
国家图书馆数字资源 唯一标识符规范和应用指南

国家图书馆出版社

本书编委会

主编：孙坦 宋文 贺燕

编写：刘峥 周静怡 毛军 梁娜 刘细文 王昉 李欣 龙伟 李志尧
杨东波 延卫平 翟喜奎 韩超 曾燕 刘明杰

目 录

第一部分 国家图书馆数字资源唯一标识符规范.....	8
1 引言.....	8
2 范围.....	9
3 规范性引用文件.....	9
4 术语和定义.....	9
5 CDOI 体系框架结构.....	11
6 语法规则.....	12
6.1 一般规则.....	12
6.2 CDOI 前缀.....	13
6.3 CDOI 后缀.....	14
6.4 CDOI 的显示.....	14
6.5 CDOI 的字符集与编码规定.....	14
6.6 CDOI 名称的推荐长度大小写规定.....	15
7 命名规则.....	15
7.1 范围.....	15
7.2 粒度.....	16
7.3 唯一性.....	16
7.4 持久性.....	16
8 元数据规则.....	16
8.1 一般规则.....	16
8.2 对 CDOI 系统元数据的要求.....	16
9 管理规则.....	17
9.1 CDOI 系统的管理体系.....	17
9.2 CDOI 系统中的角色和任务.....	18

10	解析规则.....	19
10.1	一般规则.....	19
10.2	简单解析.....	19
10.3	多重解析.....	19
10.4	反向解析.....	19
11	扩展规则.....	20
11.1	CDOI 名称扩展规则.....	20
11.2	CDOI 与其它标识符方案.....	20
第二部分 国家图书馆唯一标识符应用指南.....		22
12	国内外唯一标识符及其应用现状.....	22
12.1	国外唯一标识符应用.....	22
12.1.1	URI、URN 和 URL	22
12.1.2	SICI、BICI 和 PII	22
12.1.3	DOI	23
12.1.4	PURL	24
12.1.5	INFO URI	25
12.1.6	TagURI.....	26
12.1.7	Handle	27
12.1.8	OpenURL.....	28
12.1.9	美国国会图书馆.....	29
12.1.10	大英图书馆.....	30
12.2	国内唯一标识符应用.....	30
12.2.1	CALIS	30
12.2.2	CADAL.....	32
12.2.3	NSTL.....	32
12.2.4	中国医学科学院.....	34
12.2.5	万方数据.....	35

12.2.6	清华同方.....	35
12.2.7	上海图书馆.....	36
12.3	专业领域唯一标识符应用状况.....	36
12.3.1	数字出版.....	36
12.3.2	学术出版物.....	36
12.3.3	数字权益管理.....	37
12.3.4	电子图书.....	37
12.3.5	开放获取.....	38
12.3.6	数字科研.....	38
12.3.7	远程教育.....	39
12.3.8	电子政务.....	40
12.3.9	电子商务.....	40
12.3.10	数字文化.....	41
13	国家图书馆信息资源和应用系统唯一标识符应用现状.....	42
13.1	国家图书馆数字资源基本状况.....	43
13.1.1	国家图书馆外购资源.....	43
13.1.2	国家图书馆自建资源.....	43
13.2	国家图书馆数字资源唯一标识符的状况.....	44
13.2.1	国家图书馆外购数字资源的唯一标识符.....	44
13.2.2	国家图书馆自建数字资源的唯一标识符.....	45
13.3	国家图书馆应用系统唯一标识符的应用情况.....	46
13.3.1	自动化系统.....	46
13.3.2	信息门户.....	47
13.3.3	自建系统.....	48
13.3.4	网络搜索引擎.....	48
13.4	小结.....	48
14	国家图书馆唯一标识符实施.....	49

14.1	概述.....	49
14.1.1	目的和作用.....	49
14.1.2	CDOI 与唯一标识符.....	49
14.1.3	CDOI 标识的对象.....	50
14.1.4	元数据.....	50
14.2	唯一标识符服务系统技术框架.....	51
14.2.1	Handle System 体系架构	52
14.2.2	国家图书馆唯一标识符服务系统体系框架.....	54
14.3	唯一标识符解析服务.....	54
14.3.1	什么是解析.....	54
14.3.2	CDOI 的解析过程.....	55
14.4	唯一标识符生成与注册.....	56
14.4.1	总体流程.....	56
14.4.2	唯一标识符生成与注册过程中的用户角色.....	58
14.4.3	唯一标识符生成与注册过程.....	58
14.4.4	唯一标识符生成的例子.....	59
14.5	国家图书馆唯一标识符名称分配规则.....	61
14.5.1	唯一标识符分配的基本原则.....	61
14.5.2	各种特殊对象的唯一标识符分配.....	61
14.5.3	唯一标识符后缀分配规则.....	62
14.6	唯一标识符的维护.....	63
14.6.1	重复控制.....	63
14.6.2	对象移动.....	63
14.6.3	对象删除.....	64
14.6.4	唯一标识符维护.....	64
14.7	唯一标识符对相关系统的要求.....	64
14.7.1	对唯一标识符服务系统的要求.....	64

14.7.2	对国家图书馆元数据规范和集中元数据系统的要求....	68
14.7.3	对国家图书馆数字资源服务系统的要求.....	68
14.8	国家图书馆唯一标识符系统运行管理.....	69
14.8.1	国家图书馆唯一标识符管理体系.....	69
14.8.2	国家图书馆唯一标识符管理中心.....	69
14.8.3	国家图书馆唯一标识符注册用户职责.....	70
14.8.4	国家图书馆唯一标识符注册用户划分方案.....	71
附录 A (规范性附录): CDOI 元数据		72
A1 CDOI 描述型元数据		72
A2 CDOI 管理型元数据		72
附录 B (规范性附录): 结构类型词汇		72
附录 C (规范性附录): 资源类型代码		73
C1 标码规则		73
C2 扩展规则		73
C3 资源类型代码主表		74
附录 D (规范性附录): 文献情报机构代码.....		76
D1 编码规则.....		76
D2 部分文献情报机构代码.....		77
D3 机构变更.....		78
附录 E (资料性附录): CDOI 名称中的保留字		79
引用文献.....		82
参考文献.....		85
后 记.....		86

第一部分 国家图书馆数字资源唯一标识符规范

1 引言

《国家图书馆数字资源唯一标识符规范》是国家数字图书馆工程标准规范项目研制成果之一。

本规范主要参考了 ISO/CD 26324 《信息与文献—数字对象标识符》(Information and documentation—Digital Object Identifier (DOI))，部分内容也参考了国际 DOI 联盟编制的《DOI 手册》(The DOI Handbook)、科技部科技基础条件平台专项资金项目成果《数字资源唯一标识符解析系统应用规范》。

本规范由国家图书馆提出，委托中国科学院文献情报中心起草。

数字对象唯一标识符系统是实体（也称为对象）持久标识的基础设施。本规范中的数字对象唯一标识符指的是实现实体标识的完整体系。DOI 是数字对象唯一标识符 (Digital Object Identifier) 的英文首字母缩写，CDOI 是中国数字对象唯一标识符 (Chinese Digital Object Identifier) 的英文首字母缩写。本规范的主要目的是建立中国国家图书馆数字资源唯一标识符系统，同时也兼顾了建立中国数字对象唯一标识符体系的需要，以便将来有可能能在此规范的基础上发展中国数字对象唯一标识符规范。

实体被赋予 CDOI 名称，CDOI 名称将持久地用来标识该实体，通过 CDOI 名称可以持久地链接到该对象的当前描述信息、对象的存放地信息或如何发现并获取该对象的有关信息。

CDOI 系统的应用主要包括：持久唯一地标识实体；管理实体的存放地址和获取方式；管理实体的元数据；促进电子事务处理等。

CDOI 系统提供网络环境下可解析的、持久的、可语义互操作的标识机制，能够在其上提供各种服务和事务处理。CDOI 名称能够解析到关于对象的各种信息，如 email 地址、对象的其他标识符、对象的权限描述等。也可以被解析到多个对象相关值，如通过多重解析机制链接到对象的多个 URL 地址。

CDOI 标识对象的元数据信息是 CDOI 系统的一个有机组成部分。实体用元数据进行描述，核心的元数据模型和各种元数据方案的映射，是实现对象之间互操作的基础。

2 范围

本规范规定了国家图书馆数字资源唯一标识符系统（简称 CDOI）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。

CDOI 名称用于标识国家图书馆的数字资源、物理资源、抽象资源（如作品、概念）、虚拟网络资源以及元数据资源。通过国家图书馆给资源对象赋予的唯一标识符，可以实现国家图书馆资源对象的永久标识和持久链接。

CDOI 是对象的数字标识符，而不仅是数字对象的标识符。CDOI 与 DOI 兼容，即在语法规则、命名规则、解析规则方面遵从 ISO/CD 26324 标准，但 CDOI 系统是独立于 DOI 的系统，不受 DOI 系统管理规则的约束。

CDOI 系统不试图替代现有的标识系统，如 ISO TC46/SC9 的 ISBN，ISSN，ISAN，ISRC 等，也不试图替代其他已被共同接受的标识系统。如果对象已经用其他标识符标识，可以将这个标识符的名称字符串集成到 CDOI 名称中，除非该标识符的注册授权有约束或注册授权之间的集成机制不能达成一致。

本规范不规定唯一标识符的实现技术，中国国家图书馆数字对象唯一标识符的实现技术路线在本书第二部分国家图书馆唯一标识符规范应用指南中说明。

3 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其随后所有的修改（不包括勘误的内容）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 2260 中华人民共和国行政区划代码

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

RFC 3986 Uniform Resource Identifier (URI): Generic Syntax

RFC 3629 UTF-8, a transformation format of ISO 10646

4 术语和定义

下列术语和定义适用于本规范。

4.1

CDOI 名称 CDOI name

在 CDOI 系统内定义一个唯一对象（标识对象）的字符串。CDOI 名称可以

由遵循 CDOI 语法规则的字母数字字符组成。

4.2

CDOI 注册代理 CDOI registration agencies

是中国唯一标识符管理中心指定的机构,提供 CDOI 名称前缀的分配、CDOI 名称的注册服务, 以及提供必要的设施供注册者注册和维护元数据。

4.3

CDOI 解析 CDOI resolution

提交一个 CDOI 名称到网络服务, 接受返回的一条或多条与标识对象相关的信息(如对象的 URL 地址或元数据)的过程。解析过程包括一步或多步映射操作。解析结果可能返回标识对象的实例, 也可能不返回。多重解析则以预定义的数据结构和格式返回与标识对象相关联的多条信息。

4.4

CDOI 语法 CDOI syntax

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

4.5

CDOI 系统 CDOI system

CDOI 名称作为计算机可识别的标识符, 在名称分配、解析、标识对象描述、管理等方面的应用部署。

4.6

持久性 Persistent

不受时间期限影响而存在, 能够脱离标识符分配者的直接控制被应用在各种服务中。

4.7

对象 Object

CDOI 系统范围内的实体。实体的形态可以是抽象的、物理的或数字的, 这些实体在内容上可能具有相关性(例如人、资源和协议)。由一个专门的 CDOI 名称标识的特定对象是该 CDOI 名称的标识对象(referent)。

