中国图书馆数字文献资源长期保存现状调查[1]

张玫[1, 2]、李麟[1, 2]、张晓林[1]、刘细文[1]

1 （中国科学院国家科学图书馆，北京，100190）

2 （中国科学院研究生院，北京，100049）

［摘要］ 调查了中国 40 余家主要图书馆在数字资源长期保存的态度、安排以及对保存模式的考虑等情况。调查发现数字资源长期保存已经成为多数图书馆非常关心的问题，多数图书馆对长期保存持积极态度并已开展保存实践，大部分图书馆倾向于采用合作保存模式，但仍有一些图书馆处于观望状态。

［关键词］数字资源　　长期保存　　图书馆

［中图分类号］G250　　［文献标识码］A

Attitudes and Practices for Digital Preservation by Chinese Libraries

Zhang Mei[1, 2], Li Lin[1, 2], Zhang Xiaolin[1], Liu Xiwen[1]

1. (National Science Library, Chinese Academy of Sciences, Beijing, 100190)
2. （the Graduate School of Chinese Academy of Sciences, Beijing，100049）

[Abstract] The paper reports a survey on the attitudes, arrangements, and operational model considerations of more than 40 major Chinese libraries on digital preservation. It revels that digital preservation becomes an urgent concern for these libraries, most libraries take an active approach to the issue and many have already embarked on preservation efforts, and most are positive in collaborative preservation systems, though a few still remain a wait-and-see attitude.

[Key words] digital resources, long-term preservation, libraries

1、调查背景

随着信息技术的发展与互联网的普及，数字资源已发展成为科研与教育领域的主流资源，在许多图书馆馆藏中接近甚至超过印本资源。但是，由于数字资源对技术应用环境的高度依赖性和许多数字资源的使用权采购机制，其可靠利用容易受到自然灾害、技术失效、经济条件和法律限制等因素的影响，使得数字资源的长期保存面临严峻的挑战[1]。目前，数字资源的长期可靠利用已经成为各国图书馆高度关注的问题，许多国家和机构已经积极开展了这方面的努力，我国图书

[1] 本调查受国家科技图书文献中心资助。

馆界也积极进行数字资源长期保存的研究与试验，部分图书馆正在积极促进国家数字资源长期保存体系的建立。

数字资源长期保存体系存在多种模式，包括机构独立保存、第三方委托保存和合作保存。机构独立保存是由保存机构在拥有对被保存资源的使用权的条件下，获得相关权利人（如出版商）许可后进行长期保存，被保存资源的服务范围也大都限定在保存机构内的原有用户。例如大英图书馆（British Library，BL）作为英国法定的出版物保存机构，根据本馆的馆藏策略确保馆藏的永久可获取性，保存范围已经扩展到电子出版物，保存内容的访问服务主要限于大英图书馆自身范围[2]。荷兰国家图书馆（Koninklijke Bibliotheek，KB）的 e-Depot 服务[3]、也属于独立机构保存模式。第二种保存模式是第三方保存，由资源提供方（如出版商）与资源采购方（如图书馆）以外的第三方机构承担委托保存职责的模式，通常由大型数据中心或商业机构或大型图书馆提供第三方保存服务[4]，可共享委托机构的专业技能和设施等来减少各个资源采购方的长期保存投入。Portico 即是具有代表性的第三方保存机构，出版商与 Portico 签订保存许可协议来以保证 Portico 能对资源进行摄入、保存、必要迁移和经授权提供服务的权利，图书馆与 Portico 签订保存许可协议来保存自己所采购的资源并获得在触发条件下对资源的访问权利[5]。另外，英国的艺术人文数据服务（Arts and Humanities Data Service, AHDS）[6]等也都属于第三方保存模式。再一种保存模式是合作保存，由多家机构共同承担长期保存任务，各成员通过合作实现保存风险及保存成本的分担。斯坦福大学图书馆的 LOCKSS 可作为合作保存模式中的最具代表性的例子，参加馆可利用 LOCKSS 技术从出版商处收集相应的保存资源、在本地创建缓存站点，通过和其它参加馆的 LOCKSS 缓存站点一起形成整体保存资源，保证保存内容的完整性[7]。此外，德国的数字信息长期保存合作发展项目（Kopal）[8]、也属于合作保存的模式。

