

引用格式: 李景, 王益谊, 靳佳蕊, 等. 文献数字资源长期保存的理论与实践研究 [J]. 标准科学, 2026 (2):58-68.
LI Jing, WANG Yiyi, JIN Jiarui, et al. Practice and Research on Long-term Preservation of Literature Digital Resources [J]. Standard Science, 2026 (2):58-68.

文献数字资源长期保存的理论与实践研究

李景 王益谊 靳佳蕊* 申明爽 田方

(中国标准化研究院)

摘要:【目的】数字资源长期保存是图书情报领域持续多年的研究热点和重要课题, 它的研究成果对数字资源的保存、权益和使用具有深远影响和重要意义。【方法】采用文献调研法和比较归纳的方法, 调研了发达国家和我国在数字资源长期保存方面的重要研究项目。【结果】通过对7个国外项目和3个国内项目的对比分析, 列出当今全球通用的数字资源长期保存技术框架和相关技术标准, 并总结了中国标准化研究院国家标准馆多年来在标准文献数字资源长期保存方面的科研实践活动。【结论】如果仅以项目的形式开展长期保存研究, 则在可持续性方面存在隐患。需要从法律层面、政策层面和标准层面补足短板, 这样才能真正意义、实质性推动科技文献(含标准文献)数字资源长期保存战略的实施。

关键词: 长期保存; 数字资源; 标准文献

DOI编码: 10.3969/j.issn.1674-5698.2026.02.008

Practice and Research on Long-term Preservation of Literature Digital Resources

LI Jing WANG Yiyi JIN Jiarui* SHEN Mingshuang TIAN Fang

(China National Institute of Standardization)

Abstract: [Objective] Long-term preservation of digital resources has been a research hotspot and important topic in the field of library and information science for many years. Its research results have a profound impact on and significant importance for the preservation, rights, and use of digital resources. [Methods] This paper uses literature research and comparative induction methods to investigate important research projects in the long-term preservation of digital resources in developed countries and China. [Results] Through comparative analysis of seven foreign projects and three Chinese projects, a globally recognized framework and related technical standards for long-term preservation of digital resources have been listed and described. The scientific research and practical activities of the National Library of Standards in the long-term preservation of standard literature digital resources from 2008 to 2025 are also summarized. [Conclusion] If the long-term preservation research is only conducted in the form of projects, there are potential risks in

基金项目: 本文受国家科技图书文献中心(NSTL)科技文献信息专项项目(项目编号: 25K06S); 中国标准化研究院基本科研业务费项目“标准珍贵馆藏三维数字化保护及利用技术研究”(252024Y-11797)资助。

作者简介: 李景, 博士, 博士后, 研究馆员, 研究方向为标准文献、本体技术。

靳佳蕊, 通信作者, 硕士, 助理工程师, 研究方向为标准数据治理。

terms of sustainability. The gaps should be filled at the legal, national policy, and standard levels in order to truly and substantially promote the implementation of the national long-term preservation strategy for scientific and technological literature (including standard literature) digital resources.

Keywords: long-term preservation; digital resources; standard literature

0 引言

在图书情报工作中,长期保存(Long-term Preservation)通常是指数字资源的长期保存(Digital Resource Long-term Preservation)。数字资源长期保存是对数字形态资源进行持续维护的活动,旨在解决数字信息因技术过时、载体退化等风险带来的脆弱性问题。其目标包括维持信息的真实性、可访问性和可解释性,涉及技术与管理双重层面。

1 长期保存和一般保存的区别与联系

与长期保存相对应的,是图书情报机构对文献资源的一般保存行为。二者之间存在以下的区别和联系。

1.1 目标

长期保存旨在确保数字信息永久或尽可能长期维持可访问性、真实性和完整性,目标是解决技术过时、载体退化等问题,保障未来用户对信息的存取需求。

一般保存通常指短期维护,如数据备份或存储管理^[1],更多关注当前使用状态,缺乏对未来技术演变的预见性规划。

1.2 实施

长期保存采用多层级技术架构(物理层介质保护、逻辑层格式标准化、语义层元数据维护)。例如使用惰性气体封装光盘、定期格式迁移等技术延长存储介质上的信息寿命。

一般保存仅涉及基础备份(如定期拷贝数据),缺乏系统性迁移和格式转换机制。

1.3 工作流程

长期保存涵盖全流程管理,包括风险评估、制度规划、知识产权处理等。例如通过区块链记录操作轨迹,确保信息可追溯。

一般保存主要关注的是存储环节,可能会忽略技术迭代对信息可用性的潜在影响。

2 标准文献进行长期保存的必要性

科技文献是国家战略资源。对科技文献数字资源进行长期保存,早已被科技部、国家图书馆、中国科学院、国家科技图书文献中心(NSTL),中国高等教育文献保障体系(CALIS)等国家级图书情报机构和科技文献联盟列为战略规划中的重要内容。

