

●徐 坦 孙 坦

数字图书馆服务登记系统数据模型设计*

摘 要 在现有服务多维度描述思想的指导下,以 IESR 数据模型为基础,设计了数字图书馆服务登记系统数据模型。它包括服务、业务、机构、资源和方法 5 个核心实体,使数字图书馆服务能够被有效地发布和发现。由 DLSR 数据模型向 UDDI 数据模型的映射,可以为 DLSR 的建设提供一种基于 UDDI 注册中心的直接的解决方案。表 1。图 2。参考文献 8。

关键词 数字图书馆 数字图书馆服务登记系统 数据模型 DLSR UDDI

分类号 G250.76

ABSTRACT According to the existing multidimensional description of services, based on an IESR data model, the authors design a data model for digital library service registry, which includes the five core entities of service, operation, structure, resource and method, and enables digital library service to be effectively released and discovered. A mapping from DLSR data model to UDDI data model can provide a direct solution based on UDDI registry center for the development of DLSR. 1 tab. 2 figs. 8 refs.

KEY WORDS Digital library. Digital library service registry. Data model. DLSR. UDDI.

CLASS NUMBER G250.76

数字图书馆服务登记系统(Digital Library Service Registry, DLSR)在发布与发现网络上可重用的数字图书馆服务方面发挥着重要作用。2006 年召开的 NSF/NSDL DLSR 专题会议给出的定义是:数字图书馆服务登记系统允许机器或人发现可以访问的数字图书馆服务,定位这些服务,并获得进行服务连接的相关配置信息^[1]。近几年出现了数个致力于 DLSR 研究与开发的项目,主要包括由 JISC 资助的 Information Environment Service Registry (IESR),由 NSDL 资助的 OCKHAM Registry,以及由我国科技部所资助的开放式资源与服务登记系统(Open Framework for Service Registry, OFSR)等。

在整个 DLSR 的设计与建设过程中,数据模型的设计是十分关键的一个环节。它不仅关系到整个系统的数据存储与更新性能,而且决定了服务提供者如何对数字图书馆服务进行描述,是一种规范化服务描述的过程。而这种服务描述方案的选择在某种意义上决定了服务使用者能否有效发现服务,以及构筑在服务调用与组合基础上的数字图书馆联盟的形式、功能,甚至是成败。本文在开发 OFSR 系统实践的基础上,通过分析相关服务描述方法及与数字图书馆服务有关的要素,对 DLSR 的数据模型进行探讨。

1 相关数据模型分析

1.1 多维度服务描述

DLSR 的核心任务是有效揭示数字图书馆服务,

而相关问题在一些标准化服务的描述和共享使用过程中早已提出,例如登记遵循 Z39.50 协议及其派生协议服务的 Explain、Explain-Lite、WebClarity 和 ZeeRex 等。另外,电子商务领域中的 UDDI(Universal Description, Discovery, and Integration)和 ebXML,以及分布式计算技术 CORBA、JINI、JXTA 和网络技术中的服务发布和发现机制则是更大范围内对服务描述的探索与实践。通过横向比较以上各方法的服务描述模式^[2],我们可以发现一个通用的做法是,在服务描述过程中除了揭示服务本身的基本信息(如名称、描述信息、供检索用的关键词等),还对与服务相关的实体信息进行揭示(如操作数据源、相关服务、管理实体等)。这样不仅能够帮助服务发布者更全面地描述服务,而且可以帮助他们从多角度发现服务并选择最合适的服务,体现了以使用需求为导向的思想。本文所设计的 DLSR 数据模型也将延续这一全方位、多维度描述服务的方法,设计多个概念实体以揭示数字图书馆服务。

1.2 IESR 数据模型的不足

事实上,作为第一个并最具影响力的 DLSR 系统,它在 2004 年就已经提出了由机构、资源与服务 3 个实体共同组成的数据模型^[3]。三者的关系概括地说是:一个机构拥有一个或多个资源;一个机构管理一个或多个服务;一个资源被一个或多个服务访问。并且该模型已被另外一个主要的 DLSR 研究项目 OCKHAM Registry 所采纳。但是通过分析,我们可以

* 本文是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(2004DKA40680)研究成果之一。

发现它存在两个主要问题。

(1)未独立出方法实体。在 IESR 数据模型中,将对于服务实现方法的相关描述作为服务实体的属性。在实际服务登记过程中,这一做法造成了数据冗余,增加了数据更新异常的可能性。一个实际的例子是,对于使用某一特定规范的多个服务,该规范要被重复注册,而且当规范有所调整后,各服务描述要分别做出相应的更新。而从另一个角度,随着数字图书馆服务实现技术方法体系的日趋完善,各种标准规范的制定,以及增加系统间互操作性的需求,越来越多的服务将采用已有的技术方法。因此,虽然方法与服务之间仍然有着密切的联系,但是将它作为独立的实体进行描述与组织,既可以提高系统性能,又可供新登记的服务引用,从而降低服务描述的复杂性。

