

# 南京图书馆：数据驱动 共建智慧

## 一、建设背景

当前，信息化正在开启以数据深度挖掘和融合应用为主要特征的智能化阶段，大数据、云计算、移动互联网、物联网、人工智能等新一代信息技术的广泛应用推动着社会向智慧化发展，也深刻影响着图书馆智慧化发展方向。继 2015 年国务院《关于促进大数据发展的行动纲要》颁布以来，大数据在公共文化行业中的研究和应用逐步得到推进。2017 年，公共文化服务大数据应用文化部重点实验室在北京成立。2019 年，国家科技部等六部门印发《关于促进文化和科技深度融合的指导意见》的通知，对文化大数据体系建设提出了明确要求，指出公共文化机构应建设“物理分散、逻辑集中、政企互通、事企互联、数据共享、安全可信的文化大数据体系”。为贯彻国家大数据战略，2015 年江苏省颁布了《江苏省大数据发展行动计划》，江苏省文旅厅在“十三五”规划中明确提出“加强公共文化大数据采集、存储和分析处理，提高公共文化服务的针对性和有效性”，要求加强文化大数据体系建设，推进包含公共图书馆大数据服务平台在内的文化各领域大数据平台建设。这些规划和政策文件为全面推进图书馆大数据建设提供了政策保障和方向指导。

近几年来，我国许多图书馆都在建立“大数据”系统，这些系统基本上是在单馆若干应用系统基础上的数据统计和数据可视化展示系统。从大数据的定义、技术特点和价值意义来看，这些都不能算是真正意义上的大数据和大数据分析。

众所周知,大数据以规模性 (Volume)、高速性 (Velocity)、多样性 (Variety)、价值性 (Value) 为基本特点。图书馆大数据也应该遵循大数据基本规律,具备这些基本特点。图书馆大数据在数据规模上应该是多应用系统、多个馆、全区域甚至是跨区域、跨行业的全量数据;在技术层面上应该具备面向大数据处理的数据层、算法层、应用开发层的大数据技术特点,需要涉及数据采集、数据存储、数据分析、数据处理、数据治理、系统实现、系统运维等各个功能系统;在应用价值方面,大数据应该针对不同的人群 (行业管理层、图书馆决策层、业务建设层、读者服务层) 和智慧功能系统提供针对性的精准化数据服务,应该能以新的角度和新的手段全方位、全视角展现图书馆运行和服务的演化历史、当前状态,全局态势和细微差别,能归纳图书馆要素发展和要素相互作用的客观规律,预判图书馆发展趋势和未来状态。在 2018 年之前这种真正意义上的图书馆大数据平台系统在国内还是空白,为此,有必要开展图书馆大数据建设,建设目的是为智慧图书馆体系建设提供强有力的数据驱动,为图书馆事业高质量发展提供全面有效的数字服务。

全面的计算机管理、网络化应用和数字化服务是大数据发展的客观基础和现实条件。近 10 年来,随着社会经济的快速发展,江苏省公共图书馆事业也同步快速发展,信息化数字化水平普遍提升。2017 年全国第六次公共图书馆评估定级中,全省 109 家参评馆中的 100 家被评定为一级馆,总数和比例均名列全国行政省份第一,全省 109 家公共图书馆均实现了图书馆业务管理系统,并普遍建立了数字图书馆。南京图书馆作为江苏省省馆、全省公共图书馆的中心馆,在省文旅厅的指导和支持下秉持“统筹统建、共建共享”的理念,在全省新技术应用、平台搭建、资源服务等方面积极作为、勇于创新,先后创建了“江苏省公共图书馆联合参考咨询网”“江苏省少儿数字图书馆”“江苏图书馆云服务”“百馆荐书全省共读”等一批具有全国影响力的省级公共数字文化服务平台,建立起“省厅指导、省馆牵头、全省参与”的协作协调机制,形成了全省聚力合作、共同发展、全省一盘棋的良好氛围。这些为全省公共图书馆大数据建设打下了坚实的现实基础。

## 二、发展历程

从 2015 年开展图书馆大数据应用基础研究到现在，江苏省公共图书馆大数据建设经历了 3 个发展阶段。

第一阶段为前期基础研究和探索阶段（2015—2017 年）。2015—2016 年南京图书馆承担并完成了全国文化信息资源共享工程“十三五”发展规划重点方向课题“公共文化服务大数据的采集与分析研究”，该课题以公共图书馆等公共文化服务机构为主要研究对象，对大数据的来源和构成、采集与处理、分析与应用进行了初步的研究，形成了“公共文化服务大数据的来源、采集与分析研究”报告，相关论文发表在《图书馆建设》2015 年第 11 期上，课题研究成果为开展大数据服务平台建设提供了理论基础。2017 年，南京图书馆被列入“公共文化服务大数据应用文化部重点实验室实践基地”，开始调研并起草“江苏省公共图书馆大数据采集与分析系统建设方案”。

第二阶段为基础平台开发和数据中心初建阶段（2018—2020 年）。2018 年 2 月“江苏省公共图书馆大数据采集及分析系统建设项目”方案论证会召开。方案提出了“拟采集全省公共图书馆业务数据、数字资源服务数据、馆情指标数据、公共数字文化工程数据，初步建立全省公共图书馆大数据中心，并在此基础上实现数据统计分析、数据查询展现等数据应用功能”。来自北京、上海、江苏的专家在对方案内容基本肯定的同时，指出江苏省大数据建设是国内首家省级规模的公共文化大数据项目，是一个复杂的、庞大的、长期的系统工程，需要分期建设，第一期项目宜实现部分数据的采集处理和统计展示，工程名称可定为“江苏省公共图书馆系统公共服务大数据应用示范工程”，经过与文旅厅进一步研究协商，最终将工程名称定为“江苏省公共图书馆大数据服务平台”。2018 年 6 月，项目一期“江苏省公共图书馆大数据采集、存储及提供系统”全面启动，2019 年 11 月底完成了业务管理系统、数据采集管理系统、馆情数据填报系统、数据统计展示系统的软件开发和全省 75% 的公共图书馆数据采集工作。2020 年 11 月，在一期项目的基础上开始建设全省关键指标排名发布系统、多维度统计数据服务系统、馆情数据查询系统、读者活动数据采集系统、读者流量数据采集系统，数据

采集规模进一步扩大和完善，全省大数据中心初步建成。

第三阶段为全面建设和深化服务阶段（2021 年至今）。2020 年底，在大数据服务平台和数据中心建设取得初步成效的基础上，“图书馆大数据应用江苏省文化和旅游重点实验室”（以下简称“重点实验室”）获得江苏省文化和旅游厅首批重点实验室认定。2011 年 4 月 28 日，南京图书馆举行了“图书馆大数据应用与智慧服务研讨会暨重点实验室揭牌仪式”，开启了江苏省图书馆大数据全面建设的新阶段。重点实验室承担着图书馆大数据建设的研究、开发、管理和服务等各项工作，其成立为图书馆大数据建设提供了更有利的发展条件和更好的发展前景，特别是在智慧图书馆体系建设中将能更好地发挥大数据驱动作用。

重点实验室由江苏省文化和旅游厅主管，以南京图书馆为依托单位、南京大学信息管理学院为共建单位，设主任一名主持实验室工作，成立学术委员会对实验室进行学术指导，下设综合管理组、应用研究组、数据管理组和实践基地负责开展具体工作。组织架构如图 3-8 所示。

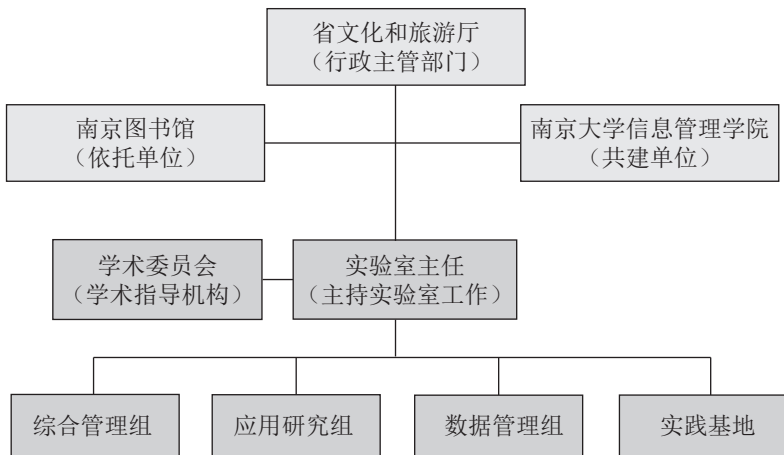


图 3-8 图书馆大数据应用江苏省文化和旅游重点实验室组织架构图

### 三、建设内容

江苏省公共图书馆大数据建设具有全方位、综合化和长期性等特点，根据三

年来的实践，建设内容可概括为：策划组织协调、应用基础研究、软件体系搭建、数据中心建设、数据应用服务、业务人才培养等 6 个方面。

### 1. 策划组织协调

江苏省公共图书馆大数据建设是一项全省性的工程，需要得到各级文化主管部门的重视和支持，需要省、市、县各级图书馆的全面参与和乡镇、村基层综合文化服务中心的积极配合，宣传发动和组织协调工作十分重要。江苏省文化和旅游厅高度重视这项工作，2018 年 11 月向全省下发《关于开展江苏省公共图书馆大数据服务平台建设的通知》，提出了建设江苏省公共图书馆大数据服务平台、建立大数据中心的目标任务，明确了“大数据服务平台是一个全省参与、共享成果的重大公共文化建设项目。项目由省文化和旅游厅主导建设，南京图书馆具体负责平台搭建和运行维护。全省各地文化主管部门组织本地区各级图书馆对接南京图书馆，开展大数据服务平台建设和共享服务”。从 2017 年起，每年两次的全省公共图书馆公共数字文化培训班都把大数据建设作为重点培训内容，对大数据建设的意义、平台功能、服务项目进行宣传，对各个图书馆的工作提出具体要求。

