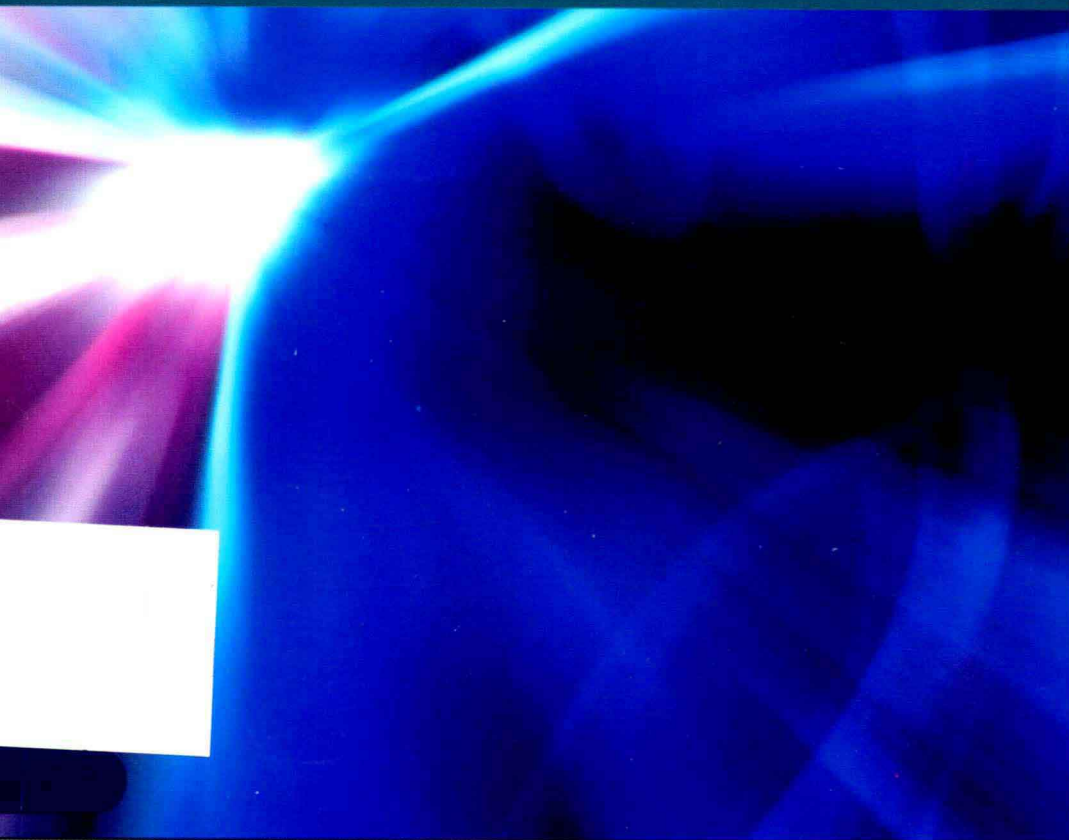


国家数字图书馆工程标准规范成果

国家图书馆网络资源 元数据规范和著录规则

赵亮 周晨 主编



國家圖書館出版社
National Library of China Publishing House

国家数字图书馆工程标准规范成果

国家图书馆网络资源 元数据规范和著录规则

赵 亮 周 晨 主编



 國家圖書館出版社

图书在版编目(CIP)数据

国家图书馆网络资源元数据规范和著录规则/赵亮,周晨主编. —北京:国家图书馆出版社, 2013.7

(国家数字图书馆工程标准规范成果)

ISBN 978-7-5013-5050-6

I. ①国… II. ①赵… ②周… III. ①数字图书馆—书目数据—数据管理—规范—中国
②数字图书馆—书目数据—著录规则—中国 IV. ①G250.76-65

中国版本图书馆 CIP 数据核字 (2013) 第 070215 号

责任编辑:高爽

书名 国家图书馆网络资源元数据规范和著录规则

著者 赵 亮 周 晨 主编

出版 国家图书馆出版社(100034 北京市西城区文津街7号)

