

# 国家图书馆存储体系架构研究与设计\*

王乐春, 季士妍  
(国家图书馆, 北京 100081)

**摘要:** 国家图书馆经过十多年信息化建设, 已经建立相对完善的三层架构存储网络体系, 能较好支撑图书馆业务服务和资源的保存利用。随着数字资源迅猛增加, 数字资源的重要性和珍贵性不断提升, 灾备中心和战略储备库建设已提上议事日程。本文根据图书馆业务对存储体系提出的新目标, 分析得出目前存储体系面临的三个挑战(系统挑战、资源挑战和长期保存挑战); 依据国家图书馆职责定位和未来发展方向, 设计出以主存储为中心的服务体系结构、多级多地战略级灾备体系结构和永久保存体系结构, 为实现国家图书馆“两地三中心”的数字资源存储和服务业务要求提供技术支撑。

**关键词:** 国家图书馆; 存储体系架构; 分级存储; 灾难备份; 长期保存

**中图分类号:** G250.76

**DOI:** 10.3772/j.issn.1673-2286.2017.3.006

## 1 引言

国家图书馆作为各类信息资源法定保存机构, 肩负着记录信息时代印记和传播人类文明的重要使命。国家数字图书馆结合数字资源加工、保存和服务的实际情况, 围绕数字资源生命周期, 构建现代化、职能化的三层架构存储网络体系, 在严格规范的工作流程下实现海量数字资源的可靠保存和长效利用。

为分担数据潜在风险, 确保重要数据的安全性和关键业务的连续性, 国家图书馆目前正在建设同城灾备中心。以“保护为主、安全第一、藏用结合、传承发展”为基本原则, 积极筹备建设国家文献战略储备库, 将更好地防范和抵御地震、火灾等灾害的影响, 实现异地灾备, 战略储备库的建设更将成为数字资源永久保存和保障的战略基地。随着数字图书馆的不断发展, 同城灾备中心和战略储备库的建设, 需要重新梳理、规划和设计国家图书馆存储体系, 为国家图书馆在数字化信息时代传承社会记忆和人类文明, 开启数字资源长期保存的新纪元, 提供统一、稳定、高速和可靠的存储支撑基础环境。

通过对国家气象局信息化体系<sup>[1]</sup>、中国银联二代系统<sup>[2]</sup>和文化部信息中心存储体系<sup>[3]</sup>的调研, 跟踪最新存

储技术的研究成果<sup>[4-5]</sup>, 以及参阅现行国家和国际的资源<sup>[6]</sup>、存储、灾备<sup>[7-8]</sup>和长期保存<sup>[9]</sup>的相关标准规范, 针对国家图书馆此类以数字资源为核心、信息化发展已初具规模的单位, 要升级完善作为关键基础设施的存储体系无现成的解决方案, 必须从业务特点、业务需求、业务目标入手, 研制适合自身系统、资源和服务特点的存储体系。国家图书馆存储体系的建设是一项周密的系统工程, 需要进行业务体系、技术体系、标准体系、体系结构、制度体系和管理运行机制等多方位研究。本文仅就国家图书馆存储体系的体系架构展开论述, 旨在通过对国家图书馆存储体系的研究, 为实现国家图书馆“两地三中心”的数字资源存储和服务业务要求提供技术支撑。

## 2 国家图书馆存储体系面临的挑战

依据国家图书馆功能职责定位和未来发展方向, 要求存储体系为依托其的系统提供统一、高效、专业、稳定的存储空间服务能力, 要求存储体系为数字资源提供海量、稳定、安全、透明、方便的数据存储环境, 要求能为数字资源永久保存提供适应性广、管理简便、功能完备、标准化和自动化程度高的长期保存能力。

\* 本研究得到国家图书馆科研项目重点课题“国家图书馆数字资源存储体系研究”(编号: NLC-KY-2014-02) 资助。

## 2.1 系统挑战

国家图书馆经过多年的数字图书馆建设,实现了涵盖整个数字资源生命周期的系统体系结构构建,目前应用系统共139个,涵盖数字资源采集、文献数字化、数字资源加工、存储、搜索,以及发布与服务的全过程。这些应用系统超过半数对外提供读者服务,其中44%为互联网读者提供服务,9%为到馆读者提供服务;在系统备份方面,有备份的应用系统占87%,超过70%的系统采用热备方式,约25%的系统采用冷备方式;备份数据量在10G—99G的系统占比最多,1G—9G和100G—1T次之,分别有22个和14个系统。通过分析,发现国家图书馆存储体系支撑的业务应用服务有三个显著特点。第一,支撑应用系统多。存储体系支撑图书馆内部业务及对外服务的各系统,包括Aleph、发布与服务、大数据应用分析等28个在线业务系统。第二,系统重要性不同。存储体系所支撑的系统重要性各不相同,既有图书馆内部核心业务系统,也有用户直接使用的对外服务,还包括覆盖全国的图书馆行业应用系统,存储体系实现了对核心系统、关键系统、重要系统和一般系统的全覆盖支撑。第三,系统对存储的服务要求不同,如存储量、I/O、安全性、可用性等。

