

ICS 01.140.20
A 14

C A D A L 项 目 标 准

CADAL 20602—2012

数字资源访问规范

Digital Resources Access 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	21
引言	22
1 范围	23
2 规范性引用文件	23
3 术语和定义	23
3.1 数字对象唯一标识符	23
3.2 DOI 注册代理机构	23
3.3 统一资源名称	23
3.4 统一资源定位符	24
3.5 通用资源标识符	24
3.6 互联网数字分配机构	24
3.7 CADAL 数字资源唯一标识符	24
3.8 开放链接	24
4 数字资源访问规范	24
4.1 数字对象唯一标识符解析规范	24
4.2 开放链接	26
参考文献	28
表 1 OpenUrl 返回值	27

前 言

《数字资源发布标准规范集》包括以下 3 个方面的内容：

- 第 1 部分：数字资源发现规范；
- 第 2 部分：数字资源访问规范；
- 第 3 部分：数字资源传输规范。

本标准是其中的第 2 部分。

本部分的制定依据了 GB/T 1.1—2009 的要求，同时参照了 RFC 3305、RFC 1738、RFC 3652、RFC 3651、RFC 3650 和 NISO Z39.88。

本部分是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。

本部分起草单位：数字图书馆教育部工程研究中心 CADAL 项目门户组。

本部分起草人：张鹏、张寅。

引 言

数字资源是以数字形式发布、存取、利用的信息资源，通常涉及计算机技术、通信技术及多媒体技术。

数字资源的访问规范多种多样，本规范对数字资源访问进行了规范化，以方便数字资源的访问、管理和应用。

本规范是基于数字图书馆对资源组织管理的需求，以及 CADAL 数字资源 DOI、开放链接的访问实践编制的。

本规范提供了面向数字图书、图像、书法的数字资源访问规范，充分说明了符合 CADAL 项目要求的数字资源访问规范。

数字资源访问规范

1 范围

本规范确立了 CADAL 项目数字资源访问的规范。

本规范适用的数字资源对象包括：图书数字资源、图像数字资源、书法数字资源、特种数字资源。

本规范适用于 CADAL 项目数字资源的访问。

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。

RFC 3305 *Uniform Resource Name*

RFC 1738 *Uniform Resource Locator*

RFC 3652、RFC 3651 《DOI 的解析协定》

RFC 3650 《DOI 架构》

NISO Z39.88 *Open URL 1.0*

3 术语和定义

3.1 数字对象唯一标识符 Digital Object Unique Identifier 缩写：DOI

数字对象唯一标识符是一套识别数字资源的机制，涵括的对象有视频、报告和书籍等。它既有一套为资源命名的机制，也有一套将识别号解析为具体地址的协定。

3.2 DOI 注册代理机构 DOI Registration Agencies 缩写：DOI RA

负责处理 DOI 事物的代理机构，目前 DOI 有 10 个注册代理，在中国的代理机构为中国科技信息研究所，由下属公司万方数据代行职责。

3.3 统一资源名称 Uniform Resource Name 缩写：URN

统一资源名称是一个用于标识某一互联网资源名称的字符串。该种标识允许用户对网络中的资源通过特定的协议进行交互操作。由 RFC 3305 文件定义。

3.4 统一资源定位符 Uniform Resource Locator 缩写: URL

URL 指因特网上标准的资源的地址。由 RFC 1738 文件定义。

3.5 通用资源标识符 Uniform Resource Identifier 缩写: URI

Web 上任意资源都可以由一个唯一的通用资源标识符进行定位。

3.6 互联网数字分配机构 The Internet Assigned Numbers Authority 缩写: IANA

IANA 是负责协调一些使 Internet 正常运作的机构。IANA 分配和维护互联网技术标准(或者称为协议)中的唯一编码和数值系统。

3.7 CADAL 数字资源唯一标识符 CADAL Digital Object Unique Identifier 缩写: CADAL-OID

CADAL-OID, 即本标准制定的数字对象唯一标识符命名方式。

3.8 开放链接 Open Uniform Resource Locator 缩写: OpenURL

OpenURL 定义了一种在 Web 服务之间传递信息的机制, 是一套应用于 Web 上超链接的标准陈述语法, 由一组已定义好的标识组成。网络上的各个数据库出版商只要遵照此机制, 就可以解析信息提供源(source)所传送的要求, 而信息提供者也可经由此规范, 对服务提供者送出深度链接服务的要求。

