

ICS 01.140.20
A 14

C A D A L 项 目 标 准

CADAL 20803—2012

资源检索协议规范

Retrieval Agreement Standard

第一稿

2012-05-08 发布

2012-05-09 实施

CADAL 项目管理中心 发布

目 次

前言	69
引言	70
1 范围	71
2 规范性引用文件	71
3 术语和定义	71
3.1 资源服务器	71
3.2 超文本传送协议	71
3.3 文件传输协议	71
3.4 元数据	71
3.5 描述项	72
3.6 万维网	72
3.7 因特网工程任务组	72
3.8 元数据收割	72
3.9 都柏林核心元数据	72
4 资源检索协议规范	72
4.1 资源服务器提供的服务	72
4.2 元数据收割	73
4.3 元数据存储	73
4.4 数据检索	94
参考文献	96
表 1 DC 描述项	74
表 2 题名	76
表 3 主要责任者	77
表 4 主题词/关键词	78
表 5 次要责任者	79
表 6 资源描述	80
表 7 日期	81
表 8 资源形式	82
表 9 资源类型	83

CADAL 项目标准规范汇编(五)

CADAL 20803—2012

表 10	资源来源	84
表 11	语言	85
表 12	资源标识	86
表 13	时空范围	87
表 14	出版者	88
表 15	相关资源	89
表 16	权限管理	90
表 17	编号	91
表 18	存储路径	92
表 19	书页数	93
表 20	存储服务器 ip	93

前 言

《数字资源服务协议规范和接口标准集》包括以下 4 个方面的内容：

- 第 1 部分：元数据互操作协议规范；
- 第 2 部分：数字对象交换协议规范；
- 第 3 部分：资源检索协议规范；
- 第 4 部分：服务集成标准规范。

本标准是其中的第 3 部分。

本部分的制定依据了标准化工作导则第 1 部分(GB/T 1.1—2009)。

本部分由数字图书馆教育部工程研究中心提出。

本标准由 CADAL 项目管理中心归口。

本标准起草单位：数字图书馆教育部工程研究中心。

本标准主要起草人：尹彦飞、张寅、洪鑫。

引 言

本标准充分分析和研究了 CADAL 目前资源检索的流程和规范,了解了资源检索在整个数字资源使用过程中的作用,结合当前资源检索的各种规范而提出。

本标准规定了资源服务器所具有的基本的服务和功能,定义了资源服务器上发现和获取数字资源元数据的基本途径,提供了对数字资源元数据的存储策略,以及对已有数字资源的元数据进行检索的方法。

资源检索协议规范

1 范围

本标准确定了数字资源服务器必须具有的服务。

本标准适用于发现数字资源中的元数据。

本标准规定了数字资源元数据仓库所具有的描述内容。

本标准提出了对数字资源进行检索的基本策略。

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，仅所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(包括所有的修改单)适用于本文件。

*The Open Archives Initiative Protocol for Metadata Harvesting
Information Retrieval (Z 39.50): Application Service Definition and Protocol Specification*

3 术语和定义

本标准界定的术语和定义,以及下列术语和定义仅适用于本文件。

3.1 资源服务器 Resource Server 缩写: RS

资源服务器是用户存放数字资源的服务器。

3.2 超文本传送协议 HyperText Transfer Protocol 缩写: HTTP

超文本传送协议是一种详细规定了浏览器和万维网服务器之间互相通信的规则,通过因特网传送万维网文档的数据传送协议。

3.3 文件传输协议 File Transfer Protocol 缩写: FTP

文件传输协议是互联网中用于文件双向传输的协议。

3.4 元数据 Metadata 缩写: MD

元数据是对数字资源进行描述的属性的总和。

3.5 描述项 Description Item 缩写: DI

描述项是对数字资源进行描述的属性,例如数字资源的名称、主题等。

3.6 万维网 World Wide Web 缩写: WWW

万维网是以 Internet 为依托,以超文本标记语言 HTML(hypertext markup language)与超文本传输协议 HTTP(hypertext transfer protocol)为基础,向用户提供统一访问界面的 Internet 信息浏览系统。

