

# 谈数字资源生命周期与长期保存政策

李文生<sup>1</sup>

(中国科学院文献情报中心 北京 100080)

**摘要:** 数字资源的飞速增长给长期保存工作带来了巨大的挑战, 长期保存政策的制定是有效实现数字资源长期保存的基础, 本文从数字资源的生命周期探讨各个阶段在制定长期保存政策及实现长期保存需要考虑的问题, 最后分析了各个阶段所处的法律和经济环境对长期保存的影响。

**关键词:** 长期保存; 生命周期; 数据存储; 风险评估

Digital Resource Lifecycle and Long-term Preservation Policy

Li Wensheng

(Library of Chinese academy of sciences ,Beijing 100080,China)

**Abstract:** Implementation of preservation depends on the long-term preservation policy. This paper firstly analyses the stages of the digital resource lifecycle and discusses the consideration of the development of long-term preservation policy, finally analyses the legal and commercial environment surrounding the digital resource which influences the long-term preservation.

**Keyword:** Long-term Preservation; Lifecycle; Data Storage; Risk Assessment

## 1 引言

数字信息资源飞速增长给我们带来了种种挑战: 数字信息呈指数级的增长, 快速甚至瞬间消失。我们怎样长期保存与访问这些数字信息? 解决这些问题我们不得不去认识数字资源从产生、使用到保存各个阶段与长期保存之间的相互关系。无论是OCLC<sup>[1]</sup>、JISC<sup>[2]</sup>还是NA<sup>[3]</sup>与澳大利亚国家图书馆的数字保存政策<sup>[4]</sup>的制定都是基于以数字化形式存在的资源, 更多的是集中选择已有数字资源进行长期保存。笔者认为数字资源的长期保存从数字资源的产生便要制定相应的政策, 而且贯穿于整个生命周期。

本文首先分析数字资源的生命周期及生命周期各阶段与长期保存之间的关系, 最后分析了法律和经济方面对长期保存的影响。

## 2 数字资源的生命周期及各阶段长期保存的相关政策

组织机构要实施数字资源的长期保存首先要调整在数字资源生命周期内每个阶段的管理活动, 因此组织机构的整个工作流程都要包含数字保存计划, 而且对新技术的发展具有较高的柔韧性。图 1 是 Cornell 大学的调研报告<sup>[5]</sup>说明长期保存政策的缺乏和其他给数字资源保存带来的威胁情况对比。

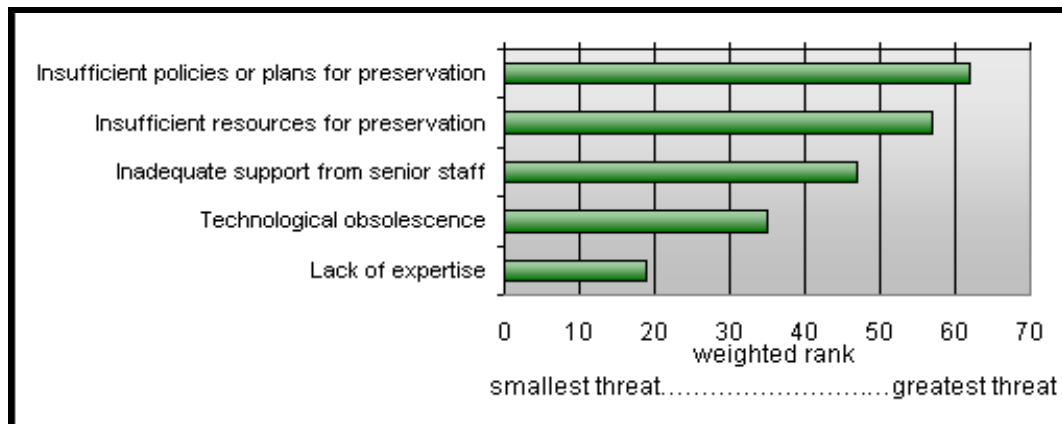


图 1 Cornell Survey of Institutional Readiness, 2003-2005<sup>[5]</sup>

<sup>1</sup> 作者简介: □李文生(1968-),男,博士研究生,主要从事电子政务、网络信息管理技术与信息系统研究

虽然数字资源长期保存的提出才有十年左右，但是我们由图 1 可以看到，经验的缺乏并不是给长期保存带来最大的威胁的因素，缺乏长期保存的政策和计划远远高于技术的过时和经验缺乏带来的威胁。只有制定切实可行的长期保存政策才能保证对数字资源进行有效的长期保存，而长期保存的政策应该贯穿于数字资源生命周期的各个阶段，包括数字资源的产生、管理、存储、保存和使用等等。