对数字资源长期保存的态度以及对保存模式的选择，一直受到研究者关注。2006 年春，美国图书馆与信息资源理事会（CLIR）与研究图书馆协会（ARL）面向北美高校图书馆馆长，通过问卷调查及电话访谈的方式调查其对长期保存的看法，发现图书馆当前从出版商处获得的许可并不足以保证他们对电子期刊的长期利益；单个图书馆很难依靠自己的力量实现保存的需要；尽管保存实践已在开展，但仍有许多数字文献不在保存范围内；此外，各高校图书馆在长期保存上是机会与挑战并重[9]。2008 年 1 月，Portico 就美国大学图书馆的电子期刊长期保存现状向 1371 位馆长发放问卷，其中回收 186 份，调查结果表明，长期保存的重要性在研究型图书馆界已经形成共识，但对其紧迫性及具体实现方式的理解上还存在明显分歧。多数图书馆采取观望态度，其中还有部分图书馆表示将根据其

他机构的行动再做决策[10]。同样是在 2008 年初，Portico 联合“学术与专业学会出版商协会”（ALPSP）面向全球 400 多家出版商发放问卷，了解他们当前长期保存的运作策略及具体实践。调查回收率超过 65％，调查发现绝大部分出版商都意识到长期保存的重要性，参与保存活动的意愿也十分强烈；几乎所有的出版商都期望保存领域能出现最佳实践或行业标准；在涉及保存责任机构时，相当数量的出版商希望由其他机构、特别是国家图书馆来承担；同时，大多数出版商非常清楚“长期获取”与“长期保存”的区别，他们也表明自己对长期获取负有职责[11]。

2. 调查过程

为了有效推动国家数字文献资源长期保存体系的建设，需要准确了解拥有数字文献资源的机构在长期保存上的态度、原则、战略和实践，包括在选择长期保存模式上的考虑。为此，在国家科技图书文献中心（NSTL）的资助下，我们对中国采购和使用数字文献资源较多的图书馆进行了调查。

本次调查从 2008 年 8 月起到 2008 年 10 月，调查对象包括国家图书馆、数字文献资源比较丰富的部分省市公共图书馆、985 高校图书馆、主要国家级专业图书馆。虽然本次调查并不是对全国范围进行随机抽样，但覆盖了主要的图书馆，调查结果有较好的揭示意义。本次调查发放问卷 63 份，回收 44 份，回收率 69.8%；回收的有效问卷为 44 份。

表 1：调查对象的类型

机构类型	数量	比例
专业图书馆	13	29.5%
公共图书馆	8	18.2%
大学图书馆	23	52.3%
总计	44	100%

本次调查集中为四大问题，即：数字资源的基本情况、数字资源长期保存的安排、对不同保存模式的要求与顾虑，以及长期保存的具体实践。为了深入了解有关情况，在调查后期，对 6 家主要图书馆的负责人进行了面谈或电话访谈。

3. 调查结果

3.1 数字资源的发展情况

数字资源长期保存的紧迫性与数字资源在机构中所处的地位密切相关。我们对这些机构在数字资源建设方面的现状与趋势进行了调查，具体情况如表 2 所示。从现有的主体文献资源形式来看，在调查的 44 家机构中，超过九成的单位

（93.2％，41 家）目前主体文献资源的印本和数字形式大体相当，而其余 3 家单位则仍以印本资源为主。从文献资源的发展趋势来看，76.7%的机构将发展印本资源与发展数字资源置于相同地位，20.9%的机构将优先发展数字资源，仅有 1 家机构表示仍将发展印本资源放在首位。若按机构类型划分，部分专业图书馆和大学图书馆选择优先发展数字文献资源。从表 2 可以看出，印本资源与数字资源共同发展的策略目前仍然占主要地位。

表 2：我国图书情报机构的主体文献资源发展状况

		专业图书馆		公共图书馆		大学图书馆		总计	
		数量	比例	数量	比例	数量	比例	数量	比例
现有主体文献资源形式	印本资源为主	2	15.4%	1	12.5%	0	0%	3	6.8%
	印本资源与数字资源共同发展	11	84.6%	7	87.5%	23	100%	41	93.2%
文献资源发展趋势	优先发展印本资源	1	8.3%	0	0%	0	0%	1	2.3%
	印本资源与数字资源共同发展	8	66.7%	8	100%	17	73.9%	33	76.7%
	优先发展数字资源	3	25.0%	0	0%	6	26.1%	9	20.9%

