

●孙坦 周静怡 宋文

数字图书馆开放式资源与服务描述规范^{*}

摘要 为了实现一站式服务,对网络化的信息资源与服务逐层进行集成化组织和揭示,提供公共登记服务,需要建立开放式资源和服务登记系统,制定相应的开放式资源和服务描述规范。开放式资源和服务描述规范设计原则和方式是:以服务方式描述所有资源和服务;多角度描述服务;基于 UDDI 的 OFSR 的描述规范。表 2。图 3。参考文献 8。

关键词 开放式资源与服务登记系统 数字图书馆 资源 描述规范

分类号 G250.76

ABSTRACT To realize one-stop services, there is a need for the integrated organization and description of networked information resources and services, a need for open framework of service registry and a need for specifications for the description of open resources and services. Principles and methods for the design of specifications for open resources and services include describing all resources and services as services, describing service from multiple perspectives, and OFSR description specifications based on UDDI. 2 tabs. 3 figs. 8 refs.

KEY WORDS Open resource and service registry system. Digital library. Specifications for resource description.

CLASS NUMBER G250.76

“科技信息资源与服务集成揭示系统建设”项目是为了建立一个一站式的服务系统,对网络化的信息资源和服务逐层进行集成化组织和揭示,并提供公共登记、开放检索、信息整合、深度链接等服务,形成一个开放式的、可以不断扩充的共享服务平台。开放式资源和服务登记系统(OFSR, Open Framework of Service Registry)是其子项目,目标是制定开放式的资源和服务描述规范,建立分类体系,开发基于 UDDI 的登记注册系统。

对于任何信息系统,描述规范是其利用的基础。规范地进行描述,并建立一整套利用这种描述的机制(例如注册、发布、检索),对于资源和服务内容的揭示、发现、评估、整合和利用具有重要意义^[1]。对资源和服务的描述可以帮助对资源和服务进行管理,帮助用户更好地发现资源和利用服务^[2]。

DC CD AP 通过简化的方式来描述、发现、选择和管理资源集合。

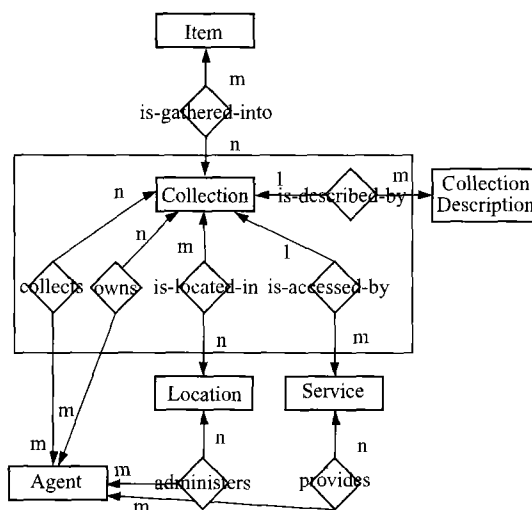


图1 DC Collection 实体关系示意^[5]

1 相关描述规范与系统

1.1 都柏林核心资源集合描述应用纲要

都柏林核心资源集合描述应用纲要(Dublin Core Collection Description Application Profile, DC CD AP)是由都柏林核心资源集合元数据委员会资源集合描述工作组推出的标准草案,已成为正式版应用纲要^[3]。

DC CD AP 是基于 An Analytical Model of Collection and their Catalogues^[4],同时在该模型的基础上增加了服务实体。DC CD AP 实体包括:Collection(资源集合)、Item(资源内容)、Location(资源定位)、Agent

^{*} 本文得到国家科技基础条件平台“科技信息资源与服务集成揭示系统建设”项目(编号2004DKA40680)资助。

(机构)和 service(服务)。除了 Collection 与 Collection Description、Collection 与 Service 之间的关系都是一对多,其他实体之间的关系都是多对多。实体和关系的所有属性(包括 Identifier, Qualified Name, Type of Term 等)都是通过 DC 元数据进行描述。

1.2 RSLP 资源集合描述

RSLP 资源集合描述(RSLP Collection Description, RSLP CD)是英国研究支持图书馆计划(Research Support Libraries Programme)之一,其目标是通过机读方式,可扩展地描述该计划中的所有资源集合。目前已经形成了非常成熟的资源集合模型和资源集合描述,它通过结构化的元数据属性集合,不仅实现了对 RSIP 项目的所有资源集合描述,还可用于博物馆、档案馆和图书馆等众多主题的物理资源和电子资源的描述。