全省组织协调的重点和难点是各馆馆情数据填报和业务系统数据采集工作，各个图书馆十分关注读者数据隐私、各类数据安全、统计分析结果用途等核心问题。为使各图书馆能够放心提供数据，省文旅厅在通知中明确了“大数据分析结果仅用于本省各级文化主管部门和图书馆掌握馆情、科学决策、改进工作、提升服务的目的，不作为评比考核、评估定级的依据。大数据中心采集的源数据、清洗后的标准数据仅用于本平台服务，不支持其他任何目的和用途的数据复制、下载和调用。数据采集、传输、存储和利用等各个环节均采取严格的数据安全保护措施”。同时南京图书馆还与全省所有公共图书馆签订《江苏省公共图书馆大数据共建共享协议书》，就大数据的提供、管理、使用和服务达成协议，明确了双方的责任和义务，保证了双方的权益。通过一系列措施使得全省数据采集工作得以顺利进行，到目前为止，全省 117 家图书馆除少数图书馆因技术原因暂时提供不了数据外，所有的图书馆都能正常提供数据。

## 2. 应用基础研究

江苏省图书馆大数据建设是一项开创性的工作，为保证图书馆大数据建设的科学性、先进性和实用性，需要对数据源、数据采集和清洗、数据分析和服务等各方面进行全面而系统的研究。如果说2016年完成的全国文化信息资源共享工程“十三五”发展规划重点方向课题“公共文化服务大数据的采集与分析研究”是对图书馆大数据的初步的研究，那么依托重点实验室开展的研究将是全面而系统的研究。重点实验室一方面积极组织申报与国家大数据相关的科研课题，如2021年承担了省文旅厅立项的“数字资源服务大数据采集与分析研究”课题；另一方面，自主设立开放课题，并以江苏省图书馆学会和南京图书馆的名义联合向社会公开招标。自主设立的课题以图书馆大数据建设实际需求和难点难题为导向，以应用于大数据软件体系和数据中心建设、数据统计分析服务、智慧图书馆体系建设等为目的，具有很强的应用价值和现实意义，2021年度设立了如下两大类16个课题。

第一类应用基础研究类课题重点研究图书馆大数据建设和治理、大数据统计分析、大数据融合应用等方面的方法、标准和技术问题，具体设立了图书馆保障条件大数据采集指标体系研究、图书馆活动系统数据采集指标研究、图书馆读者流量和热点数据采集指标研究、图书馆外部环境影响因素及数据源研究、数据统计分析方法在图书馆大数据统计分析中的应用研究、数据智能分析方法在图书馆大数据统计分析中的应用研究、数据可视化技术在图书馆大数据统计分析中的应用研究等7个课题。此类课题拟采用面向全省发布和专家评审的方式确定符合条件的课题承担人。

第二类数据分析研究类课题侧重于大数据统计数据的深度挖掘和灵活运用，将全省公共图书馆大数据的5亿条数据开放出来，组织发动图书馆行业从业人员和研究人员基于大数据开展图书馆业务研究，具体设立了公共图书馆读者结构及阅读行为分析、公共图书馆读者活动分析、公共图书馆馆藏结构及文献利用分析、公共图书馆保障条件分析、公共图书馆服务效能分析、公共图书馆文献供需平衡与文献采访规则分析、公共图书馆服务体系分析、公共图书馆少儿服务工作分析、图书馆发展综合性调研报告等9个课题。此类课题拟直接委托省内具备条

件的公共图书馆或科研单位承担。

应用基础研究的全面开展将促进大数据项目建设的质量，同时也活跃了全省公共图书馆科研氛围，提升了图书馆工作人员的科研能力。

### 3. 软件体系搭建

开发和完善图书馆大数据应用软件体系是大数据建设的核心任务，软件体系必须包含各种数据源的数据采集、存储、清洗、展示、统计、分析、接口服务等一系列功能系统，江苏省公共图书馆大数据软件体系建设采用总体规划分步实施的方针，计划用 4—5 年的时间分阶段进行功能开发和迭代完善。2019 年底完成第一期软件项目，2021 年 7 月完成第二期软件项目，2021 年 9 月开始第三期软件项目开发。

#### （1）主要技术路线

软件体系采用市场占有率和认知度最高的 Hadoop 分布式系统基础技术架构，以 HDFS 和 MapReduce 为核心的一整套数据存储和计算处理工具，很好地解决了图书馆大数据项目实施的技术问题。采集的数据以 HDFS 分布式文件系统为基础，存储于 HBase 高性能、面向列分布式数据库中；利用 Hive 对抽取的各系统数据内容进行计算，通过 Impala 提高大数据的计算效率；采用 Spark 进行内存级的数据批量处理；应用 Spark Streaming 与相关业务系统进行衔接，保证大规模流式数据处理系统的高效能力。对存储的大数据根据业务主题进行多维度的数据分析与挖掘，对资源、读者、行为以及馆情指标数据进行关联分析、个性分析和多维数据分析。第一、二期软件系统技术路线示意图和软件系统结构示意图如图 3-9、图 3-10。

#### （2）数据采集和清洗

数据采集和清洗工作由一组针对各种不同数据源开发的数据采集系统完成，第一、二期项目的数据源包括图书馆业务管理系统数据、馆情填报数据、图书馆活动数据、读者流量数据等四类（图 3-11）。

业务数据采集系统承担着采集全省 117 家图书馆的不同系统和版本的业务管理系统数据的任务，这些系统包括但不限于力博、图创、汇文、图星、ALEPH 等，采集系统可选择多种采集方式（全库采集、批量采集、增量采集）和灵活的



采集策略（数据对象、间隔时长、起止时间等）进行采集。采集系统采用数据库连接、ETL 数据抽取方式获取数据，先将采集到的数据存储在 Hive 数据仓库中的镜像库中，镜像库中的数据结构与源数据完全相同，且能实现数据同步；再从镜像库中通过 SQL 查询提取所需要的数据存入 HBase 的一个临时文件中（TMPA）等待清洗；清洗的过程是一个将来自不同图书馆不同系统的数据实现规范统一的过程，如馆藏复本与标准书目元数据的比对挂接、读者属性数据的规范统一、识别和建立同一人不同记录的关联、读者借阅信息的规范统一等；清洗后的数据暂存在各个图书馆的汇总中间库中，最后将所有图书馆中间库数据汇总到大数据总库。数据总库拥有能够包含和兼容所有系统数据的数据结构，保证各系统数据的无损转换和汇总。采集系统为满足更多维度和更复杂的大数据统计分析需求，提供更精准、更全面的数据服务。

