

【开放获取】

美国图书馆和信息资源委员会关于 机构知识库的普查报告（摘要）

----MIRACLE 项目的研究发现

Karen Markey, Soo Young Rieh 等著 谷秀洁¹译

(1 北京大学信息管理系 100871)

引言

MIRACLE 项目 (Making Institutional Repositories a Collaborative Learning Environment project, “构建机构知识库的协作学习环境”项目) 是美国博物馆及图书馆服务事业局的研究项目, 项目组设在密歇根大学。《MIRACLE 项目的研究发现》描述了 2007 年 2 月美国图书馆和信息资源委员会进行的美国全国范围的机构知识库普查情况, 报告全文共 181 页。本文主要根据其报告的总结性摘要编译而成。研究人员通过电话访谈机构知识库工作人员、深入调研 5 家机构知识库、开展用户调查、实地检索机构知识库等方法比较全面地揭示了美国机构知识库建设的现状和经验。

1. 为什么要研究机构知识库?

相当多的学术档案是原生数字资源, 并且一些学术作品被制成数字格式后就没有实体的、可以拿到手的副本。数字学术资源数量急剧增长, 使得怎样组织、获取和永久保存成为十分迫切的问题。学术机构的应对措施是通过建立机构知识库 (IRs) 来管理机构的研究团体创造的数字学术成果。建立机构知识库需要投入大量的资金、人力和技术。因此, 如果学术机构之间能够取长补短、分享经验、构建模型并且传授好的做法将是非常有益的。这种分享也可以为那些还没有进行机构知识库建设的学术机构理清思路。

2. 为什么要普查美国的机构知识库 (IR)?

以前有关机构知识库的调查重点关注已经建立 IR 的学术机构或者愿意尝试新技术的针对机构内某一特定群体建立机构知识库的学术机构。为避免重复, MIRACLE 项目的研究人员希望广泛撒网、填补空缺。美国学术机构的机构知识库普查包括了还没有加入 IR 建设行列的组织。广泛的包容性使我们确信能够获得大量的决策者构思、计划、试行或者实施机构知识库的行之有效的实践、政策和做法, 也可以让我们清晰地了解为什么一些学术机构开始全面进行 IR 的建设。

3. 谁参加了普查?

MIRACLE 项目的研究人员联系了 2147 家学术图书馆的领导和高层管理人员, 其中只有 446 家参加了普查, 回复率为 20.8%。其中, 236 家 (占 52.9%) 单位还没有计划建设 IR (简称“没计划”, NP), 92 家 (占 20.6%) 刚开始计划 (简称“计划中”, PO), 70 家 (占 15.7%) 已经积极地计划并尝试 IR (简称“试验单位”, PPT), 48 家 (占 10.8%) 目前已经有 IR 运行和使用 (简称“运行单位”, IMP), 问卷填答者中 74% 是图书馆馆长, 8% 是副馆长, 10% 为馆员。

4. 什么样的教育机构建设或不建设机构知识库?

根据高等教育机构卡内基分类表 (CCHE) 分析, 大多数研究性大学已经计划、试验或者运行机构知识库了, 而回答“没计划”或者“计划中”的单位多数是硕士或本科院校。

5. 谁负责 IR 的计划、试验和执行?

如果假定工作人员专职从事 IR 建设, 并且还是 IR 委员会的成员, 那么在“试验单位” (PPT) 和“运行单位” (IMT) 中, 图书馆员承担了 IR 试验和系统执行的主要责任。资金通常来自图书馆。典型的办法是将费用消化在图书馆日常运行成本当中。

在“没计划” (NP) 单位里, 下面的工作也在进行。馆长带头与教务长、首席信息官、教员、档案员商议有关资金、技术、专家意见、可能的贡献者、使用者以及数字馆藏问题。从试验到执行阶段, IR 委员会成员变得越来越不介入项目, 仅仅让图书馆“拎包” (holding the

bag)。

6. 什么是有用的研究活动？

不同程度介入 IR 建设的人员对 IR 相关信息有不同的喜好，特别是对那些与自己差不多的机构的成功实践感兴趣。然而，这种需求并不比其他投资活动更优先。

测试一个以上的程序包非常重要。本次调查的受访者当中，16%测试了一个以上的程序，几乎 3/4 的“计划中”机构要测试 IR 软件。测试软件有利于开发必要的专门技术、评估系统软件、估计执行成本。本次统计中，多数“计划中”机构下一步会扩大调研；多数“试验单位”下一步是运行 IR 系统软件；“计划中”机构和“试验单位”里，只有很少的机构（约 10%）可能会中止建设 IR 的努力。

