

开放存取资源长期保存政策的调查与分析¹

黄如花 胡文琴

(武汉大学信息资源研究中心, 湖北武汉, 430072)

摘要: 文章分析了国际上著名的数字保存领域相关的政策指南中关于开放存取资源长期保存的有关条款, 调查了澳大利亚的开放存取项目 ARROW 中各机构库的长期保存政策制定情况, 并选取国外代表性机构库 ALADIN Research Commons、Nottingham Eprints 和 Griffith Research Online 的长期保存政策进行对比分析, 提出了 OA 资源长期保存政策的发展路向。

关键词: 开放存取资源, 长期保存, 保存政策

开放存取 (Open Access, OA) 从 20 世纪 90 年代被提出以来, 很快引起了多个国家的重视, 一直是图书馆学界关注的焦点之一, 在理论与实践领域均进展迅速。数字资源的长期保存问题也一直是国外图书馆学界关注的热点, 但专门针对开放存取资源长期保存的政策却不多, 只是穿插于各类型数字资源长期保存的政策中。作为长期保存的有力保障, 制定合理完善的长期保存政策是非常重要的。为了促进 OA 运动的持续发展, 发挥 OA 资源的学术价值, 国家及机构各个层面都应对 OA 资源的长期保存高度重视, 加强政策引导。

1 数字保存政策中对 OA 资源长期保存的关注

出于对数字资源保存的重视, 国际上已经颁布了一些数字资源长期保存方面的政策, 一些国家的国家图书馆也制定了该国的数字资源保存政策。作为网上可以免费下载的学术资源, OA 资源的保存一般只是夹杂在数字资源或网络资源保存政策中, 很少被直接提及。然而, 结合 OA 资源的特点, 可以从这些保存政策中找到对 OA 资源长期保存的关注。下面选取数字保存领域几个有代表性的政策指南, 分析其中有关 OA 资源的内容条款。

保存政策是有效实施数字资源长期保存的管理上的保证, 而在数字保存领域权威的标准指南中, 对于保存对象, 许多基本的选择标准都是首先保存那些容易丢失的、与未来有相关的、有重要价值的数字资源, 特别是原生数字资源和重要的科学数据。

表 1 数字保存政策中保存对象选择标准与 OA 资源的联系

数字保存领域相关的政策指南	发布时间	发布者	对保存对象的规定条款	OA 资源的对应特点
《永久保存世界记忆: 关于保存数字化信息的联合声明》	2002 年 6 月	国际图联与国际出版协会 (IFLA/IPA)	“日益增长的仅以数字形式出版的信息如同传统印刷出版物一样重要, 同样具有长期留存的文化价值和历史意义” ^[1]	大多数 OA 资源为原生数字资源
《保存数字遗产宪章》	2002 年 11 月	联合国教科文组织 (UNESCO)	“显然应优先保存原生数字资料” “鼓励公立和私立大学和研究机构确保研究资料得到保存” ^[2]	

¹ 本文系教育部人文社科重点研究基地重大项目“数字信息时代的图书馆管理研究”(07JJD870221)和“新世纪优秀人才支持计划(NCET)资助”项目成果之一。

《保护澳大利亚网络资源——创建者与出版者指南》	2002 年 6 月	澳大利亚国家图书馆 (NLA)	“资源类型中应包括具有持久的研究性，被公众使用的信息资源，如研究论文。” ^[3]	中，大多具有重要的学术性
-------------------------	---------------	-----------------	---	--------------

此外，在一些著名的数字资源长期保存项目的原则方针中，OA 资源符合被优先保存的对象特征，如美国国会图书馆致力于网络信息保存的“密涅瓦”项目（Mapping the Internet Electronic Resources Virtual Archive, MINERVA），其收集策略主要遵循的标准有“满足研究人员当前以及未来的信息需求”、“学术内容”、“丢失危险度”等^[4]。澳大利亚著名数字保存项目“保存与存取澳大利亚网络文献资源”（Preserving and Accessing Networked Documentary Resources of Australia, PANDORA）在选择资源时遵循的标准中也提到“是否有对应的印本”、“联机出版过程是否有质量控制”、“是否被公认的索引服务商索引”、“有研究价值的权威出版物拥有选择的最优先权”^[5]等。