## 2.1 数据的产生

数据的产生通常涉及到设计与实施阶段，此时考虑长期保存政策对研究、项目设计、信息系统的设计和开发或选择软件工具都大有裨益。创建数字资源的目的是有千万种，同时涉及到各种利益相关者。数字资源的产生可以是本身就以数字形式存在的也可以是传统资源的数字化，这些数据可以在有限的时间段内产生，也可以是动态或者连续的产生。某些情况下是数字化的资源和传统介质存在的资源相结合的形式产生的，比如有些资源通过超链接连接到其他的资源上。

不论是处理的过程还是资源的形式都影响到方方面面的决策，比如选择和决定数据资源的成本、智力内容、结构、格式、压缩、编码、描述信息的粒度与层次、使用资源时版权归属和其他法律经济条件。数据产生的方法与数据存在的形式直接影响到数据的管理、使用、维持和将来的保存。而这里所有的或大部分的标准将决定这些资源对于其创建者或基金组织的有用性，是否适合初始目的。

不管是个人还是由组织机构创建的数据，大量的不同的利益相关者影响其产生的过程，基金组织、出版商和软件的开发者都会影响甚至决定产生的过程。因此，为了能制定出可行的政策与指导性的纲领，需要在数据生产者的内部或外部经常保持沟通，当然还有与数据的使用者和其他利益相关者之间沟通，为了对资源进行管理、保存及将来的使用而考虑其生成及被记录时的情况。

## 2.2 数据的管理与保存

数据的管理与保存涉及到许多利益相关者，他们都有不同的功能和角色。这些功能和角色可能有时间的限制也可能没有时间限制，都直接或间接的参与到了数据的管理和保存的过程中去。数据一旦由生产者创建，通常会由创建者管理进行短期的保存，这些数据也可能转移到其他部门或机构处理，支持长期的存取。这些功能可以由部门内部完成，不过这样的话他们的组织角色就扩展到了长期保存上；另外一种方法就是这些功能由其他机构、机构的档案室、图书馆和其他学术档案馆完成，创建者提供自己的内容给他们，同时明确版权。

此外，数字资源可能作为资源集合构建或资源集合管理的一部分产生，这些作为数据创建过程的补充，同样应该遵守相似的标准。经过许可，通过拷贝、对其他机构单位的资源进行镜像扩充现有的资源集合或者产生新的集合。由于信息持久动态的变化，单个机构可能并不适合进行长期保存。这种情况下，数据的创建者负责更新和维护自己的数字信息，在法律上强迫资源的创建者或他们自觉遵守外部组织制定的长期保存标准和程序。对资源进行周期性的采样和备份及时地提供给档案馆进行存档。

资源的创建者在现行的项目或业务处理中使用的数据可以根据远程管理的标准进行管理和保存，创建者必须遵循这些标准和程序，并由外部机构进行监督。这种情况下一旦这些数字资源不再被创建者使用，对这些数据进行筛选来存档。创建者正在使用的数据的备份可能是拒绝任何的访问也可能在一定的时期内有限的访问。

这些进行存档的机构我们往往称之为“数据银行”(data bank)<sup>6</sup>，是这些资源远程管理的协约人，他们经常同这些数据合法拥有者达成协议，制定出相应的条款和标准，在合同的范围内管理资源。

### 2.2.1 数据的获取、维持或处理

资源的获取涉及到很多方面的决策，如收集的政策、选择或拒绝某些标准、采样的方法、收集的水平、维持的周期、这些资源是部分还是全部要在保存之前进行处理，长期保存政策的选择，哪些资源会被纳入到永久性保存中，或哪些排除在永久性保存之外，还有哪些应该是远程管理来处理的。它还将涉及到数据的评估，资源创建决策决定或影响了这些评估，如格式、使用的结构、数据质量和连续性、现有的元数据和文档，以及这些资源的权益等。资源获取的决策定义了资源的集合，直接影响这些资源的分类、管理、对终端用户的可存取性。例如，美国斯坦福大学实施的LOCKSS(Lots of Copies Keep Stuff Safe, 多备份资源保存)是一种基于Java技术的开放性源码分布式系统，其目的就是解决电子期刊长期保存与利用问题，以其Peer-To-Peer分布式的保存策略在欧美引起极大关注，系统应用的是成熟的对等网组网技术和源代码开放策略<sup>7</sup>。