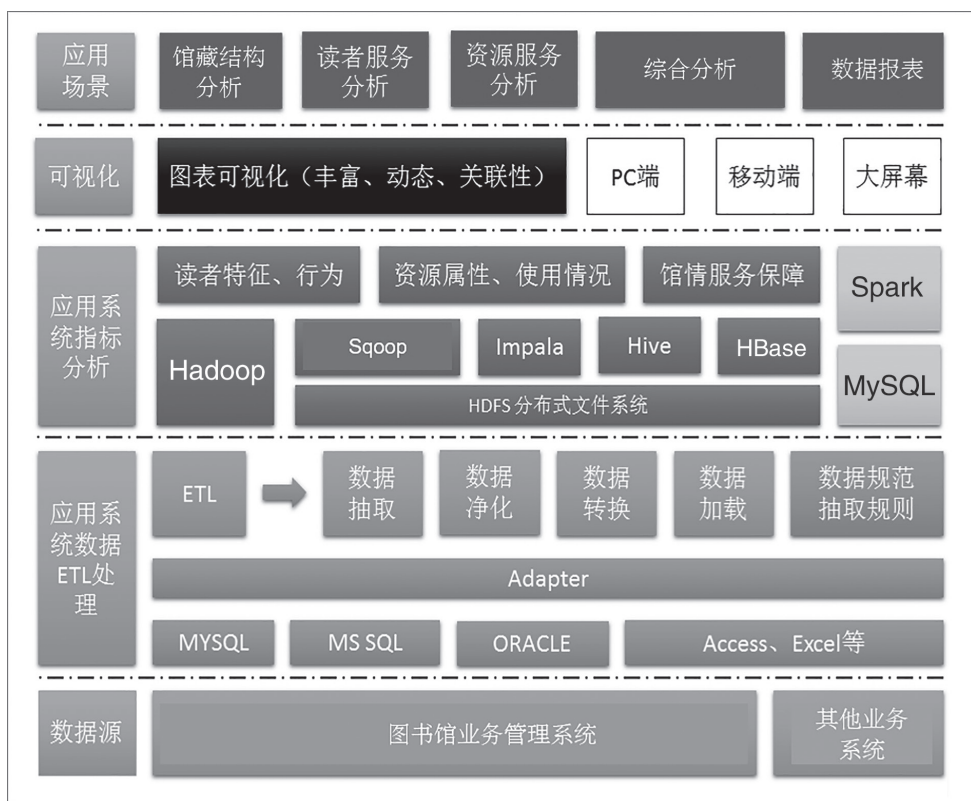


图 3-9 第一、二期软件系统技术路线示意图



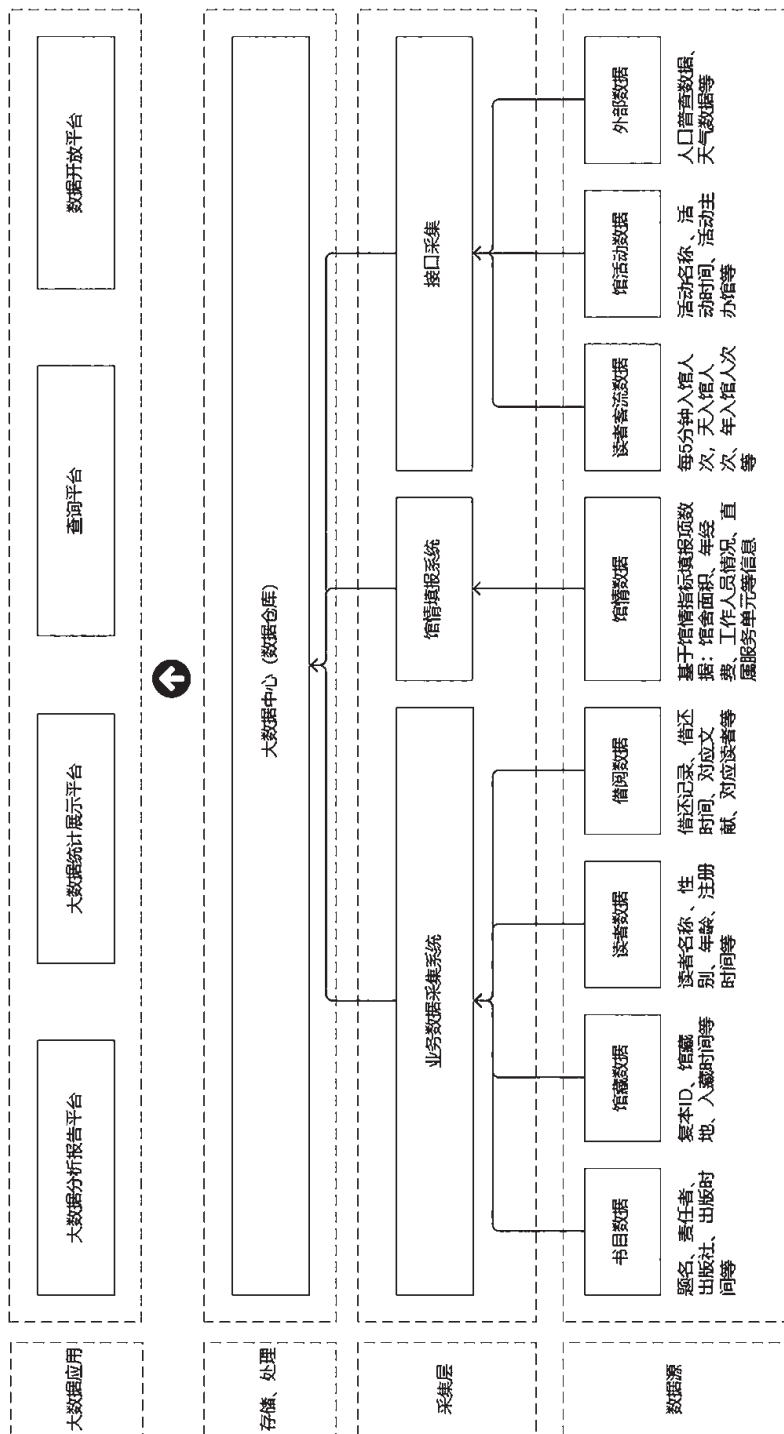


图 3-10 第一、二期软件系统结构示意图

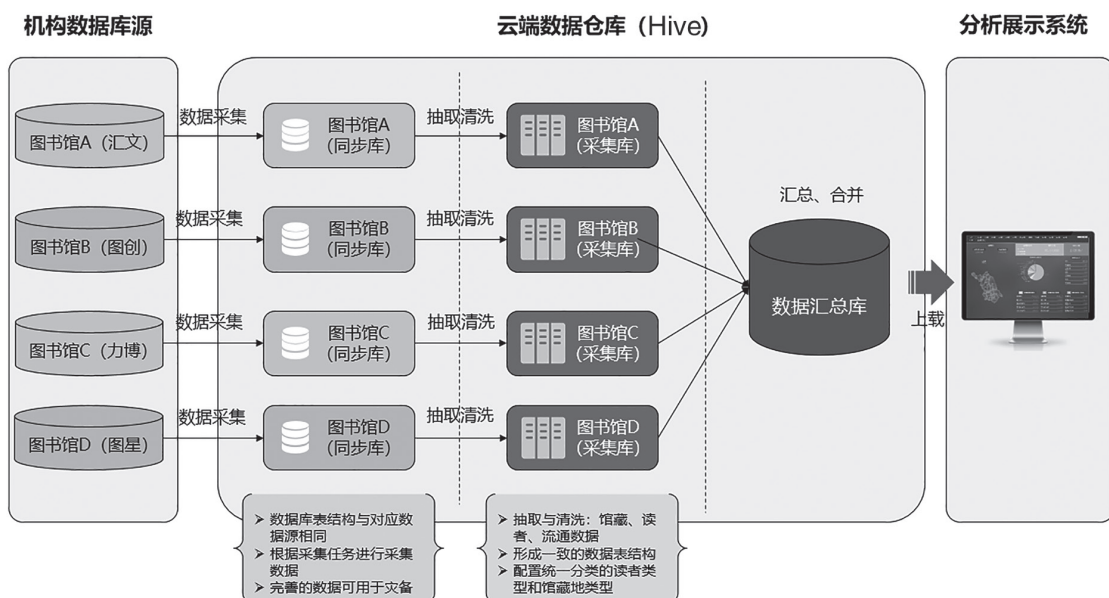


图 3-11 业务系统数据采集和清洗示意图

馆情填报系统一共设置了综合信息、详细信息、直属服务单元信息、乡镇街道基层图书馆、村级社区村图书室、合作分馆和流通点等六大类 392 个填报指标项，是一种全层次全方位摸家底式的情况调研，每年 3 月份前发动组织一次由 117 家省、市、县（区）公共图书馆，1200 多家乡镇（街道）图书馆参与的全省填报工作，15000 多个村级社区图书室的数据填报在县级图书馆指导协助下由乡镇图书馆完成。系统包含用户管理、角色管理、部门管理、菜单管理、定时任务、参数管理、字典管理、文件上传、系统日志等功能。系统基于主流开源的 **Spring Boot** 来搭建开发环境，系统提供的业务功能服务统一以异步请求接口方式实现，便于实现多终端界面的展现和适配，同时支持提供灵活的权限控制，可控制到页面或按钮，满足绝大部分的权限控制需求，安全层面提供完善的 **XSS** 防范及脚本过滤，杜绝 **XSS** 攻击。

客流量数据采集系统从多种客流量系统中实时采集读者流量数据，对数据进行清洗，实现规范统一。客流量系统的数据结构相对比较简单，但数据采集的实时性要求比较高，采集频率以秒为单位。江苏省公共图书馆主要使用了两种客流量系统，每个系统都以云服务模式部署在省馆大数据中心，数据采集和清洗简单而便捷。