2 OA 仓储长期保存政策制定现状

2002 年布达佩斯开放存取计划（Budapest Open Access Initiative, BOAI）提出开放存取的实现途径主要为两种：一种是创办开放存取期刊，另一种是自行存档，即建立开放存取资源仓储（Open Access Repository），即 OA 仓储。^[6]构建 OA 仓储不仅是实现 OA 的有效途径，也是保存 OA 资源的必要措施。在开放存取资源仓储名录（Directory of Open Access Repositories, OpenDOAR）2006 年所做的一项调查中发现，大约有三分之二的 OA 仓储没有制定关于其数字资源的数据提交、资源长期保存、资源再利用等方面的政策。^[7]为此，OpenDOAR 通过研发增加了政策工具（Policies Tool）模块，其中的资源保存政策工具（Preservation Policy Tool）可以帮助我们调查和分析 OA 仓储的长期保存政策。

2.1 OA 仓储长期保存政策制定的总体情况

OpenDOAR 的资源保存政策工具综合 OA 仓储多方面的信息，如 OA 仓储是否在开放文档先导（Open Archives Initiative, OAI）^[8]注册、是否遵循开放文档先导组织的元数据采集标准协议（The Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH）、是否采用开放档案信息系统参考模型（Reference Model For an Open Archival Information System, OAIS）、在 OA 仓储网站上有无专属页面公布其资源保存方面的政策信息等，将 OA 仓储的长期保存政策制定状态分为 5 种类型：不知道（Un-known）、未声明（Un-stated）、未明确定义（Un-defined）、未明确声明（Un-clear）和已定义（Defined）。其中 Un-known 状态是指该 OA 仓储没有公开声明任何关于其资源长期保存的信息；Un-stated 是指存在有关的政策的信息，但没有具体的政策；Un-defined 是有一些相关的政策，但无系统明确的政策体系；Un-clear 指关于政策的信息比较模糊；Defined 则表明对 OA 仓储资源的长期保存制定了较全面系统的政策（见图 1）。

