

文章编号:1673-0062(2008)04-0092-04

## 基于元数据的特色数据库构建和知识组织

梁冰,唐文杰,万建军,李兰英

(南华大学图书馆,湖南衡阳421001)

**摘要:**本文就学术图书馆的专题数据资源建设进行了探讨.结合大学图书馆“核”特色资源的建设案例,对涉及的元数据方案、知识组织、服务对象等进行了分析,最后就特色资源的可持续发展提出了几点看法和建议.

**关键词:**数字图书馆;资源;建设;元数据;可持续发展

**中图分类号:**G250.74      **文献标识码:**B

## Characteristic Database Construction and Knowledge Organizing Based on Metadata

LIANG Bing, TANG Wen-jie, WAN Jian-jun, LI Lan-ying

(Library, University of South China, Hengyang, Hunan 421001, China)

**Abstract:** This article discussed the construction of data resource in the academic library. It explored how to build the “nuclear” cases by analyzing the scheme of metadata, knowledge structure, and service objects, etc., and presented some suggestions and options about the sustainable development of special resources in the academic library.

**Key words:** digital library; resources; construction; metadata; sustainable development

数字资源因使用方便,无时空限制,有海量信息等特点在图书馆馆藏中的分量越来越重.同时电子信息资源的出版使图书馆藏建设的概念产生了变化,由原来的文献资源建设逐步转变为信息资源建设,使图书馆的资源基础突破了传统的“馆藏”局限而扩展到了整个网络,从而使可提供服务的资源极大的丰富了<sup>[1]</sup>.不容乐观的是我国的数字资源生产尚处于一种无序竞争的状态,无论是期刊数据还是电子图书各大公司之间的产品都有大量重复,资源质量参差不齐,与国外数据提供商之间那种泾渭分明、资源互补的有序运作产

生了极大反差.给用户的选择使用造成无谓的浪费和经济损失.同时一些有价值、较专业也是读者急需的特色资源因为无利可图而没能进行数字化生产,或者是因为制作成本和市场价位较高,难以在整个图书情报界推广普及,导致许多图书馆的数字资源建设都只能在低水平重复<sup>[2]</sup>.特别在高校图书馆,馆藏资源的服务对象首先是在校的师生,资源的建设应紧密结合学校的专业设置和特点,资源的内容除了公共需求外必须考虑专题特色资源建设.目前一些专业性相对较强的数据资源难以按需购进,馆藏现有文献资源未能整合和

收稿日期:2008-09-02

作者简介:梁冰(1956-),男,山东威海人,南华大学图书馆高级工程师,硕士.主要研究方向:网络工程与网络安全.

转化成数据资源,使得资源体系不完整,资源共享也就没有现实意义.因此高校数字图书馆按照学科发展需要,通过自建特色数据资源来满足学校发展并为社会提供服务已取得共识.

## 1 自建资源 and 特点

数字图书馆的建设是以数字资源建设为核心,高校特色资源库的建设是根据学校专业设置特点,通过对馆藏资源、网上资源及外协单位资源的有目的收集整理后进行加工、开发的专题资源库.资源内容首先为校内用户群体,还要为社会相关人员提供服务.

自建数据库可以丰富馆藏资源,在建设时应注意贴近用户需求,彰显内容特色.需要外购的资源规模和内容应该以可持续发展的目光予以审视,充分考虑资源经费和后续投入问题,清醒冷静地根据读者的需求和经济实力加以落实.同时还应考虑数字资源使用的技术条件、网络运行环境及技术人才的培养.自建的特色资源库必须

具备以下特点:1)资源库中的内容明确.2)符合国际、国内及行业的相关标准.3)经过数字化制作加工,使各种相关内容的印刷资源成为全息的数字化资源纳入数据库.4)资源的内容可以允许多种不同的形式.5)可以提供完整的检索服务.6)具备较好的资源推荐和正确的信息导航等服务功能.