活动数据采集系统从多种读者活动系统中采集活动数据，对数据进行清洗，实现规范统一。图书馆活动特别是阅读推广活动已成为图书馆的主要工作和核心业务，活动的类型包括线上线下的阅读推广、培训、讲座、展览等。江苏省公共图书馆使用省馆统一提供的活动管理系统，目前大数据采集和分析还比较简单，下一步将对活动的类型、主题、责任者、方式、地点、环境条件、参与对象、时间（时长）、内容资源、规模、效果、评价等信息进行全面数据采集和多维度统计分析。

### （3）数据统计展示系统

系统基于图书馆填报数据与业务系统采集数据，从资源、读者、流通与综合 4 个方面，对全省、地市以及单馆进行多维度数据统计和可视化展示。系统结合实际业务对存储在 HBase、Impala、MySQL、SolrCloud 等各数据服务中的结构和关系进行定义，以满足业务数据模型层各个应用的需求。系统通过数据接口从大数据总库中提取数据存储到 HBase 中，再使用 Spark 将数据传输到 Impala、SolrCloud，Impala 提供平台的数据分析任务，SolrCloud 提供系统的全文检索任务，MySQL 数据库用于用户和权限管理。地市主管部门和各区县级以上图书馆均可使用独立账号进行访问（图 3-12）。

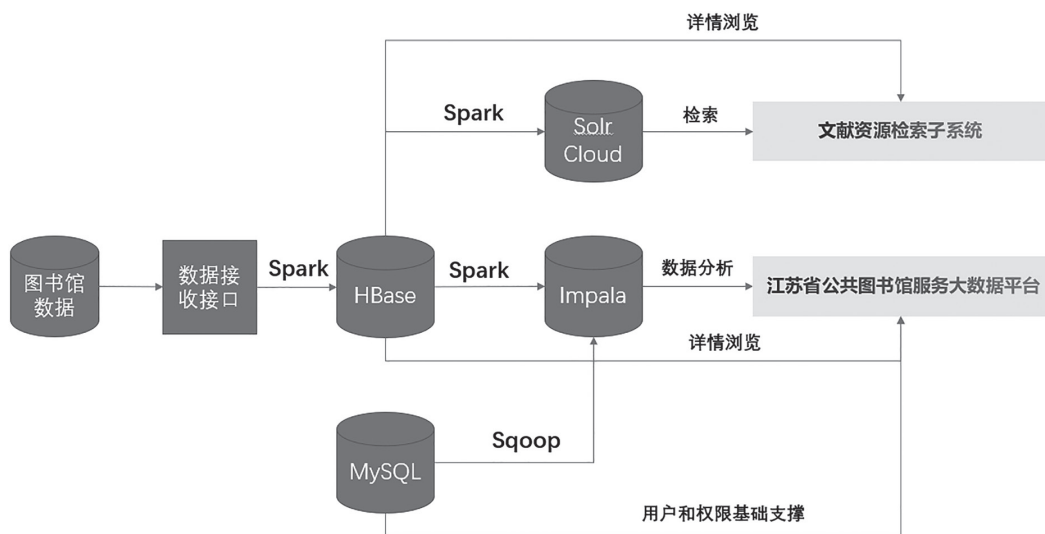


图 3-12 数据统计展示系统技术路线示意图

系统可通过省、市、县、区、乡镇/街道、村/社区六级行政区划查询分析

图书馆发展状况、保障条件及服务效能情况。对于图书馆用户，可选择本馆、合作分馆、基层分馆作为研究对象，进行详细业务数据分析。分析粒度向下可钻取到各级分馆、馆内服务单元、流动服务单元，也可细化到各馆藏地。针对图书馆业务，平台重点进行了读者分析、资源分析、流通分析、综合分析、特征画像，并提供标签化设计及管理。

资源分析包括各图书馆资源发展、资源结构、资源标签的分析，并在资源特征分析的基础上，进行资源个体画像和群体画像。结合读者对资源的访问行为以及借阅行为进行历史分析，提供以资源的访问量、借阅量、借阅率、热点资源为代表的资源访问汇总分析结果，为资源采购、资源服务布局和资源优化调整提供一定的数据依据。系统对个体读者数据进行了属性信息、行为信息、活跃程度、读者价值分析，从读者的性别、年龄、籍贯等自然属性，读者类别、读者状态等图书馆属性特征，以及资源阅读特征、阅读偏好、检索行为、借阅行为等行为信息数据进行分析挖掘，进行了个体用户画像和群体用户画像。系统采用数据推荐与挖掘算法，根据用户类型判断相关文献资源的吸引力、关注度。根据读者行为找到其他相似行为的读者群体开展分析，并根据业务主题范畴对资源、业务进行多维度的分析。

此外，分析系统设计了标签化体系，可对读者、资源通过数据分析形成的结果，进行自定义标签管理。系统还支持精细化标签处理，即基于读者属性标签与资源属性标签进行关联分析后，工作人员总结出更为精细化的标签加入隶属于某个标签的下级体系中，以进一步发现读者和资源的需求和特征。

访问地址：<http://47.110.141.197:8080/site/login>。

#### （4）大数据统计报告服务系统

统计报告是我们了解、掌握图书馆当前状况和发展变化的主要依据，是分析、研究图书馆业务规律、发现图书馆存在问题和解决方案、进行科学决策的重要参考。统计报告的主题内容十分丰富，如综合性的事业发展报告、读者阅读行为报告、少儿读者的服务报告、馆藏及利用报告等。为满足全省公共图书馆业务统计的多样性需求，大数据二期软件项目建立了一个多维度数据统计报告服务系统。系统支持按照图书馆机构维度（全省、地市、多馆、单馆）、时间维度（月、季、年或任意月份时间范围）、内容主题维度（馆藏数据、读者数据、流通数据

等主题)进行总量统计、趋势统计、分类统计和对比排名统计,并以月效能排名报表、年度事业发展报告等定期发布方式和根据全省各级行政主管部门、各图书馆个性化需求按需定制的方式为全省提供统计报告服务。

统计报告主要是由若干主题的统计报表组成的,图书馆大数据统计所涉及的统计项较多,生成一个完整的全省性统计报表往往需要耗费较大的系统资源和较长的时间,为了快速而准确地形成所需的统计报告,系统在分析了图书馆业务统计实际需求的基础上,预先设计定义了近 200 多个统计表,形成了一个庞大的统计表池,每个统计表有相应的统计说明、展示方式(文字、表格、图形),支持多种机构对象、时间区间和统计类型选择,支持对基本统计单位数据进行预统计和缓存处理。系统根据具体工作目的和主题需求从统计表池中选择所需的统计表、设置多个统计表的排列顺序,定义报告名称、报告头和内容概述,从而建立一个统计报告模板,将统计模板与需要统计的图书馆对象以及时间维度进行组配,从而形成一份完整的统计报告。系统支持以 html 页面格式、文本格式在线阅读或下载输出所需的统计报告(图 3-13)。

数据处理流程包括前期数据整理和后期数据处理,前期数据整理是将各项数据源数据进行清洗去噪后汇总至数据中心 Hive 库中,对各项数据(包括业务数据、填报数据、读者流量数据、读者活动数据)按统计要求进行结构调整和关联合并。后期数据处理是数据计算平台基于 Hive 清洗之后的数据,同步传输至高效查询的 ClickHouse 列式数据库中,在 ClickHouse 中计算相关指标并进行数据分析,并将结果性数据缓存至 MySQL 中用于后续其他查询使用或页面展示。ClickHouse 能够在大规模数据下实现亚秒级数据查询统计,保证快速提供统计结果。

#### (5) 馆情查询服务系统

系统基于馆情填报数据和业务管理系统采集数据,提供全省公共图书馆详尽的馆情信息和基本服务数据的查询服务(图 3-14),同时可以选择馆藏、面积、读者量、外借量等主要指标值接近的图书馆进行对比(图 3-15)。系统将各图书馆指标体系以及业务数据进行规整,提炼出查询类型条件,用户可快速选择地区、馆舍面积、读者数、藏书量等查询范围条件得到查询结果。

访问地址: <http://47.110.94.2:10063/lib/index>。

[illegible]

图 3-13 统计报告模板定义样例



## 江苏省公共图书馆馆情查询服务

### 江苏省公共图书馆大数据服务

Big data service platform of Jiangsu Public Library



#### 南通市图书馆

市级 一级馆

地址：崇川区崇文路2号

官方网站：http://www.ntlib.org.cn

主建筑面积（平方米）：28000

馆部总量（册）：1507129

年外借量（册次）：642622

微信：南通市图书馆

定位：全部图书馆功能

注册读者总数（人）：231674

年举办活动（次）：238

#### 功能区

开放公共区域

读者活动区域

数字多媒体服务

大型活动场地

书库

公共设施

公共休闲区

儿童服务区

行政办公、内部业务

“大厅”