## 2.2 资源挑战

存储体系除实现对应用系统提供存储空间支撑外,还需完成图书馆数字资源的存储和管理。国家图书馆在进行数字资源建设时采取“中文为主、外文为辅,通用性资源外购为主,特色资源自建为主”的方针。截至目前,外购数据库共计259个,包括中文数据库126个、外文数据库133个;馆藏特色资源数字化全文数据超过2.3亿页;征集资源地方志、民国文献、家谱、非物质文化遗产和历史文化资源等特色资源,存储量共计21TB;网络资源采集实现网络导航2.1万项,采集超过5万个政府网站信息、315种中文电子报纸及80个专题信息(中国学、开放存取、奥运会、航空航天、日本地震核泄漏、汶川地震3周年紀念等)。从存储体系角度看,这些数据资源主要有三个重要特征。第一,需要保存数据量大。近年来,国家图书馆通过各种途径不断扩大自身数字资源馆藏量,涵盖文本、图片、图像、音频、视频等多种形式;截至2015年年底,国家图书馆数字资源建设总量达1 160.98TB(如网络资源77.28TB、自建资

源969.69TB),发布总量达815.72TB,约占数字资源建设总量的70%。第二,数据复杂多样。存储的数字资源既包含依托于各系统的数据,也包括各数据库管理的数字资源及大量基于文件的裸数据;数据内容、存储格式、使用方式也各不相同。第三,数据对存储要求特殊。与服务系统相比实时性要求不高,与档案系统相比真实性要求不高,但这些数据重要性高,因此要考虑多种存储介质、同城灾备、异地灾备和长期保存等因素,以整体保障数据的存在性和可用性。

## 2.3 长期保存挑战

美国宇航局国际空间数据系统咨询委员会于1999年发布《开放档案信息系统》(*Open Archival Information System, OAIS*)参考模型,OAIS在总结前人探索实践的基础上,提出一组抽象的数字资源保存功能模型和信息模型,该模型于2003年被接收为ISO标准(ISO 14721)。

在国际上,法国在源迅公司帮助下,借助开源软件Fedora开发OAIS兼容的分布式存储与长期保存系统SPAR;IBM帮助荷兰开发项目的成果DIAS已经发展到V2.0,并在德国的互操作长期保存系统项目KOPAL中发挥作用;新西兰与ExLibris公司合作,构建国家数字遗产档案系统NDHA并已投入运行,ExLibris借此将其OAIS兼容产品Rosetta推向市场。在国内,清华大学图书馆针对开源软件开展长期保存项目研究,国家档案馆也面向其电子化档案内容进行长期保存前期研究,武汉大学联合多家机构开展长期保存大范围研究,中国科学院文献情报中心在长期保存领域开展了卓有成效的探索实践。

总体而言,数字资源长期保存领域经过近些年努力,长期保存成熟度进一步提高,有更多实践成果可供借鉴。随着长期保存标准和最佳实践的不断推出,如何通过长期保存材质、设备、系统和灾备体系真正实现数字资源永久保存,成为保存体系面临的新课题。

## 3 国家图书馆存储架构的服务体系研究

### 3.1 服务体系内涵与要求

国家图书馆存储体系提供的服务主要涵盖系统支持和资源支持两方面:一是为各系统提供统一、大型、高速、稳定的存储空间;二是为数字资源存储提供存储环境。

存储空间管理,即在存储体系上为业务系统划分适当的存储空间,以保障业务系统的数据库、应用软件、中间件等。要求存储体系根据业务系统的性能要求、数据库要求、数据读取要求等,分配适合的存储空间。对于需要部署数据库、系统软件的应用系统,需要分配高性能、高转速的存储盘;对于需要部署备份数据、系统关联性应用的系统级需求,可分配低转速、稳定性高的存储盘。

存储环境管理包括依据数字资源的数据类型、数据量、数据存储要求、数据发布要求、数据提取要求等数据特点,制定不同的存储保存策略;为不同数字资源提供相应的保存环境,完成数字资源的规整、清洗、封装、入库、提取、解封、转换、出库等保存和提取工作,以及日常数字资源的比对、校验、抽检、迁移等维护管

理工作,保障数字资源被妥善保管。

### 3.2 服务体系结构设计

基于国家图书馆现有系统和资源状况,参照《国家图书馆“十三五”规划纲要》发展方向,结合国家图书馆主存储中心现状,及经费、场地、人员等实际问题,坚持“统一规划、统一建设、集中管理”的原则,根据服务体系业务需要,设计国家图书馆存储服务体系结构(见图1)。该存储服务体系与应用系统体系独立建设,采用IP SAN存储架构实现存储网,内部实现在线、近线和离线分级存储,有效支持系统空间服务,并为数字资源存取、资源发布和资源服务等提供透明、可靠的资源存储环境。