7. 被访者对于 IR 系统软件包的体验如何？

被访者在测试和执行阶段首选的是 Dspace 软件。关于软件运行的时间，“运行单位”的受访者中 52.1%选择 12 个月以下；27.1%选择 13-24 个月；4.2%选择 25-36 个月，16.6%选择 36 个月以上。IR 系统的功能是令人满意的，但是其用户界面，包括受控词汇搜索和规范文件控制需要认真地修改。除了 PDF 文件外，其他格式的文档不能保证永久保存。提高 IR 系统的保存功能是系统发展的首要任务。已经运行 IR 的受访者说，更强的保存功能是他们迁移数据的主要原因。关于日报功能（date），受访者选用的系统评估方法多数还是简单计数（simple counts），而多数 IR 系统可以自动生成管理报告。

8. 测试和运行阶段的内容？

测试和运行阶段的 IR 规模都很小。大约 80%的测试库和 50%的运行库都不超过 1000 份数字文档。只有 4 个（8.3%）测试库和 7 个（19.4%）运行库有超过 5000 份文档。IR 的年龄和大小没有关系。测试库和运行库包括各种文本、数字和多媒体文件。传统的基于文本的文档类型是分支机构教师和学生的主要研究成果，也是机构知识库文档的主要内容。

9. IR 政策方面有哪些进展？

“运行单位”有 60%回答说有完整的 IR 政策：包括（1）可接受的文档类型；（2）决定谁有资格向机构知识库投稿；（3）定义馆藏；（4）限制获取机构知识库内容的规定；（5）识别元数据和规范文件元数据的创建者；（6）决定什么是可以接受的内容。即便是回答“正在起草政策”或是“尚未制定政策”的机构也有许多与 IR 相关的活动。

当机构公开 IR 时，没有必要急于制定完备的 IR 政策。抱着“等等看”的态度，过一段时间评估一下发生了什么，然后再固化已有的政策，根据需要添加新的条款是最有利的行动方案。

10. 谁来投稿？以什么频率？

被授权向 IR 存储的人是机构学习社区的典型成员——教员、馆员、专家、档案员、本科和研究生。为科研和教学任务提供便利的教员（例如，出版、新闻服务、教务人员、中心机房的员工）比较少地得到授权。当问及 IR 的主要供稿人都有谁时，只有“试验单位”一致认为 60%来自教师。这个比例在“计划中”单位和“运行单位”分别是 48.1%和 33.3%。“试验单位”回答一致的原因大概缘于他们在计划和测试阶段就与教师“一对一”地接触较多的事实。实际上，三种机构都认为机构知识库的工作人员与每个早期的供稿人“一对一”接触是吸引 IR 内容的最成功的方式。其他成功的方式还有“口碑”传播、亲自拜访教职工以及在院系、教师大会上展示 IR。

受访者表示，为机构知识库增添内容是非常困难的。在运行 IR 的机构，工作人员希望能从制度上授权，要求机构各研究团体的成员向 IR 存储一定的文件。问及为什么人们会向 IR 贡献内容时，回答“提高学者的名誉”和“卸下从研究到传播的重任”的占有较高的比例；回答“提高本机构排名”的分值并不高。

11. 建设机构知识库有哪些好处？

受访者对建设 IR 的 14 个好处进行了评级。除了 2 项之外，其他选项的等级都很高。（图 1）说明建设 IR 的好处不只两三项。也有可能是填答者认为，现在就确定某项收益高于另外一项还为时过早，因为 IR 本身还“没有成形”。一旦 IR 在各种高等教育机构中变得更加普遍，对这个问题的回答可能就有所不同了。最终可能只有一两种益处会占据主导地位。“没计划”机构的填答者特别关注 IR 的好处。这样他们就能够以此支持自己的论点，使本单位决策者下决心支持 IR 计划。