调查还发现，在主体文献资源的发展策略上选择优先发展数字资源的机构中，目前已开展长期保存实践的比例达到 66.7%（对比于所有机构开展长期保存实践的 59.1%）。由于样本数量很少，我们还不能做简单的统计推论，但可能从侧面反映出注重数字资源的机构对实施保存活动较其他机构略显积极主动。

3.2 对长期保存的安排

（1）对保存范围的安排

鉴于机构对不同类型数字资源的长期保存可能会有不同的考虑，因此，本调查将数字资源划分为自建资源、采购资源及网络资源，以便了解各机构对上述类型数字资源的保存安排。调查的 44 家机构均表示，对本机构的自建资源将全部保存；对采购资源来说，20.5%的机构表示会全部保存，70.5%的机构将有选择的保存，仅有两家机构表明不会保存该类资源；对网络资源的保存来说，70.5%的机构表示会有选择性的保存，在余下的 13 家机构中，除有两家没作选择外，其余 11 家机构都明确表示不会保存网络资源。

（2）对保存机构的选择

超过半数（52.5%）的机构选择图书馆作为数字文献资源长期保存的责任机构，其中一家机构明确指出，这里的“图书馆”是指“国家级的公益型图书机构”；

认为应该由出版商来保存数字资源的机构数量占被调查机构总数的 7.5%，而27.5%的机构认为应该由除图书馆与出版商以外的第三方机构来承担长期保存的职责。此外，还有 12.5%的机构表示目前尚不清楚长期保存的职责应该由哪类机构来承担。从不同类型对象看，75%的公共图书馆倾向于由图书馆来承担保存职责，而在被调查的大学图书馆中，选择图书馆作为保存责任机构的比例与选择第三方机构作为保存责任机构的比例各占 42.9%。

（3）对保存权限的安排

54.5%的被调查机构考虑获得资源的永久使用权（perpetual access），13.6%的机构考虑获得资源的永久备份权（perpetual copy），而超过四成（40.9%）的机构则考虑获得资源的永久备份权及服务所需的关键元数据；此外，还有少部分机构（11.4%）对保存权限的安排仍在考虑中。

在上述保存权限的获取上，多数（75.0%）机构表示准备在与数据提供商签订的采购合同中对期望获取的保存权限加以规定；期望通过签订单独的保存协议或非正式的许可来获得保存权限的机构数量只占很少部分，二者加起来所占的比例还不到 10%；而剩余 18.2%的机构表示目前在该问题上还没有明确的答案。

3.3 对不同保存模式的要求

（1）对保存模式的偏好

对数字资源长期保存的机构保存、第三方保存和合作保存等模式，本次调查邀请各机构按选择倾向程度对三类保存模式加以排序。调查结果如图 1 所示，其中，每类保存模式下的三项直方图分别代表对该类模式倾向度较高、倾向度较低及倾向度介乎二者之间的机构数量。