3.7 因特网工程任务组 Internet Engineering Task Force 缩写: IETF

因特网工程任务组成立于 1985 年年底,是全球互联网最具权威的技术标准化组织,主要任务是负责互联网相关技术规范的研发和制定。当前绝大多数国际互联网技术标准均出自该组织。

3.8 元数据收割 Metadata Harvesting 缩写: MH

元数据收割是通过对数字资源的检索来获取数据资源对应的元数据的过程。

3.9 都柏林核心元数据 Dublin Core 缩写: DC

都柏林核心元数据于 1995 年 3 月在美国俄亥俄州第一届元数据会议上提出,并以会议举行地都柏林市命名。

4 资源检索协议规范

4.1 资源服务器提供的服务

资源服务器是数字资源的存储服务器,一方面数字资源的元数据需要通过使用资源服务器提供的服务来进行获取;另一方面,当数据资源的使用者在使用指定的数据资源的时候,需要通过数字资源服务器所提供的服务来获取对应的数据资源。本标准规定的数字资源服务器所具有的服务有以下两种:

4.1.1 HTTP 服务

超文本传输协议(Hypertext Transfer Protocol, HTTP)是分布式和协作式的,作为超媒体系统应用之间的通信协议,是万维网交换信息的基础,由互联网工程任务组(Internet Engineering Task Force, IETF)制定的国际化标准。

数字资源服务器通过 HTTP 服务使得获取对应的图书的流程变得简洁,通过简单的文件目录,利用浏览器就可以得到对应的数字化的资源。本标准确保在数字资源服务器中的各个数字资源有唯一的目录结构,彼此之间不存在联系。

4.1.2 FTP 服务

文件传输协议(File Transfer Protocol, FTP)是用于 Internet 上的控制文件的双向传

输。同时，它也是一个应用程序(application)。基于不同的操作系统有不同的 FTP 应用程序，而所有这些应用程序都遵守同一种协议以传输文件。

数字资源服务器通过 FTP 服务来为资源的使用者进行数字资源的传递功能，在使用者对数字资源进行请求时，资源服务器使用数字资源的唯一的存储路径在获取数字资源的同时又将其传递给该用户。

4.2 元数据收割

元数据收割主要在数据资源服务器上发现新资源，并且将新资源的元数据进行收集。本标准定义了两种不同模式的收集方法。

4.2.1 主动收割法

主动收割法是通过在资源服务器上安装元数据收割软件，使其能够在数据资源出现变化时及时得知变动，并对变动的数据资源重新进行收割，并将这些收割到的元数据发送到元数据的存储仓库，从而完成对数字资源的收割操作。

4.2.2 被动收割法

被动收割法是将元数据收割器安装在资源服务器之外，这样通过定时或者是认可操作的方式来对资源服务器中的所有资源进行统一的收割，并将收割到的元数据存储到数据仓库中去。被动收割法不能够获取到发生变动的资源，所以必须一次性地对所有资源进行收割。另一方面，由于元数据收割器安装在资源服务器之外，所以对数据资源的拥有者来说比较友好。

4.3 元数据存储

通过元数据收割工具获得的数字资源的元数据将会存储在数据库中，本部分对数据库存储做了规定，确定了存储的描述项的种类，并对每个数据资源的描述项做了详细的解释。

4.3.1 描述项定义的方法

经过讨论研究，本标准使用 DC 的 15 个核心元素的描述来定义数字资源的属性。根据 ISO/IEC 11179 标准，本元数据规范的定义中所有元素有四个属性是相同的。四个元素分别是：

- 版本：3.0；
- 注册机构：CADAL；
- 语种：中文；
- 最大使用频率：不限。

以下的元素定义省略了这四个相同的属性。子元素也参照上述方法定义，并在注释项说明其父元素。元素定义方法见表 1。

在表 1 中所列的项目中，元素及元素修饰词的英文名称需遵循以下书写规则：

CADAL 项目标准规范汇编(五)