标准文献作为科技文献的重要组成部分,所有对于科技文献数字资源长期保存的必要性,对于标准文献的长期保存同样适用。

2.1 维护标准信息真实性

数字资源易受技术更新影响,若不进行长期保存,部分内容可能因格式不兼容或载体老化而无法读取。元数据(如创建时间、责任者等信息)可确保数字档案的来源可追溯^[2],避免信息失真。

2.2 保障标准信息可获取性

数字资源依赖特定软硬件环境,若存储介质或软件不再支持,可能导致资源无法访问。长期保存需确保数据在不同技术环境下仍可读取,例如通过迁移至稳定格式或封装完整依赖环境。

2.3 应对技术更迭风险

硬件和软件快速迭代可能使现有存储设备或格式失效。例如磁带、光盘等传统介质存在物理退化问题,需通过迁移至更稳定载体(如硬盘或

数字胶片)延长寿命。

2.4 保障法律与合规要求

根据《电子档案管理办法》和《中华人民共和国档案法》,电子档案需与传统档案具有同等法律效力。标准文献数字资源的长期保存是履行市场监管义务的基础。例如,司法记录需永久保存(司法记录中包含执法记录,而执法记录又会涉及某强制性国家标准GB XXXXX,如果该标准内容不能长期保存,则会影响司法记录的保存时间)。

2.5 应对自然灾害和不可抗力

2006年,台湾海峡地震,致使海缆断裂,影响国内科研人员上网查询和获取国外科技文献。该事件令NSTL、中国科学院等单位意识到了数字资源长期保存的必要性。对数字资源进行长期保存,能够应对自然灾害和不可抗力对资源实体馆藏带来的毁灭性破坏。建立地理分散的备份站点,确保单一灾害(如洪水、火灾、地震、台风等)不会同时破坏所有副本。例如,采用磁带、硬盘与云端结合的混合备份模式,兼顾成本与恢复效率。

3 国外调研

3.1 电子文件档案馆(ERA)

1997年,美国国家档案与文件署(NARA)开发的电子文件档案馆(Electronic Records Archives, ERA),用于获取、管理、保存和使用美国联邦政府的永久电子文件。ERA基础系统(ERA1.0系统)是美国首个数字档案馆项目,实现了电子档案的长期存储和开放利用等基本功能。ERA目前是美国联邦政府的核心数字档案管理系统。为了更好地履行法定职责,NARA于2015年正式启动ERA2.0项目。ERA在2018年已推出核心功能模块,2020年完成了所有系统功能开发。该系统功能已基本实现。

3.2 多备份资源保存系统(LOCKSS)

由美国斯坦福大学图书馆主持的LOCKSS(Lots of Copies Keeps Stuff Safe)是多备份资源保存系统的简称,项目于1999年启动实施,获得了美

国国家自然科学基金、Sun公司和Mellon基金的资金支持。LOCKSS底层基于JAVA的开放源代码,是分布式数字资源保存系统,具有投入成本较低、保存内容较多,保存效果安全可靠的特点。LOCKSS依照OAIS模型框架设计了一个同时拥有通用功能、可扩展功能和互操作功能的模型,并通过建立出版商—图书馆、图书馆—图书馆的互动平台,提供电子期刊保存和快速访问服务,支持授权图书馆从出版商网站直接下载数字资源,从而实现本地保存,并向用户提供持续服务。

3.3 美国国家数字信息基础设施和保存计划(NDIIPP)

由美国国会图书馆主导,于2000年12月启动实施的NDIIPP(National Digital Information Infrastructure and Preservation Program),是美国国家数字信息基础设施和保存计划的简称。NDIIPP联合了部分美国的大学图书馆、信息机构等共同开展。NDIIPP包括8个子项目,涵盖的数字资源范围广泛,并构建了美国国家层面的数字资源长期保存体系。NDIIPP的示范作用和借鉴意义影响深远,堪为标杆。

NDIIPP的特点包括:(1)非常关注数字资源版权问题,强调各个业务流程的主体应承担各自在数字资源长期保存中必须承担的责任与义务;

(2)NDIIPP并没有花很大代价去开发一个全新系统,而是提供了优化的功能集合,让图书馆对此进行评估比较,并由此产生成本较低且更加实惠的技术方案。

3.4 加拿大数字保存计划(SDPP)

加拿大国家图书档案馆(Library and Archives Canada, LAC)作为加拿大政府档案和具有历史价值的出版物与文件的永久存储库,大范围地获取、保存和管理加拿大文献,并提供对其数字资源的持久访问。LAC于2017年11月30日发布的《数字保存计划战略》(Strategy for a Digital Preservation Program)作为加拿大国家层面数字保存计划的重要组成部分,建立了一套符合ISO 16363标准的可以全面运作的数字保存计划。

3.5 欧盟CASPAR项目

CASPAR是欧盟第六框架计划资助下的一个综合项目,于2006年4月1日正式启动,致力于研究、实施和传播基于OAIS参考模型(ISO 14721)的数字保存创新解决方案。

