

数字图书馆推广工程资源整合平台构建

韩萌

(国家图书馆, 北京 100081)

摘要: 数字图书馆推广工程是我国公共数字图书馆建设与发展的国家级支撑工程。为满足我国公共图书馆在数字环境下的资源整合与服务需求, 数字图书馆推广工程对我国公共图书馆数字资源元数据进行仓储与整合, 在全国范围内构建各地图书馆广泛参与的分布式推广工程资源整合平台。该平台一方面面向终端用户群体, 为其提供公共图书馆资源检索与知识发现专业化一站式服务; 另一方面, 面向我国各地公共图书馆, 通过API对接、页面部署、云部署、分布式部署等方式为专业化数字资源整合服务提供系统平台支撑, 最终实现全国公共图书馆文献信息资源的集中揭示、分级调度与服务。

关键词: 数字图书馆推广工程; 公共图书馆; 数字资源整合平台

中图分类号: G250.76

DOI: 10.3772/j.issn.1673-2286.2018.02.002

数字资源整合是一种数字资源优化组合的存在状态, 是依据一定的需求, 对相对独立的各数字资源系统中的数据对象、功能结构及其互动关系进行聚类、融合和重组, 重新结合为一个新的有机整体, 形成一个效能更好、效率更高的新数字资源体系^[1]。通过整合的资源系统具有集成检索功能, 是一个跨平台、跨数据库、跨内容的新型信息资源体系。数字资源整合主要包括数据整合和功能整合两种形式^[2]。在图书馆数字资源整合工作方面, 主要涵盖门户整合、业务流程整合、应用整合和信息整合等内容^[3]。近年来, 随着数字图书馆发展的日益成熟, 数字资源总量的不断提升, 图书馆资源建设已不再局限于传统形式的文献建设, 而是在数据整合的基础上逐步转向信息整合与知识整合服务, 以元数据仓储、资源整合和平台构建为重点, 搭建新型数字资源服务体系^[4]。

近年来, 国外图书馆开展了广泛的跨机构、跨文化、跨平台数字资源整合应用实践。其中, 比较有代表性的是欧洲数字图书馆 (Europeana) 和美国公共数字图书馆 (Digital Public Library of America, DPLA) 等。Europeana通过数字资源建设、技术框架建设、服务体系搭建和项目运作管理, 在欧洲领先的研究型高校支持下, 已经拥有集博物馆、档案馆和图书馆为一体的资源平台, 为欧洲公众提供一站式浏览和检索服务, 在现代化、网络化数字资源建设与服务方面发挥越来越

重要的引领和示范作用^[5]。截至2015年底, Europeana已有37个国家的3 500个机构参与, 集聚4 800余万条元数据。DPLA将美国图书馆、博物馆和档案馆的数字化文化遗产及相关资源进行汇集与整理, 免费提供用户服务。截至2016年4月23日, DPLA可供检索和查阅的数字文化遗产及相关资源总量已经超过1 331万项^[6]。此外, “世界图书馆项目”“OCLC WorldCat计划”等, 尝试运用云计算模式开展全球范围内的数字资源整合服务。我国图书馆界的数字资源整合服务在高校图书馆应用较广。自1998年, CALIS通过三期建设项目, 建立面向图书馆联盟的信息服务协作网络, 推出CALIS资源统一检索平台。该平台采用Hadoop分布式模式和主从架构, 极大提升我国高校图书馆的数字资源整合、发现与获取服务能力^[7]。

1 我国公共图书馆数字资源整合服务现状

1.1 现有建设成果

经过二十余年的发展, 我国公共图书馆陆续开展数字资源整合服务实践工作, 建成一批区域性公共图书馆数字资源整合服务平台。国家图书馆的文津搜索系

统有效地整合国家图书馆各类数字资源,对其进行多维度聚合展示,不断提高文献揭示效率,截至2017年10月底,文津搜索系统元数据仓储量达2.2亿条;上海图书馆的上图发现系统,通过元数据和联邦检索结合的混合检索技术,为用户提供高质量、低成本的学术资源发现和共享服务;浙江网络图书馆通过统一认证的方式实现对浙江省内各公共图书馆持卡读者的数字资源统一检索、书目检索、原文传递、馆际互借、知识导航等服务;广东省文献资源共建共享协作网为省内公共图书馆、高校图书馆和科研机构提供数字资源统一揭示服务;贵州数字图书馆构建贵州省公共图书馆数字资源整合服务平台,为全省公共图书馆读者提供统一服务。

