

ICS 01.140.20

A 14

C A D A L 项 目 标 准

CADAL 10301—2012

数字对象唯一标识符规范

Unique Digital Object Identifier Specification

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	147
引言	148
1 范围	149
2 规范性引用文件	149
3 术语和定义	149
3.1 CADOI 名称	149
3.2 CADOI 解析	150
3.3 CADOI 语法	150
3.4 CADOI 系统	150
3.5 持久性	150
3.6 数字对象	150
3.7 唯一标识	150
3.8 元数据	150
4 CADOI 体系框架结构	150
4.1 总体架构	150
4.2 语法和命名规则	150
4.3 元数据规则	150
4.4 解析规则	151
4.5 管理规则	151
4.6 扩展规则	151
5 语法规则	151
5.1 一般规则	151
5.2 CADOI 前缀	151
5.3 CADOI 后缀	152
5.4 CADOI 的显示	152
5.5 CADOI 的字符集和编码	152
5.6 CADOI 名称的推荐长度与大小写规定	152
6 命名规则	153
6.1 范围	153
6.2 粒度	153

CADAL 项目标准规范汇编(三)

CADAL 10301—2012

6.3 唯一性.....	153
6.4 持久性.....	153
7 元数据规则	153
8 管理规则	153
9 解析规则	154
9.1 一般规则.....	154
9.2 数据缓存和过期.....	154
10 扩展规则.....	154
附录 A(规范性附录) CADOI 元数据	155
参考文献.....	156
表 A.1 CADOI 元数据	155

前 言

《CADAL 项目数字对象标准》分 5 个部分,由 5 个标准组成。

——第 1 部分:CADAL 10301—2012 数字对象唯一标识符规范。

——第 2 部分:CADAL 10302—2012 数字对象标识与命名规范。

——第 3 部分:CADAL 10303—2012 复合数字对象描述和管理规范。

——第 4 部分:CADAL 10304—2012 数字对象阅读规范。

——第 5 部分:CADAL 10305—2012 CADAL 数字对象唯一标识符解析系统应用规范。

本规范为第 1 部分。

本规范主要参考了 ISO/CD 26324 *Information and Documentation—Digital Object Identifier (DOI)*、*The DOI Handbook*、RFC 3986 *Uniform Resource Identifier (URI): Generic Syntax*、RFC 2141 *URN Syntax*、《数字资源唯一标识符解析系统应用规范》、《国家图书馆数字资源唯一标识符规范和应用指南》等。

本规范的某些内容可能涉及专利,本规范的发布机构不承担识别这些专利的责任。

本规范由西南政法大学图书馆、浙江大学图书馆提出。

本规范由大学数字图书馆国际合作计划(CADAL)项目管理中心提出并归口。

本规范起草单位:西南政法大学图书馆、浙江大学图书馆。

本规范主要起草人:江波、蔡文彬、毛一国、郑江平。

参加本规范研究工作的还有刘景宇、傅天珍、赵惠芳。

引 言

数字对象唯一标识符(Digital Object Identifier, DOI),是实现数字对象永久标识和解析的完整体系。

CADAL 数字对象唯一标识符(CADAL Digital Object Identifier, CADOI)是实现 CADAL 数字图书馆数字对象的唯一标识和解析。

本规范的主要目的是建立 CADAL 数字资源唯一标识符系统,同时兼顾建立中国大学数字对象唯一标识系统(China Academic Digital Object Identifier)。

CADOI 实现对 CADAL 海量资源的唯一标识和解析,同时开放给 CADAL 成员馆、国内其他大学图书馆及所有需要建立数字对象唯一标识系统的机构和个人使用。

任何一个数字对象一旦被赋予某一 CADOI 名称,则该名称将被持久地用来标识该对象,通过该 CADOI 名称可以获取该对象的相关信息,并永久地链接到该对象的 URL 地址。

CADOI 系统提供网络环境下可解析的、持久的标识系统,实现数字对象的永久唯一标识和解析,并在此之上提供相关的服务。

CADOI 不负责数字对象的元数据管理、存放管理、权限管理。

数字对象唯一标识符规范

1 范围

本部分规定了 CADAL 数字对象唯一标识符系统(CADOI)的体系框架、语法规则、命名规则、解析规则、管理规则和扩展规则。

本部分所指的数字对象,包括 CADAL 建设中加工制作生成的文本、图片、音频、视频等各种类型的数字对象。其解析结果表现形式为一个合法的 URL 地址。

通过给数字对象赋予一个 CADOI 唯一标识符,可以实现对其的永久标识和持久链接。

CADOI 在语法规则、命名规则、解析规则等方面参考了 Information and Documentation—Digital Object Identifier (DOI),但与其有一定差别。CADOI 独立于 DOI 系统,不受 DOI 相关规则约束。通过预留的可扩展接口,CADOI 可以实现与 DOI 系统的兼容。