4.8

唯一标识 Unique identification

有且仅有一个标识对象被该 CDOI 名称指定。

4.9

元数据 Metadata

元数据由关于对象的陈述组成，有助于发现、确认、获取、管理、评价和保存对象。

5 CDOI 体系框架结构

5.1 CDOI 标准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。CDOI 系统的体系框架结构图见图 5-1。

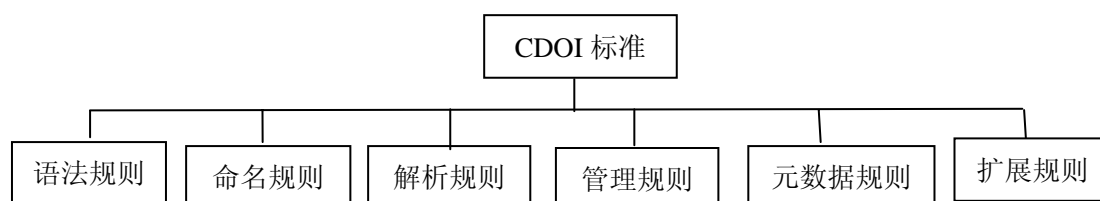


图 5-1 CDOI 系统的体系框架结构

5.2 语法和命名规则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。

5.3 解析规则规定从应用系统中的唯一标识符出发，通过唯一标识符的解析系统，到达应用系统目标对象的唯一标识符解析的流程和机制。

5.4 管理规则规定唯一标识符名称分配体系中相关各方的角色和任务，唯一标识符的管理和使用政策等方面内容。

5.5 唯一标识符的元数据规则是对各类对象的统一描述规范，元数据提供对对象的识别、确认，提供用户从对象的元数据角度检索对象的唯一标识符。

5.6 扩展规则说明唯一标识符前缀的扩展原则、唯一标识符与其他标识符的关系。

6 语法规则

6.1 一般规则

6.1.1 CDOI 是为达到 CDOI 系统的目的而建立的一个无任何含义的字符串。一个 CDOI 名称标识唯一的一个数字对象。CDOI 语法规则规定了组成 CDOI 的字符的形式和顺序。

6.1.2 CDOI 由以下三部分组成：

- a) 前缀元素；
- b) 分隔前后缀的向前斜线 “/” (‘\0x2F’);
- c) 后缀元素。

6.1.3 前缀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。

6.1.4 后缀是本地命名授权，在同一前缀命名空间下，后缀名应是唯一的。

6.1.5 CDOI 必须是唯一的。CDOI 的唯一性通过全球唯一标识符注册中心、中国唯一标识符管理中心、机构或地区注册代理等逐级命名授权机制保证。通过注

册授权和注册代理的管理方式,保证逐级授权的权利是唯一的,从而保障了 CDOI 的唯一性,其管理授权机制可详见第 9 章 管理规则。

6.1.6 CDOI 名称应遵循 URI 语法规则。为了使唯一标识符能被浏览器处理,CDOI 名称中不宜采用 URI 中的保留字等特殊字符。

6.1.7 CDOI 名称字符串以及前后缀元素的长度都没有限制。在实际应用中,CDOI 名称字符串可能受到相关通讯协议的限制。

6.1.8 CDOI 是不区分大小写的,可以是任何 ISO/IEC 10646 字符集中的任何字符。

6.2 CDOI 前缀

6.2.1 CDOI 的前缀是命名授权。命名授权由多个非空的命名授权段来构成,每个段由字符“.”(‘\0x2E’)分隔。

示例:已有命名授权“abc”,需要在其下建立子一级命名授权“xyz”,则新的命名授权应为:“abc.xyz”。

6.2.2 CDOI 的前缀由二个层次的命名授权段组成:

- a) 第一个命名授权段是中国唯一标识符命名授权;
- b) 第二个命名授权段是行业/系统/地区/机构命名授权。在本规范的附录 D 中,编制了部分文献情报机构的编码,可作为文献情报机构分配唯一标识符时的二级命名授权段编码。例如,国家图书馆编码是“011001”。

前缀构成如图 6-1 所示。

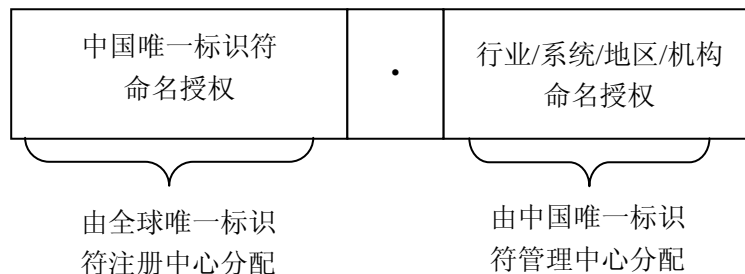


图 6-1 唯一标识符前缀构成

6.2.3 CDOI 一旦分配,CDOI 名称字符串,包括它的机构代码元素和资源类型

元素将不再改变。

6.3 CDOI 后缀

6.3.1 CDOI 后缀跟在前缀元素之后，与前缀之间用斜线“/”分隔。如

“cdoi.011001/123456”

6.3.2 在同一个前缀元素下的每个后缀都应是唯一的。

示例：一个具有前缀为“CDOI.011001”，后缀为“123456”的 cdoi:cdoi.011001/123456

6.3.3 唯一的后缀可以是一个顺序号，也可采用注册者使用的另一个系统的标识号（如 ISBN，ISSN，ISTC），参见第 11 章扩展规则。

6.4 CDOI 的显示

6.4.1 当 CDOI 在屏幕上显示或打印时，CDOI 名称前放置一小写的“cdoi:”，其中冒号为半角。示例：CDOI 名称“cdoi.011001/12354”显示为：

cdoi:cdoi.011001/12354

6.4.2 小写字符串“CDOI”的使用应遵循 URI（Uniform Resource Identifier）的表示规范（如“ftp:”和“http:”）。

6.5 CDOI 的字符集与编码规定

6.5.1 除了在前缀中要求保留的字符以外，唯一标识符可使用 ISO/IEC 10646 字符集作为 CDOI 构成的基本字符集，并规定 CDOI 字符串使用 UTF-8 规则编码。

6.5.2 CDOI 名称可通过 Web 浏览器或浏览器插件解析，此时 CDOI 名称应遵循 IETF 的 URI 表示指南标准。

示例：CDOI 名称“cdoi.011001/12354”，解析服务器地址是 <http://cdoi.gov.cn/>，CDOI 名称表示为活动超链接是 <http://cdoi.gov.cn/cdoi.011001/12354>

6.5.3 由于唯一标识符经常需要有浏览器来处理，在命名唯一标识符前缀和后缀的时候不宜使用保留字符、不可打印显示的控制字符。如果不能避免，需要网页浏览器或 Http 代理服务器对唯一标识符进行解析时，使用十六进制编码该字符。

十六进制编码的唯一标识符的转码工作一般由解析该唯一标识符的浏览器、浏览器插件或应用程序来完成。CDOI 中不宜使用的保留字见附录 E。

6.6 CDOI 名称的推荐长度与大小写规定

6.6.1 CDOI 名称字符串以及前缀和后缀元素的长度都没有限制。

6.6.2 当 CDOI 在使用 URL 表示和经由 HTTP 协议传输时，URL 长度要受到客户端浏览器或服务器端能处理的最大字符长度限制。为避免 CDOI 字符串在该种情况下被浏览器截断，CDOI 名称的长度不宜超过目前所知客户端浏览器或服务器端能处理的最大字符长度限制中最小的长度 2048 个字符、并减去实际路径中的字符数。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，CDOI 名称推荐的保守长度不宜超过 1793 个字符。

6.6.3 唯一标识符的解析系统应是不区分大小写。

示例：test/abc 和 Test/abC 应作为同一个标识符。

7 命名规则

7.1 范围

7.1.1 任何书后当需要将一个对象作为一个被识别的实体时，可给这个对象分配一个 CDOI 名称。

7.1.2 注册代理可规定更具有约束性的 CDOI 分配细则，以满足一些特定服务的需要。当注册代理有特殊说明的 CDOI 分配细则时，这些细则应与整个 CDOI 系统规范兼容，但是这些细则不宜作为 CDOI 规范的一部分。

7.1.3 CDOI 名称能分配给任何类型、格式、载体的对象。CDOI 名称分配的主要对象包括但不限于：文本文档、数据集、录音带、图书、照片、连续出版物、音频资源、视频资源、音视频资源、软件、抽象作品、艺术作品等，以及在管理这些对象过程中的相关实体，如许可、相关方等。

7.1.4 CDOI 名称不应替代对象的其他标识符方案，其他标识符仍然是该对象的

标识符并发挥相应的标识作用。

7.2 粒度

7.2.1 CDOI 名称能在对象的任意粒度或抽象层次上进行分配。

7.2.2 例如能给作为抽象著作的小说、这部小说的一个特定版本、这一版本的特定一章、一个段落、一张图片或一条用语，以及小说每个特定的载体实现，或者是注册者认为合适的任何其他粒度独立分配 CDOI 名称。

7.3 唯一性

7.3.1 在 CDOI 系统中，一个 CDOI 名称唯一的标识一个对象。

7.3.2 尽管每个被标识对象应有且仅有一个 CDOI 名称，但实际上可能会发生一个对象有多于一个 CDOI 名称的情况。一个对象被两个不同的注册机构分配 CDOI 名称，注册机构之间应进行有效合作以提供统一的元数据记录和一致的 CDOI 名称。

7.4 持久性

7.4.1 CDOI 名称应不受时间限制。

7.4.2 CDOI 名称和它的标识对象之间的联系不应随对象权限的变更或对象管理责任的变更而改变。

8 元数据规则

8.1 一般规则

CDOI 名称注册时，注册者应同时注册 CDOI 名称所标识的对象的核心元数据信息。

8.2 对 CDOI 系统元数据的要求

8.2.1 用于描述 CDOI 标识的对象的元数据应能描述各种类型的知识资产。

8.2.2 应具有与其他不同类型的知识资源的复杂元数据的互操作能力，能将其他

元数据映射到 CDOI 系统的元数据。

8.2.3 具有与国际、国内相关元数据标准的互操作能力，实现更大范围的知识共享与互操作。

8.2.4 任意的描述粒度。能在需要的时候在任何粒度层次上标识一个实体。

8.2.5 分配授权。元数据的来源能被确认和识别。

8.2.6 CDOI 系统元数据方案见附录 A。

9 管理规则

9.1 CDOI 系统的管理体系

9.1.1 CDOI 由中国唯一标识符管理中心（Chinese DOI Center，简称 CDC）进行管理。

9.1.2 在 CDC 统一管理下，建立行业、系统、地区或机构的注册中心（注册代理），负责唯一标识符的注册。CDC 体系框架见图 9—1 所示。行业、系统、地区或机构注册中心作为注册代理，发展唯一标识符的用户，负责唯一标识符的分配与注册。比如出版行业可以成立一个行业的注册中心，负责出版行业各出版社出版的资源的唯一标识符的分配和注册服务。也可以成立机构注册中心，负责本机构的唯一标识符的分配与注册，比如国家图书馆唯一标识符注册中心。

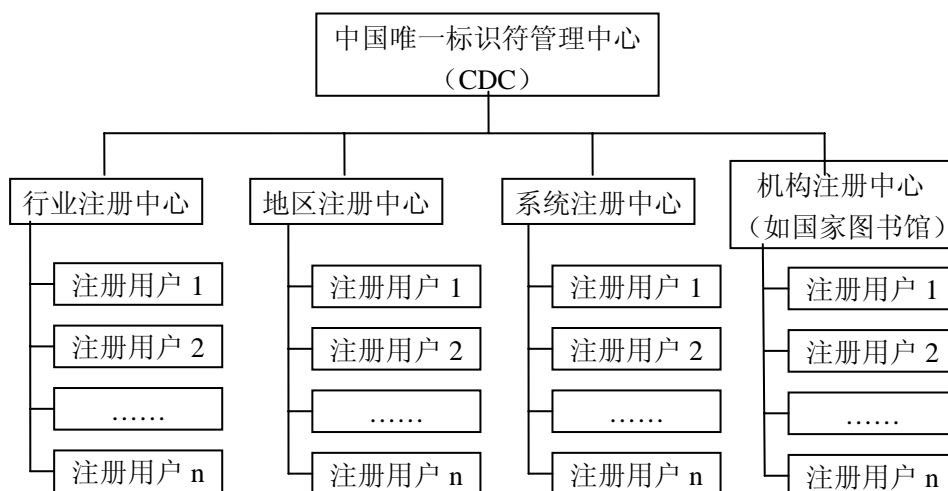


图 9—1 中国唯一标识符管理框架

9.2 CDOI 系统中的角色和任务

CDOI 系统由中国唯一标识符管理中心（CDC）统一管理，CDC 作为注册授权，指派行业注册中心、地区注册中心、机构注册中心等注册代理为 CDOI 注册提供服务。