1.2 建设面临的问题

随着用户对多元化、深层次数字资源需求的不断提升,我国公共图书馆在数字资源整合服务方面的短板日益凸显,主要表现在三方面:一是资源整合仅限于特定馆藏,数量有限,主要以数据整合和信息整合为主;二是元数据标准不统一,不支持跨类型、跨机构、跨地区的资源检索服务;三是资源“信息孤岛”现象严重,系统技术标准封闭,兼容性差,导致公共图书馆共享服务的协同效应难以发挥。国家图书馆于2016年11月~2017年4月,通过问卷调查、网站访问和工作人员访谈等方式对全国市级以上公共图书馆数字资源整合服务情况开展调研。调研内容主要包括公共图书馆资源整合系统情况、资源发布系统情况、资源整合内容情况和未来资源整合规划等。调研结果显示,我国68%的市级以上公共图书馆尚未开展数字资源整合检索服务。目前统一检索的服务内容主要为商购资源,自建资源整合服务水平较低。

2 推广工程资源整合平台框架

数字图书馆是图书馆在互联网环境下的新业态。为进一步加强公共数字文化建设,文化部、财政部于2011年5月正式实施“数字图书馆推广工程”(以下简称“推广工程”)。国家图书馆作为推广工程的资源建设中心、技术保障中心、管理服务中心,具有统领全国各地公共图书馆开展数字资源建设、系统平台搭建、数字图书馆服务推广、标准规范应用的职责,并为各地公共图书馆开展数字图书馆建设提供技术指导与支持。在数字资源

建设方面,推广工程建设分级分布式数字资源库群,实现数字资源无障碍共建共享^[8];在数字资源整合方面,文化部提出“公共图书馆年度已建资源的发布服务率不低于85%,2014年实现已建资源的全部登记,新建资源于每年年底完成登记”^[9]。经过多年发展,依托推广工程数字资源联合建设项目,我国公共图书馆数字资源拥有量明显提升,数字资源结构日趋合理,覆盖全国公共图书馆的分级分布式数字资源库群基本形成。因此,有必要在全国范围内构建一个各地图书馆广泛参与的分布式数字资源服务平台,通过统一标准、开放技术、共享服务,提升公共图书馆数字资源关联度,真正形成海量、有序、易用的数字资源库,推动我国公共图书馆文献信息资源的集中揭示、分级调度与服务。推广工程于2016年启动我国公共图书馆数字资源整合项目。国家图书馆负责该项目的组织实施,将我国公共图书馆数字资源元数据充分汇集,集中改造,有序整合,构建关联关系,通过搭建资源检索平台对其进行发布,并面向全国各地公共图书馆提供服务,最终实现我国公共图书馆数字资源一站式资源发现。推广工程资源整合项目重点集中在元数据仓储、资源整合和平台搭建三方面。首先,借助前期推广工程自建数字资源元数据仓储项目,与推广工程其他来源元数据有效集成,构建推广工程元数据仓储;其次,分阶段完成推广工程资源元数据的清洗改造和有序整合;最后,通过推广工程资源整合平台实现全国公共图书馆数字资源检索服务的无缝集成,逐步实现我国公共图书馆数字资源的一站式资源发现。

2.1 元数据仓储

元数据是推广工程资源整合的基础。从建设来源看,推广工程数字资源可分为公共图书馆联建资源、推广工程商购资源、国家数字图书馆征集资源与自建资源等;从服务范围看,可分为面向全国服务资源和面向区域服务资源。截至2016年底,推广工程面向全国公共图书馆提供的资源总量达到145TB;我国公共图书馆数字资源总量接近10PB。推广工程资源数量庞大、来源广泛、服务分散,导致资源元数据异构特征明显。在资源类型方面,包括电子图书、电子期刊、电子报纸、电子图片、音视频以及多媒体资源等;在资源发布方面,国家层面拥有包括推广工程资源发布服务平台、政府信息公开整合服务平台、国家数字图书馆数字资源发布服务平台等,地方层面的发布系统更是不胜枚举。据统计,我