CADAL 20803—2012

(1)元素英文名称,每个单词首字母均小写;

title、creator、description

(2)元素修饰词英文名称如由几个单词组成,则第一个单词首字母小写,第二、三及之后每个单词首字母大写,且书写时与前一个单词之间不空格;

dateCopyrighted、dateSubmitted、hasFormat

(3)元素标识与元素修饰词标签用英文名称,每个单词首字母均大写。

Title、Creator、Alternative Title、Date Copyrighted

表 1 DC 描述项

命名空间	元素所在的命名空间			
名称	元素中文名称(英文名称)			
标识	元素标识			
定义	定义元素的含义,用于确认著录的内容			
必备性	说明该元数据是“必备”,还是“可选”			
可重复性	说明该元素是否可以重复出现,Yes—可重复;No—不可重复			
属性	说明该元素所具有的属性,属性说明由以下几部分组成			
	标签	必备性	名称	参见
应用层次	数字文献包括中文图书、西文图书、期刊、学位论文等			
元素修饰词	描述该元素是否具有下属子元素,描述方式如下			
	标签	必备性	名称	参见
编码修饰词	描述该元素信息的描述遵循何种编码规则			
	标签	名称	参照	
数据类型	字符型、数字型、日期型、结构型			
指定值	指该元素所具有的值列表			
缺省值	指元素默认值			
注释	对元素的进一步说明			
著录信息源	元素著录内容的信息来源			
著录细则	元素著录内容与方法的细则规定			
DC 映射	与 DC 元素的映射			

4.3.2 描述项分类

数字资源的描述项分为两种类型：一种是数字资源本身具有的描述项；另一种是为了系统的使用而添加的描述项。

数据本身具有的属性包括：

- a) 题名——title
- b) 主要责任者——creator
- c) 主题词/关键词——subject
- d) 次要责任者——contributor
- e) 资源描述——description
- f) 日期——date
- g) 资源形式——format
- h) 资源类型——type
- i) 资源来源——source
- j) 语言——language
- k) 资源标识——identifier
- l) 时空范围——coverage
- m) 出版者——publisher
- n) 相关资源——relation
- o) 权限管理——rights

为了系统的使用而添加的描述项：

- a) 编号——id
- b) 路径——path
- c) 书页数——page
- d) 存储服务器 ip——ip

4.3.3 描述项详解

本部分对 4.3.2 中所介绍的元数据的 19 个描述项进行详细的介绍(见表 2~表 20)。

表 2 题名

命名空间	cadal			
名称	题名 (title)			
标识	Title			
定义	由创建者或出版者给予数字文献的正式公开名称			
必备性	必备			
可重复性	YES			
属性	—			
应用层次	数字文献			
元素修饰词	标签	必备性	名称	参照
	Alternative Title	有则必备	交替题名 (alternative)	[DCMI]
编码修饰词	—			
数据类型	字符型			
指定值	—			
缺省值	—			
注释	交替题名的英文名称为 alternative。数字文献的题名可以由若干部分组成,如正题名、其他题名信息,在 Title 中著录正题名、其他题名信息,在 Alternative Title 中著录正题名的交替题名			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Title			

表 3 主要责任者

命名空间	cadal			
名称	主要责任者 (creator)			
标识	Creator			
定义	创建数字文献内容的主要责任者			
必备性	有则必备			
可重复性	YES			
属性	标识	必备性	名称	参照
	Role	有则必备	角色 (role)	[OEBPS]
应用层次	数字文献			
元素修饰词	标签	必备性	名称	参照
	Institution	可选	机构 (institution)	
	Discipline	可选	专业 (discipline)	
编码修饰词	标识	名称		参照
	Creator Identifier	著者标识号 (creatorID)		[特定机构提供]
数据类型	字符型			
指定值	—			
缺省值	—			
注释	著录数字文献内容的主要责任者			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Creator			

表 4 主题词/关键词

命名空间	cadal			
名称	主题词/关键词 (subject)			
标识	Subject			
定义	描述数字文献主题内容的受控或非受控词汇			
必备性	可选			
可重复性	YES			
属性	标识	必备性	名称	参照
	Xsi: Type	可选	编码类型	[W3C]
应用层次	数字文献			
元素修饰词	—			
编码修饰词	标识	名称		参照
	CT	汉语主题词表		
	CCT	中国分类主题词表		
	CLC	中国图书馆分类法		
	SKC	四库分类法(北大)		
	LCCAS	中国科学院图书馆分类法		
	LCSH	美国国会图书馆主题词表		[DCMI]
	MESH	医学主题词表		[DCMI]
	DDC	杜威十进分类法		[DCMI]
	LCC	美国国会图书馆分类法		[DCMI]
UDC	国际十进分类法		[DCMI]	
数据类型	字符型			
指定值	—			
缺省值	—			
注释	用于描述数字文献的主题内容特征,包括受控主题词、关键词、分类号等			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Subject			