CASPAR是一个涵盖科学、文化和创意方面的重要数字资产联盟,也是数字信息保存领域的商业合作者和世界领先者,由包括联合国教科文组织、欧空局、格拉斯哥大学等在内的共17家单位组成。为应对可能存在的风险和挑战,CASPAR成立了由与保存数字信息有利害关系并分享、支持项目使命的个人和组织组成的全球网络,包括数字资源创造者、馆长(博物馆、美术馆、档案馆等)、服务供应商、记忆机构、研究人员和用户。

发展至今,CASPAR已经形成了相对完整的功能模型、工作流、信息打包方法等,在数字信息保存实践中发挥着重要作用。欧空局总部的数据保存、法国现代音乐研究所和法国国家视听研究院音乐表演的数字信息、联合国教科文组织在保护文化遗产方面的文化数据等都有CASPAR组件、系统发挥作用的影子。

3.6 保护和存取澳大利亚网络信息资源项目(PANDORA)

澳大利亚国家图书馆主持的PANDORA(Preserving and Accessing Networked Documentary Resources of Australia)项目,是保护和存取澳大利亚网络信息资源项目的简称。该项目始于1996年,旨在建设开发含有需要长期保存资源的数字信息资源归档系统。收录对象主要包括政府出版物、部分教育出版物、电子期刊、网站等。

3.7 德国数字资源长期存储专家网络(NESTOR)

由德国国家图书馆承担主持、德国多家大学图书馆和信息机构一起参与的NESTOR(Network of Expertise in Long-term Storage of Digital Resources)项目,是德国教育和研究委员会的资助项目。NESTOR非常重视用户宣传,注重培养大众数字资源长期保存意识(信息素养教育的重要内容),联合国家公益信息机构、科技行业等机构,

共同组建德国国际战略联盟,协同开展数字资源长期保存活动。

4 国内调研

4.1 NDPP

NSTL从建立之初就充分认识到建设国家数字文献长期保存体系的重要战略意义。2013年7月,经科技部批准,中国科学院文献情报中心(原中国科学院图书馆)作为牵头单位,组织开展国家数字科技文献资源长期保存体系(National Digital Preservation Program, NDPP)建设工作,以NSTL主要成员单位和少数重点大学(中国科学院文献情报中心、北京大学图书馆、中国科学技术信息研究所)为核心进行NDPP的建设工作,如图1所示^[3]。NDPP按照“国家主导,联合参与,可靠管理,公共服务”的原则,重点对国际重要且由国家财政经费斥资订购的数字化科技文献资源进行长期保存,建成国家保存中心网络系统,并逐步形成可靠的长期保存管理和运行机制。这是因为10年前数字资源就已经成为科技文献的主流资源,NSTL充分认识到建设NDPP的重要战略意义。

随着数字化、网络化的普及,数字文献资源已经成为科技领域的主流信息资源,学术期刊和会议录已经形成以数字出版为主的形态,学术专著也迅速走向数字出版,同时多数开放出版期刊以数字版为其唯一正式出版形态(e-only)。在我国科研教育机构中,数字文献的使用已经远远超过印本文献。主要科研教育单位均已将数字文献作为自己的主流科技文献资源,并不断削减纸本文献订购。同时,与传统的文献资源相比,数字资源需要特别的监护和管理。

经过两年的建设,NDPP于2016年初步形成了“1+3”架构的国家分布式协作保存网络,2016年之后逐步进入稳步发展期,直至2023年5月,长期保存中心正式成立并承担NDPP新阶段的建设任务和发展目标。

在建设阶段(2013—2020年),NDPP就制定

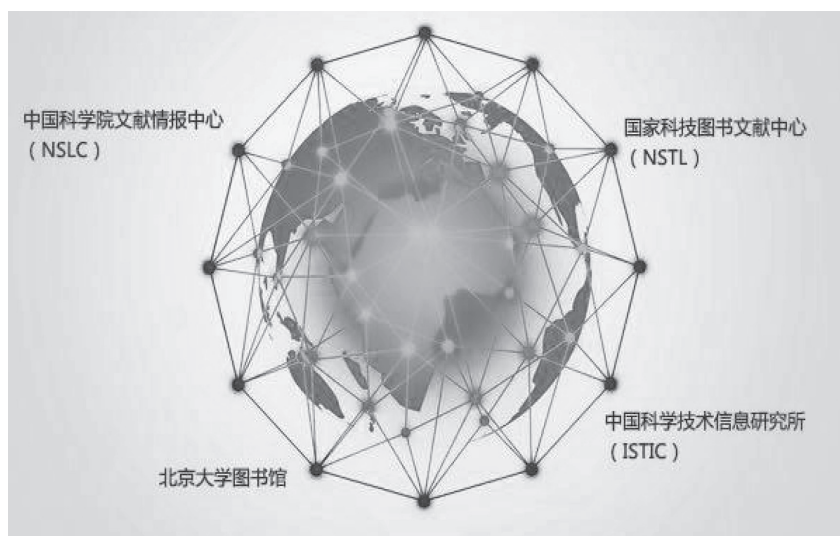


图1 NDPP项目建设最主要组成单位

了长期保存体系建设规划,开展了技术、标准、制度规范等研究,开发出实验性系统,并建设了示范节点,在对重要资源进行长期保存的同时,完善各项规范,构建协作保存网络,开展非文本数字资源的保存研究工作(见图2)。现阶段,NDPP持续扩展保存范围和规模,更加注重加强标准和规范建设,更加关注资源的有效利用,并提出了更高的战略目标。