国75.0%的省级公共图书馆和超过56.5%的市级公共图书馆都开展了数字资源发布服务^[10]。在服务权限方面,包括互联网访问、统一用户系统远程访问、数字图书馆专用网络访问、局域网访问等。

推广工程自实施以来,逐步构建全国公共图书馆自建数字资源元数据仓储。截至2016年底,元数据仓储数据量达200余万条,收录范围已从公共图书馆数字资源逐步向公共文化机构数字资源拓展。为确保推广工程资源整合服务的最终效果,在前期公共图书馆自建数字资源元数据仓储的基础上,国家图书馆基于Dublin Core并结合我国公共图书馆实际需求,对推广工程商购资源、国家数字图书馆征集资源与自建资源等元数据进行清洗整合,形成推广工程数字资源元数据仓储,确保异构资源的有效集成。

2.2 资源整合

由于推广工程数字资源的海量性、分散性与异构性,在整合过程中,一方面要分析并构建关联关系,实现对各资源的有效整合;另一方面,要保证系统间的互操作性,以完成不同服务权限的资源整合。在元数据加工中,既要保证元数据整合规则的科学与易用性,又要保证延续性与拓展性。基于上述目标,国家图书馆制定了一套包括元数据格式规则、查重规则、映射规则、清洗规则、检索规则、显示规则、可视化展示规则、检

索结果分面设置规则、智能推荐规则、知识关联规则等在内的标准规范。

以上述标准规范为指导,国家图书馆首先开展元数据的查重、清洗、转换等工作,形成统一格式的元数据;其次,根据发布服务要求对数据进行批量处理,增加字段信息,如增加对象数据的用户访问权限信息、内容形式与媒体类型信息等;最后,开展数据审核与校验,同时为实现资源整合服务的专题性,进行相关专题元数据的加工处理。

2.3 平台构建

在元数据仓储与整合的基础上,国家图书馆搭建应用于我国公共图书馆数字资源元数据发布服务的推广工程资源整合平台(以下简称“平台”)。平台基于tomcat开发,通过对Spring、SpringMVC和Hibernate框架的整合,提升平台的开发与部署效率。在数据库方面,平台采用Mysql进行数据存储,通过云数据中心实现统一搜索和知识发现;调用云数据中心提供的接口,利用Web API实现与地方图书馆原有资源检索平台的对接。平台采用多层立体的模块化结构,由云数据中心、数据管理系统、数据存储系统和数据服务系统4个子系统组成,与其他系统共同完成推广工程数字资源的检索与利用(见图1)。其中,云数据中心为核心模块,各地图书馆可以按需订制其他系统。

