

国家图书馆数字资源长期保存 数字迁移实践*

季士妍

(国家图书馆, 北京 100081)

摘要: 文章在介绍国内外数字资源长期保存领域以及数字迁移方面的实践和研究进展的基础上, 分析了国家图书馆数字资源长期保存的情况, 提出数字迁移的必要性。文章详细介绍了国家图书馆数字迁移项目实施的流程、数据整理的规范、数据存储的组织结构、数字迁移的质量控制。相信对于数字迁移项目的介绍, 可以为其他单位开展相关工作提供一定的参考意义。

关键词: 国家图书馆; 长期保存; 数字迁移

中图分类号: G203

DOI: 10.3772/j.issn.1673—2286.2015.02.004

1 引言

随着网络技术和信息技术的发展, 数字资源逐渐成为信息生产、信息传播和信息利用的主要信息资源。图书馆作为信息收集和服务的主要机构, 应该重视数字资源的大规模存储与长期保存问题。在数字资源长期保存工作中, 保存介质的不稳定性和快速淘汰性, 保存环境和保存策略的主观性和调整性, 都对数字资源可以被永久访问和持续利用提出了挑战。因此, 图书馆在实施数字资源长期保存工作的实践中, 要采用合理的技术手段和策略来保障长期保存的数字资源的长期可用性、可表现性、可理解性、可识别性和真实性^[1]。根据OAIS (Open Archival Information System) 参考模型的长期保存规划, 为维护数字对象的长期可生存能力、可呈现能力和可理解能力, 数字迁移被认为是一种有效的保存策略^[2]。本文基于国家图书馆的数字资源长期保存工作的具体实践, 对数字迁移策略进行分析和研究。

2 数字迁移概述

2.1 数字迁移的概念

数字资源长期保存指采取一定的管理策略、技术策略, 对数字资源进行长期的存储管理, 保障数字资源的长期生存能力、可呈现能力和可理解能力^[3]。在数字资源长期保存工作中, 可以采用的保存技术包括更新、迁移、备份、仿真、数据恢复与数据考古、风干、规范化^[4]。其中, 数字迁移是较为广泛使用的一种保存技术策略。

针对“数字迁移”较为权威的定义, 是美国保护与存取委员会 (CPA) 与美国研究图书馆小组 (RLG) 组建的数字信息归档特别工作组 (CPA/RLG工作组) 在其报告中给出的定义^[5]: 数字迁移是指随技术变化定期地转换数字资源的一种处理过程, 它使数字资源从一个硬件、软件的配制向另一个硬件、软件配制转换, 或是从旧的计算机技术向新一代的计算机技术转换。数

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

字迁移的目的是在面对不断变化的技术时,保留数字资源的展示、检索、操作和使用能力。CPA/RLG工作组认为,数字迁移是数字归档的一个重要功能。根据此定义可以看出,当数据格式过时、存储环境变化、技术发展时,使用数字迁移策略是一个比较适合的方法。

2.2 数字迁移实践调研

针对数字迁移的研究,国内外的国家图书馆、大学图书馆以及信息保存机构均进行了多方面的研究与实践。在理论研究方面,国外相关机构开展的较早,研究的较为全面和系统,并且在研究理论的同时,也更重视实践的指导性和适用性研究。而在国内,大量的研究着眼于理论、政策研究。

在实践研究和经验分享方面,国外相关机构开展的实践项目较多,有很多项目是通过大学图书馆与信息保存机构进行合作,进行了较为深入的研究和实际操作实践,有的机构也研制出较为适用的系统,可以对类似的实践工作给予指导和借鉴。相应地,国内相关机构在实践中的经验较少,特别是在长期保存工作中真正实施数字迁移的实践经验很少。

经过调研,从数字迁移项目的指导性、实践性、适应性等角度考虑,在长期保存工作中可以具体参考和借鉴的项目有以下3个:

① 美国康奈尔大学图书馆在CLIR(The Council on Library and Information Resources)的资助下,进行了常见数字资源迁移风险的调研,并给出了图书馆数字迁移的实践指导,总结了数字迁移各步骤的辅助工具,同时,还对图像文件和数字文件的操作指导了具体的方法^[6]。