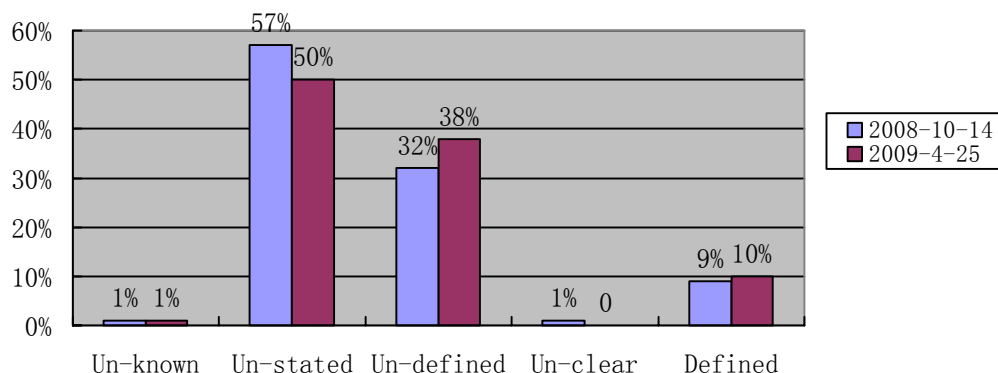


图 1 OA 仓储保存政策制定状态的历时对比

如上图所示, 从 2008 年 10 月 12 日的统计数据与 2009 年 4 月 25 日时间差距约为半年的统计数据进行分析可以发现, Un-stated 与 Un-defined 类型是 OA 仓储保存政策界定情况的主流, 而政策制定状态的变化也主要集中在 Un-stated 与 Un-defined 上, Un-stated 比例下降, Un-defined 比例上升, 下降与上升的比例相差仅 1 到 2 个百分点, 呈现出此消彼长的趋势。从二者的区别来看, 与 Un-stated 相比, Un-defined 类型的 OA 仓储中已经存在一些资源保存的有关政策, 这种由 Un-stated 向 Un-defined 过渡的势头是良好的趋势, 预示着 OA 仓储整体上正逐步向 Defined 状态靠近。另外, 从 Defined 类型来看, 全球已定义保存政策的比例增加了 1 个百分点。综上可见, 明确定义了长期保存政策的 OA 仓储是呈增长趋势的, 但增长比较缓慢。

2. 2 OA 仓储建设项目中 OA 资源保存政策的制定情况

澳大利亚是十分重视 OA 仓储建设的国家, 在其 2003 年启动的高等教育信息基础设施项目(Systemic Infrastructure Initiative, SII)中, OA 仓储的构建就开始被作为项目建设重点提升至国家信息基础设施建设的层面, 其保存和提供大学或其他科研机构的学术成果, 以改善澳大利亚学术研究成果的交流和保存机制。本文选取作为 SII 子项目之一的“向全球公开的澳大利亚研究仓储在线”(Australian Research Repositories Online to the World, ARROW)^[9]项目为例, 对其 13 个项目成员大学的机构库的长期保存政策展开了调查。

表 3 ARROW 项目高校机构库长期保存政策的定义状况

高校名称	机构库名称/参与项目	规模 (items)	OAI 注册	保存政策 定义状况	统计时间
澳大利亚中央昆士兰大学	aCQUIRe	3323	有	Un-defined	2009-03-10
	ADT Program	24	无	Un-stated	2008-04-30
澳大利亚拉筹伯大学	La Trobe Univ.-ADT Program	24	无	Un-stated	2008-04-30
澳大利亚麦考瑞大学	Macquarie Univ. Research Online	331	有	Un-defined	2008-08-07
澳大利亚莫纳什大学	ARROW Repository	3568	有	Un-stated	2009-03-10
澳大利亚莫道克大学	ADT Program	/	无	Un-stated	/

澳大利亚昆士兰科技大学	ePrints Archive	8881	有	Un-defined	2008-06-02
澳大利亚旋宾科技大学	Swinburne Image Bank	3357	无	Un-stated	2008-07-07
	Swinburne Research Bank	678	有	Un-stated	2008-08-07
	SUT-ADT Program	83	无	Un-stated	2006-10-03
澳大利亚巴拉瑞特大学	UB Library - ADT Program	33	无	Un-stated	2008-04-30
	UB Research Online	1551	无	Un-defined	2009-03-11
澳大利亚纽卡斯尔大学	ADT Program	3041	无	Un-stated	2008-10-24
	NOVA	2506	有	Un-defined	2009-03-10
澳大利亚新英格兰大学	e-publications@UNE	920	无	Un-defined	2008-08-07
澳大利亚新南威尔士大学	UNSW@ADFA	3893	有	Un-stated	2009-03-10
澳大利亚阳光海岸大学	Coast Research Database	2685	无	Un-defined	2008-08-07
澳大利亚西悉尼大学	UWS Digital Theses	1854	无	Un-stated	2008-05-15
	UWS Research Repository	5677	有	Un-defined	2009-03-10

统计结果显示：在 13 个大学的 19 个 OA 仓储中，不论数据库规模大小，都没有明确定义其资源的长期保存政策。界定状态都集中在 Un-stated 与 Un-defined 两种类型，处于 un-stated 状态的 11 个，处于 Un-defined 状态的 8 个，在 OAI 注册的 OA 仓储仅占 42%。成员高校中有的大学建造了不止 1 个 OA 仓储，有的大学还参加了其他数字项目，如中央昆士兰大学、拉筹伯大学、巴拉瑞特大学、纽卡斯尔大学和旋宾科技大学还是“澳大利亚数字论文项目”(Australian Digital Theses Program, ADT)^[10]的成员机构。