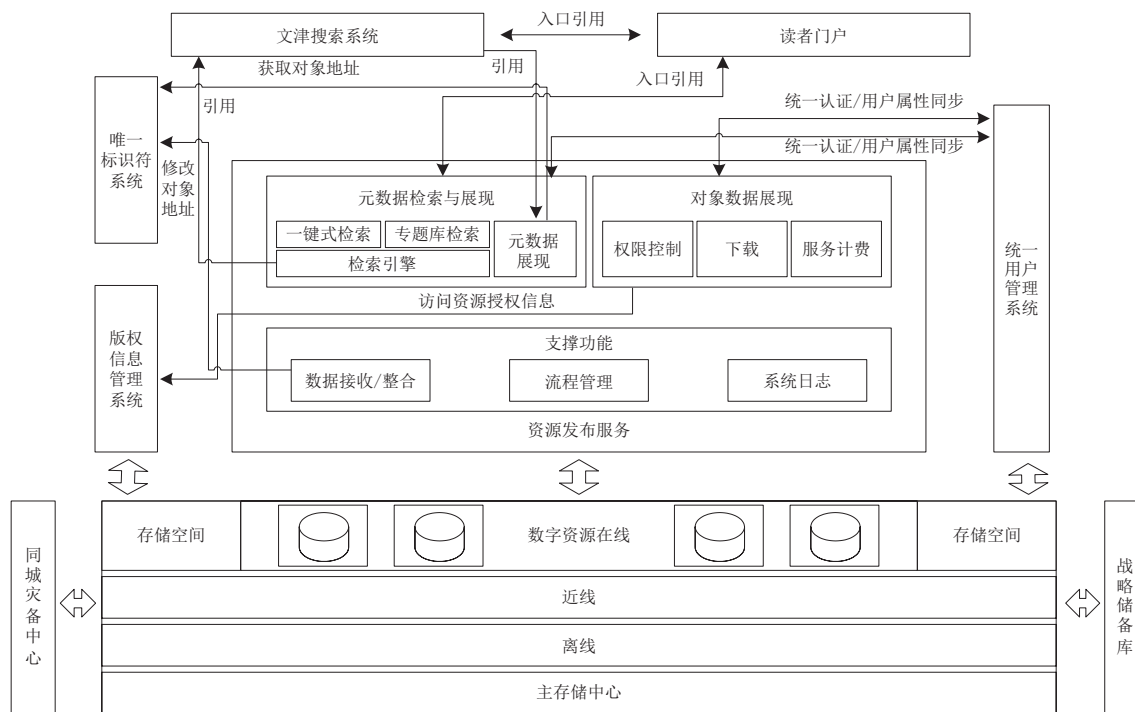


图1 国家图书馆资源服务体系示意图

该服务体系实现了应用体系与存储体系分离。应用系统所需的存储空间分两种:一种是系统运行自身所必需的存储,包括CPU的Cache、缓存、内存以及服务器自带的小磁盘存储;另一种是系统所依托存储体系提供的存储空间。该服务体系为50余个系统提供存储服务,这些业务系统既包括图书馆核心业务系统、Aleph系统、办公自动化系统、资源发布与服务系统及20余个商购数据库系统,还包括网络资源采集、大数据分析平

台、数字资源长期保存等业务系统。该方式可有效提升资源统一管理效率、降低管理成本,也为信息整合与数据共享奠定坚实的存储基础。

采用光纤导向器和IP SAN构成独立的国家图书馆存储网,所有存储设备形成统一的存储池向各系统和应用提供服务,该存储网可支持更多的系统和设备,以确保实现高性能和高速存取,独立存储网使存储数据备份不再占用LAN带宽,并且存储域的容量可独立扩充。



数字资源存储采取“在线、近线、离线相结合的三级存储”策略。需要实时在线生产的数据、提供数字资源实时服务的业务数据,采用“在线”存储服务;不需要实时在线服务的数据(包括互联网采集到的资源、作为中间数据临时保存的数字资源等),采用“近线”存储服务;需要永久保存的数字资源、备份的近线/离线资源,采用“离线”存储服务。在主存储系统中,采用高转速、性能高的硬盘(SAS、FC硬盘)或者固态硬盘支持“在线”数据存储;采用性价比高的低转速硬盘(SATA硬盘)支持“近线”数据存储;采用磁带库、光盘等离线存储介质支持“离线”数据存储,从而达到资源服务的合理和高效。

同城灾备中心根据主存储中心业务系统的数据规模、服务实时性要求、数据更新频率、业务系统数据库规模等因素,选择与之匹配的存储设备实现同城数据级灾难备份。

战略储备库根据主存储中心业务系统的重要性、复杂性,实现核心关键系统的系统级灾难备份。

### 3.3 构建以主存储为核心的服务体系

目前,广泛使用的云存储技术是在存储介质、存储结构上重新构建的一种新型服务模式。国家图书馆可在自身在线、近线存储架构的基础上,通过虚拟化技术平稳过渡到分布式存储架构。

在国家图书馆数字资源存储体系中,主存储中心作为数字资源的生产中心,所有数字资源在此生产并为业务及用户提供服务;同城灾备中心为主存储中心提供安全、稳定的数据安全保障,从而确保主存储中心的业务连续性和数据完整性;战略储备库则为主存储中心提供核心系统级灾难备份与长期保存服务。三个中心共同构成以主存储为核心的服务体系。