② 美国CPA/RLG工作组^[5],负责调研与推荐能确保“数字文件未来长期存取”的方法。他们的报告中指出,能完全适应适合数字保存需要的既不是更新,也不是仿真,而是数字迁移。迁移策略包括存储介质迁移、文件格式迁移。

③ 由Stanford大学组织实施的LOCKSS项目(Lots of Copies Keep Stuff Safe)^[7]是第一个并且唯一一个应用电子资源“Purchase-and-Own”机制的项目,致力于实现电子期刊的保存与利用。该项目保存所有网络出版格式(动画、数据集、移动图像、动态图像、软件、音频、文本)以及所有类型的文献(期刊、书籍、博客、网站、扫描文件、音频、视频)。该项目在保

存原始文献基础上,采用“按需迁移”(Migration on Access)^[8]的策略,当用户需要在线阅读时,直接从原始格式转换为适合阅读的格式,并且自动保存下来,从而为用户提供最新、最好版本的学术材料。

2.2 数字迁移策略分析

根据调研,目前存在三种类型的迁移策略:存储介质迁移、文件格式迁移、按需迁移。这三种策略各自有其适用的方面,也分别适用于不同的环境及数据情况。

(1) 存储介质迁移

根据CPA/RLG工作组的定义,存储介质迁移是将数字资源从不稳定的介质迁移到更加稳定的介质上。例如,将文件从软盘存储迁移到硬盘存储,耶鲁大学图书馆在其project Open Book项目中对缩微胶片内容进行数字化等,均为采用此类策略进行的迁移。但是,对于一些非标准的数据,当它们的访问方式发生变化时,为了实现数据迁移,常常不得不改变文件结构,有时会破坏嵌套在数据库里的数据关系,甚至是通过软件来管理和解译的认证方式都会发生变化。因此,对于复杂系统中的复杂数据的迁移,存储介质迁移并不是很适用。

(2) 文件格式迁移

根据CPA/RLG工作组的定义,另一种格式迁移策略是对于海量复杂和多样性的数字资源,在不影响数据内容和功能的前提下,将纷杂多样的数字对象的格式简化成一些易操作的标准格式。在长期保存工作中,当数字资源的文件格式出现过时等危机,有文件无法存取的危险时,需要将原始的格式转化为新的版本格式;为了保证保存系统中的数字内容的长期有效性,需要将保存的数字资源格式转化为较为标准化的格式,也需要进行文件格式迁移操作。

(3) 按需迁移

此类迁移指当明确获知文件格式不合适或者已经过时,再开始进行资源格式的迁移。此类迁移策略一个比较明显的特点即为,每次迁移均以原始文件为中心,并且永久的保存原始文件及其元数据,真正实现了资源的长期保存。同时,此类迁移策略仅对部分需要的文件进行迁移,不但降低了迁移的风险和成本,而且减少了迁移的工作量和负荷,减少迁移出现的错误,是一种较为合适的迁移策略。在实践操作中,LOCKSS项目

采取该类迁移策略,当用户发起请求时浏览器不支持资源原始格式的显示时,“LOCKSS Box”将原始格式进行转换,并采用新格式为用户提供服务。

3 国家图书馆长期保存介绍

3.1 国家图书馆的数字资源长期保存整体情况

国家图书馆认识到数字资源保存的重要性、必要性及紧迫性,于2003年起开展数字资源长期保存的研究及相关工作,于2005年启动国家数字图书馆工程^[9]建设,建设了1PB企业级磁带库设备,形成了基于存储网络的数字资源保存系统,对多类数字资源进行长期保存。

在数字资源长期保存工作实践中,按照“重点保存中文数字资源;有选择地保存外文数字资源。注重馆藏元数据、馆藏数字化的特色资源数据、重要的中文网络资源,授权永久保存的中文资源数据库,以及国外重要的工具性数据库等方面的重点数字资源的保存。”的原则,根据数字资源保存和利用的不同特点,将数字资源长期保存划分为永久保存级和归档保存级,以分别满足数字资源当前与长期利用的需要。截止到2013年底,已经保存了包括数字化的馆藏特色资源(中文图书、博士论文、数字家谱、数字方志等)、国家图书馆开展的部分合作项目数字资源(敦煌IDP数据、哈佛燕京数据)等多种类型资源,共计820.16TB,8,254,084册、306,849,825页^{[10]61}。