“总服务台”

“读者休息区”

“图书馆总出入口”

“馆室出入口”

#### 基本信息

面积(平方米)：1778.55

#### 设备信息

流量采集系统：是

信息发布大屏(平方米)：1

书目查询电脑(台)：3

朗读亭设备(台)：1

#### 馆藏量相似的馆



##### 盐城市图书馆

馆藏量：154.57万册



##### 扬州市图书馆

馆藏量：145.98万册



##### 淮安市图书馆

馆藏量：120.13万册

#### 注册读者相似的馆



##### 扬州市图书馆

注册读者：18.43万人



##### 常州市图书馆

图 3-14 南通市图书馆馆情查询

<div> <h2>江苏省公共图书馆馆情查询服务</h2> <h3>江苏省公共图书馆大数据服务</h3> <p>Big data service platform of Jiangsu Public Library</p> <div> <input type="text"/> <input type="button" value="搜索"/> </div> </div>				
图书馆 2020				
图书馆类型	市馆	市馆	市馆	
评估定级	一级馆	一级馆	一级馆	
地址	扬州市文昌西路466号	崇川区崇文路2号	府西路6号	
建筑面积	21719 平方米	28000 平方米	40080 平方米	
馆藏总量（册）	1459831	1507129	1545708	
注册读者总数（人）	184311	231674	122392	
年外借量（册次）	304682	642622	400482	
年读者流量（人次）	—	—	—	
年活动次数	163	238	175	
官方网站	www.yzlib.cn	www.ntlib.org.cn	www.yctsg.cn	
读者服务				

图 3-15 馆藏量相近的扬州市、南通市、盐城市图书馆比较

4. 数据中心建设

(1) 基础设施建设

江苏省公共图书馆大数据中心建设对服务器、存储、网络、信息安全等具有较高的要求，针对大数据数据量大且不断增长、实时性强、面向全省服务的特点，需要一个可靠性强、响应速度快、存储容量大、安全性能高、可扩展性强的信息化基础设施支持的运行环境。租用云服务是最好的选择，经过比较研究，我们选择了性价比相对较高的阿里云基础设施服务，年租用费约 60 万元。基础设施情况如表 3-1 所示：

表 3-1 江苏省公共图书馆大数据中心基础设施

序号	名称	数量
1	服务器类型 1（服务分析及展示服务器） 实例规格：32 vCPU 128 GB（通用型 g5，ecs.g5.8xlarge） 系统盘：200GB SSD 云盘 固定带宽：5 M	4 台
2	服务器类型 2（数据整合及清洗服务器） 实例规格：16 vCPU 64 GB（通用型 g5，ecs.g5.2xlarge） 系统盘：200GB 高效云盘 固定带宽：20M	3 台
3	ECS-Web 应用服务器 实例规格：8 vCPU 16 GB（计算网络增强型 sn1ne，ecs.sn1ne.2xlarge） 系统盘：100GB 高效云盘 数据盘：500GB，高效云盘。	9 台
4	弹性扩展存储 OSS （服务器类型 1 和类型 2 需另外接存储保存数据）	1 批
5	安骑士	16 套
6	Web 应用防火墙	1
7	数据库审计	专业版
8	堡垒机	16 台
9	网络安全等级保护等级评定为三级	

## （2）数据建设情况

采集汇聚全域范围图书馆的各类数据是大数据建设最基础的也是最关键的工作，图书馆大数据的数据源可以包括图书馆所有相关要素的状态数据、属性数据、描述性元数据、工作数据、运行数据等。具体有：

①文献资源类数据：各种普通资源、数字资源的馆藏或者馆用信息；文献资源的元数据，如书目数据、电子资源元数据。这些通过数据采集和元数据建设获得。文献资源内容数据或对象数据不是大数据的采集对象。

②服务对象数据：图书馆读者和用户信息，包括传统的办证读者、网上注册读者、临时访问读者、服务体系内读者、全部市民卡读者（全民读者）。其中有的有详细的用户信息（如自然属性、社会属性和图书馆属性），有的只有非常简单的信息；有的有读者类型，有的没有读者类型。服务对象数据主要通过数据采集获得。

③服务运行数据：主要有图书馆要素相互作用实时交互而产生的图书馆运转运行、业务工作、开放服务等动态数据，也包含图书馆运行规则参数等数据，如普通文献读者借阅数据、数字资源读者访问数据、读者到馆流量数据、读者馆内空间流量数据、读者活动数据等。运行数据主要从运行管理系统的数据库中采集获得。其特点是时效性强、数据量大。

④保障条件数据：主要有图书馆建筑馆舍、功能空间、设备设施等数据，图书馆全体人员信息，图书馆经费信息，等等。这些数据通过填报生成。

⑤外部环境数据：图书馆的发展变化与外部环境密不可分，外部环境数据包括但不限于出版、文化、教育、气象、地理环境、人口、经济、历史等图书馆影响因素的数据。

三年多来，江苏省大数据建设不断向全面、完整、有效、及时地采集汇聚全省数据的目标努力，中心数据日积月累不断增长丰富，逐步形成了类型较全、范围接近全覆盖、数据价值较高的中心数据库。到2021年7月底数据规模达到5亿条，平均每天新增10.8万余条，年增约4000万条。其中保障条件数据已经汇集了2018年、2019年、2020年江苏省的省、市、县（区）、乡镇（街道）、村（社区）15000多个单位填报的300多万条馆情数据；普通文献资源数据、服务对象数据、

图书借阅数据等已对接全省 117 家公共图书馆中的 101 家的业务管理系统，实现了实时数据采集，已汇聚馆藏信息 1 亿余条、读者数据 2000 余万条、借阅信息 3.4 亿条；读者流量数据已对接全省 33 家图书馆的读者到馆流量统计系统，实现了实时数据采集；读者活动数据已对接了全省 72 家图书馆的读者活动管理系统，实现了实时数据采集。

但是，数据建设工作是一项长期而艰巨的任务，目前，江苏省大数据中心数据的类型和范围还需要不断完善。

2021 年，将全面实现全省 117 家图书馆业务管理系统对接和数据采集任务；继续完成全省少部分未填报的乡镇、村级图书馆的馆情填报工作，完成全省填报数据的正确性和完备性检查工作；进一步更新完善标准书目元数据数据库，完善制定藏书地点、流通类型等对象的元数据标准，发动全省图书馆对各自业务系统中的藏书地点和流通类型进行深度的标引，完成标准书目元数据库的建设工作。

2022 年，计划全面完成全省 117 家图书馆的活动系统使用 and 对接工作，并完善细化活动信息采集指标；统一实现全省各级图书馆全面使用读者到馆量数据采集统计系统。

2023 年，力争实现国内主流数字资源服务数据的采集处理和元数据建设工作。逐步实现出版、文化、教育、气象、地理环境、人口、经济、历史等图书馆影响因素数据的采集工作。

### （3）数据管理工作

对大数据中心的运行环境进行监控维护、对数据进行质量管理和安全管理是数据管理的主要任务。为此，重点实验室下设数据管理组，专门负责协调解决全省公共图书馆大数据采集对接技术问题；检查修正全省填报采集数据，确保保障条件数据的正确性和完备性；检查验证统计结果与业务管理系统数据的一致性，确保结果的合理性和可用性；监控大数据中心系统运行状态，及时调整优化运行环境；负责大数据中心的网络信息安全相关工作，确保信息数据安全和系统正常运行。

## 5. 数据应用服务

为图书馆提供大数据应用服务，以数据驱动事业发展，是大数据建设的初心和使命。重点实验室通过实际应用检验研究成果，获得新的应用需求，为深入研究提供正确方向和新的动力。

根据大数据利用方式、分析处理深度的不同，大数据应用服务可以分为三类。一是原生数据利用服务，如在各馆镜像数据库的基础上提供异地数据库容灾服务，将采集到的各馆数据进行整合和重构，形成大规模应用数据中心，在此基础上开发建立全域图书馆读者统一认证服务、文献统一检索服务、异构系统之间的通借通还系统等；二是统计数据服务，在对全域数据进行清洗汇聚和标签化的基础上，基于数据统计实现实时数据发布和统计数据服务，也包括对数据的初步的定性分析；三是数据深度挖掘和融合应用服务，这是对大数据进行深度挖掘和智能分析，得到重要的业务规律和对象画像，基于分析结果宏观上制订工作改进和发展方案，微观上以数据应用接口或微服务方式融合应用到图书馆业务管理系统和各种服务系统中。

### （1）原生数据利用服务

#### ①全省读者用户统一认证服务

在采集到的全省图书馆读者数据的基础上，建立全省统一的读者用户中心数据库，开发和提供各种读者统一认证服务接口，满足全省各馆之间资源共享或一体化服务所需的用户认证需求，满足全省图书馆统一与国家文献资源服务系统对接所需的用户认证需求。目前已经实现了与原国家图书馆推广工程统一用户管理系统的对接，全面完成原国家数字图书馆推广工程市级馆大数据采集项目的工作任务。