## 4 国家图书馆存储架构的灾备体系研究

### 4.1 灾备体系内涵与要求

国家图书馆肩负为中央和国家机关、重点科研教育生产单位、图书馆业界及社会公众提供服务的职能,实行“全年候”开馆,打破时空界限,保障“7×24小时”为全球提供服务。业务服务的连续性要求对关键业务系统的持续运作,因此必须保障支撑应用系统存储

体系运行的连续性。

为保障业务系统的备份安全、系统安全和数据安全,国家图书馆规定至少每3个月对各应用系统的主机操作系统做1次有效的全备份;此外,在主机配置及设置发生更改时必须对主机操作系统进行有效的全备份。对于涉及金融功能资产、财务等重要系统,必须保证1个月内任意一天备份数据的有效性,有条件的应保证3个月以上备份数据的有效性。

国家图书馆规定数字资源保存采取“集中保存、统一管理”原则,具体保存时采取“不同介质、多级备份、多地保存”的方式进行保存,利用光盘、硬盘、磁带等存储介质保存多个备份,以确保数字资源保存的安全性和可靠性。

### 4.2 灾备体系结构设计

业务系统、业务系统数据以及数字资源是国家图书馆的宝贵财产,为确保数字资源的安全性、服务系统的连续性,需增强存储体系的鲁棒性,在详细分析备份需求、充分考虑备份成本及效能、慎重选用适合备份技术的基础上,构建“本地备份保服务、同城备份保安全、异地备份保可用”的灾备体系。

根据国家图书馆实际情况,在二期主存储中心实现对各系统和数据的不同介质备份和本地备份;在古籍馆建立同城灾备系统,通过同步备份实现对重要应用系统的数据级灾备,通过异步备份实现对海量数字资源的存储级灾备;为有效抵抗同城发生的火灾、水灾、地震和强电磁辐射等不可抗力损坏,在承德战略储备库建立核心系统的应用级灾备并对海量数字资源实现温备份或冷备份,其逻辑关系如图2所示。

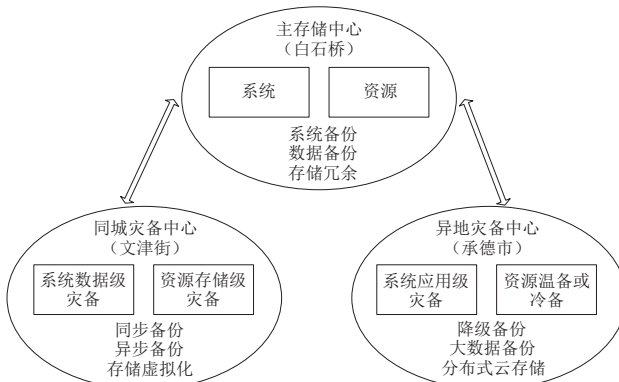


图2 国家图书馆资源灾备体系示意图

### 4.2.1 主存储中心备份体系

根据业务系统的特点和数字资源的保存要求,主存储中心的灾难备份区分为系统备份和存储备份两种业务模式。

(1) 系统备份。主存储中心支撑的业务系统包括提供业务支撑的业务系统、提供对外服务的资源服务系统及记录重要业务数据的业务系统等,系统类型多样、系统结构复杂。为保障各业务系统的正常运行,防止业务系统因软硬件故障、误操作及其他原因导致重要数据丢失,保证各系统可通过备份数据,迅速、安全地将系统和数据恢复正常状态,从而确保图书馆各项业务顺利开展。此外,业务系统要根据自身情况实现系统灾难备份。

(2) 存储备份。主存储中心的存储系统为业务系统提供安全稳定、高效快速的数据存储服务。主存储中心根据业务系统的数据存储、数据读取、数据服务实时性等因素的不同,使用不同的RAID策略,以保障存储系统提供的数据存储服务安全、稳定且能高效读取数据。在实际业务管理上,读写非常频繁的在线服务系统(如业务系统、OA办公自动化系统等)使用RAID1+0技术,在保证数据可靠性的基础上,可较好地利用阵列设备的数据读取速度,为在线服务提供较高效的数据支撑;对于需要提供实时在线资源服务、存储数据量较大的业务系统,存储系统采用RAID5技术构建存储空间。

### 4.2.2 同城灾备中心体系

同城灾备系统要对国家图书馆主存储中心支撑的业务数据、系统数据及数字资源等实现数据级灾难备份。根据业务系统的数据重要性区分为核心业务数据系统和各类数字资源业务系统,通过数字资源备份保障主存储中心数据安全。

(1) 核心业务数据的灾难备份。主存储中心的核心业务数据,包括由重要的业务服务系统和读者服务系统产生的业务数据。这些数据需采用实时同步复制的方式实现同城灾难备份。当国家图书馆主存储中心的存储系统发生故障时,核心应用系统所使用的本地逻辑卷能切换到灾备机房的存储系统上,切换在半小时内即可完成。