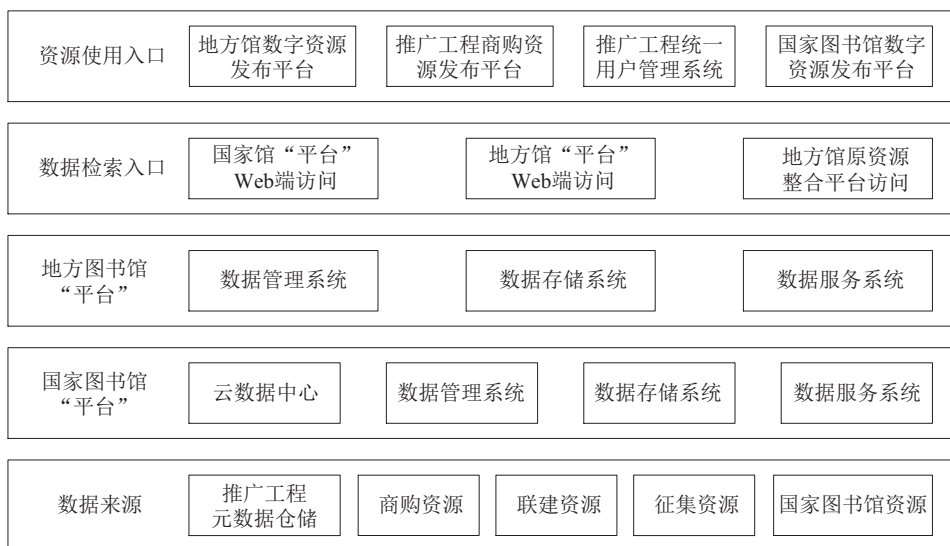


图1 推广工程数字资源检索与利用基本框架

从具体模块看,云数据中心是通过推广工程数字资源的集中存储与索引,来提供外部服务的物理位

置。云数据中心实现对平台所有数据的统一管理、存储与服务,同时可对各分数据中心的数据进行收割。云

数据中心封装各馆可能用到的数据应用和数据分析接口,保证接口调用的实时性和稳定性,进而支持各地方图书馆平台部署模式的功能实现。数据管理系统能够适应各种条件的元数据、对象数据的内容管理,通过云端数据分片获取、内部数据整合、加工流程管理、专题创建与管理以及本地数据镜像创建等功能实现对各种来源、条件元数据与对象数据的内容管理。数据存储系统为地方图书馆数据管理系统提供数据后台存储能力(包括数据的索引、备份等)外,还可实现与云数据中心的定期同步。数据服务系统根据地方图书馆用户的需求特点和数据分片,进行数据检索、分析、下载功能的构建,如一站式检索、导航、知识关联等。各地图书馆能对本馆服务系统资源进行个性化选择及对本馆服务页面的定制化修改。

3 推广工程资源整合平台功能设计

推广工程的建设目标是提升我国各地公共图书馆数字资源建设与服务能力。公共图书馆作为公共文化服务机构,满足终端用户需求是其资源建设与服务的出发点。平台作为公共图书馆数字资源的利用终端,主要包括终端用户和地方公共图书馆两类服务对象。一方面,平台面向终端用户群体,为其提供专业化的公共图书馆一站式资源检索与知识发现服务;另一方面,平台助力我国各地公共图书馆数字资源建设,为专业化数字资源整合服务提供系统平台支撑。

3.1 面向终端用户的功能设计

按照“元数据集中,对象数据分散”原则,平台为终端用户提供资源检索、知识导航、智能推荐及专题数字资源整合服务。

(1) 资源检索是平台的基本功能,包括对元数据的通用检索与高级检索。具体分为三类,其中一站式全文检索支持对不同资源类型按照不同检索点进行检索;组合检索支持按照多种筛选条件进行检索,当选择多种类型时,可自动对多种类型资源的检索点进行逻辑组合形成新的检索点;模糊检索、繁简通检与中英文混合检索等。

(2) 知识导航是平台的核心功能。平台按照资源类型、分类体系、专题等进行资源分类导航。平台提供按照关键词、时间、地区等多维度的标签云、图谱、时

间轴等可视化展示与聚类;通过URL与唯一标识符实现元数据间、元数据与对象数据间的多维关联关系;支持作品关系、主题关系等基于内容的关联关系;支持元数据与目次数据的链接。

(3) 智能推荐是平台的拓展功能。通过知识组织工具,基于访问数据统计分析,平台可实现资源检索与展示过程中的智能推荐服务。支持多个限制条件下检索结果的分面聚类;提供基于相似词、相关词以及热门搜索词等在内的搜索词智能推荐;支持基于个人访问行为的内容推荐。

(4) 专题数字资源整合服务是平台的特色功能。平台可实现基于元数据的内容整合,通过设定规则对元数据进行筛选与加工,生成用于专题资源整合的元数据集,建设用于展示专题资源的发布页面,实现文本、视频格式的资源发布。

3.2 面向地方公共图书馆的功能设计

通过API对接、云服务和分布式部署等方式,平台将助力地方公共图书馆数字资源整合服务,最终形成我国公共图书馆数字资源一站式发现服务。在功能设计上,平台充分考虑地方公共图书馆在数据管理、系统管理以及读者服务等方面业务需求。

(1) 平台可以实现对地方图书馆数字资源元数据的管理。支持通过批量导入、系统传递、OAI等模式进行数据收割;支持按照设定的元数据格式和字符集进行数据导出;支持对元数据进行批量更新或删除处理。数据管理功能可以提供定制化服务,支持对资源库按结构和元数据标准批量导入和单独管理。

(2) 平台建立完善的系统用户多级管理体系。通过建立用户多级管理体系对不同模块、不同栏目进行权限分配。用户多级管理体系按权限从高至低主要分为超级管理员、系统管理员、子站管理员和子站工作人员四级。国家图书馆具有高级管理权,各地方图书馆根据平台部署情况具有相应的管理员权限。