#### ②全省书目统一检索服务

在采集到的全省图书馆馆藏数据的基础上，建立了全省图书馆书目统一检索服务。但这个系统目前仅实现书目检索功能，还需要结合流通服务和数字资源服务拓展新的读者服务功能。

#### ③数据容灾备份服务

为各个图书馆的业务管理系统提供数据备份服务，一旦某个图书馆发生数据

损坏且不可恢复，大数据中心可提供数据恢复所需要的数据。近两年已经为省内两个县级图书馆提供了此项服务，避免了数据丢失而造成的重大损失。

#### ④异构多业务系统读者一体化服务

目前，在一个地区使用不同业务管理系统的图书馆之间实现通借通还是一件非常困难的事情，基于大数据建立的全域图书馆的读者和馆藏信息中心，将有可能实现跨地区异构系统的通借、通还、网借等读者流通服务功能。

#### ⑤评估定级数据服务

大数据中心包含了图书馆几乎全部的保障条件、服务效能数据，在图书馆许可的情况下，可以实现与评估定级系统的对接，为评估定级系统提供相关的指标数据。

### （2）统计数据服务

#### ①实时数据发布

实时提供全省公共图书馆运行服务数据，供各级主管部门、图书馆在大屏、移动终端、门户上进行大数据发布，如客流量、外借量、活动情况等数据。

● 为江苏省文化和旅游厅的江苏智慧文旅平台行业监管指挥中心提供图书馆数据发布服务。

- 全省读者到馆量实时发布（图 3-16）。
- 全省普通文献外借量实时发布（图 3-17）。
- 全省活动数据实时发布（图 3-18）。



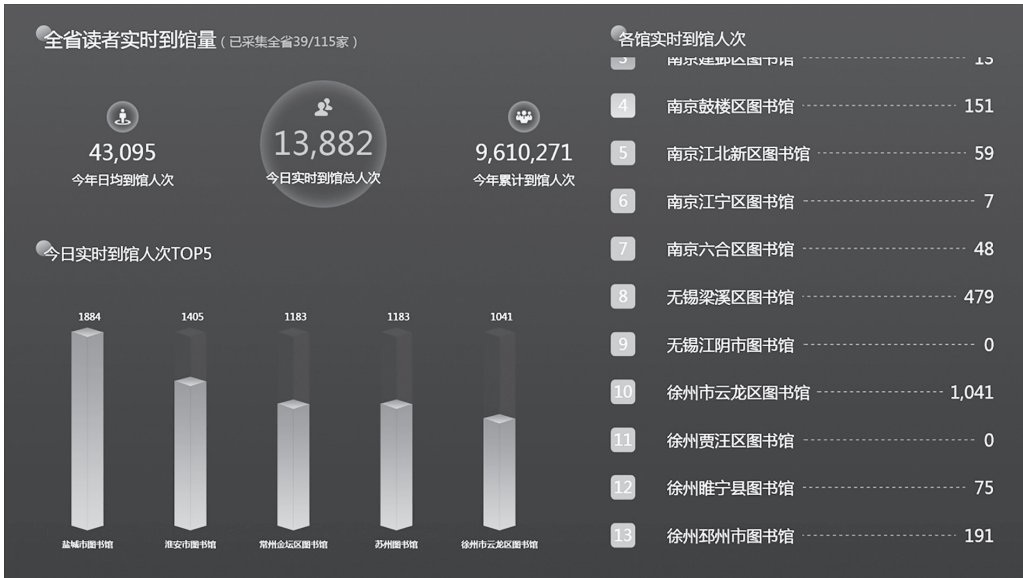


图 3-16 全省读者到馆量实时发布

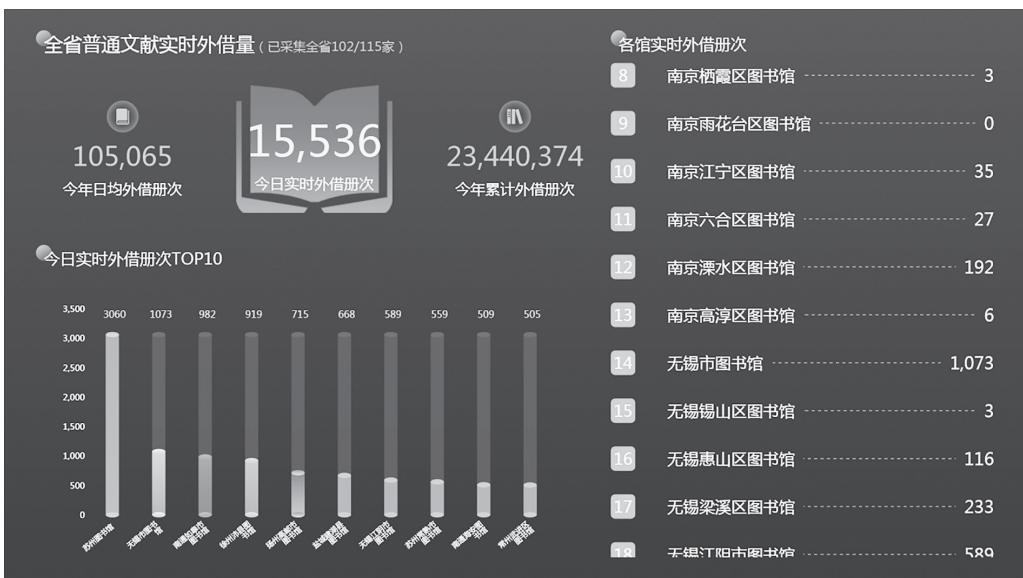


图 3-17 全省普通文献外借量实时发布

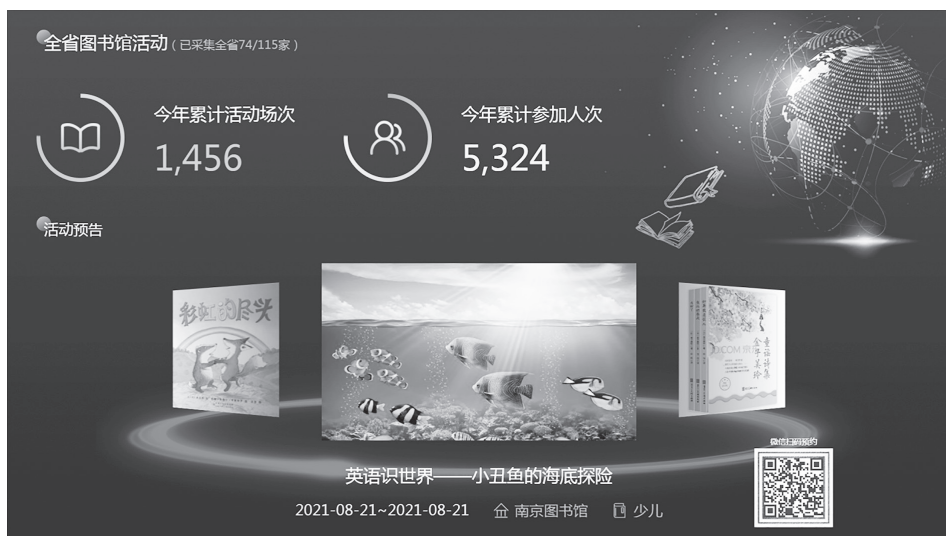


图 3-18 全省活动数据实时发布

## ②统计数据服务

● 全省每月服务数据统计。每个月在江苏图书馆云服务平台上发布上个月全省服务数据月统计报告。统计指标主要包括到馆人次、外借量、新增读者、读者活动等几个图书馆服务效能指标，按照全省各地区、各馆、地市级馆、区县级馆的月指标总量、日平均值、月环比和月同比进行统计报表和统计图展示，用户可查看和下载“江苏省公共图书馆服务数据月统计报告（2021年7月）”（图3-19）。

## 江苏省公共图书馆服务数据月统计报告(2021年7月)

报告以公共图书馆的各项服务数据为基础,对数据进行抽取和处理,进行全面立体的阐述与分析,基本客观的反映公共图书馆相关数据情况,具有重要的参考分析价值。

报告中所涉及的数据来源:业务系统数据来源 102家(点击查看名单);客流数据来源 38家(点击查看名单);活动数据来源 73家(点击查看名单);

报告目的:

《全省公共图书馆服务数据月统计报告》将每月公布全省各地区及各公共图书馆服务指标数据,包括到馆人次、外借量、新增读者、举办活动几个指标。各图书馆可以了解本馆同比、环比的服务情况对比以及在全省中的排名。