RSLP CD 的实体主要包括 Content(智力创作)、Item(Content 的具体实例,如文献)、Collection(资源集合)、Location(定位)、Agent(代理,合理拥有资源集合的机构或人)和 Administrator(管理员)等实体。RSLP CD 模型将 Agent(包括 Creator, Producer, Collector, Owner, Administrator, 位于模型的左边)从 Objects(包括 Content, Item, Collection and Location, 位于模型的右边)中区分出来。Agent 同时可以拥有很多角色和权利,如将 Content 变成 Item,然后组成资源集合,并对资源集合进行管理,决定谁可以拥有或者管理这些资源集合。

RSLP CD Schema 采用 RDF/XML 对 RSLP 资源集合元素 Collection, Location, Agent(包括 Collector, Owner, Administrator)属性的描述,以及对它们之间关系的描述,并在此基础上充分利用了元数据和 vCard 属性。

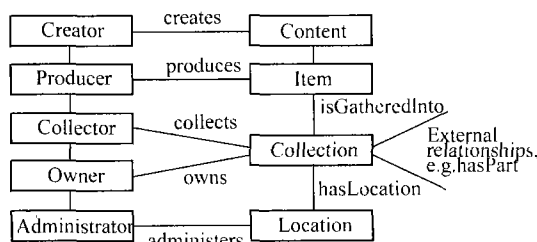


图2 RSLP CD 实体关系示意^[6]

1.3 信息环境服务登记系统

信息环境服务登记系统(Infomation Environment Service Registry, IESR)是英国联合信息系统委员会资助的项目,提供电子资源信息的登记,以便应用程序可以发现资源,提供终端用户使用。

IESR 实体包括:Collections(资源集合)、Service

(服务)和 Agent(机构,资源或服务的拥有者或管理者)。每个实体都有一组元数据对其进行描述。IESR 中用于描述规范设计是基于采用了 DC CD AP 和 RSLP CD 两种规范的融合。每一个机构拥有一个或多个资源集合,管理一种或者多种服务,资源集合的获取通过一种或者多种服务。

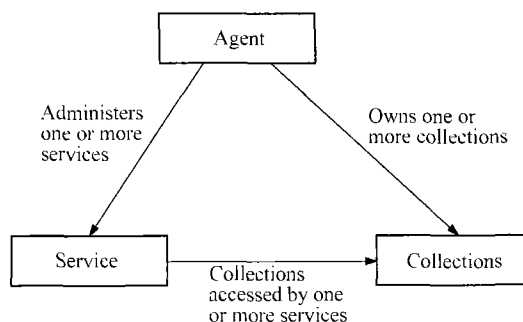


图3 IESR 实体关系

2 开放式资源和服务登记系统描述规范设计

2.1 以服务方式描述所有资源与服务

资源和服务是图书情报机构重要的核心组成,随着资源和服务的不断发展,对它们的描述越来越重要,图书情报机构需要将原有的对资源和服务的描述提高到一个更高层次^[7]。OFSR 集成国家科技图书文献中心(NSTL)、中国高等教育文献保障系统(CALIS)、国家图书馆和中国科学院国家科学图书馆的资源和服务。OFSR 中各单位参与集成的资源和服务如表1。

表1 OFSR 集成的资源和服务

机构	参与集成的资源与服务
国家科技图书文献中心	NSTL 文摘数据库;NSTL 用户订单数据库;NSTL 全文数据库;NSTL 目次查询系统;NSTL 馆藏联机联合编目数据库;NSTL 联合目录数据库;NSTL 特色文献数据库;NSTL 专家咨询知识库;NSTL 文献检索(中文/外文学位论文库);NSTL 中国预印本资源;NSTL 代查代借服务;NSTL 网络导航;NSTL 热点门户
CALIS	CALIS 联合目录;CALIS 学位论文中心网站;馆际互借系统
国家图书馆	国家图书馆馆藏目录;博士论文特色资源库
中国科学院国家科学图书馆	中科院联合目录数据库;中国科学文献服务系统;文献传递系统;参考咨询服务;中科院学位论文服务系统

