

# 基于 DSpace 构建学科知识库系统的研究与实践\*

祝忠明 马建霞 常宁 米波

(中国科学院国家科学图书馆兰州分馆 兰州 730000)

**[摘要]** 文章首先简要回顾了学科知识库建设与发展的主要模式,接着以一个面向环境与生态科学领域的学科知识库系统建设为例,给出了一种以自助提交和开放搜寻与聚集为主要手段的学科知识库建设框架,并对以 DSpace 为基础实现该框架的过程中所涉及的主要环节和问题进行了探讨。

**[关键词]** DSpace 学科知识库 数字知识库 定制

**[分类号]** TP39

## An Implementation of a DSpace-based Disciplinary Repository System

**Zhu Zhongming Ma Jianxia Chang Ning Mi Bo**

*(The Lanzhou Branch of the National Science Library, CAS, Lanzhou 73000, China)*

**[Abstract]** The paper first gives a brief review of the disciplinary repository development model. Then using an open disciplinary repository for environmental and ecological sciences as an example, it presents a framework for building a disciplinary repository depending mainly on ways of self-archiving and open harvesting. Lastly, it discusses major issues involving in the process of implementing the framework using DSpace.

**[Keywords]** DSpace, repository, disciplinary repository, subject repository, customization

### 1 前言

随着科学研究的信息化趋势日益明显和加快,科学活动和科学交流的模式和过程都正在发生着深刻的变化,科学活动的产出以数字化网络化的形式呈现和保存已成为一种普遍现象。从最早也是目前依然非常著名的由美国洛斯阿拉莫斯国家实验室(Los Alamos National Laboratory)于1991建立的粒子和高能物理领域的电子预印本系统 arXiv<sup>[1]</sup>开始,以促进学术交流和科研活动为主要目的的研究性数字知识库(digital repository)得到了科学界和图书馆界的共同重视、研究和发 展。在其后的相当长的一段时期内,这种以提供研究论文的预印本或后印本的发布和交流为主的学科或主题知识库系统的研究和建设成为数字知识库建设的主要模式,如认知科学领域的 Cogprints<sup>[2]</sup>,经济学领域的 RePEc<sup>[3]</sup>,化学领域的 Chemistry Preprint Server<sup>[4]</sup>等都遵循这种模式建立。近年来,随着开放获取、数字资源长期保藏、数字资产管理以及数字科研(e-research)等概念的兴起,使以支持机构科研活动产出的开放利用和长期保藏为主要特征的机构知识库(Institutional Repository)成为当前数字知识库研究和建设领域的热点。与早期的基于学科或主题的以研究论文的自愿和自助存档相比,一方面,研究机构一旦认识到机构知识库在机构研究和学术成果的积累、交流、管理以及促进其长期研究学术目标的实现过程中的重要作用和意义,在推进其机构知识库的建设过程中,都会实现制定明确的发展策略和政策,做出所属科研人员都必须将其科研和学术研究的产出提交到其中的规定,从而使其信息来源得到有效可靠的保障。另一方面,机构知识库的信息来源也不再局限于研究论文,而扩展到了对科学活动过程中的各种形式的产

---

\* 本文系国家自然科学基金重点项目“中国西部环境与生态科学数据中心”子项目“中国(西部)环境与生态科学知识积累平台”的研究成果之一(项目编号:90502010)

出,如研究报告、演示资料、数据、图像和视频资料、学习对象以及其他形式的数字资产等,

从而使其成为内容类型丰富多样并与科学研究过程更加息息相关的科研基础条件的重要组成部分。目前, 仅在 ROAR<sup>[5]</sup> 注册登记的大学和研究机构知识库就有 330 多个, 其发展势头和前景由此可见一斑。