更多其他馆情、业务数据情况可关注年度报告,也可登陆大数据平台进行查询。平台地址: <http://47.110.141.197:8080/site/login>。

数据说明:

1、数据来源于江苏省公共图书馆大数据服务平台。

2、存在部分图书馆数据未接入或数据源获取异常等问题。

3、存在部分图书馆统计结果环比或同比波动异常,请与省中心进一步进行数据确认与对接。

4、活动数据来源于活动管理系统。

5、数据异常问题请联系江苏智慧文旅平台图书馆接入群(群号913350954)的南图杨老师。其它问题反馈可致电025-84358029。

6、本报告数据仅供参考。

### 全省各地区读者到馆量环比与同比分析(2021-07月度)

2021-07月全省公共图书馆读者到馆量 2332244人次,日均到馆量 75233.68人次。环比上个月增加了51.65%,同比去年当月增加了3552.69%。

排名	地区	2021-07月读者到馆量(人次)	平均每日读者到馆量(人次)	环比2021-06月	同比2020-07月
1	常州	317119	10229.65	↑68.04%	0%
2	南京	306872	9899.1	↑13.96%	↑380.61%
3	淮安	272986	8806	↑28.55%	0%
4	南通	251485	8112.42	↑44.48%	0%
5	徐州	203098	6551.55	↑37.9%	0%
6	盐城	189142	6101.35	↑46.86%	0%
7	苏州	179459	5789	↑90.05%	0%
8	扬州	174490	5628.71	↑55.92%	0%
9	镇江	116030	3742.9	↑165.64%	0%
10	泰州	109016	3516.65	↑130.04%	0%
11	无锡	77910	2513.23	↑412.16%	0%
12	连云港	23491	757.77	↑64.31%	0%
13	宿迁	0	0	0%	0%
	省馆	111146	3585.35	↑22.78%	0%
合计	全省	2332244	75233.68	↑51.65%	↑3552.69%

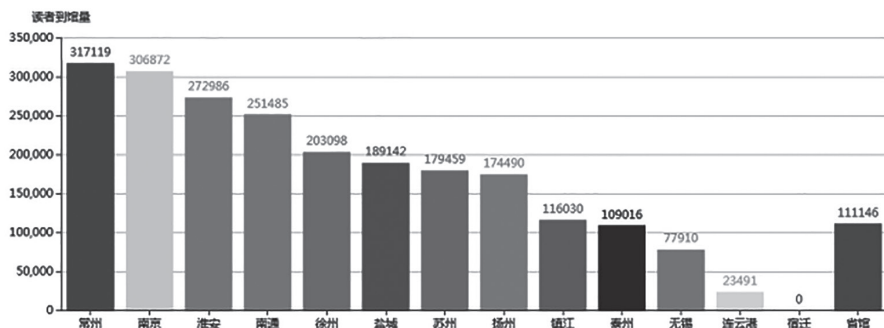


图 3-19 江苏省公共图书馆服务数据月统计报告(2021年7月)

- 全省事业发展年度统计报告。每年3月前正式发布前一年度的《江苏省公共图书馆事业发展统计报告》，报告以数据为基础，采用统计分析方法多维度展现全省公共图书馆发展状况和发展趋势。内容有全省综合统计、按地区汇总统计、按各地区分别统计、按图书馆类型统计（分为省馆、市级馆、县区级馆、少儿馆）、专项统计（阅读报告、少儿工作）。

- 按需提供各类统计数据服务。大数据重点实验室服务于全省各级文化主管部门、各级图书馆的管理人员、研究人员和工作人员，为他们提供管理决策、业务工作、学术研究所需的各种专题统计数据和分析报告。各类人员向省重点实验室提出申请，重点实验室工作人员分析用户需求，如果存在现成的统计报告数据就直接提供给用户，否则根据需求制订专门的统计报告方案，并与用户进行讨论，形成用户接受的方案，基于方案进行数据统计，把统计结果交给用户进行检查审核，如果用户对统计报告结果不满意，则要修改完善方案再进行统计形成新的结果，这个过程可能要反复多次，直到用户完全满意为止。

图3-20中这份在2021年6月江苏省公共图书馆工作年会上的“江苏省公共图书馆少儿服务统计分析”PPT报告，就是完全基于重点实验室提供的统计数据而形成的。

### （3）数据分析和融合应用服务

目前，图书馆大数据应用基本上还停留在原生数据利用和统计数据服务阶段，随着大数据分析技术的发展和智慧图书馆体系建设的提出，以数据深度挖掘为主的数据分析和融合应用的研究和实施已迫在眉睫。“十四五”期间，结合国家图书馆智慧图书馆体系建设发展目标，大数据的智慧化应用将是江苏省大数据建设的主要工作和发展方向。下面是我们的一些设想和展望。

## 我省公共图书馆少儿服务统计分析 ——基于全省公共图书馆大数据

南京图书馆  
图书馆大数据应用江苏省文化和旅游重点实验室  
吴政  
2021.06.19

## 说明

- 数据来源：全省公共图书馆大数据，全省103家图书馆的业务管理系统数据，还有15家馆未对接。
- 填报数据误差：全省114家图书馆的填报数据有较大的不确定性，部分图书馆数据明显有偏差。
- 业务系统少儿数据的甄别：无法从读者类型、注册人年龄来区分，最后从全省24466个馆藏地点中通过关键词检索及人工判断，最终找出711个少儿馆藏地点；
- 基本统计指标进行简单分析：以全省711个少儿馆藏地点的馆藏为依据，统计少儿馆藏量、借阅读者人数、外借量、馆藏学科结构、外借图书学科结构等。

## 目录

### 一、全省综合

- 保障条件
- 服务效能
- 专项分析

### 二、按馆统计

- 按地区
- 按市级馆
- 按县区馆
- 按少儿专业馆

### 一、全省综合

#### 全省少儿保障条件基本情况表

#### 1. 保障条件

序号	指标名称	指标值	单位	与总量比	与2019年比
1	服务场所面积	9.91万平方米	869平米	12.5%	增长3.60%
2	服务场所数量	487个	4.27个	4个	增长5.41%
3	图书资源数量	1274.09万册(件)	12.37万册(件)	15.44%	增长12.57%
4	经费平均投入	23.83万元	29.54万元	16.67%	29.52%
5	工作人员数量	706人	6.19人	4人	增长4.28%

说明：功能馆、经费和人员的数据来源于南京金陵文脉的全省114家图书馆2020年中旬统计报表。馆藏数据来源于103家图书馆的业务系统数据。114家图书馆包含省馆、13个市馆、93个区县馆和7个独立少儿馆。功能区面积按照建筑面积的80%取值计算。

#### 分析：

- 平均水平
  - 2020年全省114家图书馆少儿服务保障条件良好，但与对应总量占比偏低，与2019年相比各项指标略有增长。
  - 少儿功能区面积达到国家标准《公共图书馆少儿服务规范》占比10%的要求，但对于江苏经济文化发达省份，还应该有更高质量的要求。
  - 功能区数量4个以上，基本能满足亲子阅读、少儿综合阅读、少儿综合阅读、少儿综合阅读的需求。
  - 服务资源量全省少儿人均1册，达到公共文化服务体系建设人均1册的平均水平。（全国第七次人口普查：江苏少儿人口12499172人）
  - 资源年采购经费较高，按照这个经费每个馆平均每年可购买少儿资源1万册左右。
- 发展不平衡
  - 30%的图书馆的少儿功能区面积在300平米以下，最少的仅有40平米。
  - 30%的图书馆的少儿采购经费在不足5万元，最少的仅5000元，每年只能购买不到200册图书。

### 一、全省综合

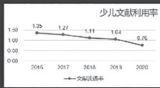
#### 2. 服务效能

#### 全省5年少儿借阅读者、外借量、馆藏量基本情况表

年份	少儿借阅读者数	少儿外借量	少儿馆藏量	总借阅读者数	总外借量	总馆藏量	少儿借阅读者占比	少儿外借量占比	少儿馆藏量占比	少儿文献流通率
2016	713763	9712712	7212052	1659012	34237866	65316767	43.02%	28.37%	11.04%	1.35
2017	747975	10707095	8429581	1740453	32964170	75311281	42.98%	32.48%	11.19%	1.27
2018	727719	11266355	10163653	1756371	33556823	85925344	41.43%	33.78%	11.83%	1.11
2019	773153	11784830	11318428	1857076	33365761	93877196	41.63%	35.32%	12.06%	1.04
2020	797523	9739416	12740907	2014392	31453378	102383570	39.59%	30.96%	12.44%	0.76

#### 分析：

- 在文献数量（12.44%）、功能区面积（12.5%）、人员（13.3%）占比都不高的情况下，少儿借阅读者达到39.59%，外借量达到30.96%的占比，说明少儿具有很高的借阅活跃度和阅读需求，公共图书馆应该为他们提供更多的资源和服务，应该把少儿服务作为读者服务的重点来抓。
- 2020年受疫情影响少儿外借量下降17.38%，文献流通率也从1.04下降到0.76。



