

·开放获取·

## 英国知识库的技术、组织模式研究报告（概要）

Alma Swan, Chris Awre 著 林颖<sup>1</sup> 编译  
(北京师范大学图书馆)

英国的信息系统联合委员会（JISC）委托项目参与者承担一项大范围研究，目的是确定可持续的技术和组织模型，以支持不同数字知识库（digital repository）面向用户的服务。与英国高等教育团体的密切相关的开放获取知识库被认为具有特别的相关性。这项研究希望能够形成一些政策来支持知识库的访问和使用，目的是建立一个国家知识库服务的基础设施或框架。

### 1. 用户需求和汲取的教训

用户及其需求如下：

✓ **知识库管理者：**他们的需求包括：帮助在其机构内建立知识库；对版权问题、建立并维护一个知识库以及知识库运营的技术问题提出建议，如数字化、文件格式、元数据结构、保存、数据揭示（例如 OAI 标准）、命名系统；访问和认证系统；知识库使用服务；提供管理维护帮助和向终端用户成员提供本地知识库服务的帮助。

✓ **终端用户：**对资源检索者而言，他们需要资源发现工具和增值内容。对内容提供者而言，他们需要在某地完成资源的存放操作、同行评议服务；他们需要与知识库管理者共同了解服务的使用和影响、技术建议或协助以及对于版权的建议。

✓ **内容聚合器：**主要需要掌握准确并且足够的元数据。

✓ **元用户（将知识库中开放获取的内容用于分析工作中的用户）：**他们需要与知识库管理者和终端用户共同了解知识库的使用统计；他们也需要用于研究评估和监控的工具。

✓ **企业方（建立知识库服务的一方，例如出版商、二次分销商、技术转移专家）：**他们需要优秀的资源发现工具、链接服务（用于识别知识库、知识库特征和知识库中存放资源的导航和定位工具）以及专业转移方面的专长。

通过对用户的需求进行分析，我们认为一个知识库服务体系从整体上主要分为 3 大层次。摄取层的服务主要是指知识库管理员和资源存放者对技术和过程的需求服务。内容聚合层则是与技术相关的元数据创建和增强服务。在聚合层之上则是输出层，这一层的服务主要通过知识库的内容实现，例如，提供特定的保存需求、研究评估和监控、资源发现、出版、同行评议的开放获取期刊(overlay journals)、元分析以及链接服务。

一些已经完成或正在进行的研究以及对英国知识库框架产生重要影响的专家，都提出了许多教训和建议，主要包括以下内容：

✓ **摄取层：**知识库在实施保存操作或获取内容的时候，由于各个机构的技术能力差别较大，知识库所提供的元数据质量也会有很大的变化。同样，知识库资源内容的数量也会有很大的变化。例如，对作者的宣传对提高研究后印本的存缴程度就是非常重要的。在这方面，IP 限制和版权依然是主要的障碍。其中一些障碍会给知识库管理者的发展进步带来严重的消极影响，这也意味着用户能够获得并使用的开放获取的资源服务依然水平较低。

✓ **聚合层：**元数据质量（即使是自带的元数据）依然是主要问题，这项研究提出的技术模型描述了在这方面的一种理想方式（如下所见）。

✓ **输出层：**特定资源发现工具是非常有用的，它帮助用户根据他们的特定需求链接知

---

<sup>1</sup> 林颖，北京师范大学，100875，liny@lib.bnu.edu.cn

识库, 尽管用户也可以以其它的方法进入知识库。如果用户知道他们要查找资源存放的地方, 他们就可以检索一个指定的知识库。如果用户要查找一个指定学科或主题, 他们也可以使用基于学科的门户或者是聚合了感兴趣的对象类型(例如移动影像或学位论文)的知识库门户。许多情况下, 用户会通过 Google 或其它 web 搜索引擎获得知识库内容, 知识库管理者和作者也需要评估 Google 和其它 web 搜索引擎对知识库内容揭示的风险。在国家层面的框架中就非常有必要考虑到这些因素。

## 2. 全国知识库规划中的备选服务和组织模型

一个国家层面联合的知识库规划的基础和需要的备选服务如下:

### 摄取层:

- 数字化服务
- 对 IP 限制和权益提供建议的服务
- 对开放获取及其资源提供建议的服务
- 提供技术帮助的服务
- 知识库构建服务
- 知识库管理服务

### 数据层:

- 机构知识库
- 对没有机构知识库的作者提供国家层面的"包罗万象的"过渡性的知识库
- 收集一次文献的学科知识库
- 收集一次文献的媒介知识库

### 聚集层:

- 元数据创建和增强服务

### 输出层:

- 访问和认证服务
- 使用统计服务
- 保存服务
- 研究评估和监控服务
- 资源发现服务
- 出版服务
- 同行评议的开放获取期刊服务
- 元分析服务
- 链接和映射服务
- 技术转换或商业咨询服务