9.2.1 注册授权——中国唯一标识符管理中心（CDC）

注册授权（CDC）提供以下的服务和功能：

- a) 负责中国大陆地区唯一标识符服务系统的稳定可靠和持续运行；
- b) 维护唯一标识符服务系统，分配唯一标识符前缀和发展行业/系统/地区/机构注册代理；
- c) 促进国内数字出版、图书情报、档案和博物馆系统的网络应用系统的互操作和规范化；
- d) 代表中国同国外同等机构进行交流与合作。

9.2.2 注册代理

注册代理提供以下的服务和功能：

- a) 在中国唯一标识符管理中心注册唯一标识符前缀命名授权段。
- b) 接受、审核各注册用户编制的资源类型代码和后缀编制规则。
- c) 依据 CDC 的规则，用唯一标识符服务系统和数据管理政策为注册用户、CDOI 命名和相关的元数据注册提供服务，CDOI 名称选择或制定 CDOI 元数据标准。
- d) 分配并验证 CDOI 系统的用户。包括获取和验证用户的信用资质。信用资质由用户名和密码组成。
- e) 当在统一命名授权下有多个注册用户给同一对象分配重复的 CDOI 名称时，协调各注册用户，尽量保证同一对象分配一个 CDOI 名称和相同的元数据注册。
- f) 当其他行业、系统或机构的注册代理机构给同一对象分配重复的 CDOI 名称时，应与 CDC 合作提供对于该对象统一的元数据记录。

10 解析规则

10.1 一般规则

10.1.1 解析是向 CDOI 系统递交一个 CDOI 名称，接受系统反馈的与 CDOI 名称所标识的对象相关的一个或多个数值的过程。

10.1.2 解析可返回对象实例，也可以返回对象的元数据信息。

10.1.3 CDOI 解析记录可包含多个对象存放地址（URL）和对象的其他信息，例如对象名称、其他标识符、描述、类型、主题、地址、与其他对象的关系等。

10.1.4 解析过程可包括一个或多个中间的映射操作。

10.2 简单解析

10.2.1 简单解析是单一的指向性解析。

10.2.2 每个唯一标识符只有一个可解析到的 URL 地址，解析结果返回对象的实例。

10.3 多重解析

10.3.1 多重解析允许一个实体被解析到该实体的多个 URL 地址、多个数据类型/值对、多个其他相关实体，如父子关系和其他各种关系。

10.3.2 一个唯一标识符解析到多个结果，用户可选择一个解析结果。

10.4 反向解析

10.4.1 反向解析是从唯一标识符对象的元数据出发，通过解析服务系统，反馈唯一标识符名称的过程。

10.4.2 唯一标识符服务系统应能注册对象的元数据并提供元数据检索功能。由于元数据检索的不唯一性，反向解析可反馈多个唯一标识符名称，供用户选择。

11 扩展规则

11.1 CDOI 名称扩展规则

11.1.1 CDOI 名称前缀的命名授权 (Naming Authority) 由多个非空的命名授权段来构成, 目前 CDOI 规范关于前缀命名授权段由两个层次的命名授权段组成:

- a) 中国唯一标识符命名授权;
- b) 行业/系统/地区/机构命名授权。

11.1.2 当前缀中第二个层次的命名授权段(由中国唯一标识符管理中心分配的注册机构代码部分)不能满足实际中对注册机构分类和唯一标识的需求时, 可对其进行扩展。

11.1.3 对命名授权进行扩展时, CDOI 解析系统应在技术上支持新的命名授权段分配方法。

11.1.4 命名授权段扩展的语法: `name-authority[.sub-name-authority*]`

11.1.5 前缀名称是一个全局性配置, 一旦一个机构的前缀名称已投入使用, 该前缀不宜随便扩展、修改。如确实需要修改, 需要在唯一标识符的分配、注册、服务等方面进行全面评估。

11.2 CDOI 与其它标识符方案

11.2.1 唯一标识符系统可将其它系统中应用的标识符方案引用到自身标识符中, 包括采用其它 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。

示例 1: 原有基于 ISSN 的标识符 1476-4687 用于标识某种期刊, 可将 ISSN 字母小写后作为后缀继承到一个 CDOI 前缀下形成一个期刊唯一标识符, 如: `cdoi.011001/issn.1476-4687`;

示例 2: 使用 ISBN 号作为 CDOI 后缀, 如 ISBN: 9787802253605, 将 ISBN 号变形引入方式可以是: `cdoi.011001/isbn.9787802253605`;

示例 3: 使用应用系统自定义的标识符作为 CDOI 后缀, 如一张博士论文 TIFF 光盘的内部编号是: BSLW040687, 可能的引入 CDOI 方式是: `cdoi.011001/bslw040687`。

11.2.2 当一个唯一标识符标识的对象同时被另一种广泛认可的其它标识符方案所标识（如 ISBN, ISRC, ISSN），应在 CDOI 元数据的标识符字段中登记该标识符。如果标识对象同时被多个其它标识符方案标识，可重复著录标识符字段。

第二部分 国家图书馆唯一标识符应用指南

12 国内外唯一标识符及其应用现状

12.1 国外唯一标识符应用

12.1.1 URI、URN 和 URL

URI(Uniform Resource Identifier 译为统一资源标识符)是所有标识因特网资源的地址和名称的通用集合。它的基本特点体现在统一、资源和标识符 3 个方面。统一保证了现有不同类型的资源标识符可以采用统一的方式使用,并为新的资源类型提供统一的标识框架;资源是能够被标识的任何对象(资源的另一种说法是对象);标识符是一个对象可以引用的名称。URI 将唯一标识符定义为符合某种语法规则的字符串。URN 是统一资源名称,它可以是某个机构和联盟对资源的统一命名,如 PURL;也可以是采用某种命名大纲制定的持久、独立于地址的资源标识符,如 RFC2141 规定的 URN 命名大纲。URN 在某种情况下可以是 URL;URL 是统一资源定位器。代表因特网上某个具体的地址,它是同某种命名大纲联系的非正式的资源名称;URI 既可以标识一个名称,也可以标识一个地址,或者同时标识两者。URI、URN 和 URL 的关系如图 12-1 所示^[1]:

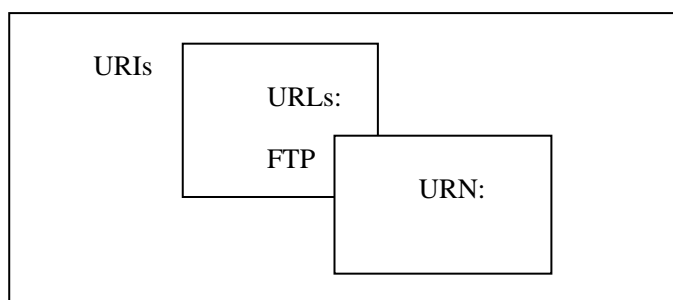


图 12-1 URI、URN 和 URL 的关系

12.1.2 SICI、BICI 和 PII

SICI (Serial Item and Contribution Identifier) 是连续出版物及其单篇文献的标识。SICI 标准是美国颁布的国家标准^[2], 目的在于对连续出版物及其各个组成部分进行唯一标识。SICI 适合标识不同层次、不同类型和不同载体的期刊文献, 并且对原有的 ISSN 进行了较好的兼容。SICI 的基本结构分为期刊物件段、内容段和控制段, 见图 12-2:

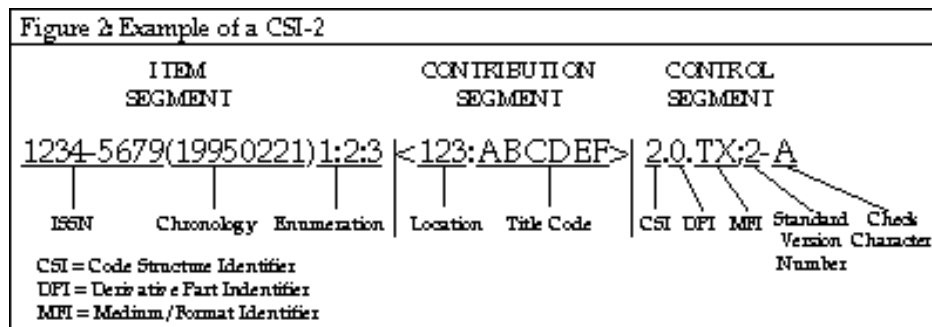


图 12-2 SICI 基本结构

BICI (Book Item and Contribution Identifier) 是参照 SICI 提出的标识图书及图书内的内容片断 (章节、前言、索引、段落) 的唯一标识符^[3]。基本结构同 SICI 类似, 只是将 ISSN 号改为 ISBN 号。

PII (Publisher Item Identifier) 出版物件标识符, 是美国科技信息出版集团 1995 年制定的唯一标识符^[4]。它的基本结构比 SICI 简单, 主要包括 ISSN、ISBN、年代和内部标识号。美国化学学会、美国物理学会、IEEE、Elsevier Science 等出版商均加入 PII。

这三种标识符详细地定义了其构成规则并由特定机构对其进行登记管理, 但是在制定时并没有为其设计一套用于 Internet 环境的分布式的解析和管理机制。由于这些标识符已经形成一套规范的标识符命名规则, 并且已经投入到实际应用中。因此, 在实际应用中可以保留原来的标识符字符串, 通过由使用该标识符的机构或系统自定义的协议和约定来实现标识符与对应资源之间的实际链接关系, 从而继承到基于唯一标识符解析的系统环境中。如在 DOI 系统中可以使用 SICI 实现参考文献链接, SICI 也可以在 OpenURL 体系中实现不同文献源的链接服务。

12.1.3 DOI

DOI 的全称为 Digital Objects Identifier, 意思是数字化对象标识符, 是由一系列数字、字母或其他符号组成的字符串, 是 DOI 系统^[5]中的唯一标识符。DOI 主要是针对因特网环境下如何对知识产权进行有效的保护和管理而产生的。在实际应用中, DOI 的发展相对于 SICI、BICI、PII 等更为成功。目前国外的多数全文数据库均采用 DOI 号码来作为文章的唯一标识符方案。1998 年 AAP (美国出版协会) 创立非盈利性组织 IDF, IDF 在 CNRI 的配合下, 制定了 DOI 标准和相应的解析系统 Handle System。DOI 从最初提供持久可靠的名称重定向服务, 发展到目前拥有上千万个已经分配并解析的 DOI 号码, 9 个 RA (DOI 注册代理机构) 和几百个使用单位, 跨越了美国、欧洲和澳大利亚。应用领域也扩展到政府部门和非英语的国家 (如韩国)。目前 IDF 提供的 DOI 服务包括: 名称解析, 基

于 INDEC 的元数据框架，连接元数据和解析技术的高级模板（AP），允许行业自治的政策框架和对新兴标准如 OpenURL 的支持。DOI 唯一标识符由前缀和后缀两个部分组成。前缀和后缀之间由“/”分隔。其中前缀是由 DOI 命名机构分配的号码，后缀是注册 DOI 的组织或个人对数字对象定义的本地标识符。一般 DOI 的注册者都通过在后缀中融入现有的唯一标识符如 ISSN、ISBN 或其他标识符来达到向下兼容。