为保障长期保存数据的安全,规避单一介质保存的风险,国家图书馆还使用光盘保存数字资源。一方面,将已经使用磁带进行长期保存的数字资源,再保存一份光盘副本,例如:数字化的馆藏特色资源;另一方面,单纯使用光盘作为介质存储数字资源,包括:图书馆购买的音像制品、电子出版物、数字化的馆藏特色资源(未提交长期保存)。截止到2013年底,国家图书馆数字资源光盘入藏数字资源总量达到77.9万盘^{[10]60},其中音像制品270,405盘,电子出版物100,500盘,数字化馆藏特色资源40.9万盘。

3.2 音像制品、电子出版物的光盘保存情况

自1987年以来,国家图书馆共入藏、保存了音像、

电子出版物资源,载体形态包括CD、MP3、VCD、DVD与EOD电子出版物等五种。这些数字资源由于存储时间较长,存在一定比例的坏盘情况,也存在由于技术发展,软件即将过时导致缺失播放软件,光盘数据无法读取的危险情况。从1987年至2010年的音像电子出版物的光盘数量如表1所示。

表1 1987年至2010年的音像电子出版物的光盘数量

时间	数量(盘)					总量(盘)
	DVD	VCD	CD	MP3	EOD出版物	
1987-2005年	13,724	56,692	37,736	183	27,075	135,410
2006-2010年	23,085	27,499	12,106	2,564	19,026	84,280

3.3 馆藏图文本数字化资源的光盘保存情况

国家图书馆从1999年起开始有计划地进行馆藏特色资源的数字化建设,数字化资源的范围包括馆藏现代文献、博士论文、古籍资源、民国时期文献等多种类型。2008年之前数字化的资源均只保存在光盘介质上,未进行长期保存处理。部分类型资源保存了2份光盘,也有部分早期数字化的资源仅有1套光盘存储,存在数据安全的隐患。2008年前入藏的馆藏文本图像数字化资源的光盘数量如表2所示。

表2 2008年前入藏的馆藏文本图像数字化资源的光盘数量

资源类型	名称	建设年份	数量(盘)	是否有备份
文本	博士论文	2002--2006	6,110	是
	民国期刊	1999、2003、2007	1,587	是
	地方志(全文影像、文本)	2002、2006、2007	27,200	否
	类书	2007	1,149	否
	中文图书	2005--	419	是
图像	老照片	2004--2006	35	否
	馆藏年画	2005--2007	41	否
	信札	2006	229	否
合计(盘)			36,770	

3.4 光盘作为单一保存载体存在的问题及隐患

由于光盘单片成本较低,存储成本低廉;一次刻录,多次读取,安全性较高;很容易实现多个拷贝,异地存储等优点,非常适合作为离线存储的介质。但是,如果单一将光盘作为存储介质,则会存在一定的问题和隐患。

(1) 光盘保存受外界的环境影响较大

温湿度、光照和灰尘颗粒污染是主要影响光盘保存的外界因素^[1]。温湿度对光盘的影响最大,温湿度的变化或长期处于非正常范围内,将对光盘的记录介质、反射层等不同层面介质产生化学反应、物理变形的作用,造成光盘读取误码率增加或信息丢失;长时间强光照射将对记录介质的燃料层等物质产生化学变化,造成光盘信息的破坏;灰尘中含有大量霉菌、粉尘颗粒等,会污染和腐蚀盘面,严重影响光盘的可读性,甚至直接造成盘内信息丢失。

(2) 光盘有一定存储量的限制

主流光盘产品中,普通DVD-R容量为4.7GB,蓝光DVD(Blu-ray Disc)容量为25GB,因此,光盘受到存储容量的限制,一张光盘只能拷贝一定量的数据。这对于拥有大数据量的图书馆来说,常常出现一本书或一卷画卷等文献资源的对象数据要分散地存储在多张光盘上,因而会造成一种文献资源的元数据与对象数据的对应关系不易管理的问题。同时,由于国家图书馆的数字化资源量巨大,目前以光盘介质存储的仓储量以“百万”计数,给相关业务管理、定期检测、数据提取、灾难恢复都带来一定的问题和挑战。