上述统计表明，OA 仓储的发展速度与其长期保存政策的制定状况的反差十分明显，机构库建设项目只关注 OA 仓储的构建与发展，而对其资源的长期保存问题不重视，这对 OA 仓储的可持续发展必将造成不利的影响。

3 OA 仓储的具体资源保存政策的对比分析

在 OA 仓储中也有一些明确定义了长期保存政策的机构，通过 OpenDOAR 的 Preservation Policy Tool 可以查询到这些 OA 仓储长期保存方面的具体政策。本文选取已定义长期保存政策的 3 个较具代表性的 OA 仓储，对比分析其 OA 资源长期保存政策的各方面规定。它们是美国华盛顿研究图书馆联盟的 ALADIN Research Commons、英国诺丁汉大学的 Nottingham Eprints 和澳大利亚格里菲斯大学的 Griffith Research Online。

表4 国外三个代表性机构库保存政策的对比

OA 仓储名称 政策关注点	ALADIN Research Commons	Nottingham Eprints	Griffith Research Online
保存的时限	无限期保存	无限期保存	无限期保存
持续的可获得性和可读性	无声明	在有需要时迁移到新的文件格式，如有可能，对没有或不能迁移的数据进行仿真处理	在有需要时迁移到新的文件格式，但不能保证一些特殊格式的文件的可读性

元数据的保存	遵循 OAI-PMH 协议	遵循 OAI-PMH 协议	遵循 OAI-PMH 协议
与其他机构合作保存	与合作伙伴多备份保存数据	与合作伙伴传递和迁移文件格式, 保存元数据	无声明
数据更改	无声明	不允许已有的记录被更改, 必要时会在原始数据旁附上勘误表	无声明
原始数据的保存	在著作权所有者要求下数据将被移除	原始数据流将被完整保存, 并附加到任何升级格式中; 在特殊情况下数据记录才会被移除	原始数据流将被完整保存, 并附加到任何升级格式中; 但在著作权所有者要求下数据将被移除
被撤销数据的保存	被撤销数据的标识符或统一资源定位符仍被永久保存; 被撤销数据的元数据不能再查询。	被撤销的数据没有被删除, 只是从公众视线里移走了。被撤销数据的 URL 记录仍被长久保存。	被撤销数据的 URL 记录仍被长久保存。
OA 仓储关闭的应对政策	假使 OA 仓储关闭, 数据库将转移至其他合适的存储机构	假使 OA 仓储关闭, 数据库将转移至其他合适的存储机构	无声明

在上述三个 OA 仓储的长期保存政策中, 比较受关注的方面有资源保存的时限、资源持续的可获得性和可读性、元数据的保存、原始数据与被撤销数据的保存、是否于其他机构合作保存等。其中英国诺丁汉大学的 Nottingham Eprints 的长期保存政策相对比较全面, 它对于数字资源保存的长期性、完整性和可利用性都制定了具体的政策: 倡导机构资源永久保存; 重视原始数据、被撤销数据、元数据的保存, 并保存数据更改情况; 通过迁移与仿真技术支持资源可持续利用; 并有与其他机构合作保存的政策和针对 OA 仓储关闭风险的应对政策。

4 OA 资源长期保存政策的发展路向

OA 资源长期保存政策是人们为了实现 OA 资源长期存取的目标而制定的方针、原则、策略和措施。目前, OA 资源的长期保存政策还很不完善, 相对而言, 现行的数字资源保存政策大多明确了数字资源长期保存的原则、规范和标准, 制定了数字资源长期存取的具体行动计划, OA 资源的长期保存政策可以借鉴其有利的部分并结合自身特点加以调整完善。从总体上来说, OA 资源长期保存政策应朝着以下三方面发展。