(2) 各类数字资源的灾难备份。各类数字资源主体

包括描述性元数据、管理元数据、保存元数据、对象数据等,还包括非核心业务数据(如大量业务系统、不需要实时在线服务的资源等),这些数据可以采用异步数据复制方式,在保证不影响核心业务数据灾备的情况下,实现主存储中心与灾备中心的数据一致。当国家图书馆主存储中心存储系统发生故障时,可使用灾备数据。

### 4.2.3 异地灾备中心体系

在承德市建设的国家文献战略储备库,针对核心业务系统,实现主存储中心与储备库的系统级“双活”,从而保障国家图书馆的核心业务系统数据安全和作业持续性,避免服务或业务的中断。此外,储备库要为主存储中心提供数据级异地灾难备份保障。

(1) 核心业务系统的应用级灾难备份机制。根据系统实际情况,在战略储备库部署核心业务系统备机,当国家图书馆主机发生故障时,可通过网络切换使储备库部署的核心业务系统备机替代主存储中心的应用系统,为读者提供正常服务。为保障核心业务系统的系统级灾备和数据备份的实时性,需采用裸光纤链接方式实现储备库与主存储中心间传输链路的稳定和高效。战略储备库系统级灾备的架构,如图3所示。

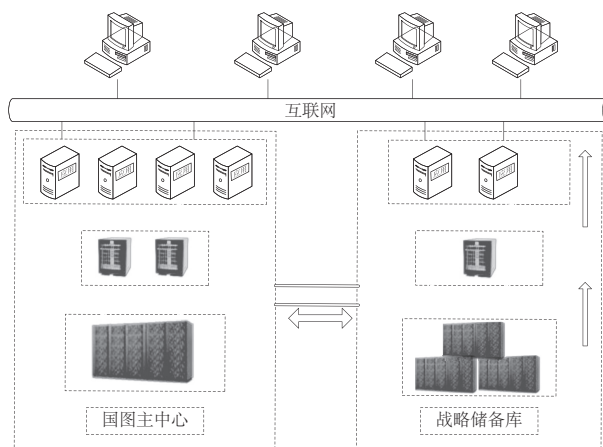


图3 战略储备库系统级灾难备份架构示意图

(2) 各类数字资源的灾难备份机制。可采用异步数据复制方式,在保证不影响核心业务系统级灾备的情况下,实现国家图书馆主存储中心与储备库异地灾备中心的数据一致。当北京的主存储中心和同城灾备系统同时发生故障时,将启动战略储备库异地灾备系统保存的数据,使严重受灾数据有机会恢复。

### 4.3 构建多级多地的灾备体系

灾备体系是保障存储基础设施可靠性、可用性的重要方式。国家图书馆存储灾备体系根据业务和服务的需要，部署“两地三中心”。其中，主存储中心为多个业务系统的正常运行提供备份支撑，也为数字资源提供安全、稳定的备份支撑；同城灾难备份中心为主存储中心提供数据级灾难备份保障，确保核心业务系统的数据安全和业务持续性；异地国家文献战略储备库灾难备份系统，实现数字资源异地保存和重要系统的系统级灾难备份保障，为主存储中心提供系统级和数据级灾难备份保障。三个中心协同工作，实现同城灾备和异地灾备协调，数据级灾备和应用级灾备并存，本地同质备份和异地降级备份并行，同步备份和异步备份互补，热备份、温备份和冷备份使用明确的多级多地的战略级灾备体系。

容灾体系的系统建设仅提供数据安全、可用的系统基础，在此基础上还需要建立完善的容灾备份运营管理体系，以保障容灾备份体系在发生灾难时能迁移、接替、保全、恢复，达到降低损失的目的。容灾体系更大的目标是保持和完善业务连续性，始终保持业务系统在面对突发性灾难事件时的持续能力，保证灾难损失在可承受的范围內。

## 5 国家图书馆存储架构的永久保存体系研究

### 5.1 永久保存体系内涵与要求

数字资源的脆弱性在业界已达成共识，针对沉淀千百年的中华文明以及最新的原生数字化成果，需开展数字资源的长期保存，以确保数字资源的可用性、可表现性、可理解性、真实性和可识别性。

根据数字资源保存和利用的不同特点，国家图书馆将数字资源保存划分为永久保存级和归档保存级。永久保存级的数据要求元数据和对象数据完整、对应关系清晰，且对象数据为加工的原始格式或引进的原始格式。该级别的数据原则上不再变动，并在三个中心保存相同的备份，分别保存于主存储中心、同城灾难备份中心、国家文献战略储备库。用于发布的数字资源以及元数据或对象数据不完整、尚不满足永久保存级条件的数字资源，在长期保存系统中为归档保存级

数据。归档保存级数据可在补充缺失数据并符合长期保存条件后，转为永久保存级数据；也可在失去保存价值后，从系统中删除。该级别的数据只在主存储中心保存1份。

数字资源长期保存涉及一系列信息记录和维护工作，由于标准、技术和实践的滞后性，目前尚无数字资源长期保存的独立系统可以满足国家图书馆长期保存的需求。而国家图书馆业务要求存储体系实现数字资源安全、稳定的永久保存。要实现永久保存业务需求，需要整合介质级、设备级和系统级长期保存应用，融合灾备体系，实现异地、异构长期保存数字资源调度，这是一项全新艰巨的任务。