DOI 技术第一次大规模的应用是 CrossRef，它是多个电子期刊出版商的同盟，由 12 个世界顶尖的 STM（科学、科技和医学）学术出版商于 2000 年 1 月成立，并于同年 9 月成为国际 DOI 基金会（IDF）所授权的第一个注册代理机构。CrossRef 自成立后，几年间不断发展壮大。目前全球近 150 家出版社加入了 CrossRef，提供数以千计的电子期刊和上百万的文献，CrossRef 同时还邀请图书馆以及其他信息服务商加入。但 CrossRef 的成员主要是出版商，无论出版商的规模大小，是商业公司还是非盈利组织，只要从事以数字方式出版学术文献的，都可以申请加入 CrossRef。目前 CrossRef 的主要成员包括：ACM、The Academic Press、The American Association for the Advancement of Science、American Institute of Physics、IEEE、Kluwer、Blackwell Science、Springer、Wiley、Elsevier Science 等国际知名出版商。

CrossRef 采用 DOI 系统实现参考文献链接^[6]。成员出版商根据 CrossRef 提供的前缀创建一个 DOI，并向 CrossRef 提交相应的 DOI、URL 和该信息单元的元数据(其中，元数据的基本集包括期刊名、ISSN 号、第一作者、出版年、卷、期、页码，此外出版商还可选择性提交附加的其它元数据信息)。CrossRef 将 DOI 和 URL 存入地址数据库(Location Database)，DOI 和元数据存入元数据库中(Metadata Database)。这样，用户通过某篇文献的 DOI 链接，可以获得该文献对应的 URL 和元数据，进而得到文献的摘要或全文。

如今 CrossRef 已经得到越来越多的出版商和学会的认可，目前已有多家数据库商的产品提供了 CrossRef 服务，如：Elsevier Science、Springer-LINK、Web of Science、EBSCO、John Wiley & Sons、IEL、Nature、ACM、IOP 等。

12.1.4 PURL

PURL(Persistent URL)是 OCLC 开发并管理的永久性名称解析系统。它的设计思路是用名称而不是 URL 来标识网络资源。因为 URL 对应的是网络上的地址，该地址是由协议名称+服务器域名+服务器本地文件路径 3 个部分组成，当服务器上的文档的位置发生变化时，该文档对应的 URL 就会无效，用户使用该 URL 访问资源就会得到“404 file not found”的错误。如果只是单独的文件转移，只需要

将原来的同名文件变成一个重定向的文件就可以保持 URL 不变，用户可以被自动转到新的 URL，但是文件位置的转移通常是按照目录成批的进行转移，为每个文件作一个重定向文件是一个十分复杂的工作。PURL 将这种重定向的功能从用户的服务器上转移并集中到域名为 purl.org 的服务器上，通过该服务器提供一种名称和 URL 的单向解析机制，即用户输入类似 http://purl.oclc.org/keith/home 的名称可以自动转换为该名称对应的 URL，用户再根据得到的 URL 去获得有关的资源。名称和 URL 的对应关系是通过名称注册用户进行维护和修改的^[7]。

PURL 并不是严格意义的唯一标识符，而是比普通的 URL 更持久的 URL。它采用一种名称解析机制和注册用户参与维护的方式来解决资源的持久链接问题。与其说 PURL 是技术上解决资源的持久链接，还不如说它采用了一种组织机制。因为 PURL 的技术很简单，只是利用了现有的 HTTP 协议上的重定向功能，而提供 PURL 服务的机构—OCLC 却可以长久的生存下去，当然，用户可以下载 OCLC 提供的 PURL 软件自己管理同自己组织相关的 PURL，这样这些 PURL 的生命周期就可以同组织的生命周期保持一致。

因为 OCLC 在 URN 的标准制定过程中扮演了活跃的角色，因此 OCLC 开发出来的 PURL 很自然的同 URN 保持良好的兼容特性。实际上只要将 PURL 作为 URN 的一个命名机构的代码就可以将其纳入到 URN 的通用名称服务框架中去。如图 12-3:

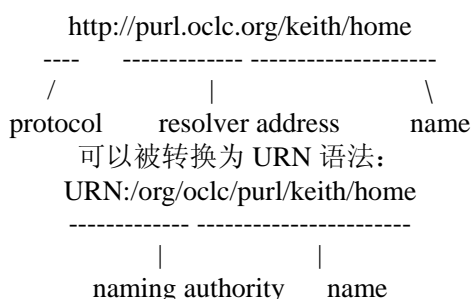


图 12-3 PURL 语法和 URN 的转换

PURL 虽然简单实用，但逻辑上并不严格。它无法解决多复本、独立性、分布式的命名和管理等多个问题。因此只是唯一标识符的权宜之计，并不能从根本上解决持久性、唯一性、分布性和可管理性等唯一标识的要求，而且 PURL 也不打算这样去做，这些问题需要更专门、更复杂的系统来解决。PURL 除了进行简单的名称到 URL 的解析工作之外，还在技术上进行了适当的扩展，如建立了主域和子域的机制，允许部分重定向，建立用户权限认证机制等。

12.1.5 INFO URI

URI(Uniform Resource Identifier, 统一资源标识符)是网络环境下的一种简

单和可扩展的标识机制。它由命名大纲(scheme)和根据该 scheme 产生的标识串两部分组成。INFO URI 是 URI 的一个子集^[8]。图书情报和出版界为了将现有的标识符机制在网络环境下整合,以有效的标识信息资产(Information assets),而现有的 URN 机制又因为技术和文化的原因未能提供公共命名手段,将已有的标识符系统如 DOI、PII、SICI 等融入到网络通用架构中,并同现有的技术如 XLink、RDF 和 Topic Maps 有效的结合以标识信息资产。

INFO URI 并不同现有的命名机制竞争,而是共同形成一个轻型的 URI 标识机制。INFO URI 作为名称空间的代表,用 URI 的格式将现有的标识符在因特网上公布,并应用于各种独立于网络应用的网络描述技术如 RDF 等。INFO URI 最大的优点就是独立标识(not dereferenceable),即它仅提供对已有的各种标识符进行整合的符合 URI 方式的标识,而不是同具体的记录和资源相关。简单的说:用户点击 INFO URI 不会有任何反应,而 http 链接则会链接到对应的网页。独立标识有如下优点:(1)简化 INFO URI 的操作,仅限于标识和比较;(2)避免在 INFO URI 上建立新的解析机制以及针对解析的管理和维护机制;(3)如果需要增加新的功能和应用,命名授权机构可以增加独立于 URI scheme。

和 INFO URI 类似的命名标识有 DATA URI 和 TAG URI。前者是在资源内部的标识,不适合网络资源;后者则是在网络名称授权的基础上建立起来的唯一标识机制。INFO 登记系统提供在公共领域发布现有系统唯一标识符名称空间的机制。在 URI 分配中它是独立标识的。它的地址是 <http://info-uri.info/registry>。该登记系统由 NISO 维护,NISO 公布了注册名称空间的政策,拥有并运行名称空间的机构提交自己的名称空间并进行登记。

12.1.6 TagURI

TagURI^[9]允许任何人用自己的方式生成因特网上唯一的标识符,且便于阅读和记忆。它也遵循 URI/URL 的语法规则。2005 年 2 月 24 日 IESG (Internet Engineering Steering Group) 接受其作为 RFC 文档。最早的 TagURI 大纲是由 Sandro Hawke 和 Tim Kindberg 分别于 2001 年独立提出的。

TagURI 目前应用在 Atom (网志数据同步格式,在 RSS 基础上)和 YAML 中。不直接采用 http 方式的 URL 的格式,如 <http://hawke.org/2001/06/05/Taiko>,原因是:首先 http 通常指网络上的某个具体的地址,而 TagURI 则要标识较抽象的对象;其次如果用户没有域名,则可以选择 email 方式作为前缀;最后 TagURI 应在一定时间内保持相对稳定,而 http 方式的域名则经常变更。TagURI 的主要目的是考虑作者如何生成唯一而持久的标识符,而没有考虑同其他人共享或复用。所以将前缀限定在域名和邮件,是简便起见。至于采用电话或者 IP 地址,

则限定过于复杂。TagURI 没有采用名称空间的方式也是因为它仅用于标识 xml 文档中的一个元素，而不是整个 xml 文件。将 RDF 方式引入 TagURI 也不如直接采用 http 方式的 URI 格式。

12.1.7 Handle

CNRI 推出的 Handle System (以下简称 HS) 是因特网上进行名称解析和管理的通用名称服务系统。HS 管理的主要对象就是 handle。handle 是一种基于名称的唯一标识符。HS 是针对目前因特网上 DNS 和 URL 对名称和地址唯一标识能力不足而提出的。它基本由解析系统和管理系统两大部分组成。解析系统把用户提供的 handle 和名字解析成同该 handle 相关的信息和值以使用户定位、访问和使用数字对象；管理系统则负责除了提供 handle 的用户对 handle 的有关信息进行编辑和修改，以维持名字同实际对象之间的关系^[10]。

HS 系统^[11]最早是 CNRI 作为 CSTR 项目的一部分在 DARPA 项目经费的资助下完成的，它的研究成果又进一步融入到后来的 NCSTRL 项目中。目前国会图书馆、美国国防科技信息中心和 IDF 等机构或组织采用 HS。HS 主要指的是包括数据结构、协议、管理机制和安全控制机制在内的整套应用系统规范。CNRI 在该规范的基础上已经开发出具体的试验用名字服务系统。该系统可以分为分布式的多个本地名称服务系统和一个叫做 Global Handle Registry 的中心名称登记系统。目前 CNRI 提供服务器端和客户端的软件，供研究人员使用测试。其中客户端软件分为 JAVA、C 的类库和 WINDOWS 平台上的名称解析器，目前最新版本是 2.1 Beta 版。

人类对计算力的需求产生了计算机，在需要一种方式进一步提高计算机计算能力的时候，多台计算机通过网络的“协作”是提高计算机计算能力的有效方式（如 C/S 模式的计算比集中式的计算能力更为强大）。现有的支撑计算机协作的机制（如 Internet）存在明显的缺陷。新机制必然带来计算能力的提升，计算能力的提升必然导致更多新的应用（如同 Internet 这种机制，极大的提高了计算能力，带来了许多新应用一样），这种机制称为网格(Grid)——动态多机构的虚拟组织中进行资源共享并协同问题解决(The Anatomy of the Grid: Enabling Scalable Virtual Organizations)。The Handle System(HS)在网格计算中应用切入点为：EPR 解析服务(EPR resolver service)：当描述网络资源的 EPR 修改以后，HS 可以提供持久的名称服务以确保最新的 EPR 被绑定。EPRs 在主机环境中更新模式有 de-activation/re-activation 两种，通过在不同机器上的重新部署、有关资源的迁移和归并等等。上层的 EPR 解析机制将在 GT4 的 WSRF 解析规范中阐明。通用的名称解析和登记服务：HS 的名称可以绑定任何层次的数字对象，包括应用程序，

如绑定 EPR 的政策并对外发布政策，将 Web 服务的元数据公开访问，对文件碎片和复制部分的全球跟踪等。

12.1.8 OpenURL

OpenURL^[12]是一种开放链接的框架，它通过所谓的链接服务器来提供链接服务。这种服务是信息源外部的、独立于信息源的。OpenURL 的两个核心特征是开放和上下文敏感（open and context-sensitive）。这也是针对目前因特网上单纯的 URL 链接的不足而提出的。这种不足主要体现在：（1）链接不是上下文相关的，并且往往是不恰当的；（2）链接严重的依赖特定的信息源，并且依赖信息提供商之间特殊的协定（如通过一种文摘数据库链接到另一种全文数据库）；（3）链接的内容往往是有限的；（4）链接难以及时更新和维护。OpenURL 是针对上述问题提出的，它采取的解决办法是提出了一种开放链接的框架，在该框架中通过建立链接服务器提供链接服务，并提出一套公共的 OpenURL 语法，允许信息源公开自己的链接接口，实现链接信息源和链接服务器之间信息的传输，从而实现异质数据库之间的互操作。因此 OpenURL 又可以看作异构系统之间互操作的详细规范。