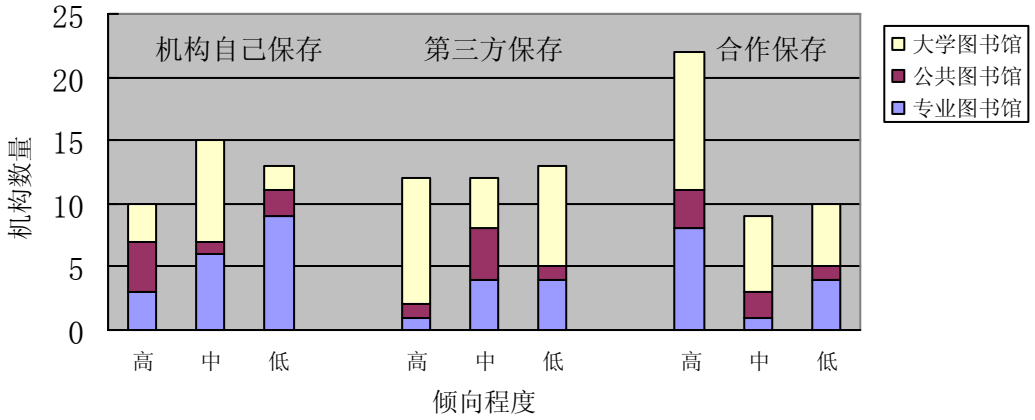

图 1：我国图书馆对不同保存模式的倾向程度

上图表明，若从整体来看，我国图书馆更倾向于采用合作保存模式，但从不同类型机构来看，公共图书馆中首选机构自己保存模式的数量略微高于选择合作

保存的数量；其他两类图书馆中选择合作保存模式的机构数量仍占多数。

此外，对承担第三方保存服务的机构类型来说，84.2%的被调查机构选择国家保存体系作为第三方保存机构；而选择与本机构属于同一体系的机构（如本单位的信息中心等）或商业机构来承担第三方保存的机构数量仅占少数。

（2）不同保存模式的要求与顾虑

鉴于不同的保存模式在服务要求以及关注问题等方面存在诸多不同，因此，本调查也将机构对服务要求和关注问题列入考察范围。

针对机构自己保存的调查结果如图 2、图 3 所示。

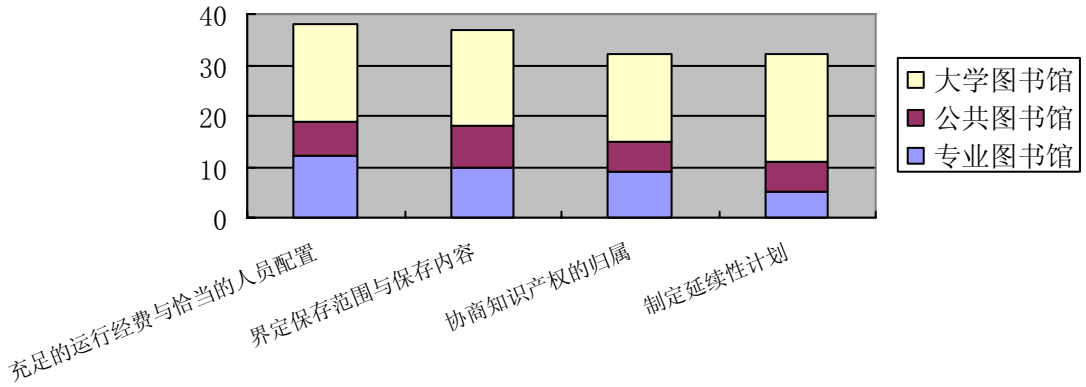

图 2：对机构自己保存的要求

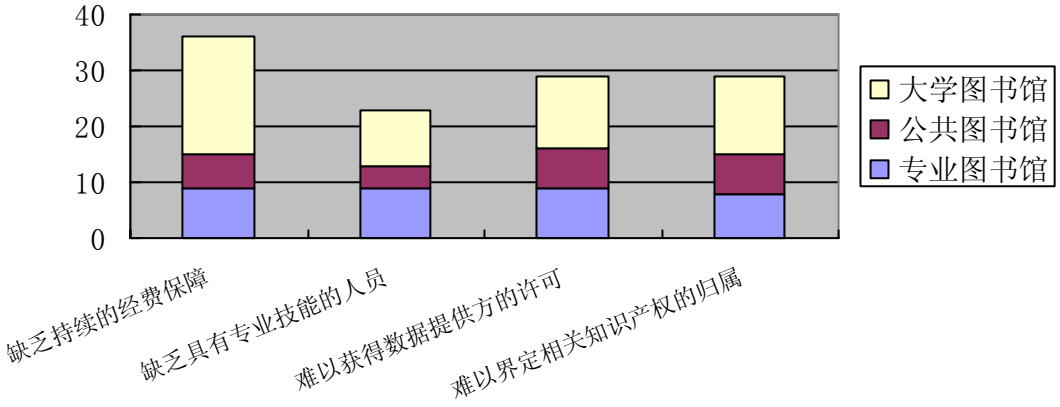

图 3：对机构自己保存的顾虑

从图 3 和图 4 可知，对机构自己保存模式的诸多要求中前两项分别是确保充足的运行经费与恰当的人员配置（90.5%）和界定保存范围与保存内容（88.1%）；对该模式的顾虑主要包括缺乏持续的经费保障（83.7%）、知识产权的归属问题（67.4%）以及担心难以获得数据商许可（67.4%）等问题。