NDPP建设项目建设成果主要包括以下3个方面:

(1) 构建了可信赖的长期保存系统。构建了遵循OAIS标准和可信赖仓储系统认证标准^[4-6]的

保存系统。该系统实现了存档资源全生命周期的管理,提供了基于订购协议的审计管理和访问控制,建立了完善的系统安全管理机制和多级防灾备份机制^[3]。

(2) 建设了全面、规范的长期保存管理及运行机制。从战略目标、保存责任与目标、权益保护与管理、可信赖系统^[4]、合作保存等方面提出国家保存体系长期保存战略框架。目前,NDPP已经形成了长期保存的可靠工作流程和配套的工作规范,确保了长期保存的可持续性和可靠性(见图3)。

(3) 实现了一定规模的国外数字科技文献资源的本土化保存。NDPP已实现Elsevier、Springer

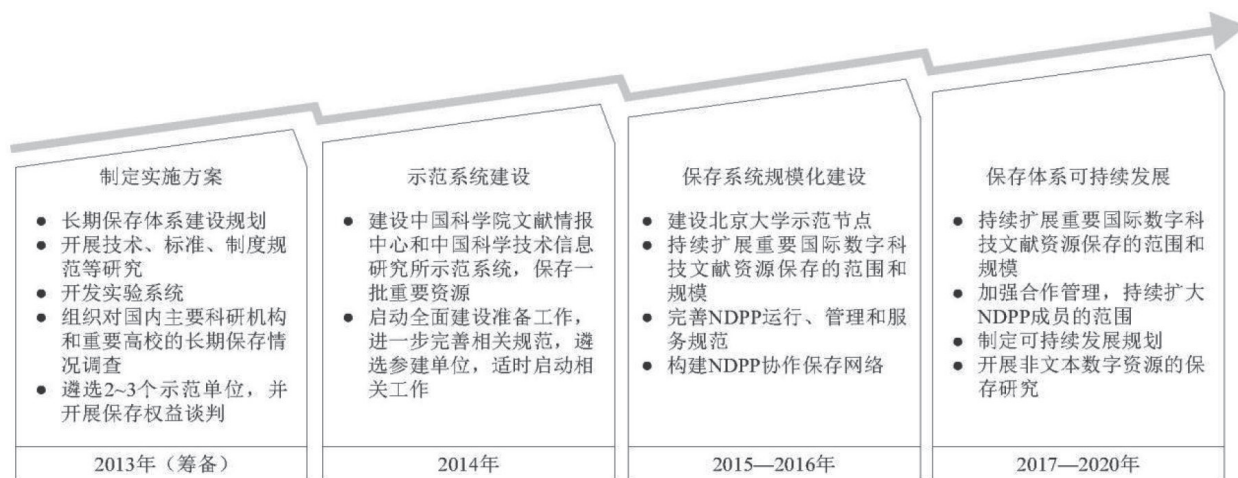


图2 NDPP发展路线图

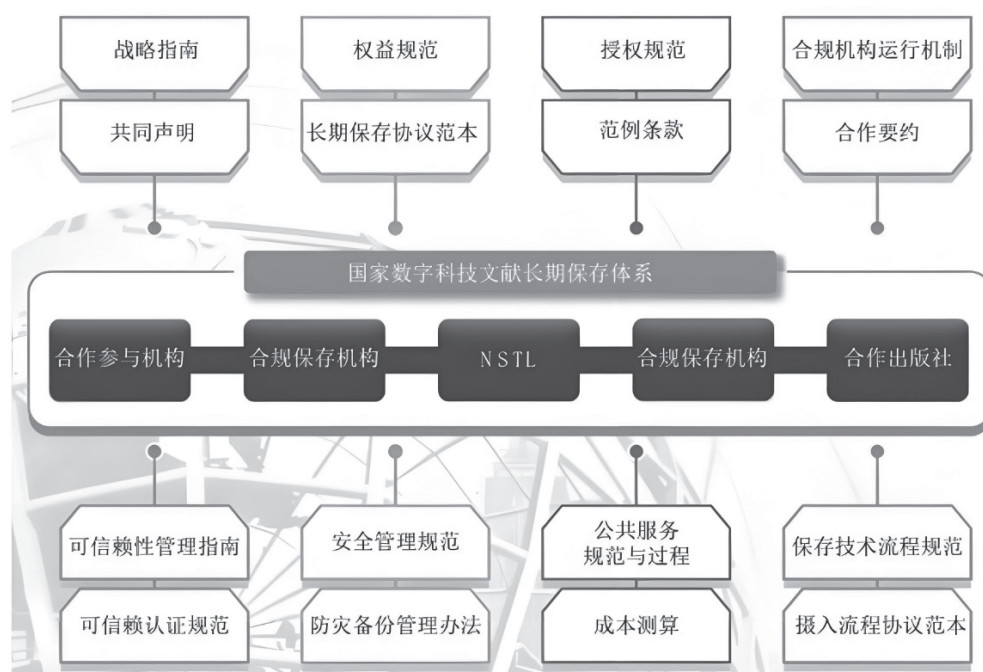


图3 长期保存管理和运行机制框架图

Nature、Taylor & Francis、Wiley等国际主要学术出版机构和重要专业学会出版数据库的自主长期保存，保存资源类型包括外文科技期刊、中文科技期刊、学位论文、事实型数据库、数值型数据库、电子书、应用软件和预印本等。

