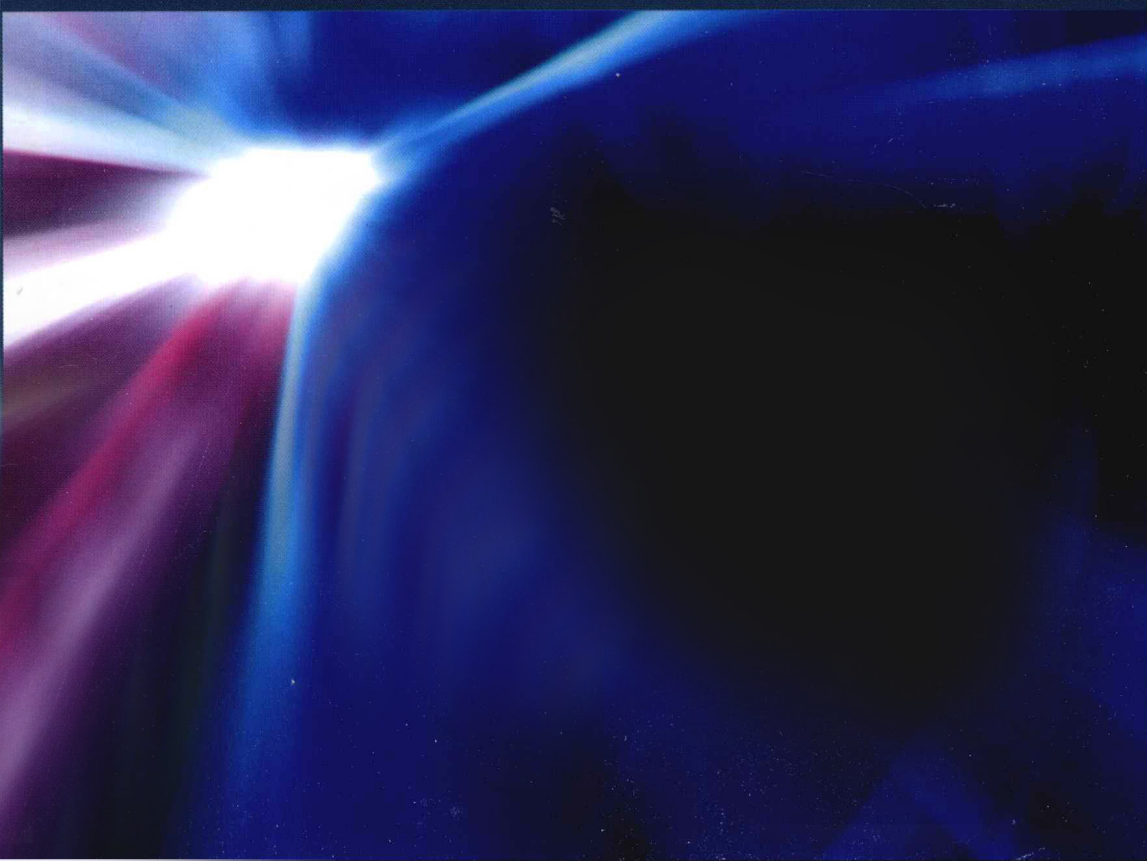


国家数字图书馆工程标准规范成果

国家图书馆管理元数据 规范和应用指南

郑巧英 王绍平 汪东波 主编



國家圖書館出版社

国家图书馆管理元数据 规范和应用指南

郑巧英 王绍平 汪东波 主编

國家圖書館出版社

图书在版编目(CIP)数据

国家图书馆管理元数据规范和应用指南/郑巧英,王绍平,汪东波主编. —北京:国家图书馆出版社,2010.10

(国家数字图书馆工程标准规范成果)

ISBN 978-7-5013-4382-9

I. ①国… II. ①郑… ②王… ③汪… III. ①元数据—信息管理—规范 IV. ①G203

中国版本图书馆 CIP 数据核字 (2010) 第 119525 号

责任编辑:金丽萍 王涛 高爽

书名 国家图书馆管理元数据规范和应用指南

著者 郑巧英 王绍平 汪东波 主编

出版 国家图书馆出版社(原北京图书馆出版社)

(100034 北京市西城区文津街7号)

发行 010-66139745 66151313 66175620 66126153

66174391(传真) 66126156(门市部)

E-mail btsfxb@nlc.gov.cn(邮购)