(原北京图书馆出版社)

发行 010-66114536 66126153 66151313 66175620

66121706(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 北京科信印刷有限公司

开本 787×1092(毫米) 1/16

印张 5.25

版次 2013年7月第1版 2013年7月第1次印刷

字数 70(千字)

书号 ISBN 978-7-5013-5050-6

定价 58.00 元

《国家数字图书馆工程标准规范成果》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：赵 亮 周 晨

编 写：刘 炜 张春景 李春明 刘 峥
赵 悦 梁蕙玮 张文静 马宁宁

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

网络资源是指能通过网络公开访问,具有网络标识,并且品质较为稳定,有一定的权威性、时效性和持久性的信息资源。在这个网络信息资源大爆炸的时代,图书馆作为人类文化的保存者,不得不考虑如何对网络资源进行保存和管理。为了满足新环境下人们对网络信息资源描述、组织和利用的需求,元数据应运而生。

随着元数据应用的不断深入,元数据应用和实践需要更具操作性和指导性的文件来规范。恰逢国家数字图书馆工程建设大规模开展,各类数字对象元数据规范的研制,成为国家数字图书馆工程建设要解决的关键问题之一。2009年9月,上海图书馆受国家图书馆委托,承担“国家数字图书馆工程专门元数据标准与著录规范——网络资源”项目的研制。

“国家数字图书馆工程专门元数据标准与著录规范——网络资源”项目的主要任务是根据国家图书馆对于网络资源描述的具体需求(资源导航、永久保存等),厘清网络资源的基本概念、其所涉及的实体类型、各类实体之间的关系、实体的属性特征以及这些属性特征之间的关系、属性的值域和范围等,并以《国家图书馆元数据应用规范》和《国家图书馆专门元数据设计规范》为指导,完成相应的调研报告、元数据规范、著录规则和样例的编制。本书即是以该项目的研究成果为基础编写而成。

项目研究成果主要参考科技部科技基础性工作专项资金重大项目的研究成果网络资源元数据规范,还详细调研 DC、CALIS 重点学科网络资源导航库元数据规范、中文元数据方案及 MODS 等标准和项目。项目的研究成果主要表现在以下几个方面:

- 1) 在分析网络资源特性的基础上,对网络资源进行范围界定和定义;
- 2) 定义了网络资源元数据元素集,并对元素及修饰词进行定义与描述;
- 3) 在分析著录对象之间关系的基础上,明确了网络资源的著录单位;研制了网络资源元数据著录规则。

“国家数字图书馆工程专门元数据标准与著录规范——网络资源”项目,根据国家图书馆网络资源元数据规范招标指南、技术需求书和合同要求,为国家图书馆数字资源建设服务。因

此,本书内容主要针对国家图书馆的实际情况。本规范也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。

编者

2012 年 12 月 20 日

目 录

前 言	1
第一部分 国家图书馆网络资源元数据规范	1
1 范围	3
2 规范性引用文件	3
3 术语和定义	3
4 网络资源著录对象的分析	5
5 元数据规范内容结构	6
6 元数据规范术语定义属性	9
7 核心元素及其修饰词定义	10
8 资源类型核心元素及其修饰词定义	35
9 个别元素及其修饰词定义	36
第二部分 国家图书馆网络资源元数据著录规则	37
1 范围	39
2 规范性引用文件	39
3 著录总则	40
4 著录细则	43
附录 A MIME 资源数据格式分类表	63
附录 B DCMI 类型词表	66
参考文献	70
后 记	71

第一部分 国家图书馆网络资源元数据规范

1 范围

本规范中的网络指的是国际互联网(Internet),也称因特网。网络资源在广义上指通过国际互联网可以获得的一切软件、硬件资源(包括路由器、打印机等),规范中所涉及的“网络资源”指的是不包括硬件资源的一切信息资源。