4.2 WICP项目

WICP是网络信息采集与保存试验项目（Web Information Collection and Preservation）的简称。WIPC是中国国家图书馆于2003年启动实施的数字资源长期保存研究项目。该项目通过试验性采集保存数据并开放用户服务，最终形成数字资源长期保存的整体方案。在实施层面，WIPC采用国际互联网保存联盟推荐的工具包进行采集试验，联合其他机构共建自动化选择性收集系统，并对采集内容实施分布式保存。

4.3 国际敦煌项目（IDP）

IDP是国际敦煌项目（International DunHuang Project）的简称。IDP于1993年启动。以IDP项目为纽带，全球多个敦煌西域文献收藏组织加入该项目，共同开展敦煌西域文献保护和数字化工作。IDP汇集了来自东方丝绸之路的线上资源，并促

进了对该地区历史和文化的了解。IDP项目于1998年正式上线。截至2011年，IDP已在英国、中国、日本、韩国、俄罗斯、法国、德国设立了8个中心，分别更新维护中、英、俄、德、日、法、韩7个语言版本的网站平台门户和数据库。

本文对国内和国际项目进行列表对比，具体见表1和表2。

5 长期保存的核心参考模型（OAIS）和相关国际标准

开放档案信息系统（Open Archival Information System, OAIS）是数字资源长期保存通常采用的核心参考模型。

OAIS参考模型由美国国家航空航天局（The National Aeronautics and Space Administration, NASA）与美国太空数据系统咨询委员会（Consultative Committee for Space Data Systems, CCSDS）于1999年联合制定，并于2003年正式成为国际标准ISO 14721《空间数据系统操作规程 开放式档案信息系统（OAIS）的参考模型》。目前

表1 国内项目比较表

项目名称	NDPP项目	WICP项目	IDP项目
启动时间	2013年	2003年	1993年
建设目标	从国家层面，从科技文献保障层面，实现对国际重要数字科技文献资源的长期保存	对数字资源采集、整理、编目、保存和利用中存在的问题进行分析研究，并提出可行性解决方案；确认目标对象，并有针对性地选取保存技术和保存策略；试验性的数据采集保存数据并为用户提供服务；提出数字资源长期保存的方案	实现珍贵敦煌古籍文献数字化，并开展长期保存，解决困扰学界多年的敦煌古籍文献资料获取难的问题。并为世界文化遗产的数字化保护提供范式和可借鉴的方案
实施策略	分层协作：以1+3核心机构（国科图、中信所、北大）主导，构建分布式保存网络； 分阶段推进：2013—2020年建设体系，2023年后持续扩展资源	流程化操作：目标确认→版权协商→采集计划→内容标准化→编目上传； 合作共建：协同多机构创建自动化采集系统，实施分布式保存	跨国协作：在8国设立中心，统一标准分工； 多维推广：展览、讲座、教育项目普及资源； 学术服务：建设专题数据库支持研究
技术架构（标准、技术等内容）	标准框架：遵循OAIS标准及可信仓储认证； 系统设计：多级防灾备份、审计管理、访问控制； 协作网络：国家分布式保存中心架构	工具链：采用国际互联网保存联盟工具（如Heritrix爬虫、Wayback回放）； 系统架构：自动化选择性采集系统 + 分布式存储	工具链：采用国际互联网保存联盟工具（如Heritrix爬虫、Wayback回放）； 系统架构：自动化选择性采集系统 + 分布式存储 数字化标准：统一高分辨率扫描、元数据规范（Dublin Core扩展）； 多语言平台：7语种网站与数据库； 协作系统：跨国联合目录与分布式镜像
主要成果	①构建了可信赖的长期保存系统；②具有一定规模的国外数字科技文献资源的本土化保存得以实现；③全面、规范的长期保存管理运行机制基本形成	①建立试验性网络信息资源保存流程；②形成分布式协作保存模式（公开信息较少）	①全球8个中心协作网络；②多语种数据库（7种语言）；③建成敦煌文献联合目录等6类专题库；④数字化10万+件文献
资金来源	财政经费	中国国家图书馆	中国国家图书馆

2025版本代替了2012版本（2012版本代替了2003版本）。OAIS参考模型为数字保存计划提供了一套标准的定义和活动，以帮助指导其开发和运行，从系统架构的层面，对数字资源长期保存系统的作用、行为、信息模型、功能和服务、数据迁移与交换接口等进行了高度的抽象和表达，还提供了设计数字资源长期保存的术语和基础模型。在CCSDS与ISO