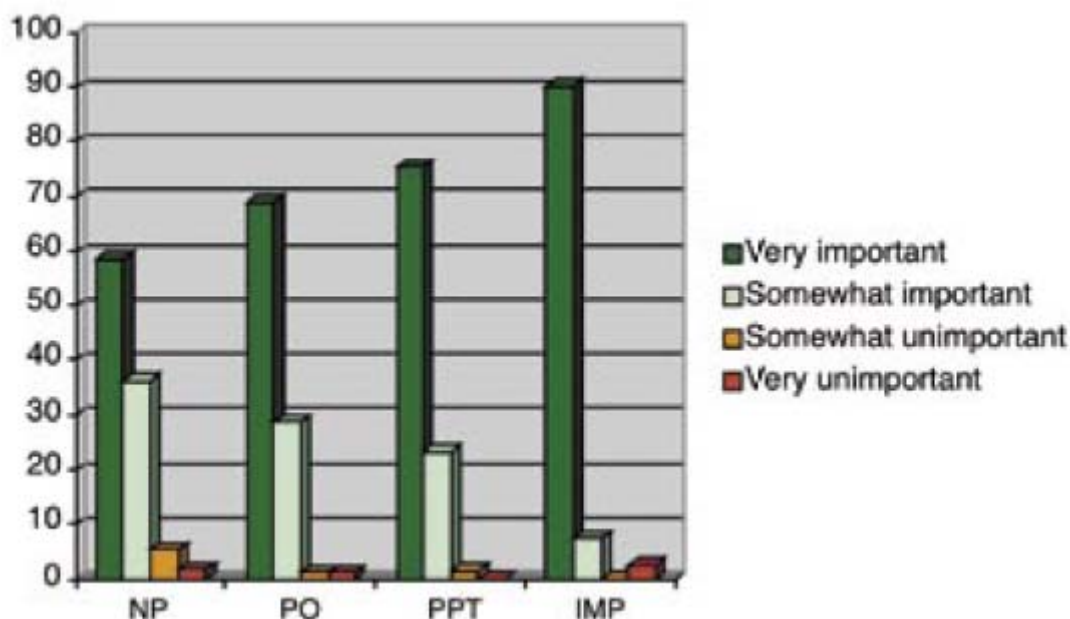


图 1 对 IR 的好处的肯定意见逐渐增加

12. 影响 IR 成功发展的制约因素有哪些？

不同阶段的制约因素是不同的（表 1）。“运行单位”考虑最多的是 IR 的供稿人和供稿内容。事实上，正是出于这种担心，他们才大力推动本单位特定类型资料的呈缴制度。“试验单位”不仅担心没内容，还操心要与其他项目竞争资金的优先权。“计划单位”考虑 IR 的维护问题，比较关心争取资金和维护的成本问题。

表 1 影响 IR 成功发展的制约因素¹

强制约因素 (1-4)	计划中单位	试验单位	运行单位
缺乏校内特定类型资源的呈缴制度（如，硕博论文、教员的预印本等）	(5)	(8)	1
供稿者不了解自己能从 IR 中得到什么好处	4	1	2
使员工确信 IR 并不违背目前的出版模式	(10)	(5)	3
供稿者担心数字资料的知识产权	3	4	4
中度制约因素 (5-8)	计划中单位	试验单位	运行单位
鼓励员工贡献内容	7	(2)	5
与其他的优先安排、项目、计划竞争资源	(1)	(3)	6
使机构学习社区的成员了解 IR	(12)	(9)	7
供稿者担心通过 IR 系统提交内容操作太复杂	6	7	8
弱制约因素 (9-13)	计划中单位	试验单位	运行单位
支撑 IR 运行的全部费用	(2)	(6)	9
供稿者不能很准确地描述特征元数据	9	10	10
数字文件长期保存的困难	11	11	11
IR 系统的数字保存能力不充分	13	12	12
缺乏 IR 系统的校内技术专家	(8)	13	13