表 5 次要责任者

命名空间	cadal			
名称	次要责任者 (contributor)			
标识	Contributor			
定义	除主要责任者之外的创建数字文献内容的其他责任者			
必备性	有则必备			
可重复性	YES			
属性	标识	必备性	名称	参照
	Role	有则必备	角色	[OEBPS]
应用层次	数字文献			
元素修饰词	标签	必备性	名称	参照
	Institution	可选	机构 (institution)	
	Discipline	可选	专业 (discipline)	
编码修饰词	标识	名称		参照
	Contributor Identifier	著者标识号 (contributorID)		[特定机构提供]
编码修饰词	—			
数据类型	字符型			
指定值	—			
缺省值	—			
注释	除主要责任者外的其他参与数字文献内容创建的责任者			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Contributor			

表 6 资源描述

命名空间	cadal			
名称	资源描述 (description)			
标识	Description			
定义	对数字文献内容相关的描述说明、附注等			
必备性	可选			
可重复性	YES			
属性	—			
应用层次	数字文献			
元素修饰词	标识	必备性	名称	参照
	Accrual Periodicity	可选	出版周期 (accrualPeriodicity)	[DCMI]
	Abstract	可选	文摘 (abstract)	[DCMI]
	Table Of Contents	可选	目次 (tableOfContents)	[DCMI]
	Fund	可选	资助 (fund)	
编码修饰词	—			
数据类型	字符型			
指定值	—			
缺省值	—			
注释	关于数字文献内容的简要描述及与文献相关的其他说明			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Description			

表 7 日期

命名空间	cadal			
名称	日期 (date)			
标识	Date			
定义	与数字文献有关的日期			
必备性	有则必备			
可重复性	YES			
属性	标识	必备性	名称	参照
	Xsi: Type	可选	编码类型	[W3C]
应用层次	数字文献			
元素修饰词	标识	必备性	名称	参照
	Date Accepted	有则必备	接受日期 (dateAccepted)	[DCMI]
	Date Available	有则必备	可用日期 (dateAvailable)	[DCMI]
	Date Created	有则必备	创建日期 (dateCreated)	[DCMI]
	Date Copyrighted	有则必备	版权日期 (dateCopyrighted)	[DCMI]
	Date Submitted	有则必备	提交日期 (dateSubmitted)	[DCMI]
	Date Issued	有则必备	发布日期 (dateIssued)	[DCMI]
	Date Valid	有则必备	有效日期 (dateValid)	[DCMI]
	Date Modified	有则必备	修改日期 (dateModified)	
	Date Replied	有则必备	答辩日期 (dateReplied)	
编码修饰词	标识	名称		参照
	ISO8601	W3C 定义的日期描述格式		[W3CDTF]
	Period	DCMI 定义时期描述		[DCMI-DCSV]
数据类型	ISO8601: YYYY-MM-DD、YYYY-MM、YYYY			
指定值	—			
缺省值	—			
注释	Date 用于注明数字文献印刷型文献的出版日期,与数字文献相关的日期均通过元素修饰词进行著录。其中,答辩日期与提交日期为学位论文专用			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Date			

表 8 资源形式

命名空间	cadal			
名称	资源形式 (format)			
标识	Format			
定义	数字文献的电子形态			
必备性	可选			
可重复性	No			
属性	标识	必备性	名称	参照
	Xsi: Type	可选	编码类型	[W3C]
应用层次	数字文献			
元素修饰词	标签	必备性	名称	参照
	Extent	可选	文件大小 (extent)	[DCMI]
	Scan Resolution	可选	扫描分辨率 (scanResolution)	
编码修饰词	标识	名称		参照
	MIME	Internet 媒体类型		[MIME]
数据类型	字符型			
指定值	Text/html(.html,.htm) Application/msword(.doc) Application/pdf(.pdf) Application/postscript(.ai,.eps,.ps) Application/powerpoint(.ppt) Application/rtf(.rtf) Application/wordperfect5.1(.wp5,1) Application/x-latex(.latex) Image/DjVu(.djvu) Image/gif(.gif) Image/.(.jpeg,.jpg,.jpe) Text/plain(.txt) Text/sigml(.sgml,.sgm) 参照 MIME			
缺省值	Image/DjVu(.djvu)			
注释	数字文献的电子形态			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Format			