### 5.2 永久保存体系结构设计

国家图书馆永久保存体系是由主存储中心的长期保存、同城灾备中心的灾备带库以及国家文献战略储备库的长期保存系统共同组成。主存储中心主要采用光盘存储和磁带库存储两种方式，通过多磁带库和光盘库构建长期保存的存储空间，实现永久保存和归档数字资源的长期保存（见图4）。在同城灾备中心购置长期保存存储设备，通过异步传输实现国家图书馆永久保存数字资源的同城灾备长期保存；在战略储备库构建满足OAIS模型的长期保存系统并构建电磁屏蔽储存室，对磁盘、磁带库进行冷备份，防止电磁辐射造成信息丢失，实现战略级防护。

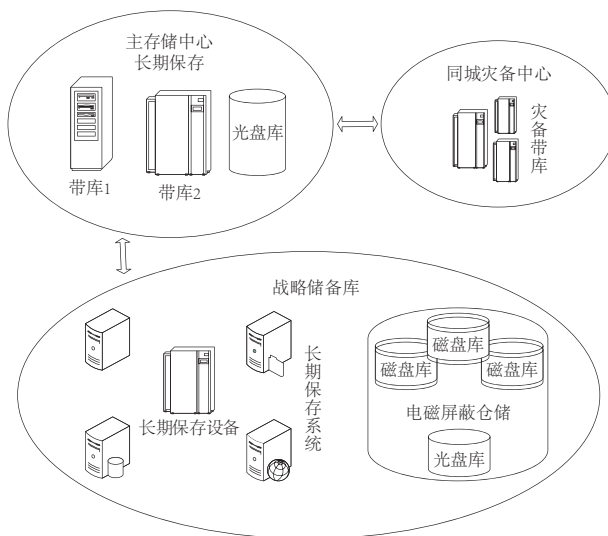


图4 国家图书馆永久保存体系示意图



### 5.2.1 主存储中心的长期保存

国家图书馆按照“中文为主、外文为辅，特色馆藏为主、引进资源为辅”的原则，重点对特色馆藏数字资源、原生数字资源、网络采集数字资源及合作建设特色文化资源等数字资源进行长期保存。在实际操作中，按照数字资源的原生性、独一性、重要性进行长期保存操作的优先级安排。

以CD、DVD光盘为代表的光记录介质具有记录密度高、容量大、随机存取、保存寿命长、稳定可靠、使用方便和价格便宜等一系列优点，适合存储需可靠保存且访问频率相对较低的数据，特别适用于大数据量资源的存储和交换。光盘成为国家图书馆实现对数字资源长期保存的重要手段，国家图书馆从长期保存介质入手，正式开始探索适合自身特点的长期保存之路。磁带具有介质稳定、价格低、占据空间小、易实现异地备份等特点，适用于大数据量记录的写入与存储，在国际上被公认为存档数据的首选存储介质，被许多政府部门推荐或指定使用。2008年，国家图书馆购置SUN STK 8500磁带库，磁带成为国家图书馆实现对数字资源长期保存的主要手段，至此，国家图书馆实现了设备级的数字资源长期保存；2010年，购置IBM3584磁带库，长期保存资源开始采用TSM存储管理软件，国家图书馆正式进入设备、资源、策略和管理为一体的长期保存新阶段。

### 5.2.2 同城灾备中心的长期保存

为保障永久保存数字资源的安全性，在同城灾备中心也进行数字资源的长期保存。在主存储中心TSM存储管理软件的统一管理下，永久保存的数字资源在主中心进行长期保存操作时，同时在同城灾备中心的长期保存系统进行保存，实现永久保存级数据在不同的存储中心完好地保存2份。

### 5.2.3 国家文献战略储备库的长期保存

国家图书馆经过十多年的长期保存工作，不断探索和实践，目前已经形成一套完整的长期保存工作流程（包括数据接收、保存前检查、数据封装、数据封装后检查、数据保存、保存后检查和数据抽检等）。在战略储备库新开发的独立长期保存系统，至少要包含图5所示的6个核心功能模块。