13. “没计划”的机构加入 IR 行列的可能性有多大？

本项目 52% 的受访者来自于没有 IR 计划的机构，主要是本科和研究生教育机构，在他们当中也有潜在的建立 IR 的雄心。这取决于计划、运行和维护 IR 的成本，以及与他们相当的机构有哪些已经开始考虑 IR。尽管没有哪项原因足以制约“没计划”单位加入 IR 的行列，但是，目前“没计划”单位还有其他需要通盘考虑的事，或者开展 IR 的资源 and 专家不足。几乎没有哪家“没计划”单位对 IR 一无所知。约有不到一半的“没计划”单位的填答者说，他们

¹以“运行单位”为参照进行排序，圆括号内表示相对偏离。

会在 24 个月内开始 IR 计划。

当问及 MIRACLE 项目将怎样支持他们时，“没计划”单位回答：学习（1）IR 的一般知识；（2）细节和专业知识；（3）最佳实践；（4）了解 IR 的好处；（5）获得基金；（6）合作机会。与研究性大学不同，“没计划”单位感兴趣是其他非研究性大学的同仁们能够提醒他们，共同分享成功的经验。这也是本项目在后续研究活动中关注非研究性大学的机会，因为哪里需求最大，哪里有关 IR 知识的缺口就越大。

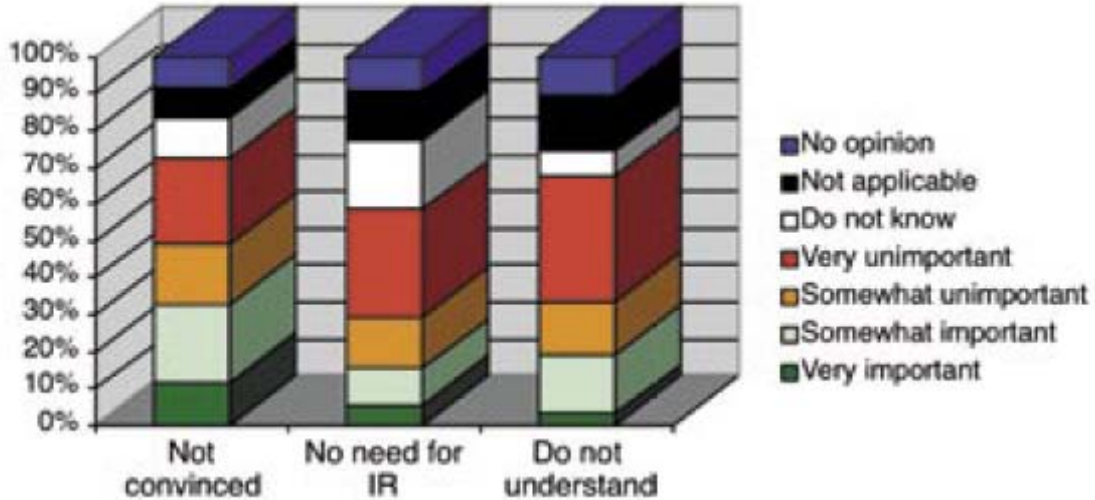


图 2 “没计划”单位不建设 IR 的原因

14. MIRACLE 项目普查的发现证实了哪些原先的发现？

本项目证实了许多原先调查的发现。即研究性大学走在前面；在 IR 的建设中，图书馆起主要作用；Dspace 软件在试验和运行阶段运用较广。

15. MIRACLE 项目有什么特别的发现？

9.3 章讨论了 13 个独特的发现，例如，档案保管人在 IR 筹建中扮演的“羞怯人”角色 (shrinking-violet role)；普查发现，被调查者对同类机构如何成功运营 IR 的相关信息具有贪婪的“胃口”；IR 能够为图书馆构建新的关系；以及 IR 对于保存功能的改进要求。

16. MIRACLE 项目结束后，IR 工作人员将面临的长期问题是什么？

9.4 章讨论了七个这类问题，例如，IR 的好处、IR 对目前出版模式的影响，以及要求学习社区成员向 IR 呈缴他们教学及科研的产品和副产品等。

编译自：

1、Karen Markey, et al. Census of Institutional Repositories in the United States: MIRACLE Project Research Findings <http://www.clir.org/pubs/execsum/sum140.html>

2、Karen Markey, et al. Census of Institutional Repositories in the United States: MIRACLE Project Research Findings. <http://www.clir.org/pubs/abstract/pub140abst.html>.

(李麟 校)