资源是服务的基础。随着文献情报机构的发展和用户需求的不断提高,静态的资源不断地演变为动态的服务,不断地为用户提供文献情报服务。因此在本项目中将每一项资源都作为服务来进行描述。四家单位提供的资源和服务可以总结为四项服务:

(1) 信息查询:通过各种资源库检索信息。包括各种查询方法和手段,比如浏览、导航、简单检索、高级检索等。如由国家科技图书文献中心提供的 NSTL 目次查询系统和 NSTL 网络导航等。

(2) 文献获取:通过图书馆员帮助,使用户获得原始文献的图书馆服务。包括原文传递服务、馆际互借服务和到馆借阅服务,不包括读者使用计算机系统直接网络全文获取和网络电子图书借阅。如由 CALIS 和中科院国家科学图书馆提供的联合目录,包括多语种书刊联合目录数据库和联机合作编目、资源共享,提供成员馆之间馆藏资源共享、馆际互借和文献传递服务;由四家单位提供的馆藏期刊图书或者相应的文摘数据库和全文资源等,提供各种期刊和图书的检索和获取;由四家单位提供的各自的学位论文及系统,提供学位论文的检索和获取。

(3) 参考咨询:通过图书馆员利用各种方式和手段为用户提供的参考咨询服务。包括虚拟参考咨询服务,电话咨询服务,电子邮件咨询服务,到馆咨询服务,科技查新服务等。如由中国科学院国家科学图书馆提供的参考咨询服务。

(4) 知识增值:通过图书馆员的工作提供增值的知识服务,包括情报研究服务,教育培训服务,展览和讲座等。如由中科院国家科学图书馆提供的情报研究服务、教育培训服务和文化传播服务等。

2.2 多角度描述服务

在本项目的分类体系设计中,面向用户,贯穿了以服务为核心、以揭示服务的内容和属性为目的的分类体系设计思想,将项目的资源和服务划分为4类服务:信息查询;文献获取;参考咨询;知识增值。并在此基础上,从多个角度对服务进行了分类:服务分类、文献类型、信息格式、服务层次、信息所属学科、服务提供者、服务提供者所在地区、信息的语言、服务的权限控制类别、服务系统的接口协议。

为了实现一站式的服务系统,对网络化的信息资源与服务逐层进行集成化组织和揭示,资源和服务在登记的过程中,可以遵循以下登记规则:

(1) 分类体系的基础——“服务分类”是必选的分类体系,在此基础上的其他分类体系可以是有则必

备,但在实际的登记过程中,仍然鼓励登记的分类体系越详细越好,这样在用户使用和发现资源和服务的过程中,可以通过多种途径发现所需要的资源和服务。如由中国科学院国家科学图书馆提供的中国科学文献服务系统(ScienceChina),其服务分类体系中可以登记到“信息查询”类,然后可以进一步对该项服务的其他属性进行登记,如文献类型可以登记到“期刊”,信息格式类型可以登记到“数据库”,服务层次类型登记到“目次”和“引文”,信息所属学科登记到“计算机、信息和一般文献”、“自然科学和数学”和“技术科学”;服务提供者登记到“专业图书馆”;服务提供者所住地区登记到“北京”;信息的语言登记到“中文”和“英文”等。用户通过任何一个发现途径,都可以很容易找到该项服务。

(2) 已经进行了集成的服务,原则上只登记一种服务分类。但是在可以对该集成服务进行明显拆分时,进一步对服务进行拆分登记服务分类。如中国科学文献服务系统,服务分类体系实际上包括3种服务:现期目次库、中国科学引文库和中国学科文献库,这样可以在将中国科学文献服务系统登记为“信息查询”的服务分类的基础上,将其集成的3种服务也分别作为独立的服务分类进行登记,同时从其他角度登记。但是在这个过程中,集成的服务与其所集成的子服务之间的关系的表现方式还有待进一步讨论。

2.3 基于 UDDI 的 OFSR 描述规范

目前主要的描述规范多采用现成的一套或者几套数字对象元数据方案为基础,加以合并、复用,并加入描述资源集合的属性内容扩展而成。虽然目前已有的 schema,如 RDF,XML 可能也提供一些较为丰富的方法,但是在描述过程中,可能有些结构并没有完全表示出来。^[8] 因此针对特定的资源和服务制定其描述规范非常必要。