然而, 从科学活动和学术交流过程主要围绕学科领域进行分工组织的模式来看, 机构知识库的兴起和发展毕竟是更多地迎合了数字环境下大学和研究机构对其科研产出进行集中管理和更高效地组织利用的需求的结果, 与学科知识库相比, 后者对于相关学科领域的科研人员来说显然更有吸引力。事实上, 由于不同的定位和应用目标, 二者之间并不存在简单的孰优孰劣的竞争关系, 而恰恰是一种相互促进的关系。机构知识库不仅为研究人员的科研产出的长期保藏和持续利用创造和提供了可信赖的环境和平台, 而且也为学科知识库的构建提供了新的发展机遇。大量的机构知识库的分散存在, 为学科知识库构建和发展提供了新的、可靠的、也将是持久的信息来源, 通过开放获取和信息聚合的手段(目前几乎所有的机构知识库都提供基于 OAI-PMH 协议的元数据开放获取和共享接口), 围绕特定的学科或主题领域对分散的机构知识库中的信息进行再发现、再组织和集成, 可以更加快速地建立起符合科研人员获取和利用学科信息习惯的数字知识库系统, 同时也将使机构知识库的内容被更广泛地发现和利用。

本文拟以中国(西部)环境与生态科学知识积累平台(Sharing Environment and Ecology Knowledge Space, 以下简称 SEEKSpace)的研究和建设为例, 并以数字知识库系统开发建设领域著名的开源软件 DSpace 为基础, 对兼具自助存档和学科信息发现集成的学科知识库系统的框架设计和开发定制过程中所涉及的主要问题进行分析 and 探讨。

## 2 SEEKSpace 的基本功能和服务框架

SEEKSpace 研究和建设的主要目标, 首先在于汇集国家自然科学基金委“中国西部环境与生态科学重大研究计划”历年来所支持的研究项目和课题的各类研究产出, 包括科学论文、研究报告、专著、演示资料、数据资料、照片和视频资料、野外考察资料、重要的中间成果等, 建立支持对这些研究产出进行自助提交、长期保存和共享利用的平台, 同时也不排除接受其他有关环境与生态科学研究活动的产出的开放存档与共享利用。其次, 通过开放搜寻获取和信息聚合的手段, 对国内外主要的研究性数字知识库(主要是机构知识库, 也包括一些学科知识库和个人知识库)所涉及环境与生态科学领域的学术信息进行搜索、发现、聚集和再组织, 方便对该学科领域主要的开放学术信息资源的集成利用。从而, 逐步发展和形成一个支持环境与生态科学领域开放学术信息的自助存档、交流和发现的学科知识库系统。以下给出 SEEKSpace 的总体功能和服务框架, 如图 1 所示。

自助提交服务、批次导入服务以及自动搜索与聚集服务提供了三种不同的内容提交或采集的途径。自助提交服务主要为西部计划项目研究人员或者其信息提交代理提供将其研究产出自主地提交到系统中的机制和接口, 这是目前绝大多数研究性数字知识库进行信息采集和积累的基本模式; 在获得许可的条件下, 环境与生态科学领域的其他研究人员也可以将其相关的研究资料以自助方式提交到 SEEKSpace 中。批次导入服务, 提供按照系统可接受的格式批量导入相关信息或数据的功能。自动搜索与聚集服务, 则主要通过基于 OAI-PMH 协议的元数据及数据的自动搜寻和获取工具的支持, 从国内外相关的开放型数字知识库中发现和采集与环境与生态科学有关的内容, 形成本地化的保藏。

元数据仓储及知识库是内容提交和聚集的结果, 提供对各种类型和格式的内容的长期保存和利用的支持。元数据仓储实现对提交和采集进来的数字对象的内容、结构以及保藏等方面的元数据描述信息的集中存储和管理功能, 知识库提供对数字对象的统一存储和管理, 并与元数据仓储共同构成支持数字对象的保藏、组织和利用等功能实现的基础。