(3) 平台实现对地方公共图书馆数字资源建设工作流程的统一管理,有效控制地方公共图书馆推广工程相关资源建设项目进度。项目管理功能可对资源建设信息进行分类管理、过程管理;按需提取信息,生成新文档;对操作行为日志定期备份、检索与导出。

(4) 平台具备完善的统计分析功能,能够从各层面反映公共图书馆数字资源的使用状况、用户关注焦

点等。支持从资源、用户等多维度进行统计分析及可视化展示；实现对相关统计指标的对比分析与实时展示；支持根据用户浏览与检索行为统计挖掘内容热点，实现智能推荐。

(5) 平台提供基于PC端、移动设备端等常见终端的页面自动适配；提供微信小程序与移动端数据检索服务，能够基于用户位置及馆藏位置进行信息排序和推荐。

4 推广工程资源整合平台部署

基于平台系统架构，国家图书馆采用多种模式开展平台的部署推广。对于已经具有数字资源整合服务的图书馆，可以通过Web API的模式实现平台接口对接，以跨库检索形式开展数字资源统一检索服务；对于尚未开展数字资源整合服务的图书馆，根据地方图书馆

的基础设施条件、数字资源建设情况和具体服务需求，由简到繁开展页面服务程序部署、站点式云部署，基于Web Service的分布式部署等模式。地方图书馆可以选择一种或几种部署模式开展平台搭建。

4.1 Web API对接

Web API是网络应用程序接口。通过API接口实现网络应用对接，并开发相关的Web应用。通过API的模式，已开展数字资源整合服务的图书馆可通过调用平台全部或部分数据，根据设定规则实现本地数据与推广工程数据的统一检索服务（见图2）。此模式效率高、成本低，并可以实现本地资源与推广工程资源的一站式检索服务。但由于我国绝大多数公共图书馆并未建设数字资源整合平台，因而部署范围较窄。

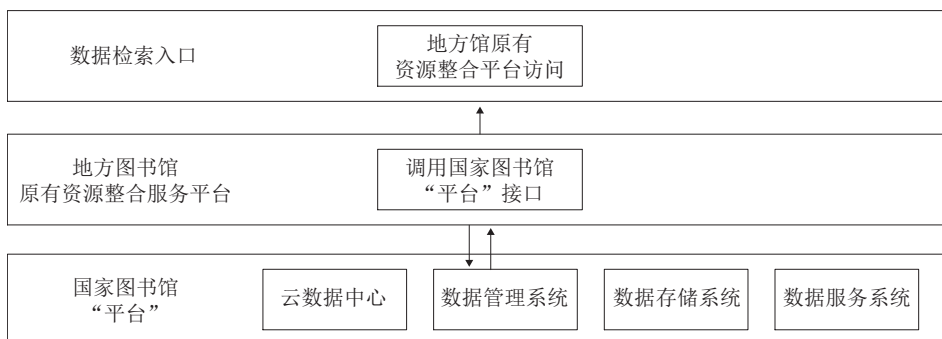


图2 Web API平台对接

4.2 页面服务程序部署

页面服务程序部署是推广工程在平台推广过程中的一项创新举措。地方图书馆可使用国家图书馆提供的Web页面程序部署工具和安装说明，制作本地个性化服务界面，获取平台资源检索服务（见图3）。此部署

模式简单，地方图书馆仅需对展示页面进行维护。同时，此模式的网页发布功能可帮助尚未建设网站的地方图书馆搭建图书馆网站，对基层公共图书馆尤为适用。但仅能实现对资源检索页面的定制，不涉及元数据的管理与整合。

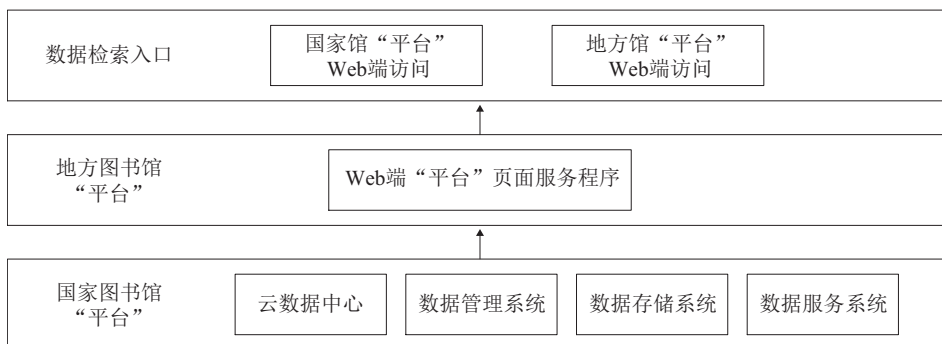


图3 页面服务程序部署

4.3 站点式云部署

站点式云部署是基于互联网和数字图书馆专用网络的平台服务利用和交互模式。在该模式下，地方图书馆不需要部署任何系统，而是通过在国家图书馆中心服务器上的云数据中心、数据管理系统、数据存储系统和数据服务系统分配相应的空间并开通服务账号，开展远

