础性的支撑体系;(2)它不仅包括硬件,而且包括软件,是硬件和软件的有机集成;(3)它具有公共和共用的服务性质^[10]。因此,政府数据开放共享服务集成平台是一个基于政务云数据中心,集成了政府数据开放、政府数据共享和政府数据整合服务等功能,以国家大数据标准规范及管理制度作为法规标准体系,以云计算、物联网、大数据技术等作为技术体系,包括各类硬件和软件设施,能提供多种应用入口,满足政府、企业和社会对政府数据多样化需求,具有基础支撑性的信息基础设施。

3.2.1 顶层设计

整个平台系统由三大结构层次和两个支持体系组成,如图2所示。政务云数据中心负责数据的收集和存



图2 政府数据开放共享服务集成平台顶层设计

储,由各大基础设施和网络系统构成;集成平台是整个系统平台的功能核心,集成了数据的开放、交换共享以及整合服务功能;应用入口则是集成平台各功能的具体实现途径,包括面向公众的门户网站入口、移动应用入口、政务大厅入口以及各级政府专用入口。整个平台系统建立在标准制度体系和技术体系两大支撑体系之上,其中标准制度体系即国家大数据标准规范及管理制度体系,包括基础术语和参考架构、平台管理指南、数据和工具标准、安全和隐私规范,目前工信部正在重点推进大数据标准体系的制定^[1];技术体系包括应用不同系统层次的关键技术,如物联网、移动互联、云存储、大数据技术等。

3.2.2 功能拓展

整个平台系统不是原有的各个系统功能的简单相加,而是在原有功能集成基础上进行了一系列的拓展。功能拓展的实现需要各种技术的配合,包括物联网技术、云存储技术、大数据技术、移动互联网技术等。

政务云数据中心的功能拓展。在集成平台中,相较于传统的收集和存储功能,政务云数据中心具有更多的实时数据预处理功能、网络数据抓取、多接口数据分类功能。由于政务云数据中心不仅收集来自各个政务部门的业务数据,还包括来自物联网设备的大量实时的数据,这就要求政务云数据中心提供更为完备的数据仓储,具备对各类流数据的预处理功能,以提高数据的质量,减少后期其他平台系统模块的数据处理负担。同时,系统平台中的政务云数据中心还具备对互

联网数据的抓取功能,能够及时准确地采集公众民情数据资源。同时,政务云数据中心还担负着向上层集成平台各个功能模块提供数据的责任,所以需要具备应对多个接口实现多种数据的分类分级功能,例如对开放数据、内部共享数据、服务数据的分流,并拥有高速的数据传输存取能力。

集成平台功能模块的拓展。平台系统的拓展性集中表现在集成平台三大功能,基于统一管理调度以及大数据技术所实现的功能拓展。数据开放功能的拓展在于依托集成平台的大数据技术,结合了数据清洗、数据脱敏、开放格式等功能,相较于传统的直接开放原始数据,集成平台实现了高质量的数据开放,对原始数据进行了更为精细的处理,包括对冗余数据的处理,对涉及安全和隐私的数据进行脱敏,对各类数据,特别是非结构化的数据进行格式化转换,采用更为开放的格式,提高数据对各个应用系统的兼容性;数据交换共享功能的拓展表现在对各个来源的数据进行关联比对、格式转换,政务云数据中心将用于政府内部交换共享的数据经过预处理后,传输至集成平台,由交换共享模块实现数据的关联比对,以便将存在重复性的各种来源数据进行核实重组,减小数据的冗余性,并进行格式转换,以便提供给政府内部各系统进行共享;数据整合服务功能的拓展主要表现在依托集成平台,实现数据的分析挖掘,包括关联分析,分析比对,对海量数据进行多维度深度分析和融合,深化对政务数据分析结果的应用,例如,政府利用大数据对社会人群进行细分,

对不同人群进行针对性服务^[12]。同时可采用可视化技术,简化复杂的政务数据,以形象生动的图像形式呈现,例如,对实时的交通数据以动态的实况分布图向公众及时展示道路交通流量状况、道路事故情况。集成平台还具备系统基础设施数据监测与统筹功能,能提供全面的数据热度统计和运行环境及状态监控。

应用入口类型的拓展。不同于以往的政务数据应用入口,平台系统具备多种类型的数据入口类型,可实现不同功能数据的应用管理。各级政府专用入口可将用于政府内部交换共享的数据以安全可靠的接口方式提供利用,主要面向政府专用网络,具备高速的数据传输速率和安全防护功能;政府数据的开放不同于传统的以门户网站为主的数据提供途径,而是采用移动互联技术,实现网站门户和移动应用的结合,以更加便捷的方式向公众提供政府数据;政务服务除了传统的由政务中心工作人员通过专门的业务系统向公民提供线下咨询,还具备让公众通过政务服务网站、移动应用端口等入口进行线上的政务信息查询、业务办理、信息反馈等。