用户服务接口以元数据仓储及知识库为基础, 面向领域用户提供基于分类或主题的

浏览、基于字段或全文内容的检索、基于电子邮件或RSS方式的内容订阅等服务功能

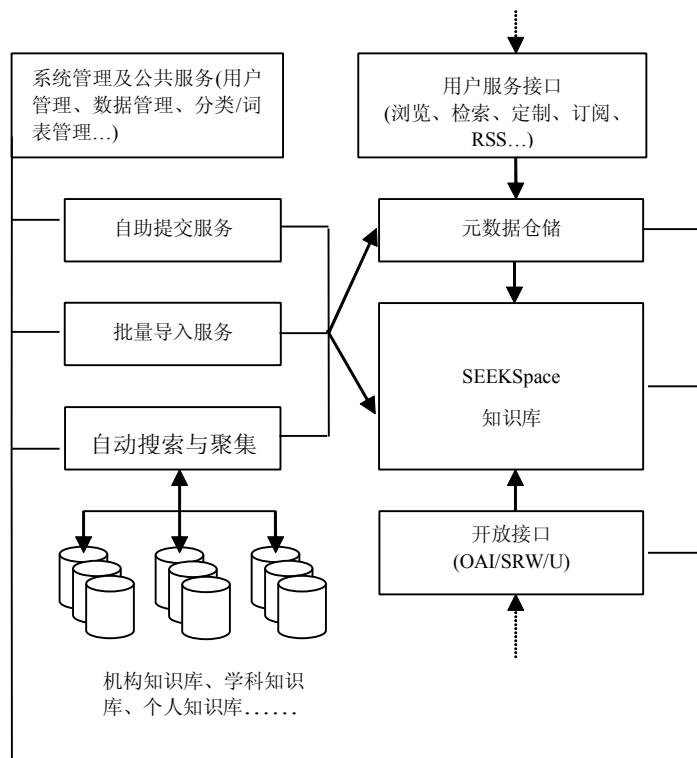


图1 SEEKSpace 基本功能框架

的实现。开放接口则提供面向机器的接口，支持OAI-PMH协议方式的元数据共享，以及支持SRW/U协议方式的检索服务，方便分布式检索互操作的实现。事实上，面向RSS服务同时也是一种面向机器的接口。

系统管理与公共服务提供整个系统运行和管理过程中的基础性的或公共性的辅助功能和服务，如用户管理提供对系统注册用户的权限的分配、审核、管理等服务；数据管理不仅提供对有关存档数字对象描述性的、结构性或保藏性的元数据的管理，也提供对存档数字对象本身的管理；分类表或词汇表的管理主要提供数字对象的描述和组织过程中的分类或主题标引的功能和服务。

### 3 基于DSpace 构建SEEKSpace 知识库系统的主要环节和问题的探讨

#### 3.1 DSpace 简介

DSpace<sup>[6]</sup>是由麻省理工学院图书馆和HP实验室合作开发的构建数字知识库系统的开源软件，其功能和服务设计上遵循了OAIS<sup>[7]</sup>参考模型，为大学和研究机构对其科研产出进行采集、描述、存储、保藏、发布和利用等多方面功能和服务的实现提供了可靠的解决方案。自2002年11月发布Dspace 1.0版以来，目前已知的有全球180多家大学和研究机构部署了基于DSpace的数字知识库系统<sup>[8]</sup>。其最新稳定版本为2005年10月发布的DSpace 1.3.2。

DSpace数据模型及其组织上试图模拟和反映研究机构本身的层级结构，按照“Community-Collection-Item”的基本模型对数据进行组织和管理。即，一个基于DSpace的系统是由若干对应于系、实验室、或研究中心等的Community组成，每个Community由若干对应于研究领域、研究项目、或技术报告集等的专题内容集合Collection所组成，而每个Collection则是由一系列相关的对应于具体的数字存档对象的Item的集合。因此，按照OAIS参考模型，Community和Collection都是属于AIC(Archive Information Collection)类型的存档对象，Item是属于AIP(Archive Information Package)类型的存档对象，是Dspace中实际地封装