表 9 资源类型

命名空间	cadal
名称	资源类型 (type)
标识	Type
定义	有关资源内容的特征和类型
必备性	必备
可重复性	No
属性	—
应用层次	数字文献
元素修饰词	—
编码修饰词	—
数据类型	字符型
指定值	古籍、图书、民国图书、期刊、民国期刊、学位论文
缺省值	—
注释	资源类型主要用于描述资源所属的类别，根据本草案的适用范围，它可以是任意形式的数字文献
著录信息源	用户定义
著录细则	参见本丛书中的《元数据著录规范》
DC 映射	Type

表 10 资源来源

命名空间	cadal
名称	资源来源 (source)
标识	Source
定义	描述资源的出处信息
必备性	可选
可重复性	Yes
属性	—
应用层次	数字文献
元素修饰词	—
编码修饰词	—
子元素	—
数据类型	字符型
指定值	—
缺省值	—
注释	可以是数字文献的全部或其中一部分
著录信息源	用户定义
著录细则	参见本丛书中的《元数据著录规范》
DC 映射	Source

表 11 语言

命名空间	cadal			
名称	语言 (language)			
标识	Language			
定义	数字文献的文字语种			
必备性	必备			
可重复性	Yes			
属性	标识	必备性	名称	参照
	Xsi: Type	可选	编码类型	[W3C]
应用层次	数字文献			
元素修饰词	—			
编码修饰词	标识	名称	参照	
	ISO 639—2	语言编码	[ISO 639—2: 1998]	
	RFC 3066	语言编码	[RFC 3066]	
	RFC 1766	语言编码	[RFC 1766]	
数据类型	字符型			
指定值	—			
缺省值	chi			
注释	建议采用规范的表达形式,如 RFC 3066、RFC 1766 和 ISO 639—2 的语种代码			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Language			

表 12 资源标识

命名空间	cadal			
名称	资源标识 (identifier)			
标识	Identifier			
定义	确认资源的唯一标识			
必备性	必备			
可重复性	Yes			
属性	标识	必备性	名称	参照
	Xsi: Type	可选	编码类型	[W3C]
应用层次	数字文献			
元素修饰词	—			
编码修饰词	标识	名称		参照
	ISBN	国际标准书号		
	ISSN	国际标准连续出版物号		
	BookID	电子书号		[CADAL]
	URI	统一资源标识符		[DCMI]
	URL	统一资源定位符		
	DOI	数字对象标识符		
数据类型	字符型			
指定值	—			
缺省值	—			
注释	用以标识资源的字符串或数字,应采用标准的编码体系,属性“xsi: type”用于指定标识类型的编码类型,其指定值为 ISBN、ISSN、URI、URL 等			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Identifier			

表 13 时空范围

命名空间	cadal			
名称	时空范围 (coverage)			
标识	Coverage			
定义	数字文献内容所涉及的时间、空间覆盖范围			
必备性	可选			
可重复性	YES			
属性	标识	必备性	名称	参照
	Xsi: Type	可选	编码类型	[W3C]
应用层次	数字文献			
元素修饰词	标识	必备性	名称	参照
	Temporal Coverage	可选	时间范围 (temporalCoverage)	[DCMI]
	Spatial Coverage	可选	空间范围 (spatialCoverage)	[DCMI]
编码修饰词	标识	名称		参照
	Period	时期		[DCMI-DCSV]
	TGN	Getty 地理名称辞典		[TGN]
数据类型	字符型			
指定值	—			
缺省值	—			
注释	时间范围的著录可采用 period 编码修饰, 空间范围的著录可采用 TGN 编码修饰			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Format			