4 长期保存的必要性及数字迁移策略的选择

4.1 单一光盘存储数据长期保存的必要性

数字资源的长期保存要实现数字资源的可用性、可表现性、可理解性、真实性和可识别性,应对技术淘汰、格式失效等问题对数据的威胁,从而确保具有长期价值的数字资源可以被永久访问和持续利用。

在国家图书馆的数字资源长期保存工作中,以光盘作为单一存储介质的数字资源,他们或者因为单一光盘存储介质的损坏造成数据丢失,或者有些资源因为技

术发展、软件升级造成缺乏读取支持的危险,或者有些资源在较早期进行数字化处理时,存在元数据缺失或部分字段缺失、元数据与对象数据对应关系不清晰等情况,这些都不利于资源的长期可用、长期保存,因此急需对其实施“数字迁移”操作,针对不同类别、不同加工时期的数字化文件完成从光盘存储介质移植到磁带中存储,同时采用多份多地多介质的保存机制,保证馆藏资源的有效性、长期性,最终实现资源的长期保存。

4.2 数字迁移策略的选择

要将国家图书馆馆藏中单一使用光盘存储的音像制品、电子出版物、自建的特色资源进行数据迁移操作,选择恰当的迁移策略至关重要。

(1) 迁移资源的介质数量多、介质种类多

需要进行数字迁移的资源包括219,690盘音像电子出版物和36,770盘数字化特色资源,迁移需要处理的光盘总数量就达到256,460盘。音像电子出版物存储使用的介质种类较多,包括CD、MP3、VCD;数字化特色资源存储的介质较为统一,均为DVD。但是,这些介质属于不同时期购买、刻录的光盘,因而其版本不同,读取方式也不尽相同。

(2) 迁移资源的存储量大、资源类型多

需要进行数字迁移的资源主要包括两部分:

① 音像电子出版物,包括视频类的DVD、VCD;音乐类的CD、MP3;电子图书、电子期刊、电子论文、专利文献等类型的电子出版物,数据存储量达到250TB。

② 数字化馆藏特色资源,包括馆藏现代文献(中文图书)、博士论文、古籍资源(地方志、老照片、年画、信札)、民国时期文献等多种类型,总量达到201,555册,16,521,500页/幅,数据存储量达到133.855TB。

在综合调研了待实施数字迁移操作的存储介质情况,分析了待迁移的资源类型、格式、数据容量的基础上,参考国外相关机构的迁移实践经验、指导意见等,国家图书馆最终采用“存储介质迁移+按需迁移”的迁移策略,不但实现了存储介质的更替与数据迁移,而且实现了从原始文件直接迁移,较好地保存了数据的原始格式,最终实现了资源长期保存。

5 国家图书馆数字迁移项目实践

国家图书馆实施数字迁移项目,对单一光盘介质存

馆藏音像、电子出版物、数字化特色资源进行介质迁移和数据规范化控制, 最终实现数字资源的安全规范化长期保存。

5.1 数字迁移操作流程

在数字迁移操作中, 首先利用光盘塔批量读取待迁移数据的光盘, 将数据复制到项目数据服务器中。然后, 根据数据迁移的整理规则进行数据整理和规范化处理, 最后将规范化数据传输至存储系统中, 完成长期保存和数据迁移。在数据规范化处理中, 采用系统规则匹配与人工识别处理相结合的方式。具体的迁移操作流程如图1所示。

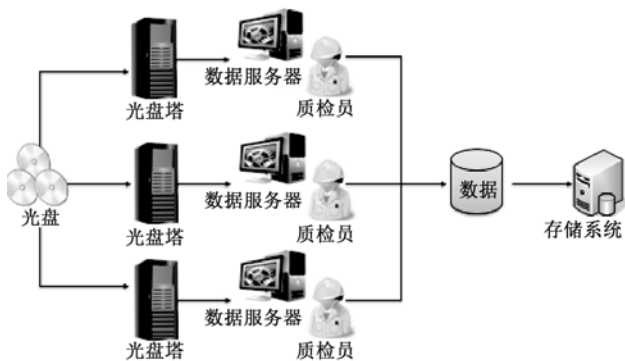


图1 数据迁移操作流程

5.2 数字迁移的规范性控制

在数字迁移操作的整个过程中, 通过规范性控制来保障数据的完整性、一致性。数字迁移与数据备份的操作, 只改变数据的存储介质, 不改变数据内容, 保证原始数据不发生变化, 保证原始的存储结构的层级关系不发生变化。当数据格式必须进行转换的时候, 要采用统一的方法和标准, 从而保证数据的一致性。