和管理数字对象的基本单元。事实上,这种抽象的层级结构数据模型,可以很好地满足和支持各种具有层级结构特征的内容的存档和管理。比如在构建学科知识库的过程中,学科知识的分类等级结构,就可以很自然地与DSpace的这种层级数据模型对应起来。

在数字对象的元数据描述过程中,DSpace采用了基于带修饰符的DC元数据来支持对Item存档对象进行描述和组织,对于Community和Collection存档集合对象仅提供一些包含简单描述性信息的元数据。此外,DSpace也提供有关保藏、来源、权限政策等管理性元数据的支持,以及提供有关Item展现、包含的文件及其关系等结构性元数据。

命名和标识数字对象对于存档系统来说至关重要,它是保证存档的数字对象可以被正确可靠地引用和参考的重要依据。DSpace采用CNRI的Handle System<sup>[9]</sup>来标识Item、Collection以及Community等存档对象。Handle System作为一种新兴的分布式的全球化命名和解析服务系统,可以为存档对象分配和赋予具有全球唯一性和持久性的标识符,并可以获得全球性的解析和定位服务,确保了所标识的对象可以在全球范围内被有效地引用和参考。

DSpace支持两种内容提交的方式。一种是基于web表单的提交方式,支持任何具有提交权限的用户将适当的内容提交到系统中来。这种提交过程往往与一定的权限政策和工作流相关联,可以灵活地支持基于各种角色和工作流的内容提交控制。一种是批次导入的方式,支持将符合DSpace SIP (Submission Information Package) 格式的数据文件批量地导入到系统中。

在面向用户的服务组织方面,支持简单的和高级的浏览及检索方式,并支持全文检索,提供基于电子邮件的信息订阅功能和服务等。此外,它还支持符合OAI-PMH协议的元数据开放共享,并支持简单的OpenURL功能和服务等。

### 3.2 DSpace 的本地化过程及方法

软件的本地化与其国际化息息相关。DSpace基于Java技术设计和开发,其国际化主要通过Java技术在支持国际化软件设计和开发方面的特性和优势,采用基于Unicode标准和可本地化独立资源文件的技术和方法来实现。下面,我们以GNU/Linux操作系统、Tomcat 5.x、PostgreSQL 7.x/8.x数据库系统为安装和运行基本环境,对DSpace的本地化定制过程中的主要环节和问题进行探讨。

(1) 基于Unicode/UTF-8的统一编码环境的设定<sup>[10] [11] [12] [13] [14]</sup>。

在典型的基于Java的Web应用环境中,用户与Web应用之间的交互可能涉及“浏览器↔Web(应用)服务器↔Web应用↔JDBC↔数据库”中的各个环节和过程,由于各个环节可能采用不同的字符编码模式,使得流经这些环节的数据或信息流经常不可避免要不同编码格式之间进行主动或被动地转换,其中任一环节编码模式的失当,都可能导致数据或信息转换的错失。当前,能够比较彻底地解决这一问题的主要方法是统一各环节的编码格式,建立基于Unicode标准的统一编码环境。DSpace即采用了这一方法来实现对多语言环境的支持和本地化的便利实现。

就上述各个环节而言,当前主流的浏览器(如IE 6.0或Fire Fox 1.5)都比较好地提供了自动检测网页中包含的编码格式信息并据此支持对不同语言内容的正确显示,并按照相同的编码格式或Web表单本身指定的编码格式对提交的内容进行编码,一般不需要用户的直接干预。

在Web服务器这一环节,由于HTTP协议只规定了传送请求的URL中出现的非ASCII字符以及一些特殊字符要以十六进制串的形式进行编码,但并没有对基于哪种字符集进行编码做相应的规定。因此,要明确指定Web服务器默认的编码格式,以保证对编码后的URL数据能够正确解析。以DSpace使用Tomcat 5.5.x同时做为Web服务器和应用服务器为例,要保证以Unicode/UTF-8方式对HTTP请求数据进行编解码,需在其配置文件server.xml中增加如下有关URIEncoding的设置:

```
<Connector port="8080" maxHttpHeaderSize="8192"
```