(2)缺乏对业务流程的描述。DLSR 建设的目标之一是促进对服务的重用,提高服务质量的同时降低数字图书馆开发运行成本。而对于服务的使用,需要在一定的数字图书馆业务流程中完成,由特定的业务目标决定所需要的功能组件(服务)和调用逻辑。以具体业务为入口的服务发现不仅能够为用户选择服务带来便利,也可以为服务的使用提供参照。IESR 数据模型虽然在服务实体中设立了服务参照这一属性以连接相关服务,但是该属性仅能够记录相关

服务,却无法从整体表现复杂的服务调用关系,更无法体现服务组合的逻辑。

2 DLSR 数据模型

在现有服务多维度描述思想的指导下,以 IESR 数据模型为基础,本文设计了如图 1 所示的 DLSR 数据模型。该模型包括 5 个核心实体:服务(Service)、业务(Business)、机构(Agent)、资源(Collection)和方法(Measures)。

“服务”是由数字图书馆提供或能够被数字图书馆使用的可重用的功能组件^[4],典型的实例有检索服务、元数据收割服务等。与图书馆数字化信息服务的概念相区别,数字图书馆服务更关注于面向服务重用的功能封装。也就是说对于例如数字参考咨询和联机目录等图书馆数字化信息服务,若对其功能进行了必要封装,并通过具体的技术方法提供可被调用的程序接口,便仍可视为数字图书馆服务,属于 DLSR 所揭示的范畴。服务实体主要包括:一个清楚具体的叙述性定义(Abstract Services),描述服务功能、处理方法,及所操作数据;一个或一组执行指针(Service Bindings),指向具体的服务实现方法,并说明执行参数;定位信息(Deployed Services),不仅要指明服务的访问点,而且要列出已知的下游服务使用者(其他服务)。

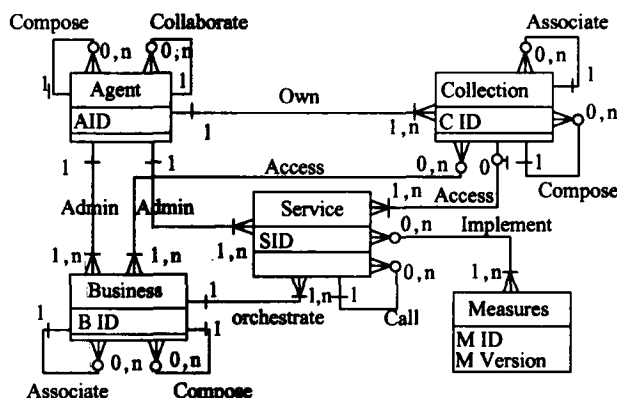


图 1 DLSR 数据模型

“业务”是数字图书馆整体业务目标的一个可识别的部分,例如发现与传递(Discovery & Delivery)业务。它作为数字图书馆服务的组织者而被注册到 DLSR。业务实体主要包括对业务目标、业务流程及所使用到的业务功能的描述。业务流程是一组业务功能的互连,以相关的资源或者信息为输入,完成某一特定的业务目标。业务功能是业务流程与具体服务的中介,用于描述功能行为、输入输出要求,甚至数据模型,将多个服务按照功能统一到一起,再经过业务流程的逻辑编排,实现通过业务实体记录对于服务

的调用逻辑与组织。

这里的“资源”是指集合类资源,它描述了服务所操作的数据源。根据处理数据的来源,数字图书馆服务可分为资源关联型服务如检索服务,和事务处理型服务如 OpenURL 解析服务。只有资源关联型服务与资源间存在访问关系。具体来说,资源实体应该包括 3 部分描述信息:基本信息,如资源集名称、类型、语种、主题词等;权限信息,如拥有权、版权、使用权等;关系信息,包括与其他资源集的联合或组成关系,以及与一个或一组服务的访问关系。

“机构”是服务、资源和业务的管理者,从管理的角度为服务提供了另外一种发现途径。任何一个服务都必须有且仅有一个管理机构,即使一个服务由多个机构共同提供,一个虚拟的机构也应该被注册到 DLSR。反之,与资源实体相同,在 DLSR 中注册的机构则可以与多个服务关联。机构实体主要记录机构的名称、介绍、联系方式,以及所管理的服务、资源与业务等相关信息。