(1) 元数据整理规则

数字迁移项目中待整理的资源均为图书馆的馆藏资源, 均使用了ALEPH500系统进行了编目处理, 故待迁移的资源的元数据均采用CNMARC (China Machine-Readable Catalogue) 的编目数据^[12]。因而, 在数据迁移中, 不需要对元数据进行任何整理和规范化处理。

(2) 对象数据整理规范

数字迁移项目中待整理的资源包括音像电子出版

物和数字化馆藏特色资源两部分, 因此对象数据的格式规范整理规则有所不同。

① 音像电子出版物: 鉴于本次数据迁移在保证介质迁移的基础上, 也要实现数据的长期可用, 则在迁移过程中, 要对此类资源的对象数据格式进行一定的迁移, 以便今后更好的保存和利用。所有音像出版物和电子出版物中的视频文件, 均制作成ISO格式镜像文件^[13], 所有音频文件均制作成IMG + .CCD + .SUB文件^[14]。

② 数字化馆藏特色资源: 目前国家图书馆的数字化馆藏特色资源均为自己进行数字化加工的产品, 无论是图书还是图片, 均采用图像方式进行数字化和存储, 遵照《国家图书馆图像数据加工标准和操作指南》的图像资源数字加工标准^[15], 长期保存级的资源对象数据均采用无压缩级的TIFF格式进行保存。因此, 在数据迁移中, 不需要对对象数据进行文件格式整理, 直接进行迁移和保存即可。

(3) 说明文件整理规范

国家图书馆对于数字化文献的加工、存储均以一类资源的一批 (可以是一批书、一批图片等) 为单位, 因此, 一种文献的数字化资源, 除了元数据与对象数据这两个必备数据外, 与之相关的说明文件是必不可少的。说明文件不但包括“标引文件”——说明元数据与对象数据之间的关联对照关系, 还应包括“光盘信息文件”——在数据迁移过程中, 以光盘为单位, 光盘存储资源的详细信息。由于数字迁移项目中两类资源的光盘内容有所不同, 因此“光盘信息文件”有所区别, 光盘信息文件——音像电子出版物的文件形式如表3所示, 光盘信息文件——数字化馆藏特色资源的文件形式如表4所示。

5.3 数字迁移的存储组织结构设计

为保证数字迁移过程中数字资源的一致性, 当数据迁移到新的存储介质后, 需要按照规范的规则进行存储。国家图书馆在实施数字迁移操作时, 充分参考原始光盘的数据存储结构, 以光盘系列为组织单位, 以元数据为组织条件, 将同一系列的所有资源整合在一起。在实际操作中, 将同一条元数据所对应的所有对象数据进行整合整理。

(1) 对象数据存储组织结构

数字迁移项目包括的数字化馆藏特色资源涵盖图

表3 光盘信息文件——音像电子出版物的文件形式

字段名	说明
光盘编号	光盘自身唯一的编号, CNMARC 文件中的Z30-1字段\$\$j+\$\$z
光盘位置编号	光盘所存放架位的架位编号
光盘条码号	条码号
ISBN号	光盘自身的ISBN号
载体类型	光盘的物理介质 (CD、VCD、DVD或其他)
存储量	光盘内文件的存储量大小
节目数	光盘内节目的数量, 主要指视频、音频文件数
是否可读	光盘是否可以被读取
是否多次读取	光盘是否经过多次或更换设备读取
不可读原因	光盘无法读取的原因
是否镜像	光盘内文件是否能够做成镜像文件
镜像文件格式	光盘的镜像文件格式
MD5码	如无法做成镜像文件, 则应该打包 且打包后生成MD5校验码
迁移时间	该光盘迁移的时间
保存时间	该光盘迁移至长期保存系统的时间, 以提交时间为准
包含于	父目录的唯一编号
备注	其他需要说明的问题