.....

```
URIEncoding="UTF-8" />
```

同样地，如果要使用基于SSL HTTP协议的方式出来处理HTTP GET/POST请求，则需要在server.xml的如下部分中加入如下有关URIEncoding的设置：

```
<Connector port="8443" scheme="https" secure="true"
.....
URIEncoding="UTF-8" />
```

如果使用其他的Web服务器为Tomcat转发JSP/Servlet请求的情况下，即仅将Tomcat做为Web应用服务器使用时，由于它不再通过HTTP协议方式而是AJP协议方式来接受其他Web服务器转发的请求，因此上述设置就不起作用了。以常见的Web服务器Apache 2.0.x与Tomcat 5.5.x的整合为例，需要对serverv.xml文件做如下的设置：

```
<Connector port="8009"
enableLookups="false" redirectPort="8443" protocol="AJP/1.3"
URIEncoding="UTF-8"/>
```

在Web应用环节的字符编码问题由于容易通过程序方式操控，Dspace已较好地解决了这一问题。其主要的解决思路和方法是，对于任何Dspace JSP文件，都在其第一行使用JSP Page指示符明确指定字符编码格式为“UTF-8”，如：

```
<%@ page contentType="text/html; charset=UTF-8" %>
```

对于Dspace Servlet，则通过在一个DspaceServlet基类中初始化Dspace Context，并设定所有请求参数的编码格式为“UTF-8”，以保证请求被读取之前都能按照UTF-8格式被正确地编码。其他所有的Servlets都必须继承该基类，从而能够以同样的编码格式进行请求的接受和处理。

在JDBC连接环节，编码问题可能因不同厂商的JDBC驱动而异。就Dspace来说，由于默认的数据库为PostgreSQL 7.3及以上的版本，其相应的JDBC驱动版本也开始将与数据库交互过程中的编码转换完全交由数据库本身解决。

在数据库环节，首先，若采用PostgreSQL 8.x以下的版本，需要在配置和编译时使用“--enable-multibyte”和“--enable-unicode”选项；从8.0版本开始，这些选项已成为默认的编译选项，因此不需要在编译阶段再配置。其次，在创建供Dspace使用的数据库时使用编码选项并声明为“UNICODE”，如“createdb -U dspace -E UNICODE dspace”，使数据库内部以UTF-8方式对数据进行编码和存储。

尽管操作系统的语言环境(Locale)并不直接影响Dspace的部署和运行，将其设定为基于Unicode的环境则有助于防止系统定制和运行过程中不必要的差错的发生，如支持JVM以及Tomcat服务器可以直接从操作系统的编码方式获得适当的默认文件编码属性，保证程序的正确编译和运行等。以Fedora Core GNU/Linux 4为例，可以通过两种方式完成语言环境的更改和设定：通过修改文件/etc/sysconfig/i18n，将其中的键LANG的值改为“zh\_CN.UTF-8”；或者在/etc/profile文件中加入一行“export LANG=zh\_CN.UTF-8”。

至此，一套支持Dspace运行的基于Unicode/UTF-8的完整的编码环境就基本构建完成了，并为Dspace后续的中文化定制和部署奠定了良好的基础。

## (2) 用户界面的中文化<sup>[15]</sup>

以Unicode编码环境为基础，Dspace主要通过利用Java技术支持的独立资源文件的方法来方便用户界面的本地化。在服务器端，它已将用户界面有关的JSP/Servlet中大部分文本信息从程序中分离出来，形成了名为“Messages.properties”的属性文件，中文化的过程中首先将该文件进行复制并按照Java有关本地化资源文件命名的格式要求命名为“Messages.zh\_CN.properties”，将其中的“键/值(key/value)”对中的值翻译为中文，这样大部分与用户界面有关的中文化工作就基本完成了。编辑过程中需要注意的是，要保证该支持中文化属性文件最终必须以UTF-8编码方式存储和编译。