表 14 出版者

命名空间	cadal
名称	出版者 (publisher)
标识	Publisher
定义	使数字文献成为可取得或可利用的出版者或颁布者
必备性	有则必备
可重复性	YES
属性	—
应用层次	数字文献
元素修饰词	—
编码修饰词	—
数据类型	字符型
指定值	—
缺省值	—
注释	用于注明数字文献所对应的印刷型文献的出版者或颁布者
著录信息源	用户定义
著录细则	参见本丛书中的《元数据著录规范》
DC 映射	Publisher

表 15 相关资源

命名空间	cadal			
名称	相关资源 (relation)			
标识	Relation			
定义	与数字文献存在某种关系的其他数字文献资源。			
必备性	可选			
可重复性	YES			
属性	标识	必备性	名称	参照
	Xsi: Type	可选	编码类型	[W3C]
应用层次	数字文献			
元素修饰词	标签	必备性	名称	参照
	Has Format	可选	格式转换为 (hasFormat)	[DCMI]
	Is Format Of	可选	格式转换于 (isFormatOf)	[DCMI]
	Has Part	可选	包含 (hasPart)	[DCMI]
	Is Part Of	可选	包含于 (isPartOf)	[DCMI]
	Has Version	可选	其他版本 (hasVersion)	[DCMI]
	Is Version Of	可选	原版本 (isVersionOf)	[DCMI]
	References	可选	参照 (references)	[DCMI]
	Is Referenced By	可选	被参照 (isReferencedBy)	[DCMI]
	Replaces	可选	替代 (replaces)	[DCMI]
	Is Replaced By	可选	被替代 (isReplacedBy)	[DCMI]
	Requires	可选	需求 (requires)	[DCMI]
Is Required By	可选	被需求 (isRequiredBy)	[DCMI]	
编码修饰词	标识	名称		参照
	ISBN	国际标准书号		
	ISSN	国际标准连续出版物号		
	BookID	电子书号		[CADAL]
	URI	统一资源标识符		[DCMI]
	URL	统一资源定位符		
	DOI	数字对象标识符		
数据类型	字符型			
指定值	—			
缺省值	—			
注释	用于揭示数字文献同其他资源之间的关联。一般用字符串或数字代码标识相关资源。如果采用规范定义的编码体系,如统一资源标识 (URI)、数字对象标识 (DOI)、国际标准书号 (ISBN) 等,则应在元素中注明编码修饰名称。能用本元素的元素修饰词表达的相关资源应采用相应的元素修饰词			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Relation			

表 16 权限管理

命名空间	cadal			
名称	权限管理 (rights)			
标识	Rights			
定义	数字文献本身所有的或被赋予的权限信息			
必备性	必备			
可重复性	YES			
属性	—			
应用层次	数字文献			
元素修饰词	标签	必备性	名称	参照
	Access Rights	有则必备	访问权限 (accessRights)	[DCMI]
	Copyright Expiry Date	有则必备	文本版权解除日期 (copyrightExpiryDate)	
	Copyright Owner	有则必备	文本版权拥有者 (copyrihgtOwner)	
	Create Centre	可选	制作单位 (createCentre)	
	Location	有则必备	收藏单位 (location)	
	Security Classification	可选	保密级别 (securityClassification)	
编码修饰词	—			
数据类型	字符型			
指定值	—			
缺省值	—			
注释	包括知识产权、版权、使用权限说明及其他相关产权。可指一个权限管理的陈述及其标识,也可指提供权限管理信息内容的服务器标识			
著录信息源	用户定义			
著录细则	参见本丛书中的《元数据著录规范》			
DC 映射	Rights			

表 17 编号

命名空间	cadal
名称	编号 (id)
标识	Id
定义	由资源服务器中的目录标示决定
必备性	必备
可重复性	YES
属性	—
应用层次	数字文献
元素修饰词	—
编码修饰词	—
数据类型	字符型
指定值	—
缺省值	—
注释	id 作为数字资源的唯一标示
著录信息源	系统定义
著录细则	参见本丛书中的《元数据著录规范》
DC 映射	—