程资源建设与服务（见图4）。站点式云部署模式减轻了地方图书馆的平台部署与运维压力，具有较强的灵活性和拓展性。但由于云服务的要求，该模式对网络要求较高，必须具备较好的网络带宽条件，适用于已经联通数字图书馆专用网络的图书馆。另外，当部署的数字图书馆达到一定数量时，此种模式会对国家图书馆的中心服务器带来较大压力，因此需要科学测算与评估。

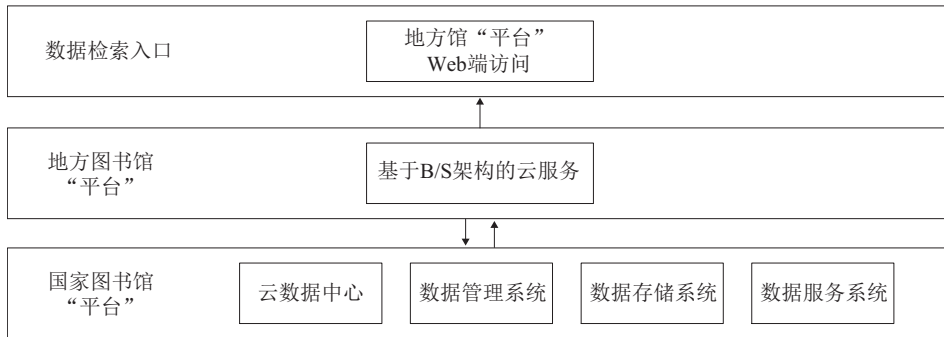


图4 站点式云部署

4.4 基于Web Service的分布式部署

Web Service是一种面向服务的应用整合技术^[11]。Web Service采用面向服务（Service-Oriented Architecture）体系架构，具有松耦合、可扩展、易集成等特点。基于Web Service的平台分布式部署，一方面因其具有通用性，可满足全国各地公共图书馆的普遍需求，各地方公共图书馆可按照通用配置进行部署；另一方面，因其具有可扩展性，各地公共图书馆可根据自身业务需求进行个性化配置和扩展。通过这种部署模式，平台采用前台页面分布、后台数据集中的模式来提供服务。推广工程云数据中心同时作为服务提供者和服务注册中心，提供数据接口和发布检索服务，为各地公共图书馆Web服务

端提供数据查询服务；各地公共图书馆Web服务端作为服务请求者，向推广工程数据中心调用查询接口，为终端读者提供Web页面浏览和数据查询服务（见图5）。在此模式下，地方公共图书馆在本地仅需要部署数据管理系统和数据服务系统，而云服务中心和数据存储系统由国家图书馆统一管理。这样可实现独立的数据管理和定制化服务，降低数据管理风险，但对地方公共图书馆网络条件要求较高。基于Web Service的分布式部署模式也可在遵循相同的数据管理标准和数据交换标准基础上，由地方公共图书馆建立独立的数据存储系统，通过定期与国家图书馆云数据中心同步，实现真正意义上的平台本地化改造与部署。