此外，Dspace还使用一些文本文件来保存某些与用户交互有关的配置、反馈及帮助信息的保存，也需要做中文化处理，以使用户能够获得完整一致的中文化Dspace的使用体验。这些文件主要包括在/dspace/config目录及其子目录下的与邮件反馈、元数据录入表单等有关文件，以及在jsp文件目录下的帮助文件

等。

目前DSpace的本地化还更多地是在多语言支持的层次,对于日期、货币及姓名等格式的本地化支持还不完善,尤其是在中文化过程中对中文姓名的处理与中文习惯明显不一致,这些都需要在中文化的过程中通过修改程序的方式进行定制处理。在用户界面的可定制性方面也没有特别的考虑和支持,更改默认的界面和风格都只能通过修改JSP程序的方式进行,并按照DSpace的建议,在其jsp目录下建立一个local子目录,将需要修改的jsp文件或者jsp目录原有的所有内容以保持其目录结构的方式复制到该目录下,再进行修改。这样做的好处,主要是为了避免在以后进行DSpace的升级过程中造成对已做的定制意外覆盖的后果。

## (2) 中文全文检索问题的解决

DSpace使用Apache Lucene<sup>[16]</sup>做为全文索引和检索的引擎,由于Lucene内置的标准文本分析器并不支持对中文内容的分析和索引,在默认的情况下DSpace也不支持中文索引和检索。因此,要使DSpace具备中文全文检索能力,首先要为Lucene包增加中文文本分析器。然后,替换有关全文索引及查询程序中的标准文本分析器为中文文本分析器。

因Lucene在全文检索领域广受关注,已经有志愿者贡献了中文文本解析器的源码,为Lucene增加中文解析器就比较简单了。下载Lucene的源码,并从Apache Lucene Sandbox<sup>[17]</sup>中选择下载合适的中文分析器的源码,按照Lucene的目录结构将中文分析器的源码组织到Lucene的代码体系中,然后重新编译通过后即得到支持中文检索的Lucene包,并用它将Dspace自带的Lucene包予以替换。

需要修改的其他程序代码主要有两处,一处是将DSIndexer类的openIndex方法中的如下语句

```
writer = new IndexWriter(index_directory, new DSAnalyzer(),wipe_existing);
```

替换为:

```
writer = new IndexWriter(index_directory, new ChineseAnalyzer(),wipe_existing);
```

另一处在DSQuery类的doQuery方法中的如下语句

```
QueryParser qp = QueryParser qp = new QueryParser("default", new DSAnalyzer());
```

替换为:

```
QueryParser qp = new QueryParser("default", new ChineseAnalyzer());
```

完成以上修改后,重新编译DSpace,就可以得到顺利支持中文全文索引和检索的DSpace。

## 3.3 Handle System 的启用<sup>[18] [19]</sup>

长期可靠的链接或标识对于数字学术资源的引用、发现、可靠定位和利用的意义是不言而喻的,因此,DSpace做为支持研究性数字知识库建设的应用平台,采用了Handle System来提供对其中的条目(Item)、专题(Collection)以及主题(Community)等数字对象和数字对象集合提供全球唯一的、稳定的标识,以保证其中存储的这些任何类型的数字对象都可以被正确地解析和定位。在DSpace框架下,启用Handle System的主要步骤为:

(1) 运行如下的命令,得到一个名为sitebndl.zip的文件。

```
[dSPACE_install_dir]/bin/dsrun net.handle.server.SimpleSetup    \\  
[dSPACE_install_dir]/handle-server
```