CADOI 遵循 RFC 3986 *Uniform Resource Identifier (URI): Generic Syntax* 标准。

CADOI 系统不试图替换现有的标识系统,包括 ISO TC46/SC9 的 ISBN、ISSN、ISAN、ISRC 等,或其他已被共同接受或实际使用的标识系统。必要时,CADOI 可以将上述系统中应用的标识符方案或应用系统自定义的标识符方案等引用到自身标识符中。

本部分不规定 CADOI 的实现技术,其相关技术问题在其他文献中说明。

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,其最新版本适用于本文件。

ISO/IEC 10646 *Universal Multiple-Object Coded Character Set (UCS)*

RFC 3986 *Uniform Resource Identifier (URI): Generic Syntax*

RFC 3629 *UTF-8, a transformation Format of ISO 10646*

RFC 1630 *Universal Resource Identifiers in WWW*

RFC 1738 *Uniform Resource Locators (URL)*

3 术语和定义

3.1 CADOI 名称 CADOI Name

CADOI 名称是指在 CADOI 系统内定义一个唯一对象的字符串。CADOI 名称可以由遵循 CADOI 语法规则的字母、数字、字符组成。

3.2 CADOI 解析 CADOI Resolution

CADOI 解析指提交一个 CADOI 名称到 CADOI 系统,接受返回的一个 URL 地址的过程。

3.3 CADOI 语法 CADOI Syntax

CADOI 语法指 CADOI 名称中字符的形式、顺序的组成规则,具体指 CADOI 前缀、分隔符、后缀的形式和特征。

3.4 CADOI 系统 CADOI System

CADOI 系统指 CADOI 名称作为计算机可识别的标识符,在名称分配、解析、标识、管理等方面的应用部署。

3.5 持久性 Persistent

持久性指不受时间期限影响而存在,能够摆脱因资源服务系统改变、升级而造成的 URL 失效,能够普遍应用于各种服务中。

3.6 数字对象 Digital Object

数字对象指数字资源管理系统中存储、管理和发布信息的基本实体,通常由标识符、数据流和元数据三要素组成。

3.7 唯一标识 Unique Identification

唯一标识指有且仅有一个标识对象被该 CADOI 名称指定。

3.8 元数据 Metadata

元数据是关于数据的数据。它描述一个具体的资源对象,并能对这个对象进行定位、管理,有助于它的发现与获取。

4 CADOI 体系框架结构

4.1 总体架构

CADOI 由唯一标识符语法和命名规则、元数据规则、解析规则、管理规则和扩展规则组成。

4.2 语法和命名规则

语法和命名规则用于说明唯一标识符名称字符串的结构组成、语法规则和名称分配规则。

4.3 元数据规则

元数据规则是对各类数字对象的统一描述规范,提供对对象的识别、确认。

4.4 解析规则

解析规则规定了通过唯一标识符解析系统,将唯一标识符转换为数字对象对应 URL 的流程和机制。

4.5 管理规则

管理规则规定了唯一标识符使用和管理政策等方面内容。

4.6 扩展规则

扩展规则说明了 CADOI 前缀、后缀的扩展原则,以及与其他标识符系统的关系。

5 语法规则

5.1 一般规则

CADOI 名称是为了达到 CADOI 系统的目的而建立的一个无任何含义的字符串。

一个 CADOI 名称标识唯一的一个数字对象。

CADOI 名称由以下 3 部分组成:

- 前缀元素;
- 分隔前后缀的符号“@”;
- 后缀元素。

前缀是本地标识,在同一个后缀空间中,前缀名应是唯一的。

后缀是命名授权,可以由分级的子命名授权段组成,每个子命名授权段之间用字符“.”(‘\0x2E’)分隔。

CADOI 名称必须是唯一的。唯一性通过后缀的唯一性和前缀在同一个后缀空间中的唯一性共同保证。

CADOI 名称的长度没有限制。为了便于记忆和使用,建议尽量简短。在实际使用中,CADOI 名称的长度可能受互联网相关通讯协议限制。

CADOI 名称不区分大小写,可以是 ISO/IEC 10646 字符集中的任何字符。

5.2 CADOI 前缀

CADOI 前缀是本地标识,可以是 ID 号、顺序号或者其他字符串,也可采用另一个系统的标识号(如 ISBN、ISSN、ISTC 等)。

本规范不对 CADOI 前缀的具体组成和含义进行规定。具体使用中,使用者可规定更具有约束性的本地标识组成和分配细则,以满足特定的需要。这些细则应与本规范兼容,但不作为本规范的一部分。