### 2.2.2 数据管理

数据怎样管理依赖于数据的产生或提供的形式（如：什么格式，何种文档类型，在什么条款和条件下）。数据产生或数据集合的选择决策以及基金和组织可利用的技术都限制了数据管理选项，也限制了数据的使用和保存选项。这些决策详述有：数据结构、格式、压缩、编码等。数据的格式（如写到磁介质上）、压缩、编码（如内部的语义与语法的描述）都决定了数据能否在不同的硬件软件平台上轻易的被访问，还决定了数据的存储、利用以及怎样使这些资源的价值最大化。

数据的描述和文档信息提供了数据的结构、内容、上下文、起源和历史信息，这样的有两种形式，一种是数据产生时出现的如用户手册、数据字典、资源转移时的文档；另外一种传统的或数字资源的目录等信息，这些信息资源的定位管理使用等。这些通常能反映出数据获取的决策（比如：提供什么样的资源文档，怎样提供，有谁提供等），也能反映出主题或部分文件标准以及数据生产者的实践等。不过这些信息也可能是暂时性的，这要有分类的人员及机构现有的专家来决定。

数据的存储涉及到是资源集合的部分还是全部被存储，是集中式存储还是跨许多站点分布式存储，这些资源集合是否与外部的数据银行达成协议，还涉及到使用何种介质和硬件平台等技术决策，还包括物理安全、刷新和存储介质的更替与意外事故的处理等。所有的这些都受限于资源的结构格式、压缩和编码；这些资源在本质上是动态的还是固定，资源权威性和完整性的维护，以及这些资源在使用和保存的重点方面都限定了数据存储决策的选择。反过来，与现有能够利用的基金和技术相关的存储决策也限制了数据的创建与选择，这些决策帮助决定怎样（是否）以及对这些数字资源存储与何种程度使用。数据存储决定了短期保存的完整性和资源的功能性，通常涉及到以下 5 个方面：①定期的检查资源的完整性、功能和连贯性；②更新存储的介质，拷贝资源，克服存储介质由于时间的长久造成的不稳定性；③将资源转移到新的存储介质或改变其存储的格式；④将这些拷贝放到不同的地点防止灾难和其他损失；⑤在每次将资源转移到新的介质之前都保留一份原始格式的数据，这样在数据恢复的时候可以进行检测和有效性验证。

利用长期保存实现数字资源的有效管理的项目有NDIIPP等项目，2000年12月美国国会建立国家数字信息基础结构和保存规划NDIIPP，并责成国会图书馆联合主要的国家机构，在数字内容存档的收集和管理方面有实践经验的团体，以及其他致力于数字格式工作的私有机构等，带动全国开展长期保存数字内容的活动<sup>8</sup>。

### 2.2.3 数据的保存

对于数据的保存方面不论是时间还是技术的变化，在策略上和决策上与数据管理的其他方面一样都要确保内容、上下文、数据的真实性、完整性等达到最少量的损失。长期保存政策的选择或使用考虑到以下 5 方面的内容：①不断的转移。数据以独立于软

件的格式保存，随着技术的变化数据要不断转移；②技术保存。数据的保存同时也要保存其存在的硬件与软件环境；③仿真。从资源的外观、感觉和行为上在每一代硬件和软件上进行仿真；④风险评估。长期保存的风险与以下情况息息相关：保存对象的内容、对象被创建时采取的格式、使内容可用的支持硬件和软件的能力、读者和内容托管者的需要、机构预算的限制以及机构管理活动的组织<sup>[1]</sup>；⑤在资源被创建和随后的长期保存对决策都有很强的依赖性，也依赖于可以利用的基金和技术。在将来使用与存取资源时同样要采用长期保存制定的策略，因此它与将来的数据的使用有很大的关系。