针对第三方保存模式的调查结果如图 4、图 5 所示。

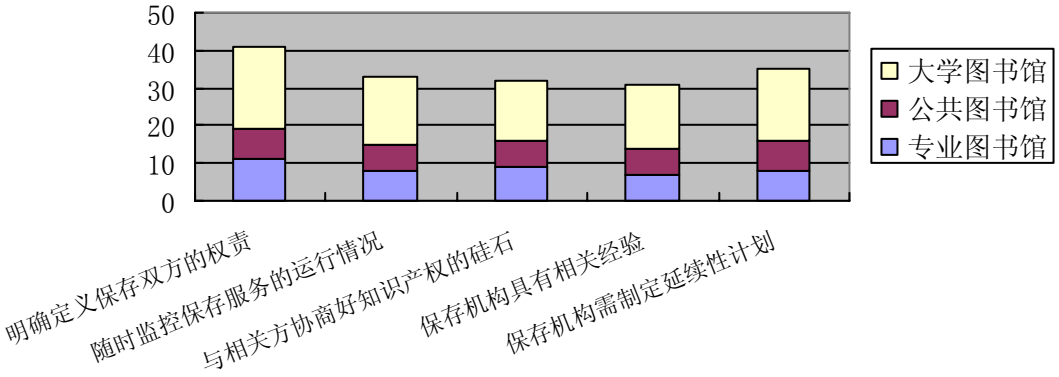

图 4：对第三方保存的要求

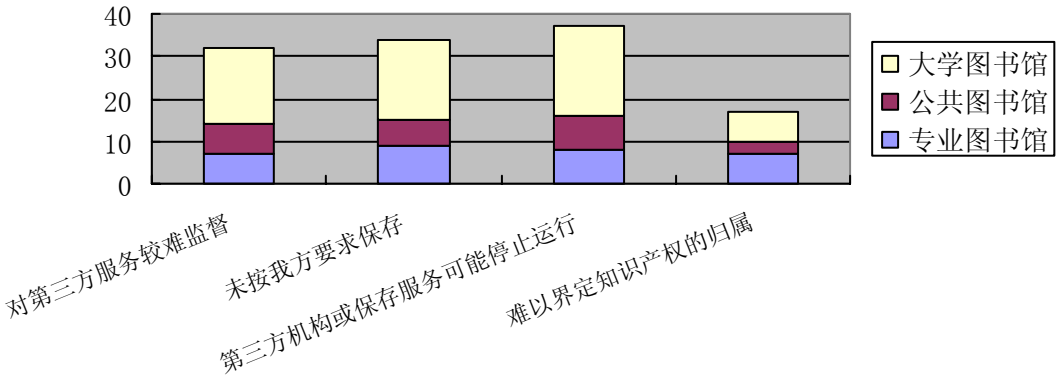

图 5：对第三方保存的顾虑

图 5、图 6 表明，对第三方保存模式，几乎所有被调查机构提出的要求中都包括明确定义保存双方的权责（97.6%），其次为要求保存机构制定延续性计划，以避免在其停止运行时丢失保存的数据(85.7%)；在对该类保存服务的顾虑方面，90.2%的机构担心第三方机构或其提供的保存服务可能存在停止运行的风险，而知识产权的归属问题较其他问题在第三方保存服务中并不明显，仅有不到一半的机构（41.5%）表示担心。