Website www.nlcpress.com→投稿中心

经销 新华书店

印刷 河北三河弘翰印务有限公司

开本 787×1092(毫米) 1/16

印张 14

版次 2010年10月第1版 2010年10月第1次印刷

字数 180(千字)

书号 ISBN 978-7-5013-4382-9

定价 58.00元

丛书编委会

主 编：国家图书馆

编委会：

主 任：周和平

执行副主任：詹福瑞

副主任：陈 力 魏大威

成 员（按姓氏拼音排名）：卜书庆 贺 燕 蒋宇弘

梁蕙玮 龙 伟 吕淑萍 申晓娟 苏品红

汪东波 王文玲 王 洋 杨东波 翟喜奎

赵 悦 周 晨

本书编委会

主 编：郑巧英 王绍平 汪东波

编 委：陈幼华 彭 佳 汤莉华 兰小媛 徐 璟 孙 华
曲建峰 申晓娟 吕淑萍 周 晨 龙 伟 贺 燕
李志尧 赵 悦 王文玲

总 序

数字图书馆涵盖多个分布式、超大规模、可互操作的异构多媒体资源库群,面向社会公众提供全方位的知识服务。它既是知识网络,又是知识中心,同时也是一套完整的知识定位系统,并将成为未来社会公共信息的中心和枢纽。数字图书馆建设的最终目标是实现对人类知识的普遍存取,使任何群体、任何个人都能与人类知识宝库近在咫尺,随时随地从中受益,从而最终消除人们在信息获取方面的不平等。“国家图书馆二期工程暨国家数字图书馆工程”是国家“十五”重点文化建设项目,由国家图书馆主持建设,其中国家数字图书馆工程的建设内容主要包括硬件基础平台、数字图书馆应用系统和数字图书馆标准规范体系。

标准规范作为数字图书馆建设的基础,是开发利用与共建共享资源的基本保障,是保证数字图书馆的资源和服务在整个数字信息环境中可利用、可互操作和可持续发展的基础。因此,在数字图书馆建设中,应坚持标准规范建设先行的原则。国家数字图书馆标准规范体系建设围绕数字资源生命周期为主线进行构建,涉及数字图书馆建设过程中所需要的主要标准,涵盖数字内容创建、数字对象描述、数字资源组织管理、数字资源服务、数字资源长期保存五个环节,共计三十余项标准。

在国家数字图书馆标准规范建设中,国家图书馆本着合作、开放、共建的原则,引入有相关标准研制及实施经验的文献信息机构、科研机构以及企业单位承担标准规范的研制工作,这就使得国家数字图书馆标准规范的研制能够充分依托国家图书馆及各研制单位数字图书馆建设的实践与研究,使国家数字图书馆的标准规范成果具有广泛的开放性与适用性。本次出版的系列成果均经过国家图书馆验收、网上公开质询以及业界专家验收等多个验收环节,确保了标准规范成果的科学性及实用性。

目前,国内数字图书馆标准规范尚处于研究与探索性应用阶段,国家图书馆担负的职责与任务决定了我们在数字图书馆标准规范建设方面具有的责任。此次将

国家数字图书馆工程标准规范研制成果付梓出版,将为其他图书馆、数字图书馆建设及相关行业数字资源建设与服务提供建设规范依据,对于推广国家数字图书馆建设成果,提高我国数字图书馆建设标准化水平,促进数字资源与服务的共建共享具有重要意义。

国家图书馆馆长 周和平
2010 年 8 月

前 言

利用元数据对信息资源,尤其是数字资源进行开发、利用、管理,已经成为信息资源建设的基本手段之一。元数据最初用于描述信息资源的外部与内容特征,以便对信息资源进行有效的检索、定位或书目控制。这一类元数据现在一般称为描述元数据,如 MARC、DC(Dublin Core)、MODS 等。遵循标准化、规范化、开放性的原则开发及应用描述元数据,增进了信息系统的互操作性与开放性,通过共享元数据,进而共享元数据所描述的信息资源,促进了信息资源的无障碍交流。

元数据并不局限于描述信息资源的外部与内容特征。人们在信息资源开发、利用、管理的不同阶段与不同层次,创建了不同功能与作用的元数据,可以从不同的角度对这些不同功能与作用的元数据加以归类。在图书情报领域通常把元数据分为描述元数据、结构元数据与管理元数据。描述元数据上文已作了介绍。结构元数据主要确定元数据的语法结构或者信息单元之间的逻辑联系,比如 HTML、XML 等置标语言以及 METS 模块中的结构元数据。管理元数据是专门对信息资源实施管理的元数据。它描述信息资源管理过程中所涉及的管理信息以及由管理信息进一步揭示的管理政策与管理机制,比如与信息资源的评估、选择、使用权限、加工等有关联的信息与相关机制。管理元数据有助于对信息资源实行自动化与网络化的科学管理,有利于信息资源管理信息的跨系统交换与共享。