例如，CAMILEON(Creative Archiving at Michigan and Leeds: Emulating the old on the New)项目的目的就是测定仿真作为一个长期数字保存方法的可行性。利用虚拟技术，该项目在证明把旧的软件和数据运行在新的机器上的可行性上取得了显著的成绩。CAMILEON计划的另一贡献就是使人们认识到了迁移方法和仿真法都有它们作为保存战略的一个空间<sup>[9, 10]</sup>。

### 2.3 数据的使用

数据产生后即可投入使用，也可能在将来很长的时间内使用。数据的使用可以是基于创建的目的，也可以是创建后二次分析等。数据的使用还涉及到数据的创建者、基金组织、出版商、机构档案馆、图书馆、学术档案馆以及用户的社团等，同时还要定义用户的需求、向用户提供哪些服务。数据产生的决策和周围的环境，数据的管理与资源的长期保存，管理的权限和经济环境，识别与调解不同利益相关者的手段等都对数据的使用产生很大的影响。

怎样将数据传递给最终用户以及用户怎样使用这些数据取决于下列因素：数据产生的原因以及产生的手段，机构相互之间合作的协议，不同机构之间数据的共享和交换，满足法律与经济环境需求的环境和程序等，怎样保存，保存到何处，需要哪些硬件和软件来存取，数据管理和保存的决策也决定着对数据长时间的使用。

澳大利亚国家档案馆将数字资源依照数字保存价值高低作了分类，将数字资源划分为三类，然后在收藏和保存数字资源时以此作为选择的一个参考标准，确定其保存的级别<sup>[4]</sup>。这三个类别是：①临时价值：在一个确定的时间段内存储，这个时间可能是一年，也可能是数字记录产生之后 130 年；②机构永久保存价值：这个记录可能对于某一机构而言有永久的价值，但对专门的长期保存机构则无收藏价值；③存档价值：这样的信息需要长期保存。

## 3 长期保存政策中经济和法律的问题

数字资源生命周期中没有哪个阶段不涉及到其周围的法律与经济环境，而且这些环境与利益相关者的组织的使命相互作用，共同影响到数字资源的生命周期和长期保存政策的实施。法律方面包括：资源的知识和财产所有权；资源附带的协约条款或者访问这些资源所需的软件与硬件环境；保证个人和机构资源的保密性；保证数字资源创建者和其他利益相关者的声望；其他在选择保存保持资源真实性的法律职责。资源是否以机器可读的形式展现或者怎样展现都受既定利益的影响；还影响到资源的使用情况，这些资源将被谁和在什么条件下怎样使用，资源怎样被记录与保存（如敏感的数据需要加密或存取限制）；这些既定利益还影响到资源的是否被合法保存，以及怎样实现合法保存。此外还涉及到权益的转移等其他机制<sup>[11]</sup>。

数字资源的长期保存涉及到的法律问题包括自身的可持续发展，也包括自身与外界的边界性问题，前者我国学者李景等人进行了初步的探讨，认为数字资源长期保存需要得到国家层面、法律层面的支持和保障，例如建立数字资源的呈缴本制度，其中国际图联发布了《电子出版物呈缴制》文件，详细阐述了电子出版物的定义与类型、电子出版物法定呈缴的立法程序、电子出版物的保存技术处理等问题，为各国建立或修改相应的呈缴法提供了指导性的

建议。其他如美国、英国、日本和加拿大等国等颁布了各自国家电子资源的呈缴法。我国新闻出版总署的《出版管理条例》和文化部正在起草的《图书馆法》都涉及到电子资源的呈缴方面的规定,但是比较分散,尚没有形成完善的体系。数字资源的长期保存除了建立呈缴制度外,必须解决版权和隐私等问题,如数字作品的复制权、数据处理权、传播权和服务使用权等问题。如果不对现有版权法律进行研究,就很容易触犯知识产权法<sup>[12]</sup>,继而影响长期保存的持续发展和利用。其次隐私权也是长期保存策略关注的重点,随着长期保存的不断发展,可能会促进第三方机构的出现,承担起长期保存的工作,也有可能一些组织由于经费和技术的问题将数字文献的长期保存任务外包给其他专门机构,第三方的介入都可能会牵涉隐私权的问题。目前国内对隐私权的关注非常缺乏,随着公民民主意识的增强,这方面的问题将日益尖锐<sup>[13]</sup>。