**资源摄取功能：**主要包括从不同资源提供方接收合格资源，按资源的不同分类完成元数据和对象数据信息包的生成、提交等工作。

**资源管理功能：**主要负责管理与对象数据相关联元数据及相关信息，包括数据格式、数据唯一性、数据标引信息等的管理；支持检索和查询，实现资源的随时查阅并能正确定位到元数据对应的对象数据位置。

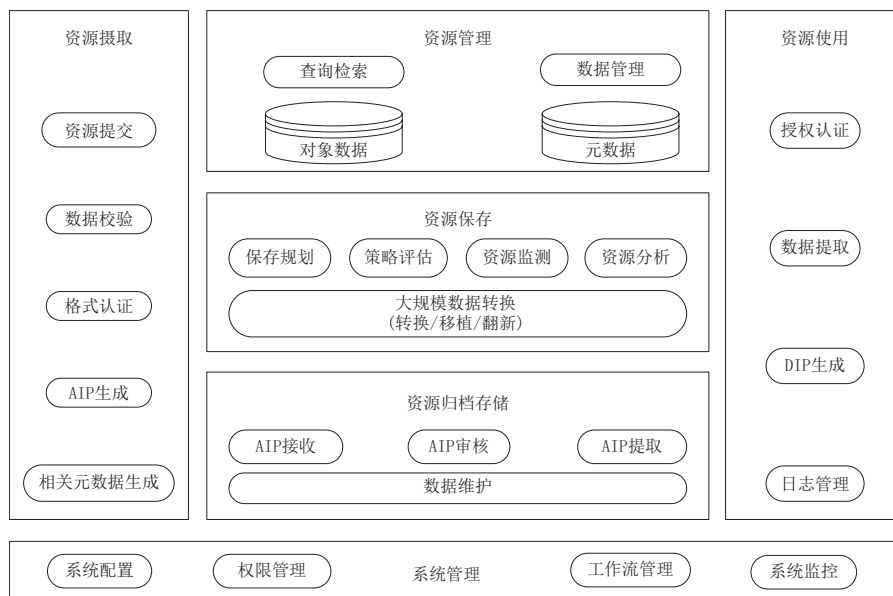


图5 长期保存系统功能模块示意图

**资源保存功能：**主要根据国家图书馆保存政策和框架，制定与保存相关的流程和策略，以完成对资源保存、数据资源迁移等方面的管理，同时完成永久保存规则和数据移植策略等方面的策略制定工作。为确保数字资源的持久访问，应提供对保存对象的例行检查策略，及时进行备份和迁移；在完成对数据移植等操作后，需对新生成的数据重新提交审核，以便归档保存。

**资源归档存储功能：**主要涉及对数据信息包的存储及按生命周期策略完成对存储资源的完整性维护。

**资源使用功能：**在资源使用前对资源的使用权限等进行确认，并检查对象数据等各类信息。根据应用需求与文本、图像及其他对象的传递规则，完成相关数据的抽取、打包等工作，以提供对外使用和保存更新并将操作历史存储到日志数据库。

**系统管理功能：**主要负责对系统整体监控和管理（包括系统配置、 workflow 管理、权限管理、系统监控等工作），该功能的目的是完成对整个系统运行的配置、管理，及用户管理等工作。

数字资源长期保存系统应兼容开放档案系统参考模型，要求能实现从提交、管理、归档，到确保数字资源长期保存和可获得性的整个流程，实现灵活的流程定制、全自动化迁移、流程全程监控并可回溯、全标准化输入与输出。此外，要求系统具有足够的开放性，以实现与国家图书馆现有相关系统的对接。

### 5.3 分布式永久保存体系资源调度

(1) 国家图书馆主存储中心的历史数据进入储备库存储中心，实现长期保存。国家图书馆主存储中心的永久保存级的数字资源，需要传输到储备库长期保存系统，实现长期保存管理。工作机制如图6所示。

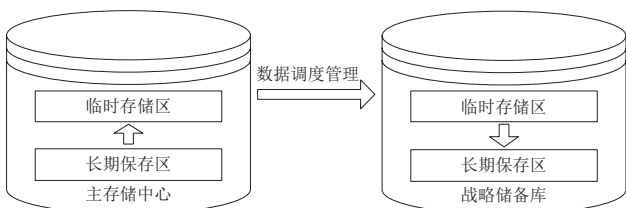


图6 主存储中心历史数据长期保存

在该业务中，首先需要将主存储中心长期保存的数字资源提交到临时存储区，然后通过储备库的长期保存管理系统的调度，将资源异步保存到储备库的长

期保存系统，再通过各级别校验以及“打包—拆包—组包”的迁移，最终完成数字资源在储备库的长期保存。

(2) 各单位提交的数据进入战略储备库存储中心，实现长期保存。储备库存储系统除支持国家图书馆主存储中心的永久保存数据外，还要对其他单位提交数据实现长期保存。合作单位通过远程复制备份的方式，提交待保存的数据到存储库主节点；主节点制定一定的管理策略，利用空闲时间将提交的待保存数据灌装到临时存储区；经过清洗、校验、打包封装后，进入储备库长期保存系统，最终完成数字资源在储备库的长期保存。工作机制如图7所示。

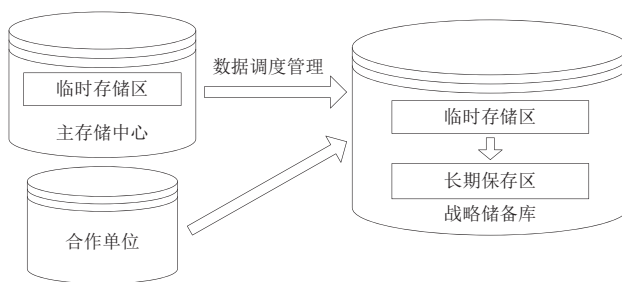


图7 合作单位提交数据长期保存

(3) 战略储备库生产的数据进入储备库存储中心、主存储中心，实现异地多备份的长期保存。储备库文献数字化系统生产的数字资源，不但要在储备库长期保存系统中实现长期保存，还要在国家图书馆主存储中心实现长期保存，最终实现数字资源异地互为备份的妥善保存。工作流程如8图所示。

