

## 基于自动登记与检索的元数据登记系统\*

梁娜

(四川大学信息管理系, 成都 610064)

张晓林

(中国科学院文献情报中心, 北京 100080)

**摘要** 本文介绍了基于自动发布和发现的元数据登记系统的特点及功能, 重点分析了 ebXML Registry 系统, 介绍了 ebXML Registry 的应用范例 GoXML Registry。

**关键词** 元数据 元数据登记系统 ebXML Registry Goxml Registry

**中图分类号** G252.7 **文献标识码** A **文章编号** 1007-7634 (2003) 06-0565-03

## Metabata Registries Based on Auto Discovery and Publish

Liang Na

(Department of Information Management, Sichuan University, Chengdu610064)

Zhang Xiaolin

(Library, CAS, Beijing100080)

**Abstract** The paper describes the characters and basic functions of metadata registries based on auto discovery and publish, emphasizes on the structure of ebXML registry.

**Keywords** Metadata Metadata registry ebXML Registry GoXML Registry

## 1 元数据登记系统和元数据规范的自动发布与发现

在传统信息环境中, 各个系统将自己的元数据定义和相关应用规范作为内部数据予以组织和管理, 甚至封闭在系统软件中, 使得其它系统难以发现这些元数据规范, 从而难以建立相应的识别处理规则。有些系统虽然也通过一定形式公布了自己的元数据规范, 但没有通过公开和持久的渠道进行公布, 无法保障元数据规范的广泛和持久。

元数据登记系统 (Metadata Registry, MR) 是对元数据的定义信息及其编码、转换、应用等规范进行发布、登记、管理和检索的系统。各个系统将自己的元数据规范通过 MR 予以发布、存储、管理和检索, 支持其他系统发现和调用这些规范, 从而支持对多元化数字资源与服务的发现、识别和集成。随着数字资源系统的日益发展和多样化, 随着元数据规范的日益发展和复杂化, 元数据规范登记系统对于开放环境下数字图书馆系统的有效利用、互操作和可持续发展具有重要意义。

MR 类型有很多, 有单一名称域 MR, 如 DCMI Registry、ROADS Registry; 有跨名称域 MR, 如

German Metadata Registry (GMR)、DESIRE Metadata Registry (DMR) 和 Schemas Registry (SR) 等; 有分布式 MR, 针对专业领域建立如地理空间 MR、教育 MR、音频视频对象 MR, 或者按元数据类型建立的 MR 如资源集合元数据、内容对象元数据、知识组织体系元数据、管理规则元数据等。

MR 还可分为人工 MR 和自动 MR。人工 MR 支持人工登记和人工检索, 人们可以将元数据规范信息 (例如元数据定义、元数据语义描述、元数据使用规范、元数据描述格式等) 通过上载模板上载到 MR, MR 根据模板上填写的信息对上载文件进行分类和组织, 并在 MR 网站上发布。检索时, 可以根据元数据的命名域、维护机构、元素、修饰符、内容编码体系、元素语义关系、元素映射关系等检索相应的元数据规范信息。MR 发展趋势是元数据规范的自动登记、发布和发现。

自动 MR 通过多个 API 来支持第三方系统自动登记和检索有关元数据规范, 元数据维护机构可以通过一定协议直接利用登记 API 上载要登记的元数据规范信息。自动 MR 通过 MR 数据模型和管理模块将各个维护机构所提交的元数据规范信息分门别类地自动登记存储、并在相互关联的信息间建立起关联机制, 提供检索 API 支持第三方系统自动

\* 本文为“中国科学院百人计划”基金资助项目

收稿日期: 2002-09-09

检索和关联验证元数据规范信息。另外自动 MR 还建立自我开放描述、在公共服务登记系统登记,支持第三方系统对自己的发现、配置和检索。

自动 MR 在数字图书馆领域尚处于探索阶段,但在电子商务和网络服务等方面已经得到很大发展,我们下面利用 ebXML Registry 为范例来说明自动 MR 的机制。

## 2 ebXML Registry 机制

(1) ebXML 机制。ebXML 是在 Open EDI 基础上发展起来的基于 XML 的开放电子商务机制。ebXML 的基本机制由图 1 描述:

①系统 A 为实现特定商务功能,向 ebXML 登记系统查询并请求商务模式(及其过程、交换信息、文件等)的标记文件。

②系统 A 利用这些标记文件建立符合 ebXML 规范的商务系统。

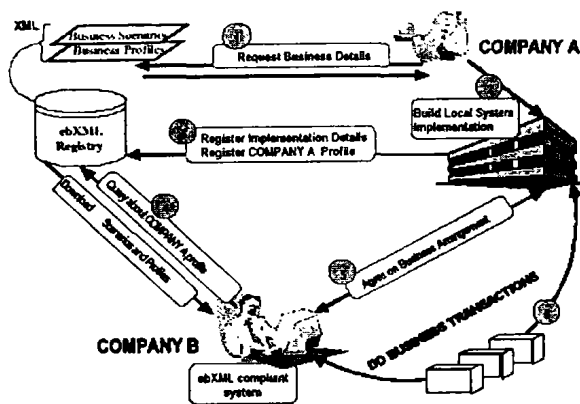


图 1

③系统 A 将自己的商务意向 (Business Profile) 提交给 ebXML 登记系统、向外界公布。

④系统 B 通过查询 ebXML 登记系统,发现符合自己商务需求的系统 A (通过其商务意向文件)。

⑤系统 B 与系统 A 利用各自的商务意向文件进行“谈判”,建立商务合作协议 (Business Agreement)。

⑥双方利用商务合作协议进行具体商务操作。为了有效实现上述机制, ebXML 定义了一系列协议,包括: ebXML Business Process Specification Schema (商务流程规则模式, BPS); ebXML Collaboration Protocol Profile (商务要约描述纲要, CPP); ebXML Collaboration Protocol Agreement

(商务合作协议, CPA); ebXML Catalog of Common Business Processes (通用商务过程目录, CBP 目录); ebXML Registry Service Specification (登记服务规则)。

(2) ebXML Registry 概况。ebXML Registry 在整个 ebXML 机制中起着核心的作用,支持分布的电子商务系统进行登记、发现和调用。它虽然不涉及数字对象描述元数据,但如果我们将关于通用商务过程 (CBP)、商务流程 (BPS)、商务要约 (CPP) 和商务协议 (CPA) 的描述也看成元数据的话, ebXML 登记系统同样是 MR, 而且由于它是基于 XML 进行自身定义、支持自动登记和检索,对于一般 MR 系统设计也有重要借鉴意义。

ebXML 登记系统所登记的对象包括: Core Components (包括 CBP 和基本商务数据元素)、商业文件格式、BPS、CPP、CPA 等。提供商业服务的商家可从登记系统检索有关的商务数据元素、商务过程、商业文件,利用这些工具构建自己的商务流程 (BPS) 和商务要约 (CPP),并将它们登记到登记系统;需要商业服务的商家可检索登记系统,发现合适的卖家,根据卖家的 CPP 制定合适的 CPA,通过 ebXML messaging service 传送给卖家,卖家接受建议的 CPA,双方可依据 CPA 配置各自的系统,从而建立电子商务互操作环境。

(3) ebXML 登记系统组成。整个体系有登记系统 (Registry) 和登记客户 (Registry Client) 组成,涉及的系统对象包括: Registration Authority (登记系统管理机构)、Registry Administrator (登记系统管理员)、Registered Users (注册用户,可被授权修改或删除特定内容)、Registry Guest (未注册用户,可能被允许阅读某些内容、但未授权进行修改或删除)、Submitting Organization (提交机构,提交被登记的内容)、Responsible Organization (责任机构,创建被登记内容)。注册用户和未注册用户都是登记客户。

登记系统和登记客户 (尤其是注册用户) 间也是遵循 ebXML 标准的电子商务对象,登记系统建立自己的系统描述 (Registry profile, 既登记系统的 CPP),在登记系统与登记客户间形成相应的应用协议 (Registration agreement, 既两者之间的 CPA)。登记系统 CPP 和 CPA 的内容包括 ebXML 登记信息模型 (ebRIM)、登记系统接口 (Registry interfaces)、登记客户接口 (Client interface)、ebXML 登记纲要 (Registry Profile) 等。

(4) ebXML 登记接口。ebXML 登记接口包括登记系统端的接口和登记客户端的接口。登记系统端的接口主要是: Life Cycle Management (LM, 生命周期管理) 接口, 提供管理登记对象的方法集合; Query Management (QM, 查询管理) 接口, 支持登记对象的查询。

生命周期管理接口支持对登记对象的全面管理, 包括登记对象的生命周期、登记对象属性、提交对象 (submitObject)、更新对象 (updateObject)、增加槽 (addSlot)、删除槽 (removeSlot)、批准对象 (approveObject)、否决对象 (deprecateObject)、删除对象 (removeObject) 等。注册用户可调用该接口的方法进行提交、修改、删除等操作。