书、期刊、论文、照片、年画, 因此, 在迁移处理中均按照文献的原始状态整理为以页/幅为单位的文件。在数据保存中采用数据打包的方式, 组织单位可以根据文献类型不同、单位不同, 将数据打包生成.tar文件。图书、期刊、论文类的资源, 以“册”为单位打包, 照片、年画类资源, 以“种”为单位打包。

(2) 迁移数据的存储结构

为了便于管理数据迁移后的保存数据, 同时遵循国家图书馆存储系统的存储规范, 在存储系统中针对不同类型的数据采用不同的存储结构进行存储和管理。为了便于长期保存数据的保存管理, 结合资源类型、项目名称、存储节点等信息, 国家图书馆定制了资源命名规则。

① 音像制品、电子出版物资源的存储目录采用四级结构。如图2所示:

第一级目录为根目录, 以项目名称+资源名称+存储节点的组合为文件夹命名。

表4 光盘信息文件——数字化馆藏特色资源

字段名	说明
光盘编号	光盘自身唯一的编号, CNMARC 文件中的Z30-1字段\$\$j+\$\$z
光盘位置编号	光盘所存放架位的架位编号
类别	资源的种类
载体类型	光盘的物理介质 (CD、VCD、DVD或其他)
数据内容	光盘内存储数据的内容名称
数据格式	光盘内文件的格式
加工年份	数字化加工的时间, 即光盘数据的制作年份
文献加工编号	光盘内文献的编号起止范围
完成册数	光盘记录清单上记录的完成册数
实际册数	光盘实际保存的册数
缺少情况	具体记录缺失何种数据
完成页数	光盘记录清单上记录的完成页数
实际页数	光盘实际保存的页数
介质数量	光盘的数量, 区分第一复本和第二复本
存储量	光盘内文件的存储量大小
是否可读	光盘是否可以被读取
是否多次读取	光盘是否经过多次或更换设备读取
不可读原因	光盘无法读取的原因
是否镜像	光盘内文件是否能够做成镜像文件
镜像文件格式	光盘的镜像文件格式
MD5码	如无法做成镜像文件, 则应该打包 且打包生成MD5校验码
迁移时间	该光盘迁移的时间
保存时间	该光盘迁移至长期保存系统的时间, 以提交时间为准
备注	其他需要说明的问题



图2 音像电子出版物资源存储目录结构

第二级目录, 以光盘系列为单位, 一个系列的光盘建立一个文件夹, 以元数据中的索取号 (Z30 \$\$j) 为文件夹命名。

第三级目录, 以单张光盘为单位, 每张光盘建立一个文件夹, 以单张光盘对应的元数据中的索取号 (Z30

\$\$j_\$\$\$) 为文件夹命名。

第四级目录,以具体文件为单位,存储具体的文件。

② 数字化馆藏特色资源的存储目录采用二级结构。如图3、图4所示:



图3 数字化馆藏特色资源存储目录结构——地方志



图4 数字化馆藏特色资源存储目录结构——博士论文

第一级目录为根目录,以项目名称+资源名称+存储节点的组合为文件夹命名。

第二级目录,以具体文件为单位,存储具体的文件。以该资源的“种”的编号+“册”的编号相结合,形成一个文件名。

5.4 数字迁移质量检测

数字迁移项目的实施,数据质量保证是一个比较重要的因素,在国家图书馆的数字迁移操作过程中采用了MD5码校验、数据清单核对相结合的方式,保证数据迁移前后的一致性、完整性、真实性。

在待迁移数据整理的过程中即产生MD5码,当数据迁移到存储系统中后,再次产生一次MD5码,将两次生成的MD5码进行比较,用以检测所存储的数字化文献的完整性和正确性。

当待迁移数据上传到存储系统后,会详细核查数据的目录结构、文件名称、文件数量是否一致,还会详细核查一批资源的种数、册数、文件数,从多方面保证保存到存储系统中的数据的完整性、一致性。

5.5 数字迁移项目总结

国家图书馆数字迁移项目,从2012年4月开始策划,2012年12月开始正式实施,历经2年多时间,目前已

接近尾声。通过该项目的策划、组织、实施,不但系统地梳理了历史悠久的数字化馆藏资源,而且通过切实可行的数字迁移将保存多年的数字资源实现了长期保存、长效利用,为今后更好的数字资源长期保存和合理利用打下了基础。