“方法”是服务的技术实现,它指定了定义服务接口的标准或规范,如 SRW/U 等,以使服务能够提供可被机器处理的规范信息。方法实体主要用于揭示具体技术方法的相关信息,如名称、版本等,并给出机器可读的规范描述 XML 文档,而对于具体的访问点和参数由服务实体提供。在 DLSR 中,一个服务可以被多种方法所实现,而一种方法除了可被多个同种类型的服务使用,还可以被用于实现多种类型的服务,如 OAI-PMH 既可用于元数据收割服务,也可以利用自己携带复杂数据对象的能力应用于资源内容获取服务^[5]。

在 DLSR 数据模型中,除了上述提到的 5 组与服务实体相关的关系,还存在 3 组实体间关系,分别是:机构与资源之间的一对多的拥有关系;机构与业务之间的一对多的管理关系;业务与资源之间多对多的访问关系。机构之间、业务之间,以及资源之间的组成与联合关系也需要通过 DLSR 进行揭示。

尽管如果将 5 个实体:机构{A}、业务{B}、资源{C}、服务{S}、方法{M}视为 DLSR 的属性,它们所构成的依赖关系 $FDLSR = \{A \rightarrow B, A \rightarrow C, A \rightarrow S, B \rightarrow C, C \rightarrow B, B \rightarrow S, S \rightarrow B, C \rightarrow S, S \rightarrow C, S \rightarrow M, M \rightarrow S\}$ 并不具有较好的模式规范化级别,例如机构对服务的一对多管理关系可以通过业务进行传递,但是 5 个实体间都可建立直接联系,而且都可以互相作为检索点。因此,本数据模型冗余的实体间关系既是对客观现实的真实反映,同时也可以减少在查询的连接时延及系统的操作复杂性。

3 向 UDDI 的映射

近年来,UDDI 凭借它成功的应用、大量成熟的产品和对 Web 服务有效地描述机制,为人们展现了另外一种服务发布与发现途径。因此,讨论由 DLSR 数据模型向 LDDI 数据模型的映射可以为 DLSR 的建设提供一种基于 UDDI 注册中心的直接的解决方案。

UDDI 的基本数据模型由商业实体(BusinessEntity)、服务信息(BusinessService)、绑定信息(BindingTemplate)、技术规范信息(tModel)和发布者断言(PublisherAssertion)5 个实体组成,它们各自的描述功能及相互关系参见表 1^[6]。其中,对于 tModel 的使

用比较复杂。在分类方面,它可被作为分类体系的命名空间(Namespace)被其他实体,甚至本身所引用,而 uncheck 类型 tModel 还构成了 UDDI 的关键词系统;作为技术指纹,它提供了定义服务的类型、协议、格式、规则等特定信息。可以说,tModel 是 UDDI 系统复杂性的源泉,但同时也提供着 UDDI 中最重要的信息,从某种意义上讲,整个 UDDI 系统都是在其基础之上构建起来的。因此,对于 tModel 的充分利用,是解决将数据由 DLSR 向 UDDI 映射的关键。经过对两个数据模型的比较分析以及注册试验,本文提出如下的映射方案,映射后的 DLSR 数据模型参见图 2。

表 1 UDDI 基本数据模型

实 体	功能及关系概述
BusinessEntity	用于描述 Web 服务提供者的相关信息,提供一个或一组服务。
BusinessService	用于描述 Web 服务信息,包括一个或一组绑定。
BindingTemplate	用于描述技术信息以及说明服务访问点和调用 Web 服务的接口构建规范,包括一个或一组对 tModel 的引用,这些 tModel 指明服务的接口规范。
PublisherAssertion	用于描述商业实体间关系。
tModel	定义技术指纹;描述技术规范。 定义抽象的命名空间引用;进行标识和分类。

(1) 实体映射。将机构、业务(拆分为业务目标、业务流程和业务功能,以提高机器可读性)、资源实体全部映射为 BusinessEntity,通过创建并引用 Entity-Type 分类体系进行区分(有效值至少包括机构、业务、资源 3 种 BusinessEntity 类型)。将服务的叙述性信息映射为 BusinessService,方法实体映射为 tModel。BindingTemplate 通过引用方法类 tModel,并提供访问点信息,用于连接服务实体与方法实体。

(2) 关系映射。将机构、资源及业务与服务之间的关系直接映射为 BusinessEntity 与 BusinessService 间的一对多的提供关系,在机构、资源及业务实体中记录相关服务的标识符即可。服务与方法之间的关系,通过 UDDI 的 BusinessService、BindingTemplate 和 tModel 的固有关系直接体现。

uddi-org:relationships^[7]是 UDDI 自带的供 PublisherAssertion 使用的经典分类用 tModel,它支持父子(parent-child)、同级(peer-peer)和等同(identity)3 种类型的实体关系。通过对它的引用,在 PublisherAssertion 实体中,将所有机构、资源与业务实体间的组成、管理和拥有关系记录为 parent-child,将合作与联合关系记录为 peer-peer。

