

3M 协办

本栏目由3M公司协办。3M公司是世界五百强企业唯一一家提供完整的图书馆资料借还流通、信息处理及图书馆馆藏文献安全保障系列产品的公司。3M在图书馆资料流通管理技术的前沿辛苦耕耘了近40年,在全世界拥有5万多个图书馆用户。

# 文献资源集成揭示的思路与实践

宋文 傅红梅 路纳新 景丽

中国科学院国家科学图书馆 北京 100190

〔摘要〕从资源建设模式和出版模式的转变对文献资源组织产生的影响出发,通过分析文献资源集成的原理和目前主要的集成模式,总结中国科学院文献资源集成揭示的内容和对象,提出中国科学院文献资源集成揭示系统的实施策略。

〔关键词〕文献资源 信息资源 集成揭示

〔分类号〕G250.7

The Research and Practice of Document Resources Integration

Song Wen Fu Hongmei Lu Naxin Jing Li

The Library of the Chinese Academy of Sciences, Beijing 100190

〔Abstract〕From the point of view of the enormous influence on document information resources organization generated by the transition of resources development modes and publishing modes, the authors analyse the type and scope of information resources needed to be integrated in CAS's document information resources integrated system. In the end, the implementary strategies of system are illustrated.

〔Keywords〕document resources information resources integrated organization

中国科学院(以下简称“中科院”)文献情报系统中长期规划的战略目标之一是建立强大的公共信息平台,通过无缝整合现有服务流程和广泛信息资源,实现充分方便的集成检索,无缝链接,动态扩展和联合服务。文献信息资源是公共信息平台的重要组成部分,文献信息资源的集成揭示是中国科学院国家科学图书馆三期创新的重要任务之一。

## 1 相关研究文献综述

笔者在LISA数据库中用“integrat\* or aggregat\*”进行检索,在用年代做限定后,分别得到如下结果:从LISA最早收录年代到1969年共检索到25篇文献;20世纪70年代共有128篇文献;80年代有386篇文献;90年代上升到826篇文献;从2000年到2006年已经达到758篇文献。笔者大致对2000年到2006年的758篇文献进行浏览,发现涉及的内容非常广泛,包括企业信息集成、数据库集成、系统集成、网络资源集成、各专业领域信息集成等。可以看出集成已经成为数字图书馆建设的主旋律。

笔者用“aggregat\* or integrat\*”与digital library、

documentation、resources进行组配检索,共检索到51篇与文献资源集成揭示相关的文献。主要讨论的问题有:纸本资源和电子资源集成揭示,电子资源的编目,各种类型电子资源的集成揭示和服务,电子资源管理系统,各类型资源的元数据集成等问题。例如:Janice Lewis Steed和Clark Nall的论文探讨在图书馆构建一个主题指南数据库,包含所有格式的资源,作为用户查找资源的入口,可以检索到图书馆的所有的资源<sup>[1]</sup>。Davies S和Wales A在论文中详细介绍了英国NHS图书馆联盟联合的电子资源建设和集成虚拟图书馆建设工作<sup>[2]</sup>。

笔者在维普中文科技期刊数据库中,共检索到200余篇与集成相关的论文,主要讨论的是关于网络信息资源整合、电子资源整合的技术、模式和方法等问题。理论和方法探讨比较多,实践研究的论文相对较少。

## 2 资源建设模式的转变对文献资源组织的影响

信息技术的发展,使信息的生产模式、供应模式发生了巨大的变化。在新的信息供应模式之上,建立全新的资源建设模式是资源整合化建设的主要内容。随着中科院国家科学

收稿日期:2007-10-23 本文起止页码:93-96 本文责任编辑:王传清

图书馆的组建,需要逐步建立文献资源整合化建设、集成化揭示新的业务模式。文献资源整合化建设主要包括以下几个方面内容:

### 2.1 资源建设渠道的整合

随着信息的生产模式、供应模式的变化,资源的来源渠道也越来越多样化。资源供应渠道除了出版商和供应商提供的资源之外,还包括开放获取资源、第三方资源等。

### 2.2 资源载体的整合

多种载体资源的并行供应,使多载体资源的整合建设和集成揭示成为迫切的需要。比如就期刊资源建设而言,就有P+E模式、E-Only模式、P-Only模式、E-first+DDP模式。在这样的资源建设状况下,将出现纸本资源和电子资源共存,相互交叉和相互补充的状况。

### 2.3 资源内容的整合

中国科学院科研机构具有地理位置分布广,学科内容相互交叉的特点。在进行资源建设中,需要从资源学科内容出发,对中科院用户的需求广度、频度、强度等进行分析和评价,针对不同需求的资源,采用不同的购买和开通模式。比如,对于主要学科的检索数据库和综合性核心全文数据库,在全院有普遍性、经常性需求,采取全院开通模式;对于面向学科领域的核心全文数据库,在大学科领域有经常性需求,采取按学科范围在相关领域的研究所开通;对于数学、天文、空间技术、海洋、能源等学科的全文数据库,在专门专业领域有经常性需求,以相关专业研究所为核心的特色分馆机制开通。

### 2.4 服务模式的整合

文献资源的服务模式包括电子资源的即查即得、电子和纸本文献的原文传递、图书文献的馆际互借等。服务模式是资源建设的延伸和补充,资源建设需要与服务模式建设相结合,才能在有限的经费情况下,最大限度地满足用户需求。

### 2.5 资源类型的整合

文献信息资源包括全文数据库、文摘数据库、期刊、图书、会议文献、专利、研究生参考资料等。中科院是国家级科研机构和国家科技人才的培养基地,用户的类型多样化,对文献信息资源的需求也各不相同,因此从文献资源类型出发的整合化建设也是文献资源整合建设的一个重要方面。

综上所述,通过中科院全院文献资源的整合化建设,将使全院文献资源在内容、类型、格式和地理位置等多纬度的分布更趋合理。在这样一个资源布局下,需要对中科院全院文献资源进行集成化揭示,才能更有效地满足全院用户的个性化需求。

## 3 文献资源集成揭示模式分析

由于集成揭示对象的不同、实现方式的不同以及达到集成揭示的目的和效果的不同,集成揭示可以有多种实现方式。

下面从集成揭示的对象、集成揭示的方式和集成揭示的深度三个方面对集成揭示系统进行分析。

### 3.1 集成揭示的对象

从集成揭示的对象来看,可以分为单件(item)集合(collection)和服务(service)。

单件:表示一种图书、一种期刊或一篇论文,对单件文献集成的最常见的例子有图书目录、期刊目录、期刊全文数据库等。现在开始越来越多的集成目录将多格式、对载体的文献集成在一个目录中,如中科院集成期刊目录集成了电子期刊和印本期刊。

集合:按照DCMI集合工作组的定义,集合是任何物理或数字对象的聚集<sup>[3]</sup>。对集合进行集成揭示的方式有:建立集合目录,也就是集合的集合,供用户查找某个特定的集成;用技术的方式实现集合的统一检索,如跨库检索系统;在对集合进行统一的描述和揭示的基础上,实现统一的检索,如英国的IESR系统<sup>[4]</sup>。

服务:表示数字图书馆提供的网络服务,一个服务可以包含一个或多个资源集合。对服务进行集成的技术包括简单的服务列表,W3C建立了一套完整技术体系,实现服务的集成,这套技术方法也可以应用到数字图书馆服务集成上。

### 3.2 集成揭示的方式

集成揭示方式可以分为:

集中式:将被揭示对象的元数据集中到一个物理数据库中。

虚拟式:实际上没有一个物理的集中数据库,只是在检索时对多个系统进行检索,将形成临时的集中检索结果进行显示,例如常见的跨库检索机制。

集中与虚拟相结合:这时各个分布系统有自己的揭示机制,但通过一定方式将这些分布系统的元数据收集到一个集中的揭示机制里,人们实际可以通过集中揭示机制来检索和获取,也可以通过对具体的分布系统本身进行检索来获取信息对象,例如OAI元数据搜索服务。

### 3.3 集成揭示的深度

从现在的实践看,集成揭示深度至少体现在三方面:

揭示信息(元数据)丰富程度:即描述被揭示对象的元数据内容是否丰富。

揭示对象调用能力:是否可以通过在元数据中嵌套的标识符(如DOI、OpenURL等)来灵活调用被揭示的信息对象。这个能力实际上还反映在多个方面,包括:能否直接调用信息对象;在调用中能否根据情景来判断调用对象和调用方法。实现情景敏感的对象调度常用的是建立调度情景的知识库,并结合OpenURL协议,实现根据用户使用情景的目标资源的动态调度。

信息对象的关联揭示能力:从一个信息对象,根据特定的关联体系,用户可以扩展到其他在知识内容上相关的信

息对象。实现从知识内容方面的关联的主要途径是建立并应用知识组织体系。知识组织体系不仅可以深入地揭示信息对象知识内容,而且能揭示与这个信息对象在知识体系上相关的其他信息对象。

#### 4 对中科院文献资源集成揭示体系的思考

出版模式和资源建设模式转变所带来的影响是巨大的,文献资源集成揭示体系建设已经成为迫在眉睫的任务。下面从中科院文献情报体系的结构特点、用户需求特点、资源建设模式特点和资源服务模式特点出发,对中科院文献资源集成揭示进行总体规划和设计。

中科院文献资源集成揭示体系需要集成揭示科学院范围的馆藏印本资源、电子资源、开放获取资源、第三方资源等可以用来为科研用户提供信息服务的资源。

中科院集成揭示体系需要提供用户灵活、高效、情景敏感的信息服务,可以为用户定制或选择合适的目标资源或服务。所以集成揭示体系在揭示文献信息资源的同时,需要揭示非文献类信息,如机构信息,用户信息;需要揭示各种信息的关联,比如全文电子期刊与全文电子数据库的关联,机构与开通信息的关联等。

中科院集成揭示体系需要对揭示的信息对象进行知识内容的深度揭示和关联,使用户可以从知识角度出发发现相关信息资源。

所以,中科院集成揭示体系建设可以从如下几个方面进行考虑:

##### 4.1 中科院文献资源集成揭示的范围

首先,中科院文献情报系统由国家科学图书馆和近百个研究所图书馆组成,是一个业务上由国家科学图书馆统一领导的图书馆联盟。中科院文献情报系统的资源和服务的整合为中科院全院的科研创新提供文献信息资源服务。所以中科院文献资源集成揭示在范围上要包括科学院各研究所图书馆收藏的印本文献资源和电子信息资源。

其次,中国科学院国家科学图书馆是国家科技图书文献中心的有机组成部分,国家科技图书文献中心的资源作为全国范围共享的资源,是可以提供全院科研人员使用的。所以,国家科技图书文献中心的资源也需要在集成揭示系统中加以揭示。

另外,开放获取资源和友好合作单位的资源可以作为全院文献资源的补充,也是提供用户服务的重要资源,集成揭示系统中也需要揭示这部分资源。

##### 4.2 中科院文献资源集成揭示的对象

从多年文献资源建设成果来看,目前需要集成揭示的资源包括:

馆藏印本资源:包括印本的中外文图书、印本的中外

文期刊、学位论文等。

全文电子资源:包括全文电子期刊、全文电子图书。

电子数据库:除全文数据库外,还有大量的检索类数据库需要集成,如剑桥科学文摘数据库。

内部资源:在科研过程中产生的大量的灰色文献,比如研究报告、项目报告、技术报告等,这些灰色文献是非常有价值的文献信息资源,也是需要加以揭示,并在一定的授权范围内提供使用的。