以 CrossRef 为主的出版商联盟正在测试 OpenURL 同 DOI 系统的兼容性。在实际应用中，可以将 OpenURL 同 DOI/CrossRef 链接系统集成^[13]。

CrossRef 是一个建立在多家出版商协作基础上的参考链接系统。出版商形成联盟，共同制定并遵守 DOI（数字对象标识符），将出版商出版的文献用 DOI 进行统一标识；接下来出版商将生成的 DOI 和对应的元数据存储在 CrossRef 公共的数据库中，其他参与联盟的出版商可以查询包含 DOI 和元数据的 CrossRef 数据库，这样出版商可以通过某篇文献中的参考文献元数据获得匹配的 DOI 标识符。出版商可以在它们出版的电子文献中的引文部分加上 DOI 方式的链接，如 <http://dx.doi.org/10.1074/jbc.M004545200>，该链接是一个可操作的链接。如图 4 所示是 CrossRef 提供的参考链接方案，该方案是建立在 Handle System 的基础上的。用户点击上述链接后，将 DOI 唯一标识符作为参数传递到 DOI 代理服务器上，DOI 代理服务器是所有这类参考链接的目的服务器，然后 DOI 代理服务器将参数中的 DOI 标识符发送到 Handle System 处理，Handle System 将 DOI 标识符解析成相关的 URL，如 <http://www.jbc.org/cgi/content/abstract/275/44/34826>，用户通过 URL 就可以直接访问到参考文献的全文。整个过程如图 12-4 的“1、2、3”。

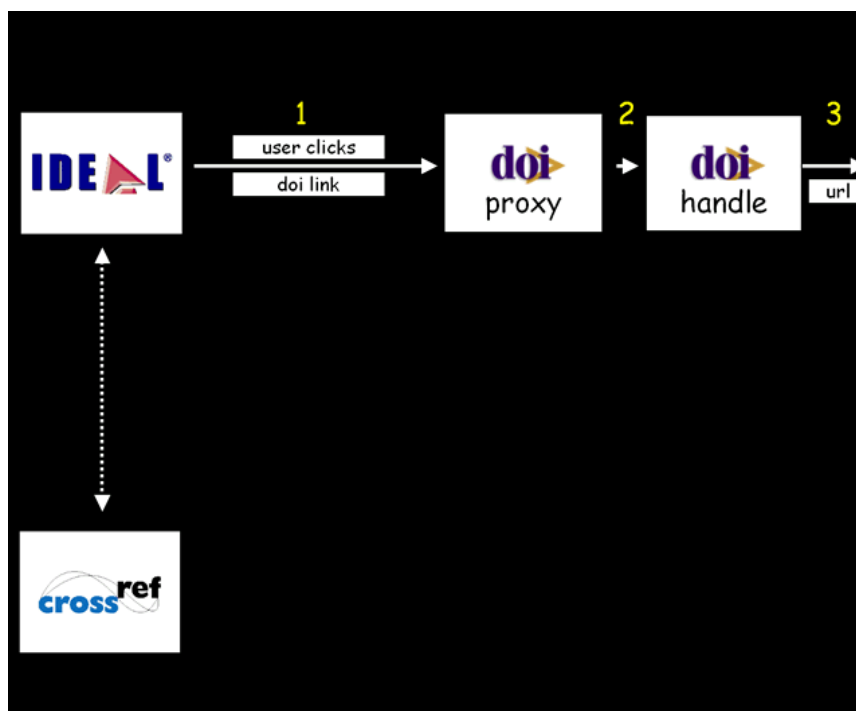


图 12-4 DOI/CrossRef 的链接方案

12.1.9 美国国会图书馆

美国国会图书馆的国家数字图书馆计划（NDLP）是美国回忆的后续项目，目标是 2000 年数字化美国国会图书馆的 500 万件藏品。NDLP 项目基于 Handle 系统实现数字化藏品的唯一标识。NDLP 对不同类型的藏品资料制定了相应的结构元数据和元对象的规范。所有数据都给定一个明确的数据类型；对所有的元数据显式编码；带有知识产权的每一个个体项都被赋予一个句柄，作为一个数字对象；元对象用于聚集数字对象，特别是版本和相关的结构元数据；句柄用于标识列在元对象中的项，元对象中的每一项都有一个句柄。

对于 NDLP 项目中数字化藏品的扫描图片，NDLP 的实施步骤如下^[14]：

(1)扫描图片的数字对象，对扫描后的高分辨率、中分辨率与低分辨率三个版本分别生成一个数字对象，创建一个元对象来描述该图片及其所有的数字化版本，其中包含三个版本各自的句柄和它们共同的元数据。

(2)每个版本的数字对象包括：关键元数据用于在网络中管理对象的元数据，包括句柄、相关的权限与许可；结构元数据包括描述字段、所有者、元对象的句柄、数据大小、数据类型、版本号、描述、存置日期等；图像本身。

(3)元对象，包括关键元数据与结构元数据。访问图片是从元对象开始，选取某个数字化版本。为了保证用户能直接访问每个版本，每个版本的数字对象都从元对象中复制一份权限和许可信息。

(4)扫描图片的句柄。最初 NDLP 给每一个数字化项赋予一个控制标识符。例 如某幅扫描图像的标识符为 3a16116r.jpg。其中“r”表示高分辨率图像，“jpg”表示 jpeg 格式。为使数字对象的标识符长期有效，NDLP 采用句柄来为数字对象提供一个唯一的、永久的、独立于位置的名字，如句柄“loc.ndllp.amrlp/3a16116”。

(5)扫描图片的保存，人工地选择待转换为数字对象的资料、人工地在相应的字段中给定元数据值；自动建立元对象及其指向其他数字对象的链接；自动将数字对象置入相应数据库中；在 HANDLE 系统中注册。

(6)对扫描图片的访问，首先查找系统通过元对象的句柄指向扫描照片，如果用户需要的是照片的概览，则低分辨率图像，如果用户没有指定照片的版本，缺省提供中分辨率图像。

例如在美国记忆项目中，对一个图片对象的描述包括：名称、创建/发布时间、描述、主题词、相关名称、载体介质、书目号、馆藏地与唯一标识符（例如 loc.gmd/g3850.ct000509）。如果该对象有副本存在，通过唯一标识符还可浏览副本对象。如果原对象是低分辨率图像，那么副本就是中分辨率图像或者高分辨率图像。

12.1.10 大英图书馆

ISSN 国际标准连续出版物号的英国中心设在大英图书馆内，负责在英国出版的连续出版物 ISSN 的分配与管理、ISSN 的使用、以及英国连续出版物书目数据送交 ISSN 国际中心数据库等工作。同时，大英图书馆也是较早加入了 DOI 组织的大型图书馆。

早在 2001 年大英图书馆就出版了数字资源长期保存手册（*Preservation Management of Digital Material Handbook*）^[15]，就数字资源长期保存为公共机构提供指导，并积极开展合作计划。该手册在管理数字资源领域具有国际权威性和现实指导意义。该手册中建议了两种数字资源标识方法：（1）永久性标识，即使数字对象的位置发生变化，也能准确定位数字对象的一种手段，例如 URN、Handles、DOI、PURL 等；（2）唯一标识，即每个数字资源都由其所属机构分配一个唯一标识，该标识号能在该机构的资源目录中标识资源对象，并用于查找或确定物理媒体和文件。这样的标识号在本机构内不得进行重新分配。

12.2 国内唯一标识符应用^[16]

12.2.1 CALIS

中国高等教育文献保障体系（China Academic Library & Information System, CALIS）在数字资源的唯一标识符规范上分别制订了《CALIS 数字对象唯一标识

符命名规范》^[17]和《CALIS 元数据命名规范》^[18]；前者是对具体的数字对象的唯一标识符定义，后者则是对元数据的唯一标识符定义。

12.2.1.1 命名原则

CALIS 在项目建设中对数字对象的唯一标识符的制订，采用了与国际命名方式结合的做法。CALIS 制定的唯一标识符（简称 CALIS-OID）符合 URN 的标准，是 URN 的一个子集：命名方式+注册机关代码+注册资源代码。其中（1）命名方式：如以 URN 方式则为 urn，DOI 则为 doi。（2）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）申请分发一代码。（3）注册资源代码则由注册单位内部自编，无一定格式，但要保证内部为唯一代号。如 URN 则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重复即可。（4）+ 为区分码，如 URN 为 "： "，DOI 为 "/" 等。

12.2.1.2 解析规范

该规范用于 CALIS 各子项目承建馆实现本地 CALIS-OID 解析接口，以供 CALIS 统一调度。图 12-5 展示了 CALIS-OID 实现体系结构。

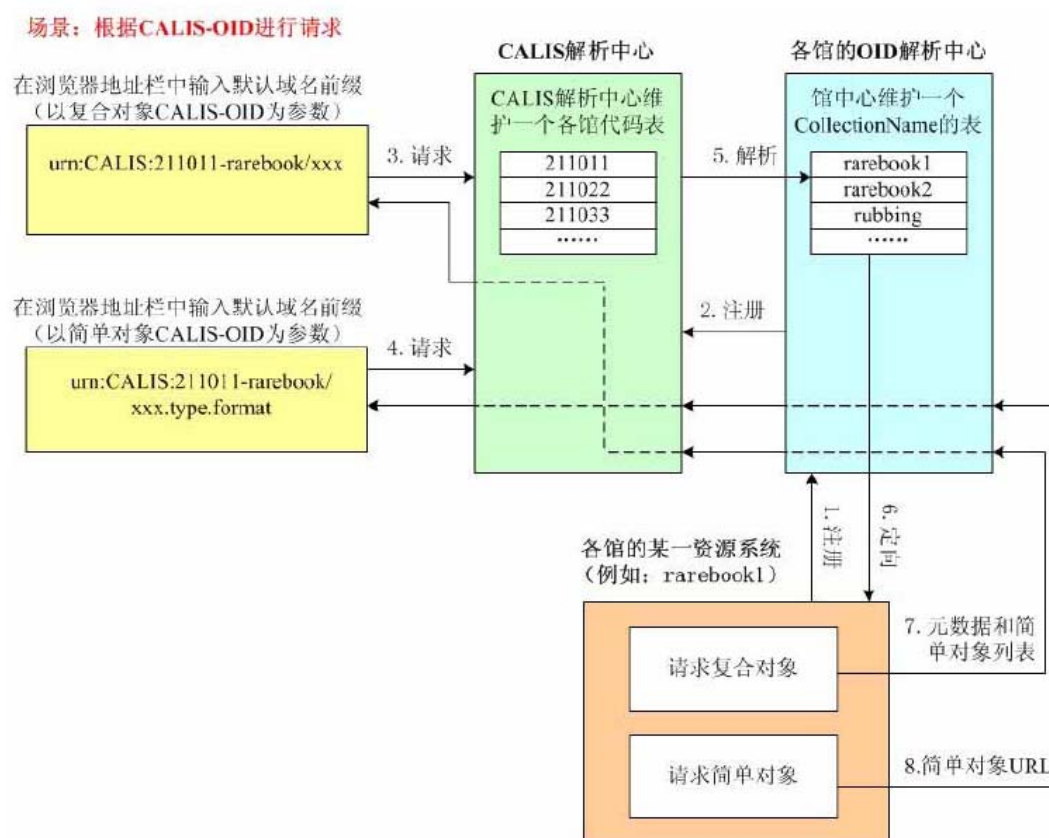


图 12-5 CALIS-OID 实现体系结构

12.2.2 CADAL

高等学校中英文图书数字化国际合作计划项目（China-America Digital Academic Library, CADAL）为便于管理，对 CADAL 项目中所有加工图书采用了统一编号，编号长度为 8 位，分为 2 段，格式为“AADDDDDD”。其中 AA 代表资源加工中心代号，DDDDDD 为 6 位流水号，如“06000001”是浙江大学图书馆制作的电子书。唯一标识符的前两位是由管理中心指定的，也就是加工单位必须获得识别号才可以加工，后面的流水号则由加工单位自行控制。目前在 CADAL 门户中是通过唯一标识符来定位资源的，所有的数字化流程中，从数据提交、质量控制、数据迁移、备份到发布，都是以这个唯一标识符来定位的。以民国图书“中国文学概说”（其唯一标识符为 11105591）为例。