4 数字资源访问规范

4.1 数字对象唯一标识符解析规范

4.1.1 数字对象唯一标识符

数字对象唯一标识符(digital object unique identifier, DOI)是一套识别数字资源的机制。发展 DOI 的动机在于补充 URI 之不足, 因为一方面 URI 指涉的 URL 经常变动; 另一方面, URI 表达的是资源所在地址, 而非数字资源本身的信息。DOI 能克服这两个问题。

DOI 码由前缀和后缀两部分组成, 之间用“/”分开, 并且前缀以“.”再分为两部分。举例明之, 以下是一个典型的 DOI 识别号:

10.1006/jmbi.1998.2354

其中的“10.1006”是前缀, 由国际数字对象识别号基金会确定。其中, “10”为 DOI 目前唯一的特定代码, 用以将 DOI 与其他采用同样技术的系统区分开; “1006”是注册代理机构的代码, 或出版社代码, 用于区分不同的注册机构。后缀部分由资源发布者自行指定, 用于区分一个单独的数字资料, 具有唯一性; 以书籍为例, 它可能是国际标准书号。发布者可以选择以何单位进行注册, 例如, 一本书可以注册单一的 DOI, 也可以依各章节分别注册, 甚至独立注册其中的一个表格或图片。

目前, DOI 作为数字对象的唯一标识符方案在管理、注册和解析方面是发展最为完善

的。但无论是为自己的资源申请 DOI, 还是申请成为 DOI 的注册代理机构, 都需要缴纳不菲的会员费或者注册费及维护费。

此外, URN 也是一种广泛应用的方案。URN 唯一标识一个实体的标识符, 但是不能给出实体的位置, 而查找位置的任务由 URL 完成, 两者合起来构成 URI。URN 主要用于 Web 系统中对资源的定位, 并非针对数字资源而制定。

综上所述, 在 CADAL 项目建设中我们决定采用 DOI 与 URN 相结合的做法来制定数字对象唯一标识符产生方案, 如下所示:

命名方式+注册机关代码+注册资源代码

- 命名方式: 如以 URN 方式则为 urn, 以 DOI 方式则为 doi。
- 注册机关代码: 如为 URN informal 方式, 则由申请机关向注册中心 (IANA) 申请分发为 urn-d (d 为数字), 或申请 URN formal 方式; 若为 DOI, 则向注册中心 (IDF 或 CrossRef) 申请分发一个代码。
- 注册资源代码则由注册单位内部自编, 无一定格式, 但要保证在内部为唯一代号。如 URN 需要提出内部编码方式给 IANA 协会审查, 而 DOI 只要资源识别码注册时不与现有的重复即可。
- + 为区分码, 如 URN 为 “:”, DOI 为 “/” 等。

不管加入哪一个网络资源组织, 其注册资源代码都要由注册机关自定, 未来只要在本方案的基础上加入注册机关代码即可成为国际唯一的识别码。因此, 不管未来盛行哪一个网络资源组织, 本方案的标识符都可以被简单快速地转换成该组织制定的命名方式。本方案具有高度的扩充性。

4.1.2 唯一标识符中的元素

基于上述方案, CADAL 制定的唯一标识符 (CADAL-OID) 要符合 URN 的标准, 成为 URN 的一个子集。

对于复杂数字对象, 其语法为:

urn: CADAL: 馆数字代码-CollectionName[. CollectionName]/ObjID

对于简单数字对象, 其语法为:

urn: CADAL: 馆数字代码-CollectionName[. CollectionName]/ObjID. type. format

说明: 命名方式“urn”应全部采用小写字母。

CADAL-OID 的长度最长不应超过 255 个字符; 复杂数字对象与其所含的简单数字对象的 CollectionName、ObjID 并无继承关系。

4.1.3 集合名称

上述命名方式中的集合名称 (又称仓库名称) CollectionName 是必备的, 且可嵌套。其中, CollectionName 采用《信息资源名称规范列表》中资源的英文简称。为避免唯一标识符过长, 取其最具体的一种分类。如在《信息资源名称规范列表》中不能找到合适的 CollectionName, 可自行分配。

4.1.4 ObjID

上述命名方式要求 ObjID 符合网络资源的命名规则：使用 ASCII code 命名，文件名称大小写不敏感，不使用“%”、“/”、“?”、“#”、“*”、“—”、“.”、“:”字符。