##### 4.3 中科院文献资源集成揭示的信息类型

为达到提供用户集成的文献信息服务的目标,笔者将文献信息资源集成揭示体系中需要揭示的信息分四个类型:信息对象的元数据描述、馆藏信息和开通信息、服务管理型信息、资源关联和组织型资源。

信息对象的元数据信息:对信息对象的属性的描述与揭示,提供用户发现特定的文献信息,包括印本文献、电子文献的元数据信息,如学位论文元数据等。

馆藏信息和开通信息:馆藏信息是指印本文献的收藏单位,馆藏信息即可以为单个图书馆文献借阅提供信息,也可以服务于馆际互借和原文传递。电子资源的开通信息即可以作为用户直接获取电子文献提供服务,也相当于印本的馆藏信息,可以提供原文传递服务。

服务管理型信息:服务管理型信息包括用户的信息,如用户识别号、姓名、用户类别、所在机构;机构信息,如机构名称,IP地址等。这些信息虽然不是直接为用户服务的文献信息,但在整个集成揭示系统中具有重要的作用。馆际互借和原文传递系统、联合目录系统、情景敏感的集成期刊目录系统都离不开这些信息的支持。

信息对象关联和组织型资源:包括知识组织体系和规范体系。知识组织体系从知识内容出发将相关资源集成到一起,规范体系从名称角度,将同一创建者的信息对象集成到一起。虽然在元数据描述中已包含了一部分信息资源间的关联,但从知识内容和信息的创建者出发进行的关联与集成是集成揭示体系建设的重要内容。

## 5 中科院国家科学图书馆集成揭示系统的实施策略

### 5.1 文献资源集成服务系统目标情景描述

文献资源集成揭示体系建设的目标是提供用户集成的信息检索和获取服务。集成服务系统要达到的目标服务情景描述如下:

5.1.1 集成检索 提供用户对图书、期刊、期刊论文、学位论文、网络资源、研究报告等各类文献的检索。集成检索系统根据用户的检索请求,需要调用各种不同的数据库,比如检索图书需要调用联合目录数据库,检索西文期刊论文可以

调用国家科技图书文献中心的期刊论文数据库。支持集成检索的数据库可以是本地的,也可以是远程的;可以是自建的,也可以是购买的、第三方资源或开放获取资源。集成检索系统应用跨库检索系统,可以实现对多个数据库的统一检索。

5.1.2 根据情景,提供用户最合适的资源和服务 检索系统查询到用户需要的资源后,可以根据用户身份和权限、资源的状况,提供用户获取资源的最佳方式。比如用户检索期刊论文,系统可以根据用户身份判断该用户是否具有对某个全文期刊的获取权限,如果用户有全文获取权限,首先提供用户全文连接地址。如果用户没有全文可以获取,系统自动判断用户所在单位是否定购该期刊,该期刊是否到馆,如果是,可以告诉用户直接到所在单位的图书馆借阅。如果用户所在单位没有收藏此期刊,可以提示用户是否需要原文传递,用户可以直接提交原文传递请求。

5.1.3 资源间的关联 用户检索到目标资源后,可以通过各种关联机制,继续查询相关资源。比如根据元数据中的相关资源描述,根据知识组织体系的关联等。

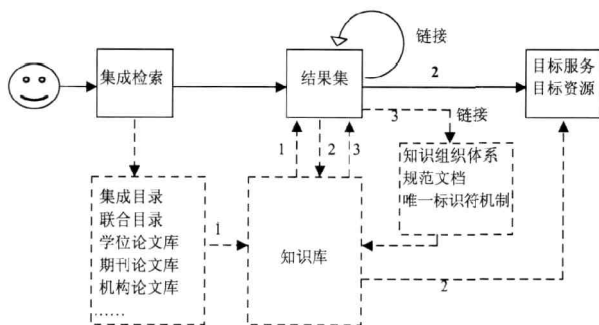


图1 集成服务流程

5.1.4 实现与外部系统的集成 集成揭示系统提供多种接口,实现从外部系统到集成系统的连接。主要应提供的接口有:Web 服务接口、Z39.50 协议接口、OpenURL 接口、唯一标识符接口等。系统可以实现连接包括从搜索引擎到集成服务系统的连接,从商业数据库到集成服务系统的连接,从友好合作单位的连接服务器到集成服务系统的连接等。