总体上说,已建设的元数据主要以描述元数据为主,缺少对管理信息的描述,不能满足有效管理以数字资源为主的信息资源(正文中简称“信息资源”)的需要。

2008 年,上海交通大学图书馆受国家图书馆委托,承担了国家数字图书馆工程标准规范项目“国家图书馆管理元数据规范”的研制,以《国家图书馆核心元数据标准》《国家图书馆元数据应用规范》为指导,完成了《国家图书馆管理元数据规范》及其应用指南的编制。本书即是以“国家图书馆管理元数据规范”项目的研究成果为基础并参考了“信息资源基础管理性元数据框架研究”课题的相关研制成果编写的。

“国家图书馆管理元数据规范”是为服务国家图书馆数字资源建设,根据国家图书馆管理元数据规范招标指南、研制需求书和成交合同研制的。因此本书主要针对国家图书馆的实际情况,兼顾在此基础上发展中国管理元数据规范的需要。本规范可直接或经修订后供其他图书馆或信息服务机构利用,但其修订必须遵循本规范的基本原则与扩展原则。

目 录

前 言	(1)
国家图书馆管理元数据规范	(1)
1 范围	(3)
2 规范性引用文件	(3)
3 术语和定义	(4)
4 管理元数据框架的构建	(5)
5 作用	(7)
6 原则	(7)
7 数据模型	(8)
8 管理元数据的语义	(10)
9 管理元数据的描述语言与语法结构	(56)
附录 A(规范性附录) 扩展的修饰词	(58)
附录 B(资料性附录) 元素与修饰词列表	(78)
附录 C(资料性附录) “国家图书馆管理元数据规范”RDFS 编码——通用元素篇	(83)
附录 D(资料性附录) “国家图书馆管理元数据规范”RDFS 编码——专用元素篇	(119)
国家图书馆管理元数据规范应用指南	(125)
1 总则	(127)
2 细则	(130)
附录 1(规范性附录) 规范档	(193)
附录 2(资料性附录) 信息资源结构的表达	(200)
附录 3(资料性附录) 相关修饰词的集中排序	(203)
附录 4(资料性附录) 通用元数据方案的链接或嵌入	(204)
参考文献	(206)
后 记	(207)

国家图书馆管理元数据规范

1 范围

本规范阐述了管理元数据的基本概念和作用,从信息资源生命周期的角度,提出了管理元数据框架的设计原则与应用扩展原则,构建了信息资源管理元数据框架及其数据模型,制订了元数据框架的数据词典。

鉴于保存元数据已经发展成一个相对独立的元数据研究领域,本规范没有把与信息资源的保存相关的保存元数据纳入管理元数据规范建立的信息资源管理元数据框架之内。如有必要在信息资源管理的整体架构中建立长期保存机制,则可采用已有的、国际通用的长期保存元数据方案。为了使管理元数据框架与保存元数据之间建立有机的联系,在本框架的相应模块设立有特定的元素^①,作为连接保存元数据的接口,供用户选用。

本规范构建的管理元数据框架、数据模型以及制订的数据词典,为国家图书馆对其信息资源建设所涵盖的采集、加工、描述、管理、服务和保存整个信息资源建设周期及其多种来源、多种类型、多种内容与多种服务方式的复杂多样的信息资源实施有效的管理,提供统一、规范且可扩展的基础性方案,着重对信息服务机构内信息资源的采集、加工、服务等基础业务环节的管理元数据作出规定,指示对这些基础业务的管理机制与管理政策进行描述的一般性方法。

2 规范性引用文件

下列文件中的条款通过本规范的引用而成为本规范的条款。凡是注明日期的引用文件,随后所有的修正案或修订版均不适用于本规范,然而鼓励根据本规范达成协议的各方研究是否可使用这些文件的最新版本;凡是不注日期的引用文件,其最新版本适用于本规范。

GB/T 7408—2005 数据交换和交换格式 信息交换 日期和时间表示法(idt ISO 8601:2000)

ISO 4217:2008 表示货币和资金的代码

DCMI 资源类型表(推荐稿),2006年8月28日 <http://dublincore.org/documents/dcmi-type-vocabulary/>