使用者必须保证本地标识在某一个后缀空间中是唯一的。

CADAL 项目作为具体使用者,前缀由 6 位机构代码和 6 位顺序号组成,中间加下划线。本地标识组成和分配细则详见《数字对象命名与标识标准规范》。

5.3 CADOI 后缀

CADOI 后缀是命名授权,可以由多个非空的子命名授权段组成,每个段由字符“.”(‘\0x2E’)分隔。

示例 1:

如已有命名授权“abc”,需要在其下建立下一级命名授权“def”,则新的命名授权为“def. abc”

一级命名授权由 CADOI 管理中心管理。

CADOI 应用机构在向 CADOI 管理中心申请一级命名授权后,可根据需要,自行增加二级及以下的命名授权。

本规范对命名授权段的数量不作限制。但为了便于使用,子命名授权段数量不宜超过 5 个。

CADAL 项目所有资源后缀均为“cadal”。

5.4 CADOI 的显示

5.4.1 CADOI 绝对格式

CADOI 绝对格式遵循 URI 要求,以 scheme(后面跟着冒号)开头,形如:[scheme:] scheme-specific-part 。

示例 2:

cadai: 123456@cadal

5.4.2 CADOI 相对格式

CADOI 相对格式也遵循 URI 格式,不以 scheme 开头,只显示 scheme-specific-part 。

示例 3:

123456@cadal

5.5 CADOI 的字符集和编码

CADOI 名称使用 ISO/IEC 10646 字符集作为基本字符集。

CADOI 使用 UTF-8 规则编码。

由于 CADOI 经常需要通过浏览器来处理,CADOI 名称不宜使用 HTTP 协议保留字符、不可打印显示的计算机控制字符等。具体参见 RFC 1630、RFC 1738 相关规定。

5.6 CADOI 名称的推荐长度与大小写规定

本规范对 CADOI 名称长度没有限制。

当 CADOI 在使用 URL 表示和通过 HTTP 协议传输时,URL 长度要受到最大字符长度限制,故不宜超过目前所知的长度限制 2048 个字符。同时,还需减去解析服务器路径的字符数。

根据超文本传输协议 HTTP/1.1(RFC 2616)3.2.1 中指出 URI 的保守长度为 255 个字符,CADOI 名称建议不超过 1024 个字符。

CADOI 名称不区分大小写。

示例 4:

123456@ZJU 和 123456@zju 均作为同一个标识符

6 命名规则

6.1 范围

任何时候当需要给一个数字对象一个永久标识时,可以给这个对象分配一个 CADOI 名称。

CADOI 名称能分配给任何类型、格式、载体的对象,只要其拥有一个有效的 URL 地址即可。CADOI 名称分配的对象包括但不限于中文古籍、民国文献、中外文图书、连续出版物、学位论文、技术报告、地方文史资料、图形图像、声像资料等,以及这些对象的某一部分、元数据等。

6.2 粒度

CADOI 名称能在对象的任意粒度层次上进行分配。

例如,可以给某套丛书,其中的某一本,某本中特定一章、一个段落、一张图片、一个表格,或者使用者认为需要的任何其他粒度独立分配 CADOI 名称。

6.3 唯一性

在 CADOI 系统中,一个 CADOI 名称唯一地标识一个对象。

但是,某一个被标识对象可能有多个 CADOI 名称。

示例 5:

12345@CADAL 和 56789@ZJU 是不同的 CADOI 标识符,但是,可能标识同一对象,其解析结果均为一个相同的 URL 地址。

6.4 持久性

CADOI 名称不受时间限制,永久有效。

CADOI 名称不受标识对象权限、URL 地址、服务规则、管理责任等变更的影响。

CADOI 名称一旦被分配,一般情况下不予删除。如果确需删除时,删除后所有对于该 CADOI 名称的解析请求均会返回给用户一个空值。

7 元数据规则

使用者在分配给对象 CADOI 名称时,应同时提交该对象的核心元数据信息。

CADOI 采用 DC(Dublin Core)都柏林核心元数据元素集作为数据模型。每个 CADOI 元数据记录包括 15 个核心元素:题名、创建者、主题或关键词、描述说明、出版者、其他发行者、日期、资源类型、格式、标识符、来源、语种、关联、时空范围、权限。

CADOI 系统中保存的元数据主要用于 CADOI 名称管理,不用于检索和对象揭示。
题名元素为必需元素。

8 管理规则

CADOI 由 CADOI 管理中心进行管理。

CADOI 管理中心提供以下服务和功能:

- 负责 CADOI 服务系统的稳定可靠和持续运行。
- 分配、管理 CADOI 一级命名授权的注册。
- 促进 CADAL 系统内外 CADOI 的应用开发和推广。
- 代表 CADOI 与国内外同类机构进行交流与合作。