图1是集成检索系统服务流程示意图,实线表示用户流程,虚线表示系统流程。

## 5.2 中科院集成揭示体系建设的实施步骤和策略

5.2.1 基本文献资源的集成揭示体系建设 建设全院统一的文献资源揭示系统,用一个统一的系统,实现全院所有单位的资源的揭示和服务。公共揭示系统应达到如下目标:实现单个图书馆自动化系统的采购、编目、期刊收登、流通等所有工作;实现电子资源和印本资源的统一和集中的揭示;实现全院的采购信息、书目信息、馆藏信息、签到信息、流通

数据、读者信息的集中统一管理;通过特殊的联盟机制,实现全院集成检索服务、馆际互借和原文传递服务、统一采购服务;在这样一个系统上,实现全院范围的全面和规范的文献资源利用的统计分析。

5.2.2 灰色资源的集成揭示体系建设 中科院每年在科研过程中会产生大量的非正式出版的文献资源,比如项目开题和结题报告、非正式发表的论文、非正式出版的会议论文、博硕士论文、博士后研究报告等。另外,对于大量的非文献类信息,如科研仪器和设备、科研试剂、实验动物、标本、科研数据、科研人员和科研机构、科研动态、科研新闻等,这些信息是科研人员在科研过程中非常需要的信息资源,也是需要揭示并提供科研人员的查询服务的。灰色资源的揭示又可以分三个部分:博硕士论文数据库。收录科学院全院的博硕士研究生的毕业论文、博士后研究报告的电子版全文,在合理的版权保护和使用制度下,提供全院范围的服务。

机构仓储。建立机构仓储系统保存机构内部的非正式发表的论文、研究报告,并在一定的条件和范围内提供服务。非文献类信息的组织和揭示。使用本体技术和语义门户技术,组织和揭示包括科研仪器设备、科研人员和机构等在内的非文献类信息。

5.2.3 规范体系建设 建立联机联合加工的规范数据库,实现全院范围的联合的规范数据库建设和维护更新。规范库应用于全院各种类型资源加工的规范,通过规范数据库实现集成服务系统中资源间的链接和统一检索。

5.2.4 知识组织体系建设 由于目前国家科学数字图书馆提供用户的资源类型广泛,各种资源依据的知识组织体系各不相同,建设集成的知识组织体系,可以实现从学科体系出发对多种资源的集成揭示和检索。可以仿照欧洲Renardus项目的经验,选取一个核心的知识组织体系,通过映射的方法,将其他知识组织体系集成到核心的知识组织体系之下。

参考文献:

- [1] Lewis JS, Nall, C. Integrating print and electronic resources: Joyner Library's PirateSource. *Acquisitions Librarian*, 2005, 16 (33/34): 107-120.
- [2] Davies S, Wales A. Integrating electronic information resources for NHS Glasgow staff at the point of need: A model of interlibrary collaboration and resourcesharing. *Health Information and Libraries Journal*, 2001, 18 (4): 229-37.
- [3] DCMI Collection Description Community. [2007-01-04]. <http://dublincore.org/groups/collections/>.
- [4] IESR: Information Environment Service Registry. [2007-01-04]. <http://www.iesr.ac.uk/>.

[作者简介] 宋文,女,1961年生,研究馆员,发表论文20篇;傅红梅,女,1967年生,副研究馆员,发表论文5篇;路纳新,女,1975年生,馆员,发表论文1篇;景丽,女,1980年生,馆员。