## 2 “核”特色资源库建设实例

### 2.1 资源建设的提出

随着未来全球经济的发展,对能源的需求必将大量增长.由于化石燃料的资源有限性和对环境的危害性,以及其他能源目前尚不能大规模为人类所利用,核能必将是21世纪人类能源的重要组成部分.我国是世界上少数几个拥有比较完整核工业体系的国家之一.根据中国核电建设项目进度设想(见表1),到2020年,核电运行装机容量要争取达到 $4 \times 10^7$  kW<sup>[3]</sup>.

表1 核电建设项目进度设想(单位:10<sup>4</sup> kW)

Table 1 Imagining of nuclear power building item plan

	五年内新开工规模	五年内投产规模	结转下个五年规模	五年来核电运行总规模
2000年前规模				226.8
“十五”期间	246	468	558	694.8
“十一五”期间	1 244	558	1 244	1 252.8
“十二五”期间	2 000	1 244	2 000	2 496.8
“十三五”期间	1 800	2 000	1 800	4 496.8

注:1、因单机容量有变化,实际开工和完工核电容量数有变化.

2、国家发展和改革委员会,《核电中长期发展规划(2005~2020年)》.

我国核电的大规模发展需要大量与核电有关的专业人才,同时又为相关行业和专业人员提供了广阔的市场空间和施展才华的机会.为实现2020年核电发展目标,国家、企业和高等院校科研院所在科研、设计、燃料、制造、运行和维修等环节,及核电设计、核工程技术、核反应堆工程、核与辐射安全、运行管理等专业领域,首先面临各类人才的培养和做好人才的储备.

南华大学长期为中国的核工业培养人才,有核反应堆工程、核技术、辐射防护、核化工与核燃料和核物理多个核专业,直接与“核”有关的专业和学位点有10余个,支撑与“核”有关的专业和学位点有20余个,已经形成“核心突出、工文渗透、核医结合”的学科体系,是办学特色鲜明的高校.学校无论从教学还是科研都需要对多年积累的技术文献加以整合,加工处理为数字资源.同时

还迫切希望能方便的了解和掌握最新的学术动态.但是长期以来由于涉“核”的技术敏感及数据资源相对较少的原因,要从一般的数据供应商得到有价值的学术数字资源一是因为制作成本太高,二是原始文献难以收集而无法直接购买.因此无论是学校的学科建设、人才培养,还是科学研究都迫切需要核专题资源库.

### 2.2 资源建设的思考

要使专题数字图书资源发挥预期的作用,首先应具备综合数字图书馆所具备的基础条件,即:1)具备有Internet网上进行通信和提供服务的条件;2)有一定规模的专题资源内容;3)有必要的手段和方法用于相关资源的索取、创建和存储管理维护;4)能为用户提供信息导航、技术咨询、探讨交流等基本服务;5)保证访问数据的可承受负载及必要的冗余;6)有可行的技术手段,对

不同的数据资源提供不同的服务.除了以上内容外,在构建专题数据库时考虑到的问题还有:1)元数据方案的设计;2)专题库所涉及的知识组织系统;3)使用专题资源的关联群体;4)提供的主要服务.

### 2.3 元数据方案设计

目前“核”相关类的数据资源由于种类繁多,所针对的对象不仅仅是纸本专著,还包括各种论文、专业标准、批文、照片、录音/录象资料及一些实物等.对于这样一个复杂的系统,系统设计的开放性、灵活性、可互操作性和可扩展性就显得非常重要,而且还要兼顾能永久保存和整个信息处理流程易管理.

“核”类数据资源的元数据设计基于以下考虑:参考其他相关的标准和成熟技术,能满足资源的保存、检索利用的实际需要.因为数据资源首先要纳入整个管理系统,因此开发管理人员是面临的第一用户,然后是专业用户,普通读者的需求可以包含在专业用户中.同时要尽量考虑与其它元数据方案的可交换性,力求做到是一个能够适用于各种与核相关的通用的元数据方案,以便向其它元数据方案映射或转换时尽可能保证语义不损失.方案基于中文核心元数据,同时拓展了个别适用元素.在结构上要具有开放性,常常要求保持平衡,不能在某一方面强调过度,例如对于限定,应该支持多种限定方式,同时对个别元素的限定级别不宜过深.最后还要注意数据元素必须够用,在能够完全实现系统需求的前提下,做到简单实用.方案尽可能采用精简的“核心”元素集,以便实现降低成本,加快实现进度.