针对合作保存模式的调查结果如图 6、图 7 所示。

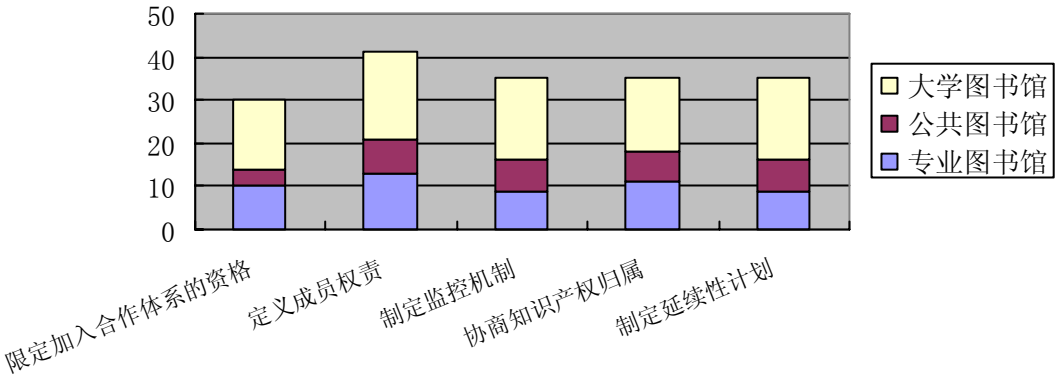

图 6：对合作保存的要求

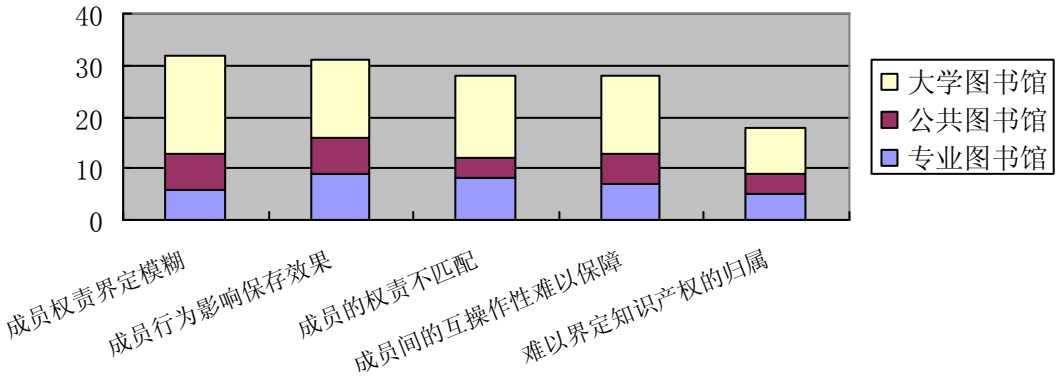

图 7：对合作保存的顾虑

对合作保存模式的最主要要求为明确定义合作体系各成员的权责（93.2%）；而顾虑主要包括：合作体系成员的权责界定不清晰（76.2%）以及成员的某些行为（如不事先告知就退出合作体系）可能会对保存效果造成负面影响（73.8%）。

3.4 长期保存的实践

为了解数字资源长期保存在我国图书馆领域的实践状况，本调查首先对被调查机构中目前已经开展了保存活动的机构数量加以统计。调查表明，在被调查的44 家机构中，59.1%的机构已开展保存实践。具体情况如图 8 所示。

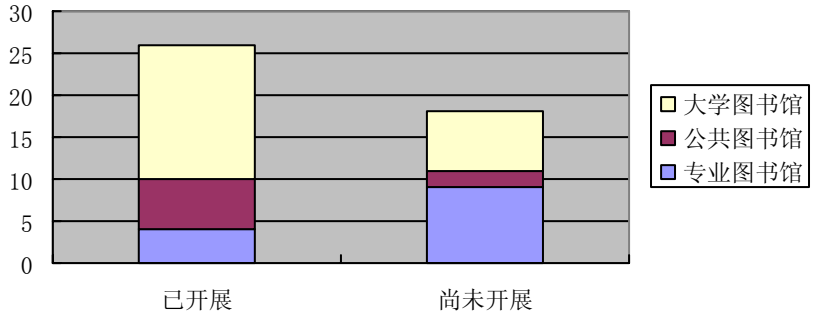

图 8：我国图书馆目前开展保存实践的情况

可看出，多数大学图书馆与公共图书馆都在开展保存实践，而专业图书馆中多数还尚未开展保存实践。下面将具体做作进一步分析。

（1）尚未开展保存实践的机构

在没有开展保存活动的机构中，66.7%的机构认为由于我国长期保存的实践较少，还需观察一段时间后再决定是否开展保存活动；把经费因素与“数字资源有相应的印本资源作为备份”列为尚未开展保存的原因的机构均有 38.9%；由于其他机构已对相同的数字资源进行保存而影响其开展保存的机构占 16.7%。此外，还有一家明确指出阻碍其开展保存活动的主要原因是缺乏机制保障。