(2) 将得到的文件发送到hdladmin@cnri.reston.va.us,CNRI在收到邮件后会很快为你的申请分配一个全球唯一的命名机构(Naming Authority)前缀。如我们申请到的Handle前缀为2239,以后所有形如http://hdl.handle.net/2239/xxx的标识将都会被解析和指向我们的本地Handle服务。

(3) 编辑[dSPACE\_install\_dir]/handle-server/config.dct配置文件,主要是更新其中与命名机构相关的配置为新申请得到的命名机构前缀。

(4) 修改[dSPACE\_install\_dir]/config/dSPACE.cfg配置文件,将“handle.prefix”选项的值更改为新分配的前缀。如: handle.prefix = 2239。

(5) 启动Handle服务:

```
[dSPACE_install_dir]/bin/start-handle-server
```

需要注意的是,如果在生成 sitebndl.zip 文件的过程中,选择了服务器认证及管理密钥,可能无法启动 Handle 服务。原因在于 start-handle-server 中将所有输出重定向到了 /dev/null,因此执行 start-handle-server 的过程中无法提供输入密钥的接口。解决的办法是,将 start-handle-server 文件中的如下的行:

```
nohup $BINDIR/dsrun -Dlog4j.configuration=log4j-handle-plugin.properties  \\  
net.handle.server.Main $handledir </dev/null >> $logdir/handle-server.log 2>&1 &
```

更改为:

```
$BINDIR/dsrun -Dlog4j.configuration=log4j-handle-plugin.properties  \\  
net.handle.server.Main $handledir >> $logdir/handle-server.log 2>&1
```

这样,在执行 start-handle-server 的过程中就可以根据提示输入密钥后,再将其放入后台作业中执行就可以了。

### 3.4 OAI-PMH 接口的开放<sup>[20] [21]</sup>

支持 OAI-PMH 协议已是当前开放性数字知识库的必备特性。DSpace 对 OAI-PMH 的支持通过采用由 OCLC 开发的 OAI-Cat 开源软件包来实现,使基于 DSpace 平台构建的知识库系统具备符合 OAI-PMH 2.0 的数据提供者接口功能和服务。DSpace 的默认编译过程中,会生成一个名为 dspace-oai.war 包,只要将其正确部署就可以使 DSpace 获得通过 HTTP 协议接受和响应 OAI-PMH 请求的能力。访问 DSpace 的典型 OAI-PMH 请求格式为:

```
http://[dspace_site_name]/dspace-oai/request?[OAI_verb_Clause]
```

目前,DSpace 只支持按照无限定词的 DC 元数据格式的元数据共享,并以形如“oai:host\_name: handle”的 URI 格式做为 OAI-PMH 记录的唯一标识符,如对于 SEEKSpace 将有形如“oai:seekspace.resip.ac.cn:2239/xxx”的 OAI 记录标识格式。如果需要对其他类型的元数据元素集的支持,可以通过创建额外的词汇转换表(Crosswalk),并对 oaicat.properties 文件进行适当配置以实现。

### 3.5 功能定制和扩展的方法

DSpace 的当前版本对灵活可扩展机制的设计和支撑能力还比较弱,暂不支持以插件(plugin)或附加件(addon)方式对系统的功能和服务进行方便地定制或扩展,而只能通过以修改程序文件的插入方式(addin)或打补丁(patch)的方式进行功能扩展。如我们在基于 DSpace 进行 SEEKSpace 技术平台的定制和开发过程中,用户界面的定制、OAI 搜索引擎、SRW/SRU 检索 API、个人研究室等附加功能的扩展都是通过直接修改相应的 JSP/Servlet 代码的方式来实现的。这也是目前大量针对 DSpace 的扩展功能和服务实现的基本模式。不过,DSpace 的新版本已开始注意加强其功能和服务扩展机制的支撑能力,考虑增加如 Plugin Manager 或 Addon Prototype 等灵活支持对 DSpace 的功能和服务行为进行裁减、或在不需要改动其核心代码及数据库模式的条件下实现功能和服务扩展的机制<sup>[22]</sup>。