本规范旨在厘清网络资源的基本概念、所涉及的实体类型、各类实体之间的关系、实体的属性特征以及这些属性特征之间的关系、属性的值域和范围等,从而建立一套网络资源的元数据规范。

本规范不对网络资源的管理信息、技术信息等做专门的模块规定。若有需求,可应用国家数字图书馆标准规范建设项目中的管理元数据或保存元数据规范所规定的元素;同时,本规范亦不规定与特定系统相关的网络资源元数据集的分析、设计与实现技术。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改版)适用于本文件。

GB/T 25100—2010 信息与文献 都柏林核心元数据集

DCMI Metadata Terms. [DCMI-TERMS]

DCMI 元数据术语集 [DCMI-TERMS]

< <http://dublincore.org/documents/dcmi-terms/> >

DCMI Namespace Policy. [DCMI-NAMESPACE]

DCMI 命名域政策 [DCMI-NAMESPACE]

< <http://dublincore.org/documents/dcmi-namespace/> >

Metadata Object Description Schema. [MODS]

元数据对象描述模式 [MODS]

< <http://www.loc.gov/standards/mods/> >

3 术语和定义

3.1 网络资源 web resource

能通过网络公开访问,具有网络标识,并且品质较为稳定,有一定的权威性、时效性和持久

性的信息资源。

3.2 DCMI-都柏林核心元数据计划 DCMI- Dublin Core Metadata Initiative

都柏林核心元数据元素集的维护机构。

3.3 元数据 metadata

关于信息资源或数据的一种结构化的数据。

3.4 元素 element

元数据集合中用于定义和描述数据的基本单元,由一组属性描述、定义、标识,并允许值限定。

3.5 修饰词 qualifier

当元素无法满足对资源对象精确描述的需要时进一步扩展出的术语。修饰词包括两种类型:元素修饰词和编码体系修饰词。

3.6 元素修饰词 element refinement

对元素的语义进行修饰,提高元素的专指性和精确性。

3.7 编码体系修饰词 encoding scheme

用来帮助解析某个术语值的上下文信息或解析规则。其形式包括受控词表、规范表或者解析规则。

3.8 核心元素 core element

使用频率高的、共性的可用于不同类型的信息资源描述的元数据元素。

3.9 资源类型核心元素 resource core element

在制订不同类型和不同资源的元数据规范时,根据资源对象的特点设计出的、相似资源共同所需的元素及修饰词。

3.10 个别元素 unique element

为某一特定的资源对象设计的,仅适用于这类对象的元素,不用于交换。

3.11 复用 reuse

在元数据应用过程中,对于其他元数据标准中已经有明确定义并适用于本应用领域的元

素的直接使用,并在使用时明确标明其地址。

4 网络资源著录对象的分析

4.1 范围

原则上,一切网络上可以访问的、具有网络标识(例如 URL、URI、DOI、PURL 等)的资源对象均可以作为著录对象。本规范参考了 OCLC Connexion、ROADS 和 CALIS 的网络学科资源导航元数据规范,并结合国家图书馆的实际应用,将网络资源的著录对象限定为:

- (1)有 URL 的网络资源,包括需要浏览器特定插件或者其他程序支持方可显示的资源,如 PDF 文档、WORD 文档、PPT 文档、各种多媒体文档等;
- (2)在互联网上可公开访问的资源;
- (3)各种学术机构的机构库资源。

4.2 网络资源的著录单位

对网络资源著录单位进行处理时,有以下四种情况:

- (1)以网站为一个著录单位;
- (2)以网站中某个频道或栏目为一个著录单位;
- (3)以具体的网页为一个著录单位;
- (4)以具体的文件为一个著录单位。