3.2.3 政府数据生命周期全程管理的实现

通过政务云数据中心的政府数据开放共享服务集成平台,能实现政务数据采集、存储、分析、利用的全生命周期管理,如图3所示。

政务云中心实现了政务数据的采集和存储,对来自PC、移动等设备终端的网络数据进行采集,并存储为社会、灾害、舆情等数据资源库;对来自各个物联网监控设备的实时监测数据进行采集,存储为视频、气象、环保、交通、位置等实时的监测数据资源库;对来自各级政府委办局内部数据库的业务数据进行分类存储,包括空间、人口、法人、建筑、经济以及专题数据等基础的政府业务数据。

集成平台实现了对政务云数据中心所存储得到的数据的进一步处理,按照数据不同的用途,对数据进行分类处理,包括对开放数据进行清洗和脱敏;对共享数据进行关联比对和格式转换、对服务数据进行融合与挖掘。

各大应用入口则实现了政务数据的利用,以多种形式和途径提供政务数据的开放、共享和服务。



图3 政务数据生命周期全过程管理的实现

4 结语

无论是政府治理的科学化、精细化,还是公共服务水平和能力的全面提升,政务大数据都还有非常大的价值潜力可以挖掘。随着政府数据组织的不断规范化、采集的不断精准化、开放规模的不断扩大化,政务大数据的服务和应用也将变得更具特色和深度。围绕政务数据的存储、开放、共享和服务,与之配套的各项基础设施臻于完善,但从政府数据生命周期的整个流程出发,构建统一的标准基础设施平台还处于起步阶段,基于政府云数据中心的政府数据开放共享服务集成平台无疑是实现这一目标的最佳途径,而政府数据开放共享服务集成平台的技术实现将是下一步研究的重点。

参考文献

- [1] Big Data, the Next Frontier for Innovation, Competition, and Productivity [EB/OL]. [2016-12-05]. http://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx.
- [2] 推进政务信息资源共享管理[EB/OL]. [2016-12-05]. <http://politics.people.com.cn/n1/2016/0920/c1001-28725276.html>.
- [3] 刘志强. 云存储推动云安防落地[J]. 中国公共安全:学术版, 2014(9):62-64.
- [4] 陈培忠. 区域科技平台建设中资源有效整合的思路与方法研究[J]. 科技情报开发与经济, 2007, 17(29):224-225.
- [5] 国务院关于印发政务信息资源共享管理暂行办法的通知[EB/OL]. [2016-12-05]. http://news.xinhuanet.com/politics/2016-09/19/c_129288099.htm.
- [6] 黄思棉, 张燕华. 当前中国政府数据开放平台建设存在的问题与对策研究——以北京、上海政府数据开放网站为例[J]. 中国管理信息化, 2015(14):175-177.
- [7] 郑跃平, 刘美岑. 开放数据评估的现状存在问题——基于国外开放数据评估的对比和分析[J]. 电子政务, 2016(8):84-93.
- [8] 黄如花, 王春迎. 我国政府数据开放平台现状调查与分析[J]. 情报理论与实践, 2016, 39(7):50-55.
- [9] 国务院印发促进大数据发展行动纲要[EB/OL]. [2016-12-10]. <http://politics.people.com.cn/n/2015/0905/c1001-27545655.html>.
- [10] 康仲如. 试论标准及标准化科技基础平台的作用及意义[J]. 轻工标准与质量, 2005(1):39-40.
- [11] 全国信标委:国家大数据标准化工作进展[EB/OL]. [2016-12-20]. <http://www.useit.com.cn/thread-14161-1-1.html>.
- [12] 丁健. 浅析大数据对政府2.0的推进作用[J]. 中国信息界, 2012(9):12-14.

【作者简介】唐长乐,男,1991年生,武汉大学信息管理学院博士研究生。

王春迎,女,1992年生,武汉大学信息管理学院硕士研究生。

收稿日期:2017-04-17

欢迎订阅

复印报刊资料《出版业》

《出版业》由中国人民大学书报资料中心编辑出版,精选国内千家报刊,荟萃国际学术精华。刊物关注国家出版工作的方针政策;反映出版业界的改革与发展;精选出版理论研究方面的优秀学术文章;研讨各类图书与期刊的编辑、营销、发行等工作以及国外出版方面的研究成果,理论性与应用性并重,是出版业内外同仁交流、收获的平台。

栏目主要有:发展战略、学术平台、业界观察、数字出版、出版法苑、书业经纬、期刊纵横、编校实务、出版史论、环球出版、探索与争鸣等。

《出版业》为月刊,大16开,72页,国内外公开发行。每期国内定价15.00元。

编辑部地址:北京市海淀区中关村大街甲59号文化大厦1301室

编辑部电话:010-82504550-2165

营销热线:010-82503412、82503440、82503029

E-mail:wsbjs@zlzx.org

网址:www.zlzx.org