## 4 结语

总之,DSpace 提供了支持研究性数字知识库系统建设的所需要的绝大部分功能,如数据摄入、元数据描述和封装、浏览和检索、数字对象的唯一标识及长期保存、数据管理、OAI 数据提供者支持等开放学术存档系统的必备功能服务支持,还提供了诸如基于 LDAP 的用户认证、基于 OpenURL 的开放链接和服务集成、以及基于 SRB 的“无限”可扩展存储能力支持等多方面的丰富功能和服务支持,而且随着版本的不断更新升级和用户群体的扩大,其技术框架、功能和服务也将日益完善和丰富。因此,在依托 DSpace 建设 SEEKSpace 平台的过程中,一方面,我们将继续加强对其现有功能的开发利用和扩展,另一方面,也将密切关注其更新和发展,及时对 SEEKSpace 的技术平台进行更新和升级。

参考文献:

- 1 arXiv. <http://xxx.lanl.gov/>(Accessed Apr. 12,2006)
- 2 Cogprints. <http://cogprints.org/>(Accessed Apr. 12,2006)
- 3 RePEc. <http://repec.org/> (Accessed Apr. 12,2006)
- 4 Chemistry Preprint Server. <http://www.sciencedirect.com/preprintarchive>(Accessed Apr. 12,2006)
- 5 Registry of Open Access Repositories (ROAR). <http://archives.eprints.org/>(Accessed Apr. 12,2006)
- 6 Dspace. <http://www.dspace.org/>(Accessed Apr. 12,2006)
- 7 Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS). <http://public.ccsds.org/publications/archive/650x0b1.pdf>(Accessed Apr. 12,2006)
- 8 DspaceInstances. <http://wiki.dspace.org/DspaceInstances>(Accessed Apr. 12,2006)
- 9 The Handle System. <http://www.handle.net/introduction.html>(Accessed Apr. 12,2006)
- 10 Tansley, R., Bass, M., Branschofsky, M., et al. DSpace System Documentation. <http://dspace.org/technology/system-docs/> (Accessed Apr. 12,2006)
- 11 Panula-Ontto T. Internationalization Guide for Java Web Applications. August, 29, 2005. <http://tomi.panula-ont.to/i18n/> (Accessed Apr. 16,2006)
- 12 Kuhn, M. UTF-8 and Unicode FAQ for Unix/Linux, December, 29, 2005. <http://www.cl.cam.ac.uk/~mgk25/unicode.html> (Accessed Apr. 16,2006)
- 13 Apache Tomcat Configuration Reference. <http://tomcat.apache.org/tomcat-5.5-doc/index.html>(Accessed Apr. 16,2006)
- 14 PostgreSQL:Documentation. <http://www.postgresql.org/docs/>(Accessed Apr. 16,2006)
- 15 同 10
- 16 Apache Lucene. <http://apache.justdn.org/lucene/java/>(Accessed Apr. 24,2006)
- 17 Apache Lucene Sandbox. <http://lucene.apache.org/java/docs/lucene-sandbox/>(Accessed Apr. 24,2006)
- 18 同 9
- 19 同 10
- 20 The Open Archives Initiative Protocol for Metadata Harvesting(OAI-PMH). June, 14,2002. <http://www.openarchives.org/OAI/openarchivesprotocol.html>(Accessed Apr. 24,2006)
- 21 同 10
- 22 Tansley, R. DSpace Architecture Progress and Ideas. February, 1, 2006. [http://www.aprs.edu.au/Open\\_Repositories\\_2006/tansley\\_architecture\\_update.ppt](http://www.aprs.edu.au/Open_Repositories_2006/tansley_architecture_update.ppt)(Accessed Apr. 24,2006)

(作者 E-Mail: [zzm@lzb.ac.cn](mailto:zzm@lzb.ac.cn))