查询管理服务接口支持多种查询方法, 包括浏览查询 (Browse and drill down query)、过滤查询 (FilterRequestQuery, 可利用各种条件对登记对象进行查询) 和 SQL 查询。登记客户可以通过该接口的方法进行查询。

登记客户接口支持登记系统与登记客户间的异步响应方法, 登记客户可请求生命周期管理接口及其相关方法进行登记、修改等操作, 也可请求查询管理接口及其相应方法进行查询。

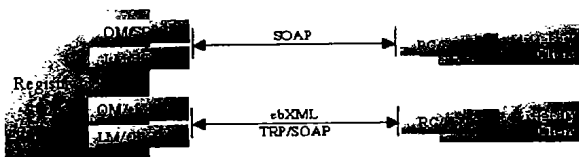


图 2

具体的接口调用被绑定在一定的通讯协议上, 例如基于 HTTP 的 SOAP 协议和 ebXML Messaging 消息服务协议 (ebMS)。一个登记系统可以实现一个或两个具体的捆绑。图 2 演示了 ebXML 登记的具体实现 (登记服务)。登记服务提供询问管理接口 (QM) 和生命周期管理接口 (LM) 对多重的协议绑定 (SOAP 和 ebMS) 是可用的。图的右半部分显示出 ebXML Registry 的 2 个不同的客户端。上面的客户端使用 SOAP 接口访问登记系统, 下面的客户端使用 ebMS 接口。客户端根据协议优先权使用登记服务中适当的接口。

(5) ebXML 登记系统的登记安全。为了保证登记内容和登记操作的安全, 一是基于数字签名进行注册用户的身份认证; 二是为注册用户建立使用控制政策 (AccessControlPolicy); 三是对传递的信息

负载进行加密。当登记一个对象时, 缺省 AccessControlPolicy 要求: 任何注册用户能发布内容; 任何登记用户不需要认证就可以访问内容; 内容所有者对由它创造的登记对象能使用所有的存取方法; 登记管理员对所有的登记对象可应用所有的存取方法。

### 3 ebXML 范例 GoXML Registry 介绍

GoXML Registry 是 Global XML 的登记模块, 主要支持跨 Web Services 和 ebXML Registry 的登记系统机制。它包括四个主要部分: Registry engine (登记引擎), 用于商务服务和过程、消息和词汇的分类、存储、管理, 基于 ebXML Registry 定义; Repository (存储库), 用于存储和检索 Registry Services 所需信息; Web-based Registry Client (基于 web 的登记客户端), 用于登记系统的结构配置、管理、内容管理、浏览和查询; A Registry Services API (登记服务 API), 对 Registry engine and ebXML Registry Services 提供程序化控制。

图 1 是 GoXML Registry 的结构图, 它的 Registry Client 和 Registry Server 由 Web Services 和 ebXML 接口相连。Registry Server 要素包括:

(1) Lifecycle Manager, 管理登记对象的整个生命周期, 从登记项的提交、分类、联系、驳回和迁移, Queue management, 队列管理, 允许跟踪提交内容的检查和通过程序, Version control, 版本控制, 帮助跟踪文档更新。

(2) User Manager, 管理注册用户登记、授权和认证。

(3) Query Manager, 由 ebXML Filter Queries 支持, 为相关对象间的浏览检索提供发现和组织功能。

(4) Repository Manager, 控制登记对象的物理存储和安全保护, 主要是通用的关系型数据库。

(5) Message Service, 该模型提供了 ebXML 消息的安全可靠传递。

(6) Administrator Workbench, 基于网络的管理机制, 作为 Registry Server 的客户, 通过启动客户端的有关工作方式进行管理。

(7) Registry Service Interface, 支持三种登记系统接口类型 (Web Services、ebXML、JAXR), 支持在 .NET 和 Java 环境下的 B2B (下转第 612 页)

个巨大的分布式信息空间。但由于因特网本身的开放性、动态性与异构性,使得用户很难准确、快捷地从 WWW 上获取有价值的数字化信息资源。研究和创建数字化期刊的元数据标准、规范和格式,有助于促进以因特网为基础的数字化期刊资源仓储的透明、公开、可扩展、互操作性、组织性和规模化建设,使分布式数字化期刊仓储组成一个单一的、虚拟的、有组织的信息集合体,实现全球期刊信息的资源共享和跨仓储无缝查找。DC 在数字化文献资源建设中发挥出的这种巨大作用,无疑会给全球信息化建设带来一场革命。作者希望我国图书情报界和计算机网络界的有关组织与专家,共同加紧推动这项重要工作。