4.1 以 OA 资源的可持续利用为长期保存的目标

保证 OA 资源的可持续利用也是获取科研人员对 OA 仓储信任度吸引其参与自我存档的重要因素。成功的 OA 仓储应该不仅能够将研究者提供的科研文献长期保存于此, 它还必须在存缴、组织和保存转移等方面的政策和技术上为长期保存提供具体支持^[11]。因此, 不少 OA 仓储的长期保存政策中在资源的保存时限方面都倡导无限期保存, 并声明将通过迁移和仿真技术确保资源持续的可获得性和可读性, 使科研人员更有动力将研究成果存储在 OA 仓储中。而关于 OA 仓储的信任度问题, 研究图书馆集团 (Research Library Group, RLG) 和美国国家档案与记录管理委员会 (US National Archives and Records Administration, NARA) 在 2005 公布的《认证可信数字仓储的审核清单草案》(Audit Checklist for the Certification of a Trusted Digital Repository, Draft for Public Comment)^[12]和其 2008 年修正与扩展版《可信数字仓储审核与鉴定: 标准与清单》(Trustworthy Repositories Audit & Certification: Criteria

and Checklist)^[13]中,都将资源的可持续利用看作决定一个数字机构库是否能被鉴定为一个可信赖的数字信息收集场所的主要因素之一。

4.2 在保存内容上重视 OA 资源的选择与评价

OA 资源类型十分复杂,不仅保存研究人员的论文资料,还保存科研过程的其他记录,如教学课件、实验数据与报告等,而长期保存问题又涉及技术应用、资金支持、数字权益保护等诸多问题,所以其数字保存过程中的选择和评价工作是必不可少的,在制定 OA 资源长期保存政策过程中应确保优先保存最有价值的数字资产的长期可获得。在资源的选择上,可以学习数字资源保存和数字资源评估领域的一些政策与方法^[14]。如在如何选择需要长期保存的数字资源方面,数字保存联盟(Digital Preservation Coalition, DPC)开发的数字资源长期保存选择工具(Decision Tree for Selection of Digital Materials for Long-term Retention)^[15]就很有借鉴性,它采用互动式的评估方式,在数字资源版本与内容选择、保存责任与权利、保存技术与费用、文件与元数据保存四个方面均制定了一系列评价细则,并在选项旁附有一些建议。中国高等教育文献保障系统(China Academic Library & Information System, CALIS)数字资源评估指标体系中关于数字资源的可持续发展的五条评估标准^[16],即数字资源在馆藏资源中的比例、可永久获取的数字资源在馆藏中的比例、长期保存数字资源的费用、公平使用与存档方法评价也值得参考。

4.3 在保存方式上鼓励合作保存与多备份保存

促进学术自由交流是 OA 运动发起的主要目的,许多 OA 项目都由多个机构共同建设,机构合作不仅有利于资源的保存,也有利于资源共享与促进跨机构的学术合作。合作保存有利于明确各机构的保存任务和责任,协调在于不同组织和利益相关者之间的关系,制定统一的行为标准,还可降低保存的成本与风险。如英国的“机构资源存取计划”(Focus on Access to Institutional Resources Programme, FAIR)^[17]的实施涉及 OA 运动的技术、法律、文化等各个方面,其发展策略和经验比较具有借鉴意义。此外,“多备份资源保存项目”(Lots of Copies Keeps Stuff Safe, LOCKSS)也可以作为保存 OA 资源的一个较实际的运行模式。虽然 LOCKSS 发起的初衷是为了实现商业电子资源的收集与保存,但随着科研环境的变化以及学术交流模式的发展,OA 资源的组织与利用在科研活动中的作用日益显现,LOCKSS 也开始在系统结构设计及功能实现上支持对 OA 资源的采集、保存与使用^[18]。