因特网媒体类型(IMT) <http://www.iana.org/assignments/media-types/>

统一资源标识符(URI):通用句法 <http://www.ietf.org/rfc/rfc3986.txt>

数字对象标识符(DOI) <http://www.doi.org/topics/je-mh-doi-030970.pdf>

DCMI 元数据术语集 <http://dublincore.org/documents/dcmi-terms/>

元数据编码及传输标准(METS) <http://www.loc.gov/standards/mets/version17/mets.v1-7.xsd>

XML 访问控制语言(XACL) <http://www.trl.ibm.com/projects/xml/xacl/xacl-spec.html>

个性化信息描述语言(PIDL) <http://www.w3.org/TR/NOTE-PIDL>

开放数字权利语言(ODRL) <http://odrl.net/1.1/ODRL-11.pdf>

互联网资源内容选择平台(PICS) <http://www.w3.org/PICS/>

^① 具体内容参见“8.2 数据词典”。

中国机读目录格式(CNMARC) 国家图书馆编. 新版中国机读目录格式使用手册. 2004
MARC21 <http://www.loc.gov/marc/bibliographic/lite/>
通用机读目录格式(UNIMARC) <http://www.ifa.org/VL/3/p1996-1/sec-umi.htm>

3 术语和定义

下列术语和定义适用于本规范。

3.1 管理元数据(Administrative Metadata)

描述信息资源管理过程中所涉及的管理信息以及由管理信息进一步揭示的管理政策与管理机制的元数据。譬如描述与信息资源的评估、选择、使用权限、加工、利用等有关联的信息与相关机制的元数据。

3.2 信息资源生命周期(Lifecycle of Information Resource)

标记信息资源发展和使用的一系列事件。在信息服务机构(如图书馆、信息研究所、档案馆)内部主要包括信息资源的采集、加工、服务等。

3.3 元素(Element)

管理元数据的基本语义单位,描述管理元数据框架内的基本实体。

3.4 通用元素(Universal Element)

在管理元数据框架内适用于采集、加工、服务等各个模块的元素。

3.5 专用元素(Special Element)

在管理元数据框架内只适用于特定模块的元素。

3.6 本地元素(Local Element)

在管理元数据框架内,特定的应用系统为适应特定的应用环境而扩充的元素。由使用者自行定义。

3.7 修饰词(Refinement)

对管理元数据语义单位的细化。是某一语义单位的属性的描述。元素和修饰词本身都可细化而拥有其修饰词。编码体系是一类特殊的修饰词,它说明元素或修饰词的取值或编码所遵循的规范体系。

3.8 通用修饰词(Universal Refinement)

在管理元数据框架内原则上可限定任何元素及其修饰词的修饰词。

3.9 扩展的修饰词 (Extended Refinement)

在管理元数据框架内推荐使用的可对特定元素或修饰词加以限定的补充修饰词。

4 管理元数据框架的构建

4.1 信息资源生命周期

本规范以信息资源生命周期作为调查、分析管理元数据的基本需求,建立数据模型,制定数据词典的基础。

从信息服务机构内信息资源管理的总体来看,信息资源经历了创建、传递、转换、利用的运动过程。图 4-1 是这一过程的流程图。

信息资源创建后或经过信息服务机构^①传递到最终用户,或直接由用户消费。就数字信息资源而言,并不排除最终用户直接利用的可能性,但在大多数情况下,都需经过信息服务的环节。这里的信息资源可以包括数字信息资源与非数字信息资源。信息服务机构一般包括政府信息部门、图书情报与档案服务部门和其他信息供应商,它们从事不同深度与广度的信息采集、加工与提供利用的服务工作。从信息资源在信息服务机构内部的运动过程来看,信息资源生命周期主要包括信息资源的加工、采集和服务等基础环节。如果信息资源直接被最终用户利用,那么它们之间只存在信息采集关系;如果最终用户通过信息服务的中介获取信息资源,那么信息资源须经过信息的采集、加工、提供利用等复杂的服务环节。