在涉及未来保存计划时，尚未开展保存活动的大部分机构（66.7%）对未来

两年内是否保存本机构的数字资源仍不确定，另外 37.5%的机构表示两年之内会采取一定的保存措施，还有三家表示两年内不采取任何保存措施。

（2）已开展保存实践的机构

已开展保存活动的机构中，从资源来源来看，自建资源是所有机构都将保存的类型，三分之二机构会保存部分采购资源，29.6%机构对网络资源也予保存。

从保存资源的类型来看，大部分机构（74.1%）的保存对象中包含科技期刊，44.4%的机构保存档案资料，而保存网页与图片的机构均有 29.6%。此外，被调查机构补充的保存对象类型中，还包括学位论文、电子图书、音频视频、PDF 报纸、数据库以及科技报告等。

从选取的保存模式来看，所有机构均采用自己保存的方式，在此基础上，有 4 家机构还额外参与了合作保存体系来保存某类资源。从采用的保存系统来看，有 4 家机构采用 Dspace，其余机构采用的系统还包括 dILAS/ MagicStore、TRS、TPI 及硬盘备份系统。

从机构是否制定专门的保存政策来看，有 17.9%已经制定了相关政策，50%机构的保存政策正在制定中。在已制定的保存政策中，包含定义保存的职责范围的占 22.2%，用于定义保存对象的筛选标准的占 18.5%，定义保存经费来源的占 11.1%，而用于定义保存工作流程的占 25.9%。

从影响保存效果的因素来看，调查结果如图 9 所示。从图中可以看出，完善的保存政策框架对保存效果的影响得到多数机构的认可。此外，还有两家机构提到领导的重视及与其他机构的广泛合作也会对保存效果产生影响。

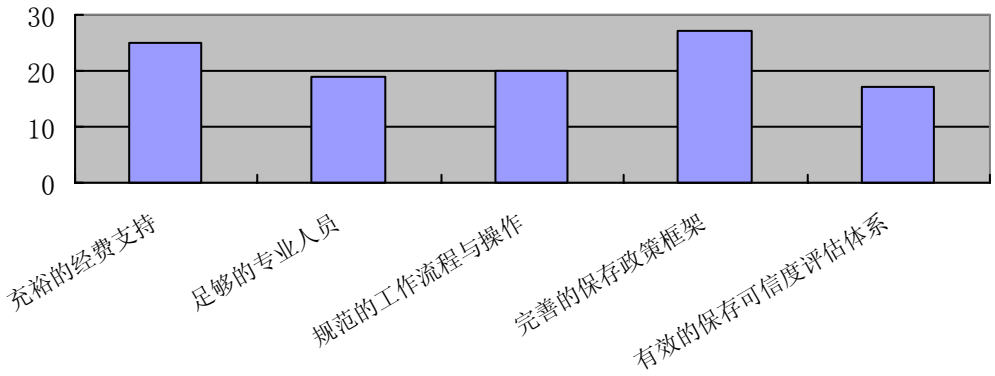

图 9：影响保存效果的因素

在涉及保存效果的评估问题上，42.9%的机构主要通过制定长期保存的评估要求，并由上级机构或本机构定期检查来保证保存的效果；17.9%的机构则采用邀请第三方机构按照特定标准来对长期保存的运行效果加以审核；还有 53.6%的机构目前尚未开展对长期保存效果的评估活动。

3.5 针对我国建立数字文献资源国家保存体系的建议

本次调查还通过一个开放问题，请被调查机构对我国建立数字文献资源国家保存体系提出建议。有 50％的机构做了回答，有关观点概括如下：

（1）增强长期保存的意识：不仅需要增强业内人士的长期保存意识，还应该加深高层领导对长期保存的必要性与重要性的认识。可通过加强对数字文献资源长期保存的国内外现状调研及相关策略规范的研究、加强数字文献资源长期保存的必要性可行性论证、积极为图书馆等机构承担数字文献资源长期保存提供理论和实践依据，争取将长期保存的职责写入制定中的图书馆法中。

（2）有必要建立国家保存体系：国家合作体系比单个机构更容易解决标准、管理、安全、关键技术等问题，避免重复投入、集中优秀人才，使数字文献资源保存的受益面最大化。国家还应支持行业间区域间的数字文献保存体系建设。

（3）制定国家保存体系的政策框架：应该从国家层面上制定长期保存体系的发展政策和指导意见，这是建立国家保存体系的基础，包含：合作体系的参与机构及其权责；制定保存相关的技术规范体系：制定认证与评估机制。