### 参考文献

- 1 <http://ifla.inist.fr/II/metadata.htm#dc>.
- 2 马珉. 元数据——组织网上信息资源的基本格. 情报科学, 2002, 20 (4): 377~378
- 3 Susan Feldman. Jesswica Milstead. Metadata Projects & Standards. Online, Vol. 34, Issue 1, 1999
- 4 Christine McGeever. Multivendor Group Approves Metadata XML Standard. Computerworld, Vol. 34, Issue 26, 2000.
- 5 张健挺. 地理信息网络资源共享的研究和进展. 地理科学进展, 1998, 17 (4): 73~77
- 6 庄育飞, 郑卫. Dublin Core: 网络资源组织与整理的新思路. 情报学报, 2000, 19 (2): 164~169
- 7 Jane Hunter. MetaNet: A Metadata Term Thesaurus to Enable Semantic Interoperability Between Metadata Domains. Journal of Digital Information. Vol. 1, Issue 8, 2001
- 8 Liu Tingyuan. Metadata Standard of Digitized Journals. The Proceedings of Digital Library — IT Opportunities and Challenges in the New Millennium. Editor — in — Chief: Sun Jiazheng. Beijing Library Press. July, 2002
- 9 王晔, 等. 基于元数据的 Web 信息检索. 情报学报, 2001, 20 (3): 309~316
- 10 Preben Hansen. User Guidelines for DC Metadata Creation. [http://www.sics.se/~preben/DC/DC\\_guide.html](http://www.sics.se/~preben/DC/DC_guide.html).
- 11 赵亮. 元数据在上海数字图书馆的应用. 中文元数据应用国际研讨会文集. 上海, 2001
- 12 胡海帆, 等. 北京大学古籍数字图书馆拓片元数据标准的设计及结构. 中文元数据应用国际研讨会文集. 上海, 2001
- 13 莫少强. DC 元数据的中文研究及其在超星数字图书馆中的应用. 中文元数据应用国际研讨会文集. 上海, 2001 年

(责任编辑: 徐波)

(上接第 567 页) 和 application-to-application 的应用集成。Windows98/NT/2000/ME/XP、Solaris、Linux 都宣称支持 GoXML Registry。

总之, 我国数字图书馆领域的元数据规范建设正在起步, 在为数不多的元数据规范研究与应用活动中, 定义信息往往没有公开发布、也不能够被公开获取。支持元数据格式的语义定义、编码体系、应用协议和实施指南等往往没有明确建立, 更谈不上公开公布, 使得元数据规范难以通过公开、可靠和标准的方法来识别、确认、验证, 造成已有元数据规范没有得到广泛的利用。随着我国数字图书馆建设的全面推进, 各个系统所引进、建立、修改或扩展的元数据规范必将迅速增加, 发现和识别其他系统元数据规范的需要也必将迅速增加。迫切需要建立一个开放、公共和可靠的元数据规范登记系统, 提高元数据规范共享的程度。

### 参考文献

- 1 张晓林. 元数据研究与应用. 北京图书馆出版社, 2002 (5)
- 2 Blanchi, C. and Petrone, J. Distributed Interoperable Metadata Registry. D-Lib Magazine, V. 7 (12), Dec., 2001 <http://www.dlib.org/dlib/december01/blanchi/12blanchi.html>
- 3 梁娜, 张晓林. 基于人工登记与检索的元数据登记系统. 大学图书馆学报, 2003 (待发)
- 4 UDDI Specifications. <http://www.uddi.org/specification.html>
- 5 ebXML Registry Services Specification v2.0, Accessed July 22, 2002. <http://www.oasis-open.org/committees/repreg/documents/2.0/specs/ebrs.pdf>
- 6 ebXML Business Process Specification Schema (BPS). <http://www.ebxml.org/specs/ebBPSS.pdf>
- 7 ebXML Collaboration Protocol Profile and Agreement. <http://www.ebxml.org/specs/ebCCP.pdf>
- 8 ebXML Catalog of Common Business Processes. <http://www.ebxml.org/specs/bpPROC.pdf>
- 9 ebXML Registry Service Specification. V2. <http://www.ebxml.org/specs/ebrs2.pdf>
- 10 GoXML Registry. Accessed July 22, 2002. <http://www.xmlglobal.com/prod/registry/>

(责任编辑: 赵立军)