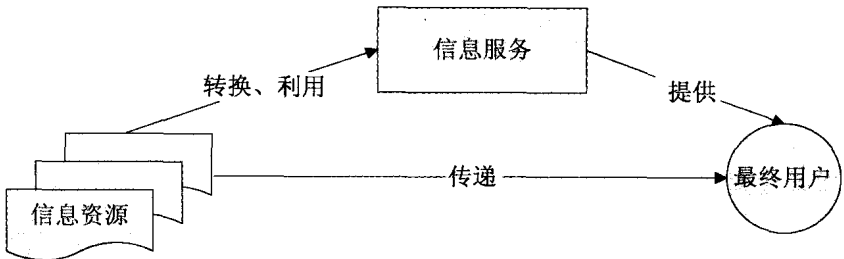


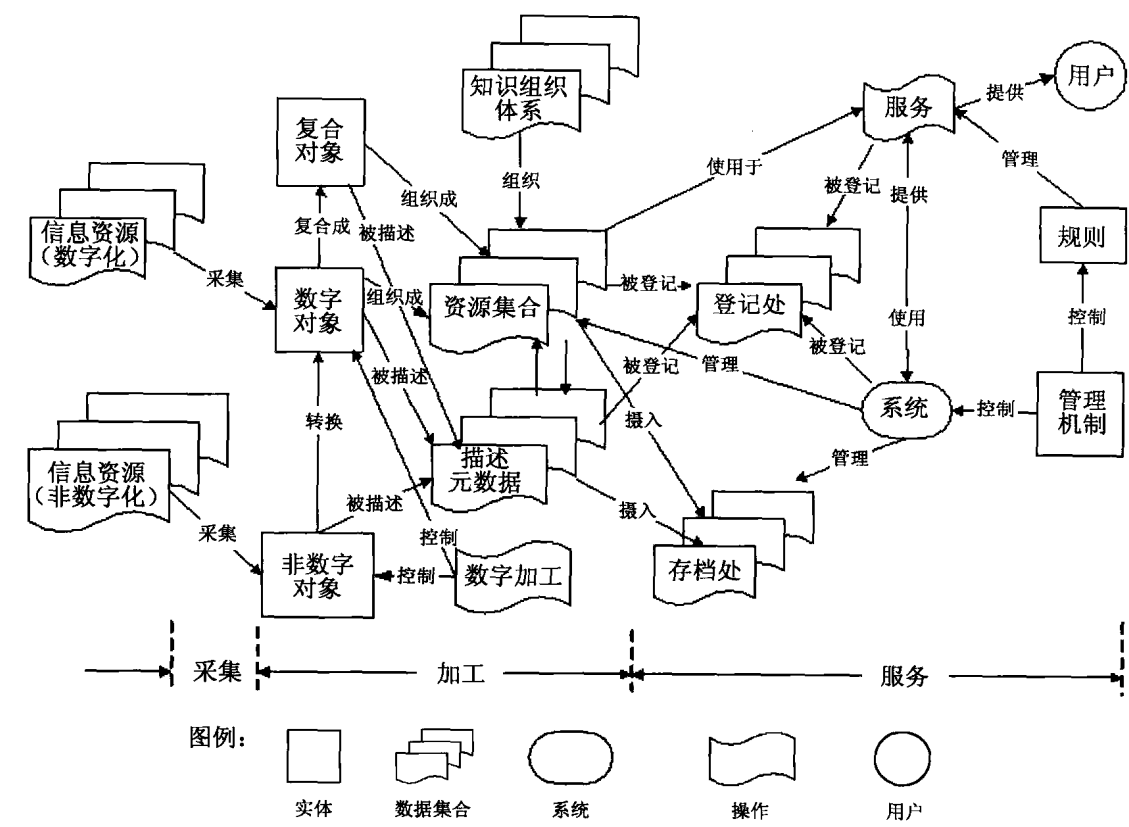
图 4-1 信息资源创建、传递、转换、利用流程图

借助“信息资源生命周期”模型,可对信息资源生命周期,尤其是信息在信息服务机构的运动过程作以下简要的剖析(见图 4-2)。

- a) 信息资源经采集后聚集为数字对象与非数字对象。
- b) 非数字对象可加工转换成数字对象。数字对象可复合成复合对象;数字对象与复合对象可进一步通过知识组织体系组织成资源集合。无论数字对象还是非数字对象,无论信息内容还是信息对象、资源集合都可由描述元数据进行描述。
- c) 资源集合(个体资源对象的集合)以及描述信息资源的元数据可以被登记、存档,并通过