的推动下，档案馆、图书馆、各专业领域的科学数据中心，以及文化、古籍等领域的长期保存项目都愿意采用OAIS进行项目顶层设计。OAIS参考模型分为功能模型、信息模型、信息包。信息包被用来在OAIS功能模块之间进行交换传递。

5.1 OAIS功能模型（Functional Model）

OAIS 的功能模型涉及6个功能模块，分别为

表2 国外项目比较表

项目名称	ERA (美国)	LOCKSS (美国)	NDIIPP (美国)	SDPP (加拿大)	CASPAR项目 (欧盟)	PANDORA项目 (澳大利亚)	NESTOR项目 (德国)
启动时间	1997年 (ERA1.0) 2015年 (ERA2.0)	1999年 (斯坦福大学)	2000年 (美国国会批准)	2017年	2006年	1996年	2003年
建设目标	获取、保存、管理并使用美国联邦政府的永久电子文件	通过分布式本地缓存系统, 确保电子期刊资源的永久访问	通过保存平台的搭建、完善保存合作框架体系, 实现具有重要历史价值数字资源的保存, 最终建设成为美国国家战略层面的国家数字资源长期保存体系	愿景目标: 到2024年, LAC已构建了基于ISO 16363标准的可持续性数字保存计划, 并能实现OAIS参考模型的所有功能	致力于研究、实施和传播基于OAIS参考模型 (ISO 14721) 的数字保存创新解决方案	建成一个包含需要长期保存资源的数字信息资源归档系统	(1) 培养公众数字资源长期保存意识 (信息素养教育) (2) 加大宣传力度 (3) 开发网络平台 (4) 完善配套技术体系, 包括信息采集技术、信息保存法律法规等; (5) 联合国家公益信息机构、科技行业信息机构, 共同组建德国国际战略联盟, 协同开展数字资源长期保存实践。
实施策略	分代迭代: ERA1.0 (1997-2015): 基础功能实现; ERA2.0 (2015-2020): 模块化重构 (DPE/BOM/DOR)。模型驱动: 严格遵循OAIS功能模型	去中心化备份: 图书馆本地缓存出版商资源; 失效容灾: 出版商服务中断时由缓存系统接管访问	最小化干预: 不替代现有系统, 提供保存功能基础框架; 多方协作: 联合政府、图书馆、信息机构共建国家保存网络	第一阶段: 信息收集。第二阶段: 计划制订。第三阶段: 计划实施	跨域联盟: 17家机构 (教科文组织、欧空局等) 分工协作; 案例验证: 在科学数据 (欧空局)、文化遗产 (教科文组织) 等场景落地应用	选择性采集: 合作机构提名高价值资源; 双系统支撑: PANDAS (协作管理平台); Trove (统一检索门户); 全民服务: 免费开放访问	平台化运作: 建立专业门户网站整合技术/法规/研究资源; 联盟共建: 联合图书馆、科研机构、政府部门组建战略联盟
技术架构 (标准、技术等内容)	核心架构: DPE (数字档案处理环境); BOM (业务对象管理) DOR (对象存储库)。标准框架: 完全兼容OAIS参考模型	技术栈: Java开源分布式系统; 核心机制: 内容完整性校验; 自动格式迁移; 分布式冗余存储	开放框架: 提供可扩展的保存功能模块集; 权责规范: 明确资源创建者/分发者/保存者的版权责任链	认证导向: 以ISO 16363可信数字仓储 (TDR) 认证为目标 OAIS落地: 实现摄取/存储/访问全流程功能	创新组件: 表示信息增强技术; 虚拟化访问服务; DRM集成系统。标准贡献: 扩展OAIS模型并反馈ISO标准	核心系统: PANDAS (基于Web的协作归档系统); Trove (国家级发现服务平台)。元数据: 专用保存元数据方案	知识门户: 在线平台提供技术指南/法律框架/最佳实践; 协作网络: 分布式任务分配与成果共享机制

续表2

ERA (美国)	LOCKSS (美国)	NDIIPP (美国)	SDPP (加拿大)	CASPAR项目 (欧盟)	PANDORA 项目 (澳大利亚)	NESTOR项目 (德国)	NESTOR项目 (德国)
主要 成果	①管理50TB+ 联邦政府电子 档案; ②服务 1000+机构用 户; ③完成两 代系统迭代 (ERA1.0/2.0)	①全球 600+图 书馆参与 网络; ② 保存数万 种电子期 刊; ③开 创“图书 馆-出版 商”协作 保存模式	①建成8大 子项目覆 盖多类型资 源; ②制定 《国家数字 保存联盟框 架》; ③推 动《数字公 平法》版权 条款修订	①发布《数 字保存计划 路线图》; ②完成ISO 16363差距 分析; ③建 立成本估算 与资源分配 模型	①开发 CASPAR核心 组件套件; ②完成科学/ 文化/创意领 域验证; ③ 贡献15+项 OAIS扩展标 准提案	①保存5万+ 澳大利亚网 站资源; ② 建成Trove国 家级检索平 台(年访问1 亿+); ③形 成9机构协作 网络	①开通专业门户; ② 发布《德国数字保存 白皮书》; ③建立跨 机构认证培训体系