(1) 请求该民国图书的详细信息的 URL 为：

<http://www.cadal.zju.edu.cn/Detail.action?bookNo=11105591>;

(2) 请求全文的 URL 为：

<http://www.cadal.zju.edu.cn/Reader.action?bookNo=11105591>;

(3) 请求该文的全文检索的 URL 为：

<http://www.cadal.zju.edu.cn/FullTextSearch.action?range=book&bookNo=11105591>。

12.2.3 NSTL

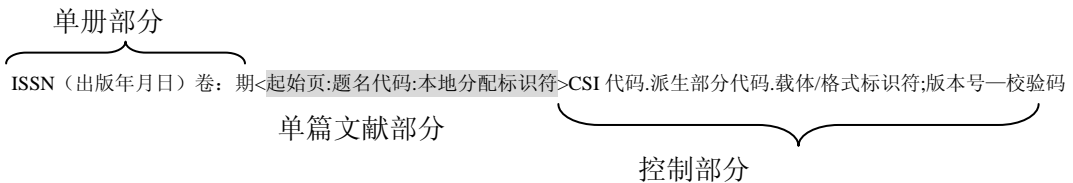
对于不同的资源类型，NSTL（国家科技图书文献中心）的工作思路是为不同的资源类型采用了不同的编码规则。其中期刊文献采用 SICI，会议文献采用 BICI，学位论文采用作者+出版年+标题（取前 100 个字符），标准采用标准号+日期形式，专利直接沿用采用专利 ID，可见下表：

表 12-1 NSTL 唯一标识符编码

类型	唯一标识符	备注
期刊文献	SICI	标题取前 6 个单词首字母，需验证
会议文献	BICI	07 年共 68108 条数据，其中 2788 条既无 ISSN, 又无 ISBN, 推荐使用订购号。
会议文献	BICI	如无 ISBN 号，以母体文献前 6 位替代
学位论文	作者+出版年+标题	需验证

类型	唯一标识符 (前 100 个字符)	备注
标准文献	标准号、日期	沿用过去的做法, id 字段中已包含标准号和年
专利文献	id	公告号
科技报告	BICI: ISBN ISSN NTIS MTI T+	套用 BICI 规则, 优先使用如下字段: ISBN, ISSN, NTIS MTIT
文集汇编文献	BICI	
文献题录类	title+page_start+hold ng_num	保留原始 id 作为唯一标识
科技成果	title+unit_code	

例如, 对于期刊文献, SICI 将每一个数据元素之间用标点符号进行分割, 可以清晰地识别出不同的数据元素所在位置。一个完整的 SICI 结构如下:



例如: 期刊: American Libraries June/July 1996 vol. 27:2

SICI: 0002-9769(199606/07)27: 6<>1.0.%X;2-1

期刊: ASIS Bulletin Dec/Jan 1994 vol.20:2;

SICI: 0095-4403(199312/199401)20:2<>1.0.TX;2-U

由于 NSTL 的外文二次文献库是由九个成员单位加工而成, 这些成员单位有各自的系统进行加工, 所以在加工篇名目次数据时, 均按照 SICI 编码格式生成篇名的唯一标识符。成员单位订购的刊有少量重复, 因此也造成了重复加工的现象。成员单位提供数据时, 中心系统要进行数据入库, 此时, 按照唯一标识符进行查重。如果不重, 则入库; 如果重复, 则查看准备入库数据的馆藏单位与库中馆藏单位是否一致, 如果一致, 则认为是重复数据; 如果不一致, 则追加馆藏

单位信息。

12.2.4 中国医学科学院

中国医学科学院根据 SICI-1996 的语法结构和描述方法,建立的中文期刊文献数字对象唯一标识符,即 Chinese Serial Item and Contribution Identifier(CSICI),采用由期刊标识段、内容标识段和控制段组成的三段式结构。CSICI 的期刊标识段由 ISSN 号码、出版年、卷、期按照一定的规则组成,具体结构如图 12-6:

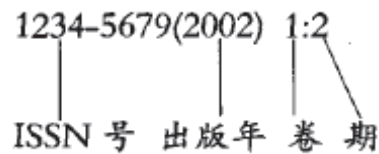


图 12-6 CSICI 结构

内容标识段由“首页码”和“具体文献标识”两部分组成,语法结构为: <首页码: 具体文献标识>。“首页码”指示该数字对象在一本期刊中的出现位置,中文期刊文献的起始页码除了常用的阿拉伯数字外,还有英文字母、罗马数字符号以及“封”、“插”等汉字,根据尽量简单并减少回溯数据不一致性的原则,采取一定的措施进行规范。“具体文献标识”用于区别首页码相同的多篇文献,可以用题目、作者等信息来表示。

一个标识符的唯一性还与管理控制性信息有关,具体包括:(1) 标识对象的分类。因为 CSICI 既可以标识期刊,也可以标识期刊中的文献,所以需要在管理信息中指示所标识对象的类别。(2) 标识文献类型和载体形式的信息。(3) 该标识规则的版本号,它对应着一系列特定的规则。(4) 校验码。为了验证标识符的合法性,需要按照特定算法规则为每一个标识符生成唯一的校验码。

每一个 CSICI 编码都必须有控制段。控制段的各个要素之间用特定的间隔符号分隔开。控制段的语法结构为:



图 12-7 CSICI 编码控制段

其中(1) CSI (Code Structure Identifier, 编码结构标识) 利用不同值表示标识对象的类别,说明标识的对象是期刊还是某一篇文章,或者这个标识符是其他系统的借鉴。这种复杂的结构不仅明确指出了标识对象类别的不同,还为

CSICI 兼容其他标识符提供了条件。(2) DPI (Derivative Part Identifier, 文献类型标识)用以指示标识对象是论文、摘要、目录还是索引等。(3)MFI(Media Format Identifier, 载体类型标识)用于标识不同的文献载体。(4) VN (Version Number) 指示生成该标识符的不同版本号。(5) 校验码。采用 Modelous37 算法对 CSICI 编码中的字符进行模运算, 保证标识符的合法性。

综上所述, CSICI 采用三段式结构, 即期刊标识段、文献内容标识段和控制段, 利用尖括号“<>”将文献内容标识段与期刊标识段和控制段分开。图 12-8 是一个完整的 CSICI 标识符的例子(“具体文献标识代码”仅做示例)。

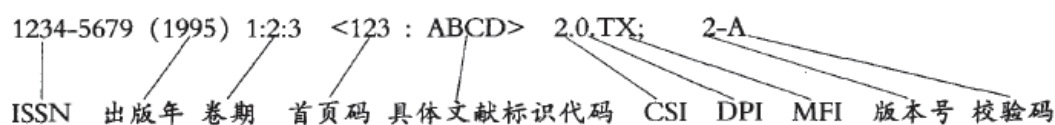


图 12-8 完整 CSICI 的图解示例

12.2.5 万方数据

万方数据是中国较大的数字资源加工商, 万方数据的期刊的唯一标识符方案采用自定义的唯一标识符体系。它的主要作法是在数字文献加工时为每篇文章分配文章唯一代码, 该代码即为数字文献的唯一标识符。万方数据的唯一标识符代码结构为由期刊代码、年、月与流水号组成, 其中“期刊代码”是万方数据标识期刊的代码, “流水号”是数字文献加工时产生的编号。

12.2.6 清华同方

清华同方在国内数字出版商中, 是较早意识到唯一标识符规范的重要性的。在其进行学术期刊进行数字化加工时制定了一套数据规范, 将“文章编号”作为文献的唯一标识。文章编号由期刊的国际标准刊号、出版年、期次号及文章的篇首页码和页数等 5 段共 20 位数字组成, 结构表示为“XXXX-XXXX(YYYY)NN-PPPP-CC”, 其中 XXXX-XXXX 为文章所在期刊的国际标准刊号, YYYY 为文章所在期刊的出版年, NN 为文章所在期刊的期次, PPPP 为文章首页所在期刊页码, CC 为文章页数, “-”为连字符, 如果是英文文献, 则文章编号的标识为“Article ID:”。一些具体的规则有:

(1) 期次为两位数字。当实际期次为一位数字时需在前面加“0”补齐, 如第 1 期为“01”。仅 1 期增刊用 S0, 多于 1 期用 S1, S2……

(2) 文章首页所在页码为 4 位数字。实际页码不足 4 位者应在前面补“0”, 如第 139 页为“0139”。文章页数为两位数字, 实际页数不足两位数者, 应在前

面补“0”，如9页为“09”。转页不计。

(3) 文章编号由各期刊编辑部给定，中文文章编号的标识为“文章编号：”或“[文章编号]”。如：文章编号：1003-2797(1997)02-0013-05 为发表在《图书情报知识》1997年第2期第13-17页(共5页)上题为《关于社会经济信息化的思考》(作者：严怡民)一文的文章编号。

12.2.7 上海图书馆

上海图书馆建设的数字图书馆项目组织了丰富的特色馆藏资源，包括上海图典、古籍善本、上海文典、中国报刊、民国图书、点曲台、科技百花园、年画家谱等特色资源库，其中收藏的中文古籍，包括刻本、活字本、抄本、稿本、校本、民国年间出版的石印本、影印本、珂罗版印本共计约129660种，家谱共收藏有约14500种、110000余册中国家谱，是国内外收藏中国家谱(原件)数量最多的单位。

目前上海图书馆尚没有唯一标识符对上述特色资源内容进行标识的应用，主要标识仍以馆藏号为主，尚不能与其它数字图书馆的同类型数字资源进行集成和整合。

12.3 专业领域唯一标识符应用状况^{[19][20]}

12.3.1 数字出版

出版业在数字化和网络化的推动下正在经历面向数字出版的全面转变。数字出版缩短了内容创造到出版物消费的距离和时间，数字化的内容和原子化的“载体”呈现多元化的分离和融合。下面从学术出版物、数字权益管理、电子图书和开放获取四个方面探讨唯一标识符在数字出版中的应用。

12.3.2 学术出版物

在学术出版物中，期刊物件和文章标识符(SICI Serial Item and Contribution Identifier)可标识期刊单个卷期以及卷期中的具体文章，和出版物标识符(PII)、数字对象标识符(DOI)和通用出版物名称(USIN)一样，它们都用来标识不同层次的出版物(年、卷、期、篇、引文等)。SICI的描述性和可读性更强，它不但兼容了ISSN号，还通过计算确定号码段，这样为文摘和目次制作、原文传递服务和图书馆自动化系统提供很多便利，由于SICI号码在期刊和文章两个层面上较好的描述文献，因此在JSTOR、FirstSearch等服务中得到广泛应用。在本文的1.2部分与1.3部分也详细描述了这类唯一标识符的规范和应用。