^① 本规范的管理元数据框架主要以国家图书馆为应用对象,但不排除其他信息服务机构采纳本框架的可能性。详见“第二部分 国家图书馆管理元数据规范应用指南”。

服务系统提供给用户。



4.2 管理元数据框架

本规范利用基础管理元数据,构建了开放的、可扩展的、模块化的管理元数据集,形成了一个贯穿信息资源生命周期中基础业务环节的管理元数据框架。

在“信息资源生命周期”模型中,信息资源从创建到被用户利用,在信息服务机构内经历了采集、加工、服务等基础环节。除了图 4-2 中列出的服务环节的管理过程外,采集、加工等环节同样受到管理过程的控制。本规范从信息资源生命周期的角度观察信息资源管理,划分出上述基础管理环节,每个环节有其相应的管理机制与管理政策,都可以用元数据进行描述,这些元数据称为基础管理元数据。

就信息服务机构而言,信息资源生命周期中采集、加工、服务等基础环节将信息资源生命周期划分成较小的阶段,一方面缩小了研究对象的范围,另一方面增加了一定范围内不同对象之间的共同点,从而形成相对独立的、具有某些共性的模块。基础管理元数据是某一特定管理环节的基本元数据,它既反映了这一管理环节管理机制的共性,又能根据更加具体的管理功能进行扩展。在开放的、可扩展的基础管理元数据逻辑框架下,各个专门的管理元数据不再是孤立的、互不相关的,而是可以相互融合的有机整体。

5 作用

管理元数据对信息资源,尤其是数字信息资源的管理政策与管理机制进行描述,是元数据在信息资源管理层面的拓展。管理元数据对信息资源采集、加工、利用过程的管理信息加以规范化、开放性的揭示,有助于用户或第三方系统方便地、无障碍地共享本系统的信息资源,有助于各个信息系统在安全机制保障下,通过管理元数据,有效地交流管理政策与管理机制,促进不同信息系统之间的互操作性。信息系统的信息资源及其集合可以通过管理元数据与具体系统的绑定来形成实际的表达机制,也可以通过修改管理元数据方便地定制、更改表达机制,还可以支持智能代理对信息资源及其集合进行自动的搜索与匹配,支持智能代理解析信息资源结构,并在此基础上集成多个信息资源集合。信息系统的管理机制与政策可以通过管理元数据与具体系统的绑定来形成实际的运行机制,也可以通过修改管理元数据方便地定制、更改管理机制与政策,还可以支持智能代理对管理政策和规则进行自动的搜索、解析与相应的处理。

管理元数据框架提供了基于信息资源生命周期的信息资源采集、加工、服务三个基本模块,应用系统可以根据实际需求,采用其中的一个或几个模块,并可以按照管理元数据框架的扩展机制,建立更为细化的、适应实际需求的管理元数据应用体系。

在数字图书馆中,可以利用管理元数据框架构建信息资源的管理平台,该平台根据应用环境划分成若干应用平台,譬如,基于信息资源业务管理流程的信息资源采集管理平台、信息资源加工管理平台、信息资源服务管理平台;基于信息资源类型的网络资源管理平台。每个平台下还可以建立若干子平台,如信息资源加工管理平台下的元数据管理平台、数字化加工管理平台等。

6 原则

6.1 基本原则

管理元数据框架的设计遵循了元数据标准规范的模块化、开放性、互操作性原则。

(1) 模块化原则

管理元数据本身是数字图书馆标准规范体系中的一个逻辑模块,对数字图书馆中信息资源的采集、加工、服务实施管理,同时依据信息资源在数字图书馆中的流程进一步划分成采集、加工、服务三个相对独立,而又密切联系的子模块。每个子模块的元素分成三类:通用元素、专用元素、本地元素。通用元素是三个子模块共有的元素,如代理(Agent)(包括用户和工作人员等)、信息资源、权限等;专用元素是特定的子模块所特有的,如信息采集模块的“采集来源”、信息服务模块的“服务请求”。本地元素是特定的应用系统为适应特定的应用环境而扩充的元素,这些元素只适于本地系统的需求,并不参与不同系统之间管理信息的交换。

(2) 开放性原则

管理元数据框架是开放的,它尽可能复用或嵌套那些被广泛接受和应用的标准规范。比如,元素“代理(Agent)”的属性基本复用 vCard 元素;“信息资源(InfoResource)”的描述可直接嵌套或链接 DC 或 MARC;数字信息资源的鉴定可嵌套 PICS 资源评鉴元数据。它实施开放扩展机制,应用单位可根据需要对原框架内的元素实行纵向与横向的扩展。所谓纵向扩展,是对已有元素或修饰词的“细化”,即根据特定的应用环境,对已有元素或修饰词的语义加以限定,派生出