摄取模块、存储模块、数据管理模块、获取模块、系统管理模块及保存规划模块。

5.2 OAIS信息模型 (Information Model)

OAIS 对信息的定义是可以用于交换的任意类型的知识, 且信息总是由某种类型的数据表示。即通过表征信息解释的数据从而产生信息。为了成功保留这些信息对象, OAIS对信息模型进行了阐明(见图4)。OAIS的信息模型通常包含了3类对象: (1) 信息对象 (Information Object); (2) 数据对象 (Data Object); (3) 表征信息 (Representation Information)。信息对象 (Information Object) 通常由数据对象 (Data Object) 构成, 并通过与表征信息 (Representation Information) 结合成为用户可以理解的内容。数据对象 (Data Object) 通常是实体数据。表征信息 (Representation Information) 通常是指将数据对象以一种具有实际意义(或语义)的方式呈现给用户。



图4 OAIS 信息模型对获取信息的阐述

信息提供者每次向OAIS提交信息, 以及OAIS每次向用户传播信息都是通过一次或多次的传

输完成的。因此, 这里所指的信息包是指能够在OAIS功能模型的各个模块间被传递和交换的信息表。信息包通常包括: (1) 内容信息; (2) 保存描述信息; (3) 封装信息; (4) 包描述信息。

展开来说, 信息包中的封装信息是一个容器, 这个容器的作用是将内容信息和保存描述信息联系、封装和识别。内容信息、保存描述信息通过封装信息构成一个整体。信息包中的内容信息通常是指被保存的原始信息对象。信息包中的保存描述信息 (Preservation Description Information, PDI) 通常是用来描述内容信息的特征, 并使其得以完整保存的必要信息。包描述信息用于描述信息包的特性和属性。信息包是包含内容信息和PDI两种信息对象的容器。这些信息包可用于构建和存储OAIS的存储, 将所需信息从信息提供者传输到OAIS, 或在OAIS和用户之间传输所请求的信息。每种功能都有不同的信息要求。信息包包含0或1个内容信息对象, 0个或更多PDI对象, 并且只与一个封装信息相关联。信息包还与0个或更多描述内容对象的包描述相关联, 以实现有效访问。

在提出信息包的概念后, 有必要对OAIS长期保存的信息包、信息提供者提交给OAIS的信息包, 以及从OAIS传播给用户的信息包加以区分。

信息包在OAIS模型中包括3种类型: SIP、AIP、DIP。(1) SIP通常是指信息提供者(Provider)交给OAIS的信息包, SIP从形式到内容都可以由Provider和OAIS协商来决策和定制。(2) AIP是符合长期保存标准的信息包。摄取功能将一个或多个SIP转换为一个或多个AIP以进行保存。AIP具有相关内容信息的完整PDI。AIP的包信息应符合OAIS内部标准。(3) DIP则可以根据用户的需求传递给用户, DIP是从一个或多个AIP中提取出来的, 用户需求是DIP的首要特征, DIP可以具有或不具有完整的PDI。

需要说明的是, 开放档案信息系统模型(OAIS)的存在并不是只为给数字资源长期保存提供模型框架。建立OAIS的初衷是提供一个共同遵循的原则和参考模型, OAIS的应用因为不同领域实际状况的差异而大有不同。即便是应用于不同长期保存项目中的OAIS模型, 因项目目的和资源范围的差异, 也会呈现出很大差异。在对标准文献长期保存项目进行设计和部署前, 需对长期保存实践项目所采用的技术和环境进行深入的研究和比较, 从而找出符合标准文献数字资源长期保存需求的设计和实施方案。

6 国家标准馆在长期保存方面的科研与实践

早在2008年, 国家标准馆就开始进行长期保存的项目研究和实践。依托项目研究完成了长期保存的文献调研, 并形成了针对标准文献数字资源长期保存的初步方案。同时制定了面向资源采集、流通和数据加工业务流程的长期保存制度(草案)。

2016—2019年, 宫轲楠在其博士后出站报告《标准文献数字资源长期保存及分析研究》中, 结合分布式策略以及OAIS参考模型, 针对标准文献数字资源的长期保存需求进行分布式长期保存系统的整体设计。分别从分布式长期保存策略、保存技术选择、具体技术方案等方面出发, 确保分布式长期保存系统符合OAIS要求以及标准馆藏机构

的联合参与, 以达到标准文献真实性、完整性和有效性的目的。

2022至2025年, 国家标准馆田方在“标准科学数据长期保存政策机制研究”项目研究中, 聚焦标准科学数据, 调研了ISO 14721、PREMIS等基础框架, 分析了国内在政策和法律层面的现存问题, 对统一协调的保存机制和保存体系建设提出了方案。该项目立足于《国家标准化发展纲要》和国家数字标准馆建设, 首次对标准科学数据的长期保存提出了保存体系建设方案(初稿)。