12.3.3 数字权益管理

数字化和网络化为电子出版物的制作和发行带来巨大的挑战，特别是廉价和非法的复制传播严重影响电子出版物市场前景。数字权益管理提出了在出版物的生产、传播和使用各个环节建立有效的管理机制来保护知识产权，这种机制的核心就是对出版物的各个层面进行有效的唯一标识^[21]。这一类的唯一标识符包括 DOI、MPEG-21、cIDF、UUID 与 GUID 等。

DOI 就是学术出版物中主流的唯一标识符机制。MPEG-21 “Multimedia Framework^[22]”是 21 世纪多媒体应用的标准化技术，它定义了可互操作和高度自动化的框架，而且考虑到了数字权益管理（DRM digital rights management）的要求、对象化的多媒体接入以及使用不同网络和终端进行传输等问题。

内容标识符论坛 cIDF^[23]为管理数字资料（文档、图片和多媒体）的版权而创立的，由 130 多个公司（主要来自日本）和机构组成，来自日本的大学、协会、服务提供商、出版商、内容提供商、软件公司和法律机构。

通用唯一标识符（UUID）^[24]是 128 位比特的数字，用来唯一标识因特网上对象或者实体。有保证的通用唯一标识符包括一个指向生成通用唯一标识符的主机的网络地址，时间戳（用来记录事务的精确时间的记录），以及一个随机生成的组件。

全球唯一标识 GUID^[25]是微软使用的一个术语，广泛应用于微软的产品中，用于识别接口、复制品、记录以及其他对象。不同类型的对象对应不同的 GUID 值。但由于安全和隐私保护的问题，GUID 限于软件的唯一标识。

12.3.4 电子图书

电子书、网上书店、数字图书馆、手机看书正在改变大众的阅读方式^[26]，从而引发图书产业向网络化全面转型。电子书发展出现了多元并存、图书馆和大众市场互为促进、出版社成立专职机构推进网络出版机制、e 纸一体化四大趋势。2006 年 7 月，国际数字出版论坛（IDPF）设计推出新的电子书通用标准：新的电子书标准格式将名为 OEBPS Container Format（OCF）。但 PDF 是使用最广泛的电子书格式，但却不是开放的国际标准，我国的方正，书生和超星也推出自己的电子书格式。在数字图书馆参与的中美百万藏书项目也定义了电子书的格式。部分依赖阅读器软件的网络阅读仍然是当今电子图书的主要状态。而唯一标识符的应用机制在这种复杂和多样的格式下也呈现不同的应用态势。

图书是现代图书馆的“立馆之本”，电子图书可有效缓解现代图书馆的空间和服务问题，但知识产权的限制一方面使图书馆对已有图书的数字化“举步为艰”，而且电子图书的数量和种类也远远不能满足广大用户日益增长的需求。加

上国家政策、行业管理的因素，图书承载了太多的标识符系统，如 CIP 编目号码、知识产权登记号码，邮发代号，ISBN 号、BICI 号码、条码号（EAN）等，再加上一个唯一标识符显然难以达到设想。但是基于网络的电子图书交易仍迫切需要能够标识交易对象的标识符系统，目前大多数出版商的作法是推荐网络在线阅读的方式，即以图像的方式提供按页阅览的图书，并通过 ISBN 号等传统号码还标识或查询，这种方法也只能是“权益之计”。

12.3.5 开放获取

开放获取是一种新的传播期刊学术信息的方法，开放获取资源主要有三种类型：机构资源库、学科资源库和开放期刊^[27]。机构资源库是大学创建的知识产品的数字化档案库，提供给校内外的终端用户利用，没有或只有很少的限制，如 Dspace@MIT。学科资源库是为了促进学科研究资料的共享与保存，如 DOAJ。开放期刊是一种网络化的免费期刊，提供给所有的用户利用，如 BioMed 等。开放获取资源的唯一标识符应用主要分两类，DOI 号码和自定义的号码。对于采用 Dspace 构建的机构知识库，由于 Dspace 内嵌了 Handle System，因此也自动拥有 Handle 号码。

12.3.6 数字科研

数字科研（E-Science）就是科学研究过程的信息化，科学家通过因特网，通过 E-Science 信息化平台环境思考、试验、协作和交流。数字科研涉及大科学装置、试验仪器、计算能力、数据库和资料库等科研基础设施共享。在数字科研领域，唯一标识符的应用存在于网格、科学数据库与机构知识库中。

网格（Grid）是计算机通过因特网协同的新机制，实现多个组织间的资源共享和协同式的科学问题求解。全球网格论坛（GGF）下属的 Globus 项目组开发的 Globus Toolkit 标准工具包是建立网格系统和开发网格软件事实的参考标准，它由安全架构、信息架构、资源管理、数据管理、通信和错误监测等部分组成。CNRI 的句柄系统（Handle System）自 2005 年也纳入 GT 中担任命名和解析网格服务（Grid Service）的功能，在现有的 Web 服务的基础上，为网格部署过程中涉及的政策、能力和密钥等 Web 服务提供元数据的标识和解析服务^[28]。这种代理服务类似 GT 中 SAML 属性的查询、WSRF 服务组的属性查询等。在 CNRI 提供的演示系统中：首先下载支持句柄解析的网格服务组件，然后就可在 Web 服务的界面下演示服务的命名和解析。

科学数据库将科学研究中利用和产生的各种数据（包括文本、数字、图谱、视频音频等）汇总，为科学家之间共享研究成果提供了便利，但是如何在不同组

织机构建立的科学数据库之间实现共享,也涉及到对科学数据进行唯一标识的问题。德国国家科学图书馆(TIB)的作法是利用 DOI 号码对科学数据进行标识的研究,试图规范科学文献的引文部分对科学数据的引用。世界气候数据中心的米歇尔(Michael Lautenschlager)主持此项目,利用 DOI 号码对从观测站、卫星等活动的科学数据进行标识,以便于研究人员引用和使用。通过 DOI 号码,研究人员希望科学数据也可以向科技文献那样被科研人员引用,并形成类似 SCI 的引文索引。中国科学院在长期的科学研究实践中,通过观测、考察、试验、计算等多种途径产生和积累了大量科学数据和资料,涵盖了化学、生物、天文、材料、等多种学科,建设科学数据库数量超过 1000 个。针对科学数据库还制订的标准规范,以统一 40 多个参建研究所的工作,并颁布“核心元数据标准 V2.0 版(2004-09-24)”利用 URI 机制对科学数据集和服务集合进行了唯一标识定义^[29]。

机构知识库(Institutional Repository,简称 IR)是搜集、存储学术机构成员的知识资源,并提供检索的数字知识库,为世界范围内的网络用户服务^[30]。DSpace 平台是开放源代码的系统,用作组织或机构数字研究和教育资料的资源库,目前在世界 150 多个大学和研究机构部署实施,包括我国的厦门大学^[31]和清华大学^[32]。DSpace 采用集中的机构服务运营模式。机构中不同的群(community),如实验室、中心、学院或系可以在系统内拥有独立的区域,群的成员能够直接通过 Web 用户界面存放内容,系统存放过程简单,同时还提供了批量条目导入程序(batch item importer)。DSpace 内嵌了 CNRI 的句柄系统本地解析服务器,可与 CNRI 的全球解析服务器通讯并获得有关唯一标识符解析服务。DSpace 采用句柄系统主要用于为对象分配全球唯一的标识符,每一个运行 DSpace 的站点都需要从 CNRI 获得一个句柄“前缀”,这样使用该前缀创建出来的标识符就不会与其它地方创建的标识符发生冲突。目前句柄被分配给群、集和条目,数据包(Bundle)和比特流(bitstream)未被分配句柄,因为条目被编码成比特的方式可能会发生改变,以支持未来的技术和设备。由于被引用的通常是条目,而不是具体的比特编码,所以它仅仅用于永久地确定和支持对条目的访问,使用户能够再从那里访问到合适的比特编码。

12.3.7 远程教育

远程教育(E-Learning)是教育机构利用资源和技术构建的教育环境,为教师和学生通过网络等数字化交互手段进行教学学习的过程。网络环境中由信息资源和信息技术构成的教育资源,也存在命名、标识和解析的问题。

我国的《教育资源建设技术规范》(CELTS-41.1 WD1.0)^[33]是配合我国现代远程教育资源建设工程而制定的资源开发指导规范,目的是统一学校开发网络教

育资源的行为，促进资源共享，其核心是按照资源类型制定资源属性标注标准，其中也涉及到唯一标识的问题。《教育资源建设技术规范》则针对具体的教育资源建设，提出具体的资源属性标准，具有实践指导意义。但是在教育资源的实践中，上述技术规范的推广仍然受到限制。

1998年美国国家科学基金会正式启动了美国国家科学、数学、工程与技术教育数字图书馆计划(NSDL)。发展为远程教育环境，向数百万学生和教师提供高质量的科学、数学、工程与技术教育资料。NSDL采用面向对象设计的Fedora数字资源框架，将数字对象分为如下部分：数字对象唯一标识符(PID)、数据发布器、系统元数据和数据元素。它淡化了资源类型，将数据和元数据统一用数据元素封装，并且形成了Fedora存储模型。在Fedora系统中，数字对象的PID是外部可见的唯一标识^[34]，被提供来进行仓储间的数据引用，这样一个仓储可以包含其它仓储里面的数字对象，仅需要提供被包含数字对象的PID即可。

12.3.8 电子政务

电子政务(E-Government)是政府机构应用现代信息和通信技术，将管理和服务集成，实现政府组织结构和 workflows 的优化重组，向社会提供规范、透明、符合国际水准的管理和服务^[35]。从政务信息公开到政府信息资源库的建设和服务，涉及到海量、异构数据库的建设和管理，目前上述数据仍然分散在不同的政府机关和部分，也存在大量的重复，缺乏有效的标识。在电子政务的标准化建设中，已经将唯一标识符纳入标准体系。如《国家电子政务标准化指南》中包括政务信息资源标识符编码方案^[36]。《政务信息资源标识符编码方案》由中国标准化研究院牵头起草，规定了政务信息资源标识符的编码结构、前段码和后段码的分配和管理机制。参照国际上统一标识方法(如DOI、IARN)和我国政务信息资源的标识需求，政务信息标识符采取前段码和后段码两部分组成，用“/”分开。前段码共5位，由10个阿拉伯数字(0-9)和26个大写英文字母(A-Z)组成，后段码长度不固定，可采用GB18030中规定的任何字符。前段码由信息资源前段码管理中心统一管理，并分配给目录管理者和相关的政务部门，后段码则由使用单位自行编码。政务信息资源标识符是针对单个信息资源，目前是唯一标识。该方案还详细规定了前段码5位的分配规则和依据，并对全国所有省市区和台湾、香港和澳门进行了统一标识。并制定了前段码分配的具体流程规范。

12.3.9 电子商务

对商品进行有效和唯一标识一直是电子商务的重点。目前广泛采用的条形码(Barcode)就是从商务活动中产生并推广的。电子商务标准代码体系包括产

品/商品代码、参与方代码、物流相关代码、资金流相关代码和信息流相关代码。

电子商务编码有专门的国际组织负责有关编码体系的制订和维护，国际上有 ISO、IEC、EU、UN/CEFACT 以及微软、OASIS 等，我国的电子商务编码体系主要有中国物品编码中心 ANCC 负责。中国物品编码系统（ANCC）^[37]是一套全球统一的标准化编码体系，是对流通领域中所有的产品与服务，包括贸易项目、物流单元、资产、位置和服务关系等的标识代码及附加属性代码，以条码符号、射频标签等可自动识别的载体承载编码信息，从而实现流通过程中的自动数据采集。