在 CADOI 管理中心的统一管理下,可在各地建立区域性的注册中心,负责一级命名授权的注册。

一旦一个一级命名授权投入使用,则不宜随便修改。如果确实需要修改,则需在分配、注册、服务等方面进行全面评估。

9 解析规则

9.1 一般规则

解析是向 CADOI 系统递交一个 CADOI 名称,接受系统返回的 CADOI 名称所标识的对象的 URL 的过程。

CADOI 解析记录有且仅有一个有效的 URL 地址,同时可包括其他元数据信息。

9.2 数据缓存和过期

各个命名授权的域解析服务器是其权威解析服务器。同时,在其他服务器上可缓存 CADOI 解析记录。

每条缓存的 CADOI 解析记录均包含一个时间戳和一个过期参数。当某条 CADOI 解析记录过期后,缓存服务器自动丢弃该记录,并向权威解析服务器申请新的解析记录。

10 扩展规则

CADOI 可将其他系统中应用的标识符方案引用到自身标识符中,包括采用其他 ISO 标准的标识符方案,如 ISBN、ISSN、ISAN、ISRC 等,非 ISO 标准但被广泛认可和使用的标识符方案,或应用系统自定义的标识符方案等。

CADOI 解析服务器可以通过解析插件,对其他唯一标识符系统(如 DOI、CDOI 等)进行解析。

同时,免费开放解析接口给其他唯一标识符系统使用。

附录 A
(规范性附录)
CADOI 元数据

CADOI 复用 Dublin Core 的 15 个元素作为核心元素集见表 A.1。

表 A.1 CADOI 元数据

元素名称	DC 元素名称	与 Dublin Core 的对映
题名	名称	title
主要责任者	创建者	creator
主题	主题	subject
附注	描述	description
其他责任者	其他责任者	contributor
时间	日期	date
资源类型	类型	type
	格式	format
资源标识	标识符	identifier
	来源	source
语种	语种	language
相关文献	关联	relation
时空范围	时空范围	coverage
权限管理	权限	rights
出版者	出版者	publisher

参 考 文 献

- [1] 贺燕. 走中国 DOI 发展之路[J]. 图书馆理论与实践, 2009, (7): 1-5.
- [2] 江波. 开放型数字对象标识系统设计[J]. 情报理论与实践, 2011, (9): 107-109.
- [3] 梁健. Handle System 在中文 DOI 系统中的应用[J]. 数字图书馆, 2008, (9): 16-20.
- [4] 龙健, 赖茂生. DOI 的兴起与我国的对策[J]. 情报杂志, 2009, (12): 159-161.
- [5] 王海英, 龚永红, 张金艳. DOI: 数字对象标识技术[J]. 情报杂志, 2004, (12): 91-95.
- [6] 任瑞娟, 刘丽斌, 濮德敏, 米佳. 中文 DOI 路在何方——从参考文献著录与 DOI 的关系探讨中文数字对象唯一标识符的发展方向[J]. 中国图书馆学报, 2010, (2): 115-121.
- [7] 赵蕴华. 中文 DOI 应用现状及建设构想[J]. 情报理论与实践, 2007, (6): 855-857.
- [8] Norman Paskin. Digital Object Identifier System[S/OL]. 2008-01-12. <http://www.doi.org/overview/070710-Overview.pdf>.
- [9] Robert Kahn, Robert Wilensky. A Framework for Distributed Digital Object Service [EB/OL]. 2008-02-20. http://www.doi.org/topics/2006_05_02_Kahn_Framework.pdf.
- [10] ISO/IEC 10646 Universal Multiple-Object Coded Character Set (UCS) [S/OL]. 2012-04-12. [http://standards.iso.org/ittf/PubliclyAvailableStandards/c051273_ISO_IEC_10646_2011\(E\).zip](http://standards.iso.org/ittf/PubliclyAvailableStandards/c051273_ISO_IEC_10646_2011(E).zip).
- [11] RFC 3986 Uniform Resource Identifier (URI): Generic Syntax[S/OL]. 2012-04-12. <http://www.rfc-editor.org/rfc/rfc3986.txt>.
- [12] RFC 3629 UTF-8, a transformation Format of ISO 10646 2012-04-12[S/OL]. <http://www.rfc-editor.org/rfc/rfc3629.txt>.
- [13] RFC 1630 Universal Resource Identifiers in WWW [S/OL]. 2012-04-12. <http://www.rfc-editor.org/rfc/rfc1630.txt>.
- [14] RFC 1738 Uniform Resource Locators (URL) [S/OL]. 2012-04-12. <http://www.rfc-editor.org/rfc/rfc1738.txt>.