7 总结与建议

标准文献数字资源的长期保存不仅是应对标准化领域数字资源脆弱性的必要手段, 更是支撑国家数字化转型、促进数字经济发展的抓手之一。对标准文献数字资源进行长期保存开展政策机制研究, 不仅有利于从国家层面在数字化时代对数字资源的生产和开发利用进行顶层设计, 推动数字资源长期保存的立法, 更能促进全社会对数字资源脆弱性的认知, 进而更有效地开展保护和研究, 在未来更好地传承人类的科技成果。

基于已有的文献综述和前期项目, 标准文献数字资源的长期保存存在以下几方面问题并附建议:

(1) 在法律层面, 缺乏对标准文献数字资源呈缴制度的立法^[7]。这使得谁是长期保存的主体, 谁应获取长期保存的权益, 谁应履行长期保存主体责任, 谁应尽到长期保存义务等问题无法得到法律认可。所以, 以科研项目来支持长期保存, 在长期保存的可持续性方面存在潜在风险, 在长期保存的主体责任方面存在很大程度的模糊性。这些问题亟待相关立法来解决, 并应在法律条文中明确长期保存的主体及其应承担的法律责任。

(2) 在政策层面, 长期保存缺乏国家的统一协调机制。这对谁保存, 谁参与保存, 保存什么, 保存多久等技术和业务流程中无法回避的问题难以从政策层面找到答案, 从而无法获取到国家政策 and 战略层面的稳定经费来源和投入。特别是对

于专业和中小型图书情报机构, 如果其数字资源无法纳入国家长期保存体系中, 一旦出现重大灾害和不可抗力, 对于固定资产(财政经费购买的数字资源和印本资源均是固定资产)和数字资源的损失将无法弥补和挽回。因此, 需要在国家层面进行顶层设计, 建立覆盖所有科技文献品种和科学数据的数字化资源长期保存的国家战略, 并出台相关政策与实施细则。

(3) 在标准层面, 目前虽然有相关的国际标准支持长期保存的技术参考模型, 但是国内仍然缺乏面向数字资源长期保存的专业的标准体系, 并缺乏覆盖长期保存各个技术环节的有针对性的技术标准。同时, 在保存地点、保存条件、硬件和软件配置, 启动保存要求、资源使用条件等方面缺乏标准要求和指南。应伴随长期保存工作和项目的实施, 有计划、有步骤地制定相关技术标准, 并逐步完善标准体系。

(4) 在应用层面, 对数字资源长期保存的效果缺乏相关的标准规定与绩效评价机制。与数字资源长期保存相对应的是, 保存的目的是否仅仅是保存, 而不考虑应用? 笔者认为, 长期保存是为

了长久的应用。不以应用为目标的数字资源长期保存, 缺乏可持续发展的内在动因。对数字资源的长期保存应充分考虑保存主体(特别是资源数量有限的专业图书馆和中小型图书情报机构)的主观能动性。有些专业图书馆, 本身具有特色馆藏资源, 从资源角度, 具备长期保存的价值。但是因为缺乏相关的激励机制, 缺乏相关的保存条件, 缺乏经费, 而无法开展数字资源长期保存工作。多年来, 在国家财政和地方财政的支持下, 各类资源和数据的产生与衍生产品层出不穷。而承担这些资源和数据保存职能的图书情报机构不应只是被动的承受者, 应该在数字资源的整个生命周期中, 成为承上启下的保存单位, 对资源进行二次开发利用的研发单位, 对社会进行宣传推广的窗口单位。同时还是承担长期保存主要责任的主体责任者和享受长期保存效果的既得利益者。只有建立相应的激励机制并有计划、有步骤地引入市场调节手段, 才能让数字资源长期保存成为良性发展的事业, 并让保存主体得到参与长期保存的红利, 维系机构持久并可持续地健康发展。

参考文献

- [1] ISO/TC 46/SC 11. Information and documentation — Trusted third party repository for digital records: ISO 17068:2017[S]. Switzerland: International Standardization Organization, 2017.
- [2] 信息技术 OFD档案应用指南: GB/T 42133—2022 [S]. 北京: 中国标准出版社, 2022.
- [3] NSTL. 国家数字科技文献资源长期保存体系建设[EB/OL].(2024-09-01)[2025-06-22].https://jtys.nstl.gov.cn/Portal/kxyj_cqbext.html.
- [4] ISO/TC 20/SC 13. Space data and information transfer systems — Audit and certification of trustworthy digital repositories: ISO 16363:2025[S]. Switzerland: International Standardization Organization, 2025.
- [5] ISO/TC 20/SC 13. Space data and information transfer systems — Requirements for bodies providing audit and certification of candidate trustworthy digital repositories
- [6] Trusted digital repositories: attributes and responsibilities [EB/OL].(2023-01-15)[2025-06-25].<https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>.
- [7] 李景. 国内外图书馆法中呈缴本制度对标准文献呈缴归档的启示[J]. 标准科学, 2010(4): 43-49.