### 一、全省综合

#### 2. 服务效能



说明：数据来源全省103个图书馆的711个少儿馆藏地点的馆藏和流通数据；借阅读者为一年内至少有过一次借阅记录的读者。

#### 分析：

- 2020年少儿借阅读者占少儿人口的6.26%，高于2.49%的总借阅读者人口占比，说明少儿读者的人口占比高于平均水平，对读者数量贡献大，但6.26%人口占比还很小，少儿读者数量仍有很大发展空间。
- 2020年少儿外借量9739416，文献流通率达到0.76，远高于0.31的总馆藏流通率。说明少儿资源的服务效能好，图书馆还需要进一步丰富少儿馆藏。

图 3-20 江苏省公共图书馆少儿服务统计分析

### ①数据分析服务

数据分析就是从大数据中找出规律的过程，图书馆大数据分析可概括为基于唯物辩证法的逻辑思维分析方法和基于现代信息技术的数据智能分析方法。逻辑思维分析方法也是当前图书馆行业最常用的方法，是在统计数据的基础上进行归纳与演绎、分析与综合、抽象与概括等逻辑思维处理，而得出一定的结论和规律；数据智能分析方法包括基于机器学习和数据关系分析的一般数据建模技术和正在迅速发展的基于神经网络的深度学习算法的高级数据建模技术。



逻辑思维分析方法对分析人员的要求较高,一方面需要有全面的图书情报专业知识和丰富的图书馆工作经验,另一方面需要有很强的逻辑思维能力,就像医生面对检查数据要给出病情诊断并提供治疗方案一样,图书馆大数据分析人员也需要从统计数据中发现规律、找出问题,并给出解决问题的建议和方案。这些工作可能包括:依据评估定级标准、现代公共文化服务体系建设标准等,帮助具体图书馆进行指标分析,寻找差距,并制订相应整改方案;帮助具体图书馆进行馆藏结构和读者需求分析,为完善馆藏结构和文献采购提供具体方案,包括各类图书的合理的种、册数量比例,复本数建议方案等;帮助具体图书馆进行服务效能分析,寻找影响服务效能的主要因素,并提出改进方法和措施。基于逻辑思维分析方法的图书馆大数据分析人员可以是专职人员,也可以是聘用人员或兼职人员,大数据应用重点实验室每年设立“数据分析研究类”课题,委托或招标确定行业专家和研究人员进行分析研究工作。

数据智能分析方法比较常用的是数量关系建模分析,如关联分析、回归分析、聚类分析等,最常用的工具是以 SPSS 等为代表的统计分析工具,这些基础的数据建模技术和工具应该在图书馆大数据分析中得到广泛应用。重点实验室将对基于神经网络的机器深度学习的高级数据建模技术在图书馆大数据分析中的应用进行研究,以期获得应用进展。

数据智能分析方法是具体的图书馆业务规律和对象画像,对象画像包含读者画像、资源画像和图书馆画像。读者画像包括读者全体、各类读者群体和读者个体的文献需求、阅读行为等情况,资源画像包含图书馆全体资源、各类资源和个体资源的馆藏和利用情况,图书馆画像是全体图书馆、各类图书馆群、个体图书馆在保障条件、业务工作和服务效能等方面的综合情况。

## ②融合应用

融合应用是指大数据分析成果在智慧系统中的融合应用,通过深度挖掘和智能分析得到图书馆业务规律和对象画像等分析成果,基于分析成果以数据应用接口或微服务方式融合应用到图书馆业务管理系统和各种服务系统中,赋予这些系统以智慧的能力。具体支持的智慧功能可以包括:文献智慧采访、文献智慧分配和调度、文献资源智慧推送、智慧参考咨询服务、智慧服务机器人、智慧图书馆



评价、服务功能空间智慧管理、工作人员岗位和值班智慧分配和调度等。

## 6. 业务人才培养

### (1) 目标与途径

大数据建设人才是关键，人才培养必须有目标、有计划、有措施，江苏省大数据建设通过跟班工作、兼职服务、课题研究和培训交流等途径为全省图书馆培养了一批具有数字思维和数据分析能力、能够从事大数据建设和应用的人才，为实现大数据在全省公共图书馆高质量发展和智慧化应用的建设目标提供人才保障。

业务人才培养的具体途径有：

①短期工作培训。从全省图书馆和合作共建单位中遴选人员到重点实验室跟班参与课题研究、数据分析等实验室工作，每年培养人数不少于 100 人，提升骨干人员的综合业务水平和研究能力。

②兼职工作实践。设立数据检查员、数据分析师等兼职岗位，从全省图书馆和合作共建单位中招聘符合条件的人员协助重点实验室开展数据审核验证、数据分析服务等工作。提升兼职人员的大数据分析能力。

③参与课题研究。一是邀请全省图书馆和合作共建单位参加或联合申报各级相关课题；二是鼓励全省图书馆和合作共建单位参与或承担实验室自主设立的开放研究课题，提升参与人员的研究能力和学术水平。

④开展培训活动。举办大数据应用培训班，采用 30 人左右小班培训学习加实践操作方式，培养学员的业务数据处理和基本应用能力。

### (2) 2021 年培训工作

2021 年 5—7 月开展了一次全省覆盖的短期工作培训。完成了全省 117 家公共图书馆大数据业务骨干培训任务，培训分为 5 期，每期时长 5 天，参与者达 150 人次。第 1 期为地级市图书馆人员培训，共培训人员 15 名，目的是为后面 4 期区县人员培训班培养助教师资力量。第 2 期至第 5 期为区县图书馆人员培训，每期 30 人以内，由地级市师资人员带队到省馆集中培训。

短期工作培训的内容包括：了解图书馆大数据应用基本理论和方法、重点实验室的主要工作内容和应用成果、各图书馆的工作内容和本地化应用；熟悉江苏省公共图书馆大数据服务平台相关系统的操作；开展馆藏地点、读者类型大数据

规范属性标引工作；对照检查本图书馆业务系统统计与大数据统计的数据一致性，填写业务统计数据一致性情况表；检查馆情填报系统数据的准确性，填写填报数据准确性情况表；核查本地区填报单位的完整性，填写本地区未填报单位情况表；检查本图书馆活动系统的利用情况，填写活动系统信息、活动上线情况、与省中心对接情况；核查本图书馆读者流量数据采集统计系统情况，填写读者流量系统信息、与省中心对接情况。

同时完成了南京图书馆与全省公共图书馆签订《江苏省公共图书馆大数据共建共享协议书》的签约工作，并建立了全省智慧图书馆 QQ 工作群。

## 7. 主要经验总结

图书馆大数据建设是一个复杂的、庞大的、长期的系统工程，需要文化和旅游行政主管部门在政策、经费、组织等方面的大力支持；工程承担单位或部门（一般是省级图书馆）要有强大的牵头组织协调能力和相关工程技术人员；全域公共图书馆要有一定的信息化建设基础，至少要基本实现业务自动化管理，要对大数据建设的作用和意义有充分的、正确的认识。

真正意义上的图书馆大数据应用在我国还处于起步阶段，相关的应用研究还不够全面和深入，对照智慧图书馆体系建设的融合应用要求还存在一定的差距，因此要加强加快图书馆大数据应用的研究和实践探索。

大数据应用软件体系建设十分重要，作为创新事物没有标准和参考对照，软件需求在不断变化，因而需要不断地实践并推进需求研究，不断进行迭代开发和完善，一支稳定的、长期的需求研究分析和开发团队十分必要。

数据中心建设是核心和关键，数量众多的图书馆和机构、多个系统数据源、异构的系统 and 不同版本，特别是个别图书馆的保护主义，使得全面、完整、及时地进行数据采集比较困难，只有通过各种方法和措施、坚持不懈开展协调工作才能保证数据采集工作顺利进行。

大数据建设的初心和宗旨是为各级主管部门、各个图书馆提供全面周到和高质量的服务，数据取之于各馆就必须服务于各馆，只有为各级行政主管部门、各图书馆提供更多的有价值的服务，才能真正获得各方持久、稳定、有力的支持。

智慧图书馆建设离不开大数据的支撑和驱动，因此需要全面加强和推进大数

据与智慧图书馆应用系统的融合应用研究和实践，需要围绕国家图书馆智慧图书馆体系建设的总体目标和工作部署做好配合和支持工作。

最后，从大数据的本质和规模价值来看，江苏省大数据建设所取得的初步成就和体现出来的价值还不是真正意义上的公共图书馆大数据，只有当全国所有的公共图书馆机构都成为数据源，建立起全国的大数据中心，那才能真正体现出大数据的价值意义。

（南京图书馆 吴政 耿健 袁勇）