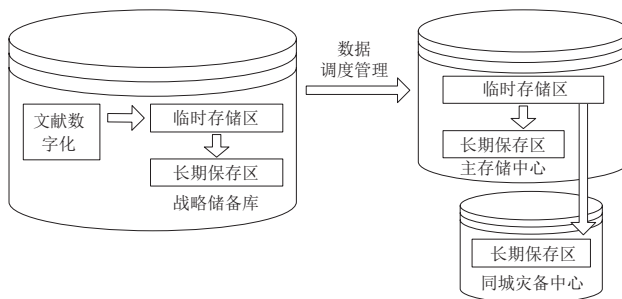


图8 储备库文献数字化数据长期保存

储备库文献数字化系统生产的数字资源，首先要统一存储于储备库的临时存储区；待验收合格后，进入储备库长期保存系统，最终完成数字资源在储备库的长期保存。通过数据调度，传输数字资源到主存储中心的临时存储区，经过TSM存储管理系统的统一控制，数据在主存储中心及同城灾备中心的长期保存区分别进



行长期保存, 最终实现“两地三中心”的长期保存。

## 6 结语

随着国家图书馆系统和资源存储服务体系、同城和异地灾备体系、永久保存体系不断完善, 国家图书馆可实现超过2 000TB数字资源的快速访问、安全存储和永久可用; 实现国家图书馆自建珍贵文献、广泛搜集各地馆藏中珍稀资源和特色文化资料的长期保存和永久利用, 实现传承文明、服务社会的宗旨。

随着互联网、大数据、移动互联网时代的开启, 人类文明的沉淀已经由图书、学术期刊作为主要载体, 转换为以互联网为主要承接载体, 使保存机构面临更艰巨的保存任务与更复杂的保存环境<sup>[10]</sup>。互联网资源如何进得来、存得下、可检索、能服务, 对国家、图书馆, 更对存储体系带来前所未有的艰巨挑战。战略数字资源保存与利用是国家信息基础设施的重要内容之一, 关系到国家的文献保护和文化遗产。作为国家重要文化机构的国家图书馆, 要参考国际通用或国家颁布的保存标准规范, 联合各机构共同实现国家文化保存和有效利用战略, 对中华文明传承、提高信息资源的国家控制力及经济社会发展具有重大而深远的意义。

## 参考文献

- [1] 国家气象局信息中心[EB/OL].[2017-02-01].<http://www.nmic.cn/web/index.htm>.
- [2] 中国银联[EB/OL].[2017-02-01].<http://cn.unionpay.com/>.
- [3] 文化部信息中心[EB/OL].[2017-02-01].[http://www.mcprc.gov.cn/sjzz/whbxxzx\\_2993/](http://www.mcprc.gov.cn/sjzz/whbxxzx_2993/).
- [4] 刘博涵, 路龙惠. 文献资源战略保存机制研究[J]. 国家图书馆学报, 2013(5):53-57,67.
- [5] 付鸿鹄, 吴振新. 分布式数字资源保存系统与技术架构研究[J]. 国家图书馆学报, 2015(2):82-88.
- [6] 肖珑, 申晓娟. 国家图书馆元数据应用总则汇编[M]. 北京: 国家图书馆出版社, 2011.
- [7] 全国信息安全标准化技术委员会. 信息安全技术 信息系统灾难恢复规范: GB/T 20988—2007[S]. 北京: [s.n.], 2015.
- [8] 全国信息安全标准化技术委员会. 信息安全技术 灾难恢复中心建设与运维管理规范: GB/T 30285—2013[S]. 北京: [s.n.], 2015.
- [9] The Open Archival Information System(OAIS) Reference Model: ISO 14721 [S/OL].[2017-02-01].<https://wenku.baidu.com/view/79eaf5a0284ac850ad02422a.html>.
- [10] 吴振新, 付鸿鹄. 数字信息资源分布式协作保存网络构建研究[J]. 数字图书馆论坛, 2016(9):43-48.

## 作者简介

王乐春, 男, 1971年生, 高级工程师, 研究方向: 数字图书馆体系结构、资源发现, E-mail: wanglch@nlc.cn。  
季士妍, 女, 1978年生, 高级工程师, 研究方向: 数字资源存储与应用。

## Research and Design of the Storage Framework of National Library

WANG LeChun, JI ShiYan  
(National Library of China, Beijing 100081, China)

Abstract: In this paper, the challenges on constructing the storage architecture for National Digital Library are analyzed in three aspects: systems, digital resources, and long-term preservation system. Based on the new requirements, we present three architectures: system service architecture, disaster backup and recovery architecture, and permanent preservation architecture. The proposed framework provides technical support for the storage and service of digital resources in National Library.

Keywords: National Library; Storage Framework; Hierarchical Storage; Disaster Backup and Recovery System; Long-Term Preservation System

(收稿日期: 2017-02-16)