如在镜像、备份等情况下，发生数字对象的拷贝情况，原则上使用相同的 ObjID。

4.2 开放链接

4.2.1 开放链接简介

OpenURL 框架是上下文相关的开放链接框架，使信息服务者与信息的提供者相分离，实现同时对不同机构的多个数据库或信息资源进行统一的检索。2001 年，美国国家信息标准组织(National Information Standards Organization, NISO)成立专责委员会 AX，制定关于 OpenURL 的美国国家标准(American National Standards Institute, ANSI) Z39.88—200X。Z39.88 协议的核心是定义了一个用于描述上下文环境的上下文对象及其传输机制。同时，为了规范对上下文对象的描述，以及建立统一的上下文对象传输方法，协议规定了一套完整的 OpenURL 框架。Z39.88 协议具有很强的开放性，不同机构、不同领域的异构资源可通过对框架中各组件元素进行注册来实现对 OpenURL 的支持。

4.2.2 上下文对象

在 OpenURL 框架中定义一种叫上下文对象(context object)的信息结构，用于对链接的上下文环境进行描述：

(1)被参考资源(referent)：在一个特定的网络上下文环境中被参考的对象资源，它是上下文对象的核心元素，如果没有所指资源，上下文对象就没有产生的必要。

(2)参考者(referring entity)：在一个特定的网络上下文环境中对其他资源进行参考的实体。

(3)请求者(requester)：对被参考者发出服务请求的实体。

(4)服务类型(service type)：对服务类型进行定义的实体，如全文或摘要等。

(5)链接服务器(resolver)：服务请求所发向的目标，连接服务器以 OpenURL 为输入，接收到 OpenURL 后对其进行分析，评估传来的元数据并动态计算出合适的链接目标。

(6)上下文对象产生者(referrer)：产生上下文对象的实体。

4.2.3 OpenURL 的传输机理

在 OpenURL 方式中，对其他资源进行参考的一方并不直接产生指向参考目标的连接，而是通过一个挂接点产生一个 http 请求，该请求就是 OpenURL。OpenURL 以 http 协议的 get/post 方法将上下文对象提交到一个第三方的链接服务器，该链接服务器接收到 OpenURL 请求后，根据上下文对象动态计算出链接的目标。为了实现这种上下文敏感的链接服务，必须解决以下两个问题：

(1)判断出当前使用者是否有这个服务的使用权。为了避免在一个没有资源使用权的上下文环境中显示出一个无效的链接点，必须首先得到用户当前上下文环境的描述，然后

根据上下文环境判断是否有使用权；对于有使用权的显示一个链接点，否则不显示。为了达到这一目的，可采用 CookiePusher 用户配置文件等多种方法来实现。

(2)对有权限的用户为每个元数据对象提供一个 OpenURL 链接。

4.2.4 函数原型

```
CInternetSession: OpenURL  
CStdioFile * OpenURL(LPCTSTR pstrURL, //文件 URL 地址  
DWORD dwContext=1, //上下文 ID  
DWORD dwFlags=INTERNET_FLAG_TRANSFER_ASCII, //标记  
LPCTSTR pstrHeaders=NULL, //发送到服务器的数据头  
DWORD dwHeadersLength=0 ); //发送到服务器的数据头长度
```

4.2.5 返回值说明

该函数的返回值根据 pstrURL 的服务类型而改变，详见表 1。

表 1 OpenURL 返回值

URL type	Returns
file: //	CStdioFile *
http: //	CHttpFile *
gopher: //	CGopherFile *
ftp: //	CInternetFile *

参 考 文 献

[1] MEALLING M, DENENBERG R. Report from the joint W3C/IETF. URI planning interest group: uniform resource names (URNs): clarifications and recommendations[J/OL]. <http://www.rfc-editor.org/pipermail/rfc-dist/2002-August/000067.html>.

[2] MASINTER L, MCLAHILL M, BERNERS-LEE T. Uniform resource locator (URL)[J/OL]. <http://www.faqs.org/rfcs/rfc1738.html>.

[3] SUN S X, REILLY S, LANNOM L, et al. Handle system protocol(ver 2.1) specification[J/OL]. <http://www.faqs.org/rfcs/rfc3652.html>.

[4] NISO Z39.88 open URL 1.0 <http://www.niso.org/standards/index.htm> 2004.

[5] IANA The Internet Assigned Numbers Authority 互联网数字分配机构[DB/OL]. <http://www.iana.org>.