因此在具体应用时,需根据应用需求界定明晰著录的粒度,比如是著录到网站层级、频道层级还是网页层级。

4.3 网络资源著录对象之间的关系

网络资源著录对象之间的关系是网状的,层层相扣,互相可以链接与跳转。超文本与超媒体链接的特点在于这种引用和链接是直接可以跳转进而直接获得资源。

一方面,上述的链接和跳转的关系并不一定等于著录对象之间内容上的关联,即网页之间形式上的超链接并不完全代表页面内容的关联,很多时候链接和跳转的资源在内容上并无联系;因此著录的侧重点在于揭示意义上、内容上的联系。

另一方面,链接虽然不能完全体现内容上的互联,但也可以体现一部分,如果到了语义网络,则链接能更好地实现页面内容上的关联。目前,资源 A 与资源 B、资源 C 等之间的链接与跳转有可能在内容上存在如下的关联:资源的不同版本之间,互相参照的网络资源之间,完整的内容分成几个部分的页面来显示,文本介绍与对应的音视频之间,同一作者的不同作品之间,等等。

鉴于上述分析,在对网络资源进行处理时,每个著录对象都当成是独立的对象处理,对象与对象之间可以链接(单向或多向链接均可),无数个对象与对象之间互相链接体现了整体的网状关系,这些关系可通过元数据语义得到揭示。

本规范不要求著录时对网络资源进行十分严格的查重,简单的查重以资源标识符、外加资源名称为标准。也就是说,当资源标识符与资源名称均相同时,才把两个资源认为是同一资源。

5 元数据规范内容结构

国家图书馆网络资源元数据规范包括核心元素 15 个,分别是名称、创建者、主题、描述、出版者、其他责任者、日期、类型、格式、标识符、来源、语种、关联、时空范围、权限。这 15 个元素全部复用自 DC;资源类型核心元素 3 个,分别是版本、馆藏信息、受众;个别元素 1 个,采集地址。其扩展原则为:

- (1)新增加的元素和修饰词须优先采用 DCMI 中采用的元素和修饰词,或者是现有的其他元数据标准中的元素和修饰词,例如 EAD、CDWA 和 PANDORA 等。
- (2)新增元素如果复用来自其他元数据标准的元素或修饰词,要求必须说明来源,使用时严格遵守其语义。
- (3)在现有元数据规范中,如果没有十分恰当的元素可供复用,允许自行扩展元素。
- (4)自行扩展的元素不能和已有的元素有任何语义上的重复。
- (5)扩展的修饰词必须遵循向上兼容原则,即修饰词在语义上不能超出被修饰词(元素)的语义。

表 1-1 网络资源元数据规范元素及修饰词列表

元素	元素修饰词	编码体系修饰词	复用标准
核心元素(15 个)			
名称			dc:title
	交替名称		dcterms:alternative
创建者			dc:creator
	责任方式		
主题			dc:subject
		美国国会图书馆标题表	
		医学主题词表	
		杜威十进分类法	
		美国国会图书馆分类法	
		国际十进分类法	
		中国分类主题词表	
		汉语主题词表	
		中国图书馆分类法	
		中国科学院图书馆图书分类法	
		学科专业目录	

续表

元素	元素修饰词	编码体系修饰词	复用标准
描述			dc:description
	目录		dcterms:tableOfContents
	摘要		dcterms:abstract
出版者			dc:publisher
其他责任者			dc:contributor
	责任方式		
日期			dc:date
	创建日期		dcterms:created
	发布日期		dcterms:issued
	修改日期		dcterms:modified
	采集日期		mods:dateCaptured
		W3C-DTF	
类型			dc:type
		DCMI 类型词表	
格式			dc:format
	范围		dcterms:extent
	媒体		dcterms:medium
		IMT	
标识符			dc:identifier
		URI	
		SICI	
		ISBN	
		ISSN	
		DOI	
来源			dc:source
		URI	
		SICI	
		ISBN	
		ISSN	
		DOI	
语种			dc:language
		ISO 639—2	
		RFC 4646	