表 18 存储路径

命名空间	cadal
名称	存储路径(path)
标识	Path
定义	数字资源的存储路径
必备性	有则必备
可重复性	YES
属性	—
应用层次	数字文献
元素修饰词	—
编码修饰词	—
数据类型	字符型
指定值	—
缺省值	—
注释	著录数字文献内容的主要责任者
著录信息源	用户定义
著录细则	参见本丛书中的《元数据著录规范》
DC 映射	—

表 19 书页数

命名空间	cadal
名称	书页数(page)
标识	Page
定义	图书数字资源的总页数
必备性	可选
可重复性	YES
属性	—
应用层次	数字文献
元素修饰词	—
编码修饰词	—
数据类型	整数型
指定值	—
缺省值	—
注释	用户描述数字图书的页数
著录信息源	用户定义
著录细则	参见本丛书中的《元数据著录规范》
DC 映射	Page

表 20 存储服务器 ip

命名空间	cadal
名称	存储服务器(ip)
标识	Ip
定义	数字资源所在的服务器的 ip 地址
必备性	有则必备
可重复性	YES
属性	—
应用层次	数字文献
元素修饰词	
编码修饰词	—
编码修饰词	—
数据类型	字符型
指定值	—
缺省值	—
注释	除主要责任者外的其他参与数字文献内容创建的责任者
著录信息源	用户定义
著录细则	参见本丛书中的《元数据著录规范》
DC 映射	Ip

4.4 数据检索

数据检索是从数据资源的使用者的角度定义了对已经获得的数字资源的元数据的检索，主要包括模糊检索、布尔检索和联邦检索三种类型。

4.4.1 模糊检索

“模糊检索”是指搜索系统自动按照用户输入关键词的同义词进行模糊检索，从而得出较多的检索结果。同义词由系统的管理界面配置。如配置了“电脑”与“computer”为同义词后，检索“电脑”，则包含“computer”的数字资源也会出现在检索结果中。

使用模糊搜索可以自动搜索关键字的同义词，提高搜索的精确性。在数字资源服务系统中使用模糊检索能够使用户按照自己的意图进行简单化的检索，同时为用户提供更加详尽的检索结果，方便用户进行判断和选择。

4.4.2 布尔检索

布尔检索是数据库检索最基本的方法，是用逻辑“或”(＋、OR)、逻辑“与”(×、AND)、逻辑“非”(－、NOT)等算符在数据库中对相关文献的定性选择的方法。

(1) 逻辑“或”(＋、OR)：用来组配相同概念的词，文献中凡含有“A”或者“B”检索词或者同时含有检索词“A”和“B”的文献均为命中文献。

(2) 逻辑“与”(×、AND)：检索时，数据库中同时含有检索词“A”和检索词“B”的文献才是命中文献。组配方式： $A \times B$ ，表示检索必须同时含有“A”、“B”两词的文章。

(3) 逻辑“非”(－、NOT)：数据库中凡含有检索词“A”而不要检索词“B”的文献，为命中文献，是用来检索中排除某些词的。

在 CADAL 中，我们提供了对以下描述项的布尔检索：

- a) 题名——title
- b) 主要责任者——creator
- c) 主题词/关键词——subject
- d) 次要责任者——contributor
- e) 资源描述——description
- f) 出版者——publisher

4.5.3 联邦检索

联邦检索使用了 Z 39.50 作为标准。Z 39.50 的目的是为了信息系统的开放互联，由于各信息系统分别采用各自的数据库软件，数据的描述格式、访问方式等都各不相同，必须为各自数据库系统建立一个抽象、通用的用户视图，将各个系统的具体实现映射到抽象模型上，才能使不同的系统在一个相互理解的、标准的通信平台上进行交互，满足互操作的需要。

允许计算机搜索远程系统上馆藏信息的一种协议，产生可进一步处理的检索结果和检索信息，主要用于检索书目信息。

参 考 文 献

[1] CAGOZE M N C, de SOMPEL H V, WARNER S. The open archives initiative protocol for metadata harvesting[R]. Technical report, Open Archives Initiative, 2008 (11).

[2] Information Retrieval (Z 39. 50): Application Service Definition and Protocol Specification[S]. National Information Standards, 1995.

[3] 因特网工程任务组. The Internet Engineering Task Force(IETF)[S]. 1985.