通用描述、发现和集成协议(Universal Description, Discovery, and Integration, UDDI) 基于现有的标准(如 XML 和 SOAP),主要包括用于描述机构及其服务注册信息结构的 XML Schema,一系列实现对 Web 服务信息进行注册和查询的基于 SOAP 的通用 API,以及在多个 UDDI 节点间实现注册信息复制的规范。

根据 OFSR 参与集成的各单位所提供的资源和服务,需要实现的是根据统一的检索和交换需要来规定统一的核心数据格式及其扩展方式。而 UDDI 刚好提供了一个基于标准的查找软件服务、调用服务和管理服务的元数据体系。作为 Web services 系列标

准的一个重要组成部分,它定义了标准的机制发布、发现基于网络的、面向服务的软件组件,目标是建立一个开放并且与平台无关的框架以描述、发现和集成服务。OFSR 需要构建一个集中式的数字图书馆资源和服务 UDDI 注册中心,它可以对数字图书馆中的

基于 Web Service 的资源和服务提供集中的注册和发现机制。

OFSR 基本沿用 UDDI 提供的四个核心数据,如表 2 所示。

表 2 UDDI 与 OFSR 核心数据

数据类型	功能概述	
	UDDI	OFSR
businessEntity	通过机构所提供的服务及其相关信息对机构的描述	提供资源和服务的机构的描述
businessService	服务的描述	机构提供的资源和服务的描述
bindingTemplate	Web service 的技术描述,包括服务访问点、协议和访问方式	服务的访问存储方式的描述
tModel	描述技术规范,进行标识和分类	服务的调用规范和技术信息的描述

在 OFSR 中,机构(BusinessEntity)包括 Name(机构名称的描述),Description(机构信息的描述),Contacts(机构联系方式的集合的描述),BusinessService(机构所包含的一个或者多个资源或者服务的描述信息/Web services 的描述),CategoryBag(包含不同方面的机构分类体系的集合),IdentifierBag(包含一个或多个 keyedReference 结构的集合,每一个结构表示一个单一的识别或识别体系)等主要元素。其中每个机构可以包括多个 BusinessService。

服务(BusinessService)包括 Name(服务名称的描述),Description(服务信息的描述),绑定(Binding-Templates,服务所包含的 Web services 技术信息的描述),CategoryBag(根据不同的服务角度对机构分类体系的集合)等主要元素。

绑定(BindingTemplates)包括 Description(绑定的描述),AccessPoint(调用 Web service 的访问点,包括 URL,e-mail 或电话),tModelInstance(每个 tModel 指向的绑定的详细规范),CategoryBag(绑定的分类体系)。

TModel 包括 Name(tModel 名称的描述),Description(tModel 信息的描述),overviewDoc(存放 tModel 相关指令的信息),CategoryBag(tModel 分类体系,如技术类型),IdentifierBag(标识 tModel,包括标识集合列表)。

3 结语

本文主要完成研究制订了描述规范设计原则,探讨了元数据格式选择原则,但是并没有元数据具体格式选择,许多细节和具体元数据格式的问题还有待进

一步研究和探讨。

本文主要研究了开放式资源与服务描述规范设计原则,探讨制订了元数据格式选择原则,但是仍存在一些细节问题,如具体元数据格式的问题还有待进一步研究、探讨和优化。

参考文献

- 1 刘炜,张晓林,曾蕾等.信息资源集合开放性描述研究.国家科学数字图书馆项目研究报告.2002.7.30
- 2,6,7 Andy Powell, Michael Heaney, Lorcan Dempsey. RSLP Collection Description. D-Lib Magazine, 2000(9)
- 3,5 Dublin Core Collection Description Working Group. Dublin Core Collection Description Application Profile. [2006-12-08]. <http://www.ukoln.ac.uk/metadata/demi/collection-application-profile/>
- 4,8 Michael Heaney. An Analytical Model of Collection and their Catalogues. [2007-02-06] <http://www.ukoln.ac.uk/metadata/rslp/model/amcc-v31.pdf>

孙 坦 中国科学院国家科学图书馆副馆长,研究馆员,硕士生导师,中国科学院研究生院教授。通讯地址:北京北四环西路 33 号中科院文献情报中心。邮编 100080。

周静怡 中国科学院国家科学图书馆馆员。通讯地址同上。

宋 文 中国科学院国家科学图书馆研究馆员,硕士生导师。通讯地址同上。

(来稿时间:2006-12-25)