业务与资源实体间的访问关系通过对 UDDI 所提供的导航机制——BusinessEntity 的 DiscoveryURL 属性^[8]的扩展应用实现。该属性最初的目的是用于记录一个指向可替换描述性信息的 URL 列表,扩展应用可使之用于保存其他 BusinessEntity 的标识符,从而将业务与被访问资源关联起来。

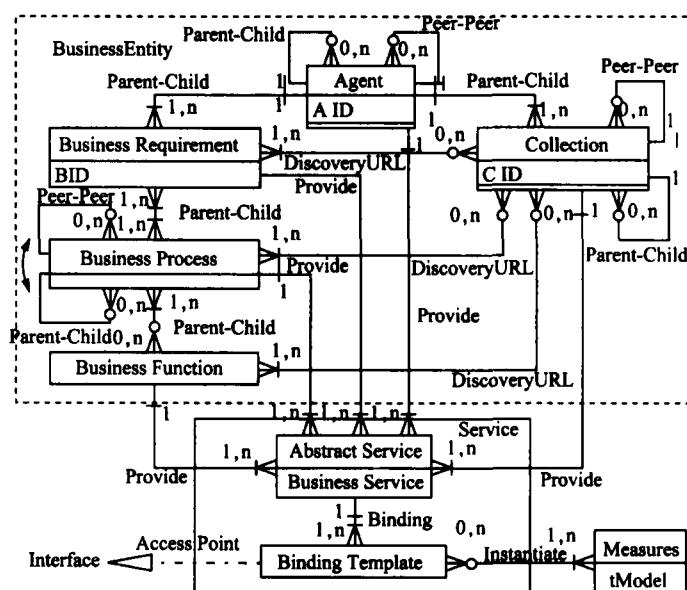


图 2 向 UDDI 映射后的 DLSR 数据模型

(3)属性映射。虽然 UDDI 对于各实体也提供多种属性,但是在用来描述 DLSR 各实体时难免会有不能满足实际使用需求的情况。对于这一问题,可以通过创建特定的分类及命名空间 tModel 来解决。通过对 tModel 进行引用,对命名空间可穷尽的实体属性给出分类,对命名空间不可穷尽的实体属性在限定范围内,按照一定的规范给出具体的值。而对于此类 tModel 的创建,则可以采用由登记系统管理机构统一创建和由服务发布者自行创建两种方式相结合的方法完成。

目前,DLSR 模型已经在 OFSR 系统中得到使用,并初步证明了其设计的合理性。本文所提出的 UDDI 数据模型的映射则更为 DLSR 的开发提供了一种基于 UDDI 注册中心的直接的解决方案。但是,服务描述的另外一个目标是机器可理解,而在实践中我们发现,目前对于资源访问权限、服务调用逻辑等描述信息仍然无法很好地达到这一要求。如何将权限控制、业务流程描述语言等方法融入到 DLSR 系统中,提高服务描述的机器可读性,以增强服务调用的自动化处理能力,从而构筑无缝数字图书馆联盟仍然是值得进一步思考的问题。

参考文献

- 1 Digital Library Service Registry Workshop. 2006. [2007-04-23]. <http://wiki.library.oregonstate.edu/confluence/display/DLSRW>

- 2 Matthew J. Dovey. (2003). JISC Information Environment Service Level Descriptions. [2007-04-23]. <http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/service-description/study/>
- 3 Amanda Hill. The Information Environment Service Registry: Promoting the use of electronic resources. Ariadne, 2004(40)
- 4 DLF Service Framework for Digital Libraries. 2005. [2007-04-23]. <http://www.diglib.org/architecture/serviceframe/dlfserviceframe.htm>
- 5 H. Van de Sompel, Carl Lagoze, Jeroen Bekaert, et al. An Interoperable Fabric for Scholarly Value Chains. D-Lib Magazine, 2006, 12(10)
- 6 UDDI Version 2.03 Data Structure Reference. 2002. [2007-04-23]. http://uddi.org/pubs/DataStructure_v2.htm
- 7 UDDI v2 Relationships Category System. 2004. [2007-04-23]. http://uddi.org/pubs/uddi_v3.htm#Relationships
- 8 UDDI Version 3.0.2. 2004. [2007-04-23]. http://uddi.org/pubs/uddi_v3.htm

徐 坦 中国科学院研究生院硕士研究生。通讯地址:北京北四环西路33号。邮编100080。

孙 坦 中国科学院文献情报中心研究馆员,博士,硕士生导师。通讯地址同上。

(来稿时间:2007-04-23)