参考文献

-
- [1]Preserving the Memory of the World in Perpetuity: a joint statement on the archiving and preserving of digital information [EB/OL]. [2009-05-31]. <http://archive.ifla.org/V/press/ifla-ipa02.htm>
 - [2]Guidelines for the Preservation of Digital Heritage [OL]. [2009-05-31]. <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>
 - [3]Safeguarding Australia's Web Resources: Guidelines for Creators and Publishers [EB/OL]. [2009-05-31]. <http://www.nla.gov.au/guidelines/webresources.html>
 - [4]赵俊玲. 美国国会图书馆网络信息保存项目 Minerva 及其启示[J]. 图书馆建设, 2005 (5): 40
 - [5]Online Australian Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia [EB/OL]. [2009-04-12]. <http://pandora.nla.gov.au/selectionguidelines.html>
 - [6]高冰洁, 赵俊玲. 国内外开放存取实践分析[J]. 图书馆工作与研究, 2008 (1): 68
 - [7]Peter Millington. Moving Forward with the OpenDOAR Directory, 8th International Conference on Current Research Information Systems, Bergen, 11th-13th May 2006 [OL]. [2009-05-31].

- <http://www.opendoar.org/tools/en/policies.php>
- [8] Open Archives Initiative. <http://www.openarchives.org/> [2009-05-31]
- [9] Australian Research Repositories Online to the World. <http://www.arrow.edu.au/> [2009-05-31]
- [10] Australian Digital Theses Program. <http://adt.caul.edu.au/> [2009-05-31]
- [11] 张晓林. 机构知识库的政策、功能和支撑机制分析[J]. 图书情报工作, 2008(1): 25
- [12] Joanne K. Patricia H. Janet Eke. Thomas G. H. Using the Audit Checklist for the Certification of a Trusted Digital Repository as a Framework for Evaluating Repository Software Applications[J/OL] D-Lib Magazine 2006(12). <http://www.dlib.org/dlib/december06/kaczmarek/12kaczmarek.html>
- [13] Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist [EB/OL]. [2009-05-31]. <http://www.crl.edu/content.asp?11=13&12=58&13=162&14=91>
- [14] 范媛静, 宛玲. 国外数字资源长期保存政策扫描[J] 数字图书馆论坛, 2006(4): 5-8
- [15] Preservation Management of Digital Materials: The Handbook[OL]. [2009-05-31]. <http://www.dpconline.org/graphics/handbook/>
- [16] 肖琰, 李浩凌, 徐成. CALIS 数字资源评估指标体系及其应用指南[J]. 大学图书馆学报, 2008(3): 6
- [17] 袁红梅, 乔冬梅. 开放存取: 英国 FAIR 计划的实践与启示[J]. 图书情报知识, 2007(7): 92-95
- [18] 吴振新, 李春旺, 郭家义. LOCKSS 数字资源长期保存策略[J]. 现代图书情报技术, 2006(2): 36

A Survey and Analysis of Long-term Preservation Policies for Open Access Resources

Huang Ruhua Hu Wenqin

(School of Information Management of, Wuhan University, Wuhan, Hubei, 430072)

Abstract: This paper analyzes representative international digital preservation policy guidelines related to open access resources, investigates the status of preservation policy for open access repositories in ARROW project, explores the specific preservation policy of three typical open access repositories including ALADIN Research Commons, Nottingham Eprints and Griffith Research Online. Suggestions for long-term preservation policies of open access resources are provided at the end of the paper.

Keywords: open access resources, long-term preservation policy, preservation policies

作者简介:

黄如花, 女, 武汉大学信息资源研究中心教授、博士生导师, 图书馆学系副主任。

胡文琴, 女, 武汉大学信息管理学院图书馆学专业 2008 级研究生。通信地址: 武汉市武昌珞珈山。邮编 430072。联系方式: rhhuang@whu.edu.cn 13720116276。