除了列表中最后一点外, 目前存在的项目或计划已在不同程度上完成, 或正在被提供, 但是还是有许多项目受到分散区域的限制, 或仅仅只是一个示例项目。如果要将这种试验或项目的操作层次提高至一个可行的国家框架层面, 那么还需要更细致的计划制定和来自 JISC 的更有力的领导。

最迫切需要完成的服务如下:

- ✓ 为没有机构知识库的作者建立"包罗万象"的中间知识库
- ✓ 全国性的资源发现服务
- ✓ 元分析服务, 尤其是能够影响将来国家层面研究评估的引文分析和书目分析服务

- ✓ 知识库使用和统计服务
- 次优先的服务则包括：
- ✓ 没有从专门服务中获益的跨区域工作的保存服务
  - ✓ 国家层面的名称规范服务
  - ✓ 国家层面的文件格式（转化）服务

### 3. 技术模型

提出聚合模型是为了支持终端用户服务的建立。这个模型根据以往的经验一般使用 OAI-PMH（元数据收割开放存档协议）作为资源采集的首选方案。但是，考虑到资源内容和其它知识库开放获取的深度，还需要进一步地考察 OAI-PMH 的能力和它可能会使用到的标准和技术。

聚合将元数据和潜在信息一起作为终端用户服务的基础，避免了与多个单独的知识库进行操作的服务。聚合增强了对数据的控制，保证了信息基础的稳定，它通过来源知识库来控制内容。经常性的聚合实现终端用户服务有效和最新的访问。更有价值的是，聚合还允许重新分解元数据或内容使其较之单个知识库能更好地适用于终端用户服务。

#### 元数据和内容

知识库终端用户服务的核心是知识库中数字资源的高质量元数据。知识库中任何格式的元数据自动生成都需要外加工过程，同时还要考虑如何鉴定元数据能在什么地方生成。元数据生成不仅需要指导在哪里能够生成一个能在知识库内部使用的丰富的元数据记录，而且这些元数据还是以外格式用于资源聚合的基础。

与元数据的揭示相比，用于聚合的内容揭示不太容易理解。在内容揭示的时候我们要知道想做什么，才能针对这些需求应用更恰当的技术，因而这就需要准确地建立模型。在决定将数字内容及相关子内容赋予标识符的作用时，建模方式也是很有用处的，它能够便于聚合工具或用户服务更清晰地识别他们的工作对象。

#### 知识库接口

OAI 服务者通过 OAI-PMH 的广泛地应用来促进知识库之间的访问。我们还可以通过其它协议增加知识库的附加值，使得服务提供者的内容聚合能够更好地关注他们会向用户提供什么样的服务。例如，RSS/ATOM 新闻反馈是源自单个知识库自身需求的小的聚合方式，而 RSS/ATOM 阅读器就是多个知识库对这些新闻反馈的聚合。Web 爬行器可聚合当前的网页信息，其中常见方式就是 web 搜索引擎。这两种方法分别提供了用于建立终端用户服务的不同的聚合途径。

#### 聚合和终端用户服务

聚合器编辑后，可以聚合元数据并将此作为元数据生成的基础，这是一种实现规模经济的更可行的元数据加工选择。通常，聚合并不直接作为终端用户服务提供，而是通过一些接口方式来创建用户服务。比如，可以通过 OAI-PMH、RSS 或 web 爬行器进行重新揭示以便实现更高层次的深度聚合。一旦一个用户服务进行了内容揭示，那么还有必要建立与来源知识库的返回链接，这样能够提供额外的功能和全文资源。

#### 建构方法

聚合模型中有三个组成部分，分别为知识库、聚合器和终端用户服务。一般而言，知识库相对独立于聚合器，而终端用户服务与聚合器的联系更密切，而目前的趋势是要分离这两部分（例如通过 web2.0 方式）。目前的一种观点是将这三个组成部分都视为服务，这样能够促进建立面向服务的体系结构，这种面向服务的方式能够最大可能地发挥各个组成部分在实现上的灵活性。aDORe 和 CORDRA 两个具体实例说明了组成部分是如何实现联系的。两项计划都实现了尽可能丰富地揭示元数据，促进了不同知识库终端用户服务的聚合和建立。。