核类元数据元素定义与著录说明:

元素列表 著录说明

1)名称 必备字段,可重复使用.是由资源创建者或出版者定义的资源名称.

2)原始创建者 非必备字段,可重复使用.原始内容的创作者,可以是个人、团体等.

3)提供者 非必备字段,可重复使用.资源来源提供者,可以是个人、团体或网络空间上的实体.

4)其它责任者 非必备字段,可重复使用,可著录多个责任者.对资源内容做出贡献,但未在创建者和提供者中列出的.

5)出版者 非必备字段,可重复使用.资源出版单位.

6)原始创建日期 非必备字段,可重复使用.资源创建时间.

7)主题类别 必备字段,可重复使用.该资源所涉及的主题类,如核武器、核工程、核电子、核防护等.

8)说明 非必备字段,可重复使用.该资源的一些必要说明.

9)主题词 必备字段,可重复使用.描述资源主题或内容的关键词、词组短语或分类号.

10)内容摘要 非必备字段,可重复使用.资源内容的文本描述,包括文献类对象的文摘或实物的内容描述等.

11)内容类型 非必备字段,可重复使用.资源的特征和类型.如会议论文、学位论文、专著、标准、批文、实物等.

12)格式 必备字段,可重复使用.资源的媒体形式,如文本、图片、声音、影象等.

13)语种 非必备字段,可重复使用.资源内容使用的语言说明.

14)关联资源 非必备字段,可重复使用.与该资源相关的其他资源,包括该资源版本是否发生变化、替代等.

15)资源创作者 非必备字段,可重复使用.数字资源创作者名称.

16)资源创作日期 非必备字段,可重复使用.数字资源创作时间.

17)版权所有 非必备字段,可重复使用.本资源版权人.

18)版权声明 非必备字段,可重复使用.本资源的使用权限说明.

19)权限级别 必备字段,可重复使用.对资源索取的限定.

20)获得方式 非必备字段,可重复使用.资源取得方式说明.如捐赠、馆藏、外购、网上检索等.

21)技术环境 非必备字段,可重复使用.数据对象加工时对计算机硬件、软件要求的说明.

### 2.4 资源库知识的组织系统分析

核技术是基于原子核科学基础知识、粒子加速与射线产生的原理和方法,利用射线与物质相互作用而产生的物理、化学或生物效应为人类各项事业服务的交叉学科领域.除核武器与核电之外的民用非动力核技术的应用涉及工业、农业、医疗健康、环境保护、资源勘探和公众安全等领域.

同时在核电项目的带动下,中国已建成了从铀矿勘查、采冶、铀同位素分离、核燃料元件制造直至乏燃料后处理、核废物处置的完整、配套的核燃料循环系统.以同位素和辐射技术为代表的核技术应用,也在人类生活和社会发展中发挥着重

要的作用.一些技术已达到国际先进水平.核技术应用作为新兴的高技术产业,正在突破传统的工业、农业和医学等应用领域,向现代化科学技术的前沿,如信息技术、生物技术、纳米技术、新材料技术和环境技术渗透,不断形成更新的交叉应用领域<sup>[4]</sup>.所以“核”专题相关的已经涉及到工、农、医、经济、管理等多交叉学科领域.