数据的产生、管理、保存和使用都处在相似于商业的环境中,这些环境影响到长期保存政策的实施。在商业环境中产生的资源都有一个商业生命周期影响数据的管理保存和使用。对资源投资的回报可能要求控制数字资源的存取,或者对系统和程序进行加密,锁定资源,让用户注册验证其身份,对使用者收费,加强权限管理等。基金与基金组织的优先权和主要目的在资源的生命周期中各有不同,都从不同的方面影响生命周期的各个阶段。目前在经济方面普遍存在的问题无法准确评估数字资源长期保存所需的成本,对于数字文献的长期管理所需的成本目前还没有明确的定义。这就使得制定的相关计划存在较大风险,同时组织也不愿在不确定的情况下采取过多的行动。由于数字文献保存处于起步阶段,对于网络信息保存的经济分析非常缺乏。包括保存活动实施成本、数字文档存储和长期保存系统成本、人员成本、文档保存方法和保存软件的开发成本以及其他间接的成本影响因素,对这些指标的成本效益加以确定也是很大的挑战<sup>[14]</sup>。

因此从长期保存可持续发展的长效机制分析,只有确保长期保存不会侵犯产生数字内容主体的版权,同时注重保护隐私权的前提下,才能健康发展;其次只有解决成本分析,探讨盈利模式,长期保存促进商业模式的发展,实现生产者和消费者共赢,数字资源的长期保存才能快速发展。

#### 4 总结

数字资源的长期保存政策的制定贯彻其生命周期的每一个阶段,这些阶段考虑其所在环境的法律要求、经济要求、业务需求、技术要求以及对历史的价值等。本文从生命周期的每个阶段概括了制定与遵守长期保存政策应考虑的问题,下一步计划根据文中的分析对长期保存项目进行调研。

#### 参考文献

- 
- <sup>1</sup> OCLC Online Computer Library Center, Inc. OCLC Digital Archive Preservation Policy and Supporting Document[EB/OL].[2011-06-10].  
<http://www.oclc.org/support/documentation/digitalarchive/preservationpolicy.pdf>
  - <sup>2</sup> JISC[EB/OL].[2011-06-11]. <http://www.jisc.ac.uk/whatwedo.aspx>.
  - <sup>3</sup> National Archives of Australia[EB/OL]. [2011-06-09].  
<http://naa.gov.au/services/digitisation-copying/index.aspx>.
  - <sup>4</sup> National Library of Australia[EB/OL].[2011-06-08].  
<http://www.nla.gov.au/digital-preservation>.
  - <sup>5</sup> Cornell Survey of Institutional readiness[EB/OL].[2011-06-06].  
<http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file3612.html#article0>
  - <sup>6</sup> Neil Beagrie, Daniel Greenstein. A Strategic Policy Framework for Creating and Preserving

- 
- 7 Digital Collections[EB/OL].[2011-06-07]. <http://www.ahds.ac.uk/strategic.pdf>  
Stanford University Libraries. Lots of Copies Keep Stuff  
Safe[EB/OL].[2011-09-04].<http://lockss.stanford.edu/lockss/Home>.
- 8 Amy Friedlander. The National Digital Information Infrastructure Preservation Program.  
[2011-4-15]. <http://www.dlib.org/dlib/april02/friedlander/04friedlander.html>.
- 9 安艳杰.国外数字保存研究的相关进展[J].情报杂志,2004(2):128-130
- 10 张艳敏,马秀峰.中外数字资源长期保存策略比较研究[J].图书馆学研究,2009(6):29-33
- 11 宛玲,张晓林.数字资源长期保存权益管理政策研究[J].图书情报知识,2006(2):33-36.
- 12 曼苏,李景,殷立新,潘薇.数字资源长期保存的标准与法律问题综述[J].标准科  
学,2010(5):52-57
- 13 朱焯.数字保存中的技术、组织及法律因素分析[J].图书馆学研究.2005(4):36-38
- 14 李阳.数字文献长期保存的非技术因素探析[J].情报探索,2009(9):59-63

(E-mail: [liws@bjcit.gov.cn](mailto:liws@bjcit.gov.cn); [liws@mail.las.ac.cn](mailto:liws@mail.las.ac.cn))