aDORe 还实际证明了 CORDRA 的许多概念，这两项研究都对将来的发展有益。

### 展望

聚合模型各个组成部分之间的交流对建立高效的终端用户服务是至关重要的。这能够加强个性化服务的建立，因为用户的需要会随着他们教育与研究的变化而变化。

## 4. 知识库服务的业务模型

知识库服务要采用一些可行的业务模型，主要集中在五个方面：

- ✓ 机构支持：适于数字化、知识库提供、在一定层次上的保存、同行评议的开放获取期刊的出版。
- ✓ 公共资金支持（例如 JISC 的高风险基金）：恰当的“包罗万象的”中间知识库的咨询服务、元数据创建和增加、资源发现、技术转移以及链接服务。
- ✓ 社会支持：适合于学科和媒体知识库的建立、使用、评估、元分析和出版服务（特别是由学术团体作为中介）
- ✓ 订阅支持：适合于访问和认证、保存以及资源发现服务
- ✓ 完整的商业化模型（包括广告、商业和效用模型）：适合于数字化、知识库建立和管理、技术咨询服务、元数据创建和增加、技术转移以及各个输出层服务（访问、认证、使用统计、保存、监控和元分析服务、资源发现、链接服务、同行评议的开放获取期刊出版服务）

最高的费用可能会发生在保存和访问、认证服务上。资源发现服务和元数据服务的花费相对次之。知识库建立和管理、数字化、使用统计、链接服务和出版服务的花费属于中等支出。咨询服务、监控和元分析服务、技术转移、特定学科和中间知识库以及同行评议的开放获取期刊服务的费用相对来是最低的。

## 5. 建议

我们向 JISC 提出了一些如何在实践中实现这一愿景的若干建议。知识库需要有力而全面的管理，尤其是交流与协调。系统各个部分并不可能有效地各司其职，这主要是因为对系统的某些部分而言，仍然需要持续地提供包干的基金。JISC 协调发展的自然选择对象，它不仅要预计最后的成果，还会通过必要的资金来影响和促使研究人员来加强整个系统。因此，作为整个计划的一部分，还要给那些私人企业发展的机会。

对 JISC 的较全面的建议：

1. 科研界应该在更高层次上致力于鼓励在所有的高等教育机构建立知识库，并且制订各种政策来保证内容的收藏。
2. 建立与知识库管理者的交流渠道，并且鼓励交流社区的建立。这可以通过已有的架构来实现，比如 UKCORR 就非常适用，两个主要的开放源码的知识库软件（EPrints 和 DSpace）都有他们的用户群来实现这种交流。这都是为了在 JISC 和所有正在运行的知识库之间建立清晰有效的交流系统来推动双向的讨论，并维护其发展。
3. 同样地，还应该建立 JISC 和现行或潜在服务提供者之间的接口或接触，这能够促使用户参与建立面向用户的服务。
4. 知识库、聚合工具、用户服务以及其它中介服务的建立应走向面向服务的体系结构，并且还应该为聚合模型建立单独的层，来使用户服务灵活性最大化，以满足用户的需求。
5. 建立终端用户服务，包括调查通过这些服务获得的信息将怎样得到应用，这有助于服务的建立，并且能将信息反馈给通过服务加以揭示的知识库。
6. 还需要自动生成元数据的手段，建议抓紧调查相关的技术和工具。
7. 要更多地关注标识符，尤其是与地址无关的标识符，需要建立必要的解决方案来增

强资源的理解和使用。

8. 建立利用 RSS，调查 ATOM 作为 OAI-PMH 的另外一种标准用于元数据和内容的聚合。这两种标准提供了一种潜在的目标知识库资源的揭示，这对特定群体的用户服务建立是非常有益。同时建议，详细考察将知识库的内容在网络搜索引擎中得到揭示，评价现有的揭示方法以及通过这一途径揭示知识库的意义。

9. 建议将来开发聚合器，建立终端用户服务，包括从开始就与知识库建立沟通。这种方式保证这种做法不会孤立化，而且还能提高聚合模型中这三个主要部分的互操作性。

10. 英国知识库想要建立理想结构的知识库服务，不可避免地就需要专项基金来支持系统的一些部分，JISC 也需要有个中长期规划。

编译自：Alma Swan, Chris Awre. Linking UK Repositories: Technical and organizational models to support user-oriented services across institutional and other digital repositories — Scoping Study Report. [http://www.jisc.ac.uk/uploaded\\_documents/Linking\\_UK\\_repositories\\_report.pdf](http://www.jisc.ac.uk/uploaded_documents/Linking_UK_repositories_report.pdf). [2006-6-21]

(初景利 校)