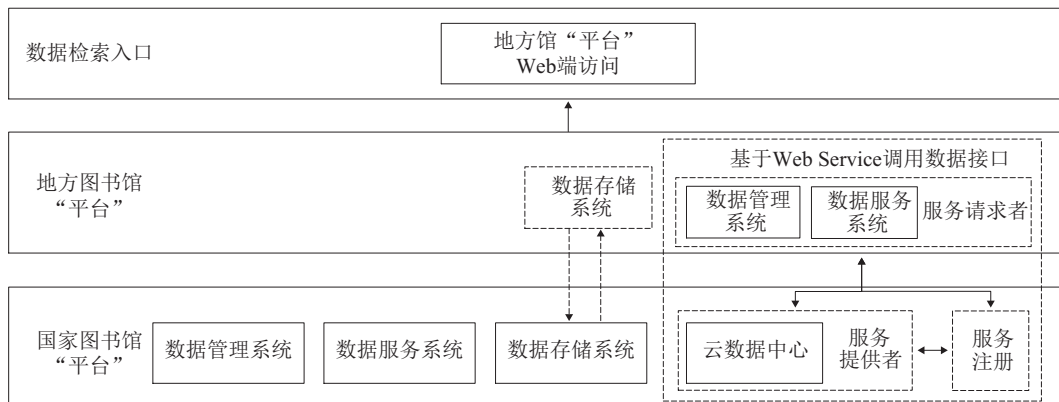


图5 基于Web Service的分布式部署

5 结语

信息时代,图书馆资源建设已经从资源加工向资源管理与服务转变。目前,资源整合服务已成为掣肘我国公共图书馆信息资源服务的短板。推广工程作为引领我国公共数字图书馆发展的支撑性工程,构建覆盖全国各地公共图书馆的数字资源整合平台势在必行。未来,推广工程资源整合服务将逐步向纵深发展,在构建我国公共图书馆元数据仓储的基础上,强化数据组织与挖掘,完善平台功能设计,向基于知识的数字资源整合延伸,采用更加易用、灵活的方式面向全国推广,最终构建我国公共图书馆资源的一站式资源检索与知识发现服务。

参考文献

- [1] 马文峰. 数字资源整合研究 [J]. 中国图书馆学报, 2002, 28 (4): 64-67.
- [2] LEYMAN F, ROLLER D. Using flows in information integration [J]. Ibm Systems Journal, 2002, 41 (4): 732-742.
- [3] JHINGRAN A D, MATTOS N, PIRAHESH H. Information integration: a research agenda [J]. Ibm Systems Journal, 2002, 41 (4): 555-562.
- [4] 李家清. 信息资源整合 [J]. 图书情报工作, 2005, 49 (8): 35-38.
- [5] 汪静. Europeana发展现状及启示 [J]. 数字图书馆论坛, 2017 (3): 46-53.
- [6] 黄文浩. 美国数字公共图书馆项目研究及启示 [J]. 河北科技图苑, 2016, 29 (3): 3-6.
- [7] 王文清, 张月祥, 关涛, 等. CALIS资源整合与云服务 [J]. 国家图书馆学刊, 2015, 24 (2): 36-42.
- [8] 文化部财政部关于实施“数字图书馆推广工程”的通知 [EB/OL]. (2011-05-26) [2017-09-11]. <http://www.ndlib.cn/cswjxz/201210/P020121023600959317575.pdf>.
- [9] 文化部关于加快实施数字图书馆推广工程的意见 [EB/OL]. (2012-09-20) [2017-09-11]. <http://www.ndlib.cn/cswjxz/201210/P020121023602263010094.pdf>.
- [10] 李晓明. 我国数字图书馆发展研究 [M]. 北京: 国家图书馆出版社, 2014.
- [11] 莫秀娟. 资源整合技术研究 [J]. 图书馆学研究, 2011 (1): 69-73.

作者简介

韩萌, 男, 1983年生, 硕士, 馆员, 研究方向: 公共图书馆数字资源建设、公共数字文化管理与服务, E-mail: hanm@nlc.cn。

The Construction of Resource Integration Platform of the Digital Library Promotion Project

HAN Meng
(National Library of China, Beijing 100081, China)

Abstract: Digital Library Promotion Project is a national program that supports the construction and development of public digital library in China. In order to meet the needs of resource integration and service of the public library in the digital environment, the DLPP collects and integrates the metadata of digital resources held by public libraries, and builds a distributed “Digital Resources Integration Platform of DLPP” among public libraries which aroused broad participation. On the one hand, the platform provide one-stop professional service of resources retrieval and knowledge discovery for end users; On the other hand, the platform provides specialized support of digital resources integration through API interface, web deployment, cloud deployment, distributed deployment and so on. Finally, the platform will achieve unified retrieval, hierarchical dispatch and service of digital resources of public library across China.

Keywords: Digital Library Promotion Project; Public Library; Digital Resource Integration Platform

(收稿日期: 2017-11-30)