6 结语

数字迁移工作是一项复杂和繁琐的工作,不但涉及策略的选择、规范的制定、质量的检测,还与数字资源长期保存的整体架构、工作机制、规则规范紧密联系。尽管国内外多家国家图书馆、大学图书馆、信息保存机构进行了多年的研究和实践,但在国内没有任何机构有实际的数字迁移工作实践可以借鉴。同时,由于不同单位长期保存的数据情况不同,保存的介质基础不同,保存的管理机制不同,即使有相关的数字迁移项目实践经验,在真正要实施一个数字迁移项目的过程中,只能是选择性的借鉴,流程性的借鉴,概念性的借鉴。因此,国家图书馆的数字迁移,是在充分分析了待迁移数字资源情况的基础上,以妥善长期保存为目标,以“存储介质迁移+按需迁移”为技术策略,通过规范化、流程化、体系化的控制实现了数字迁移的实施。

国家图书馆的数字迁移,是长期保存工作中的一个成功实践。虽然在数字迁移过程中,也发现了一些不足,例如没有完善的迁移工具,针对格式过期没有实现自动监测与预警等。今后,希望可以借助实际的长期保存工作,与业界同行多进行沟通交流、多交换工作经验与教训,集合不同行业、机构的力量,共同促进数字资源长期保存工作的完善与提高。

参考文献

- [1] GC-HD090188,国家数字图书馆长期保存元数据标准规范[S].北京.
- [2] 董晓莉. 图书馆数字资源长期保存迁移技术分析[J]. 图书馆杂志,2012(7):63-69.
- [3] 吴振新,张智雄,郭家义. 数字信息资源长期保存技术策略分析[J]. 现代图书情报技术,2006(4):8-13.
- [4] 赵敏. 浅谈数字资源长期保存的相关问题[J]. 科技情报开发与经济, 2010,20(18):98-101.
- [5] The Commission on Preservation and Access, The Research Libraries Group, Inc. Preserving Digital Information: Draft Report of the Task Force on Archiving of Digital Information[R].1995.

- [6] Lawrence GW, Kehoe WR, Rieger OY, et al. Risk Management of Digital Information : A File Format Investigation[R/OL]. [2014-11-18]. <http://www.clir.org/pubs/reports/pub93/reports/pub93/pub93.pdf>.
- [7] What Is LOCKSS? [EB/OL].[2014-11-18].<http://www.lockss.org/about/what-is-lockss/>.
- [8] How LOCKSS Works[EB/OL].[2014-11-18].<http://www.lockss.org/about/how-it-works/>.
- [9] 国家图书馆数字图书馆工程[EB/OL].[2014-11-05].<http://www.nlc.gov.cn/newstgc/>.
- [10] 国家图书馆. 国家图书馆年鉴2014[M]. 北京:国家图书馆出版社,2014.
- [11] 光盘与磁带的存储方法研究[EB/OL].[2014-11-17].<http://tech.watchstor.com/management-115037.htm>.
- [12] CNMARC[EB/OL].[2014-11-20].<http://baike.baidu.com/view/679509.htm?fr=wordsearch>.
- [13] ISO镜像文件[EB/OL].[2014-11-20].<http://baike.baidu.com/view/1829481.htm?fr=wordsearch>.
- [14] 镜像文件[EB/OL].[2014-11-20].<http://baike.baidu.com/view/928.htm?fr=wordsearch>.
- [15] 朱强,张春红,龙伟. 国家图书馆图像数据加工标准和操作指南[M]. 北京:国家图书馆出版社,2011.

作者简介

季士妍, 国家图书馆高级工程师, 研究方向: 数字资源长期保存, E-mail: jisy@nlc.gov.cn.

The Practice of the Digital Migration Project on National Library of China

Ji ShiYan

(National Library of China, Beijing 100081, China)

Abstract: The paper introduces the progress in practice and research on domestic and international long-term preservation of digital resources and digital migration, and analyses the National Library of China's long-term preservation of digital resources, points out the necessity of the digital migration. The paper introduces the digital migration project implementation process, the specification for data treatment, the organization structure of data storage, the quality control of digital migration. The introduction of the digital migration project will provide some reference for the related projects.

Keywords: National Library of China; Digital Preservation; Digital Migration

(收稿日期: 2015-01-27)
编辑: 刘伟