### 2.5 使用资源的关联群体分析

建设“核”专题资源库的目的首先是为了满足相关专业的教师、学生教学研究的需求,同时还必须向社会需要者提供资源共享.从人员分析主要有三种类型:1)学校的相关专业教师、学生,主要满足辅助教学、拓展知识结构和学术交流的需要;2)相关行业的研究人员,产品开发、生产和使用者.主要为了学术研究,掌握技术发展及产品开发和使用;3)对本知识领域有兴趣的一般爱好者,希望了解“核”学科的科普知识.从专业分析会涉及到:军事、电力生产、地质勘探、采冶、金属加工、防护与环保、能源经济以及核技术在医学、仪器、农业、生物等应用的多个领域范畴.

### 2.6 访问界面的主要服务功能

特色资源网站服务可考虑实现以下几个基本功能:1)用户身份确认决定不同的访问权限;2)数据导航,引导用户查找本馆藏以外的全球“核”数据资源;3)便捷检索,如按时间段、名称、作者名、类型、媒体形式……等组成不同的逻辑关系查找;4)新闻通告,用于新资源信息及重要事宜的发布;5)分类浏览;6)用户技术讨论,可以设立一些不同主题的内容,为有共同需求和兴趣的用户进行交流提供空间;7)服务定制,用户可通过本栏目要求得到个性化服务,如特定资源预约,将内容进行语种转换,资源推荐等.可通过 Email、QQ、MSN 完成;8)常用工具下载,对于不同格式的资源提供客户端阅读工具.

## 3 特色资源的知识与信息组织

数字图书馆建设是一项全球性的信息系统工程,因此各种资源库的发展也必须遵循客观规律,数字图书馆的可持续发展必须与经济发展状况和科技发展水平相适应,既要适应当前,又要面向未来.

### 3.1 开展全方位合作加强资源整合

随着信息全球化进程的加快,特色资源建设必须打破区域界限甚至国界,要加强与国内有关单位及国际机构间的合作,要逐步签定和遵守相关协议,促进知识的融合和资源全球共享.同时要

根据资源本身的特点,通过各种导航路径多层次、多方位地对各种信息资源进行加工、重组,不断将分散的资源积聚整合,让用户迅速得到所需要的信息.

### 3.2 多种形式并存,注重提高资源的价值

传统图书馆被视为一种社会公益事业,它的运行几乎不存在市场环境,而数字特色资源的建立和运营引入了很多商业性资源,追求的是用户和读者最大需求.对用于教育与培训的基础性文献资源,仍然可以无偿服务为主.而对服务于科学研究居于前沿的信息资源以及服务于经济和商业可转化为生产力的信息资源,可以采用商业化的运营模式.要选择有较高利用价值的资源类型进行数字化,通过有偿服务获取投资回报,追求最佳的投入产出效益,保证资源建设的协调发展.

要注意提高已有资源的价值,图书馆是一个文化典藏机构,同时也应该是一个研究开发型的机构,要利用馆内外人力和数字资源的优势,把馆员的工作价值与信息服务效益结合起来,可以利用图书馆对文献擅长的著录标引、索引组织、分析研究的特长,对文献内容进行揭示和研究,按用户要求完成资源语种的转换,使信息服务向更高级的知识服务领域发展.

### 3.3 合理合法使用资源

特色资源的建立和使用要认真研究有关法律问题,对存在知识产权问题的资源在进行数字化之前要得到知识产权拥有者的授权.在资源建设开发中,大量经增值加工的数字化资源又形成了新的版权,可采用收费服务的方式.

总之,专题特色资源的建设是信息化发展的必然要求,要力求在统一标准的基础上将有限的资金和人力用于最有利用率的资源建设中,不断探索利于发展的新思路,逐步形成数字资源建设的良性循环.

### 参考文献:

- [1] 杜薇薇.网络环境下图书馆信息资源配置研究[J].情报杂志,2007(3):143-145.
- [2] 龚景兴.“数字化陷阱”与数字图书馆建设的可持续发展[J].图书情报工作,2005(5):136-138.
- [3] 刘丹.基于元数据的特色数据库构建——以福州大学图书馆“催化化学特色数据库”为例[J].2007(6):51-53.
- [4] 徐革.特色数据库中基于本体论和元数据的知识与信息组织应用[J].情报理论与实践,2006(3):368-370.