（4）针对不同类型的保存对象，可选择不同的保存模式，例如，自建资源可考虑自己保存，如本机构不具备条件时可委托第三方保存；商业性采购的数字文献资源可由国家指定机构保存；网络资源则可以由多个机构合作保存。

4. 调查小结

本次调查初步了解了国内图书馆界对数字文献资源长期保存的基本态度，这与国外调查具有相同之处；此外，本调查还从资源类型对保存安排的影响、保存模式的选择、保存模式的要求与顾虑等方面作了进一步的探究。

从调查结果看，绝大多数机构文献构成中数字资源已经与印本资源相当，而且多数机构将采取印本资源与数字资源共同发展的策略、部分机构将采取优先发展数字资源的策略，数字文献资源长期保存的现实紧迫性已经不言而喻。同时，所有被调查机构都将长期保存自建的数字文献资源，绝大多数机构将保存或有选择地保存采购数字文献资源，超过半数的机构还倾向于由图书馆来承担长期保存职责，这些都体现了我国大型图书情报机构对长期保存的积极态度。

调查还表明，多数机构已经开展了长期保存的实践，多数机构（尤其是大学和专业图书馆）倾向于采用合作保存模式，而且在考虑第三方保存服务时绝大多数机构选择国家保存体系。大部分机构表示，期望通过参与国家合作体系，获得开展长期保存活动所需的经费来源和人才培养上的资助。这些都为我国数字文献资源长期保存的有效开展奠定了良好基础。

通过调查也看到，由于数字文献资源长期保存在我国才刚刚起步，开展机构不多，保存规模也相对有限，相关的技术、管理机制、法律权限等方面的研究与

操作还不成熟，图书馆对此还存在一些顾虑，部分机构对是否在近期开展长期保存还持观望态度，这与 Portico 调查反映出的美国高校图书馆对该问题的态度非常相似。此外，长期保存的持续经费保障和明晰的知识产权授权，合作保存中需要明确定义保存各方的权责和建立可靠的可持续计划等，都是人们关注的问题。因此，如何通过理论与实践研究，通过建立相对健全的法律授权、技术标准、管理规范、业务流程、尤其是通过可靠的示范系统，提高图书情报机构开展长期保存的信心和参与长期保存合作的积极性，还是一个亟待解决的问题。

参考文献

1 胡启恒. 加强数字资源长期保存是我国科技文献保障体系的战略任务。数字图书馆高层论坛 2008 年年会，北京，2008.11.

2 British Library. British Library: Digital Preservation Strategy. [2008-11-13]. http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/digpresstrat.pdf.

3 KB. e-Depot and digital preservation. [2008-11-13]. http://www.kb.nl/dnp/e-depot/e-depot-en.html.

4 Michael Day, et al. Co-operation for digital preservation and Curation: collaboration for collection development in institutional repository networks. [2008-11-13]. http://ils.unc.edu/digccurr2007/papers/dayPennock_paper_9-3.pdf.

5 Portico. Portico Brochure. [2008-11-13]. http://www.portico.org/about/portico_brochure.pdf

6 AHDS. The Arts and Humanities Data Service. [2008-11-13]. http://www.ahds.ac.uk.

7 LOCKSS. How It Works. [2008-11-13]. http://www.lockss.org/lockss/How_It_Works.

8 Kopal. kopal: Co-operative Development of a Long-Term Digital Information Archive. [2008-11-13]. http://kopal.langzeitarchivierung.de.

9 Anne R. Kenney, et al. E-Journal Archiving Metes and Bounds: A Survey of the Landscape. [2008-11-13]. http://www.clir.org/pubs/reports/pub138/pub138.pdf.

10 Portico. Digital preservation of e-journals in 2008: Urgent Action revisited. [2008-11-13]. http://www.portico.org/comment/porticosurveyondigitalpreservation.pdf.

11 Portico. ALPSP and Portico Collaborate on Long-term Digital Preservation Survey. [2008-11-13]. http://www.portico.org/news/102108.html.

作者简介：

张　玫：女，1982 年生，中国科学院国家科学图书馆 2006 级博士研究生，发表论文十余篇

李　麟：女，1981 年生，馆员，发表论文数篇。

张晓林：男，1956 年生，研究员，博士，发表论文多篇，编著论著若干部

刘细文：男，1965 年生，研究员，发表论文多篇，编著论著多部