从标识维度上，条形码包括一维条码与二维条码，前者无法表示汉字和图像信息，后者的功能较为强大，能通过矩形空间标识对象。具有代表性的矩阵式二维条码有：Code One、Maxi Code、QR Code、Data Matrix 等。在目前几十种二维条码中，常用的码制有：PDF417, Data Matrix, Maxi Code, QR Code, Code 49, Code 16K, Code one, 龙贝码等。QR Code 码^[38]可表示汉字及图象多种文字信息，具备保密防伪性强等优点，在日本应用广泛，用来描述产品名称、产地、生产日期、价格等信息。龙贝码(LPCode)由上海龙贝信息科技有限公司开发。与国际上现有的二维条形码相比，具有更高的信息密度、更强的加密功能、可以对所有汉字进行编码、适用于各种类型的识读器、最多可使用多达 32 种语言系统、具有多向编码/译码功能、广泛用于电子门票、电子购物券、电子报纸等方面。

无线射频识别（RFID Radio Frequency Identification）是一种非接触式的自动识别技术，它通过射频信号自动识别目标对象并获取相关数据，识别工作无须人工干预，可识别高速运动物体并可同时识别多个电子标签^[39]。RFID 标签是一种“固化”的唯一标识符，可有效的监控商品从制造到流通消费的全过程，但由于目前尚没有统一标准、RFID 标签的成本仍然较高，以及 RFID 对消费者隐私的保护仍然受到广泛的置疑，RFID 在中国的应用范围很有限。

12.3.10 数字文化

在文化教育领域，以数字地球、数字博物馆、数字科普馆、数字标本馆、数字植物园和数字遗产保护等为代表的“数字文化”正在全面加入大众的日常生活。唯一标识符在数字文化领域承担者命名、管理、标识等作用。

目前国际上著名的博物馆如美国大都会博物馆、法国卢浮宫博物馆、英国大英博物馆和我国的故宫博物院都建立了自己的网站，并通过网络公开展示部分珍品，但是对藏品的唯一标识还使用本博物馆的内部编号系统，尚未建立全球统一的博物馆藏品目录和唯一标识符系统。2006 年 HP 公司利用其 Dspace 系统参与了中国的数字博物馆项目，将 Dspace 领域引入数字博物馆的建设，并基于

Dspace 系统中内置的 Handle 系统建设数字博物馆信息库的建设中唯一标识符系统^[40]。

数字化植物园采用虚拟现实技术，面向公众开放具有三维界面和多种分辨率浏览器支持的植物物种景象与数据，并为公众提供身临其境的交互式访问平台^[41]。英国皇家植物丘园(KEW)，美国密苏里植物园等与中国的植物园都开展了数字化植物园的建设，但这些信息系统中的唯一标识符仍主要以本系统的自定义为主，尚不具备与其它数字植物园集成和整合的能力。这样的状况同样存在于全球的数字科普馆中，目前尚没有唯一标识符对各种科普内容进行标识的应用，主要标识仍以 URL 为主。

在数字文化遗产领域，各国图书馆都制定了收集与保存计划，例如澳大利亚图书馆的“潘朵拉”计划、芬兰 EVA (European Visual Archive) 项目、瑞典文化遗产工程与中国国家图书馆的中国因特网信息保存^[42]等。这些活动对因特网上的数字遗产进行长期保存，采用的标识符以 URL 为主，并在 URL 的基础上增加版本和时间标识，提供给用户使用。

13 国家图书馆信息资源和应用系统唯一标识符应用现状

中国国家图书馆是全国的总书库和文献信息服务中心，履行搜集、加工、存储、研究、利用和传播知识信息的职责。2002 年，国家图书馆启动了“国家图书馆二期工程暨国家数字图书馆工程”项目，工程建设的目标是有重点地收藏、建设和长期保存中文数字资源，并向本国以及世界其它各国提供中文数字资源服务，使中国国家数字图书馆成为世界最大的中文数字资源保存基地与服务基地。

国家数字图书馆建设任务主要包括：通过软硬件系统建设，构建支持海量数字资源采集、加工、保存、管理、服务的技术支撑平台；有重点地对馆藏特色文献进行数字化，采集与保存重要的数字信息资源库和互联网信息，建立国家级学术性数字资源长期保存中心；搭建中文信息资源服务平台，为政府机关、教育、科研、企业单位及社会公众提供信息服务，为其它数字图书馆系统提供服务支撑；开展数字图书馆相关标准规范的研制工作，初步建立国家数字图书馆标准规范体系，对中文信息处理中涉及的关键技术与标准进行研发。

建设中的国家数字图书馆工程，以海量的资源、方便的服务、先进的技术和相关标准规范为工作内容，旨在为用户提供一站式方便的文献信息服务。唯一标识符规范与汉字规范、元数据规范、知识组织规范、长期保存规范等一道构成了国家数字图书馆的标准规范体系，对全国联合目录等资源主导项目和原文传递、统一门户、开放链接等服务主导项目提供基础性、技术性支撑。唯一标识符

作为系统互操作的重要技术，将为国家数字图书馆的资源集成、服务整合和知识链接提供重要的手段。

13.1 国家图书馆数字资源基本状况

国家图书馆数字资源建设采取了“中文为主，外文为辅，通用性资源外购为主，特色资源自建为主”的方针。

13.1.1 国家图书馆外购资源

截至 2008 年 3 月，国家图书馆外购数字资源 134 个，其中中文数据库 56 个，外文数据库 78 个。类型包括期刊论文、电子图书、电子报纸、事实数据、光盘数据库、引文、专利标准、学位论文（附表 1 和附表 2）。国家图书馆还链接国家科技图书文献中心为全国开通的全文科技期刊 107 种。

13.1.2 国家图书馆自建资源

自建数字资源是国家图书馆数字资源建设中的重点，也体现国家数字图书馆特色资源和服务的核心。截至 2008 年，国家图书馆共自建总容量约为 180TB，全文影像数据超过 1 亿 1500 万页。

国家图书馆的自建数据分为 3 个层次，即元数据、特色全文数据库和主题知识数据。其中元数据主要包括来自图书、期刊、报纸、古籍、电子资源、网站信息等载体的书目数据、篇名数据、题录数据、目次数据、结构数据；特色全文数据库主要是国家图书馆具有馆藏特色的甲骨、拓片、老照片、图片、手稿、舆图、善本书等；主题知识数据是华夏春秋（中国历史、中国人物、中国文学）、中国学、民国法律法规等。

13.1.2.1 按数据库类型

从数据库类型来看，国家图书馆自建数字资源可分为：

（1）元数据库，包括书目数据库、篇名（目次）数据库、摘要书刊、规范数据库以及专题导航库。

（2）全文数据库，包括全文影像和全文文本。

（3）多媒体数据库，包括音视频数字资源库。

（4）专题知识库，包括各类型的知识性资源库。

13.1.2.2 按数据库系列

从数据库系列来看，国家图书馆自建数字资源可分为：

(1) 文献系列：包括古籍文献、民国文献、少数民族文献、学位论文、政府出版物、缩微文献、视听文献。

(2) 特藏系列：包括敦煌、甲骨、青铜器、拓片、法帖、舆图、西夏文献、年画、家谱、老照片、名家手稿、地方志、新善本、外文善本。

(3) 专题资源系列：中国文化史、中国历史人物、中国地方志、中国古代法律、参考资源。

(4) 专题导航系列：学科知识导航、行业服务导航。

(5) 网络资源系列：网站网页信息、网络专题信息。

13.1.2.3 特色资源

中国国家图书馆作为全国的总书库和文献信息服务中心，履行着全国书目中心的职责，编辑出版国家书目、联合目录和馆藏目录；存储着丰富的珍品特藏善本古籍、金石拓片、古代舆图、敦煌遗书、少数民族图籍、名人手稿、革命历史文献、家谱、地方志和普通古籍等 260 多万册（件）；国家图书馆也重视对国内非出版文献的收藏，负责全面收藏和整理我国的学位论文和博士后研究报告，集中收藏图书馆学专业资料、年鉴资料。

以这些资源为依托，形成了国家图书馆具有特色的数字资源。同时在网络时代，国家图书馆十分注重中文网络信息资源的采集与保存，如政府网站、中文 PDF 报纸、专题资源。

13.2 国家图书馆数字资源唯一标识符的状况

13.2.1 国家图书馆外购数字资源的唯一标识符

国家图书馆外购的数字资源，其生产厂商都有自己的数字资源标识方式，其中国外的数字资源生产商使用的唯一标识符技术相对较成熟，国内的数字资源生产商在唯一标识符应用上还出于起步阶段。表 13-1 是国家图书馆外购资源使用唯一标识符的情况实例：

表 13-1 国内外数字资源的唯一标识符应用情况

号	数据库名称	唯一标识符方案	例子	备注
	ACS	DOI	10.1021/bi0359527 S0006-2960(02)05949-4	文章层次
	AIP	DOI	URL: http://link.aps.org/abstract/PRB/v69/e121101	文章层次，文摘

			doi:10.1103/PhysRevB.69.121101	有解析服务器
	APS	DOI	doi:10.1063/1.1688231	有解析服务器
	ELSEVIER	DOI	DOI:10.1016/j.actaastro.2003.12.006	文章层次, 加入 CrossRef
	Springer	DOI	DOI:10.1007/s10211-003-0082-9	文章层次
	清华同方	新闻出版署试行方案	1003-2797(1997)02-0013-05	文章编号
	维普	自定义	CSI: 1000-6974(2003)25:6 <328:WJKZDHF XMZ>2.0.TX;1-U	仿 SICI
	万方	自定义	期刊名称-年-月-流水号	文章唯一代码
	方正电子书	自定义	bookid=ISBN-7-115-00000-7/TP.308	ISSN 号和中图分类号

13.2.2 国家图书馆自建数字资源的唯一标识符

国家图书馆自建的数字资源，目前没有统一的标识规则和方式，基本上是根据国家图书馆开展的各数字资源项目中的数字化加工标准对数字资源进行命名。

以新善本图书的唯一标识符命名规则为例进行说明。善本图书唯一标识符是数字化加工过程中一册图书的唯一标识，它由新善本加工号和册号构成。新善本加工号共 8 位长度，具体划分是：文献种类代码（2 位，1-2 字符位）、数据代码（1 位，3 字符位）、流水号（5 位，4-8 字符位）。

一册图书文件夹内包含若干图像文件，其文件命名规则：

(1) 封面扫描文件名为 Axxxxx_00.tif，其中 xxxxx 为数字，按原书顺序依次排序。

(2) 前附页部分指由封面之后正文之前，不含目录页的部分组成。若前附页的部分内容在封面滞后目录页之前，其扫描文件名为 Bxxxxx_00.tif，其中 xxxxx 为数字，按原书顺序依次排序。若前附页的部分内容在目录页之后正文之前，其

扫描文件名为 Dxxxxx_00.tif, 其中 xxxxx 为数字, 按原书顺序依次排序。

(3) 目录页包括总目录、分目录、目录、条目索引等内容, 其扫描文件名为 Cxxxxx_00.tif, 其中 xxxxx 为数字, 按原书顺序排序。

(4) 图书正文部分, 图书正文扫描文件名为 Txxxxx_00.tif, 其中 xxxxx 为数字, 表示正文有页号的流水号, 按原书顺序依次排序。

正文夹页: 正文中不占页号的书页为夹页, 扫描文件名为 Txxxxx_yy.tif, 其中 xxxxx 为数字, 表示夹页的前一页顺序号, yy 为数字, 表示夹页, 并按原书顺序依次排序。

一图多画幅的文件命名, 当一张图过大, 需分多画幅扫描, 文件按如下方式命名: tiff 格式保留多画幅不做拼图处理, 其扫描文件名为 Txxxxx_00zz.tif, 其中 xxxxx 为数字, 表示该图在原书的顺序号, zz 为数字, 表示该图的多画幅顺序号。发布格式需做拼图处理, 拼接后该图的扫描文件名为 Txxxxx.pdf, 其中 xxxxx 为数字, 表示该图在原书的顺序号。

既是夹页又需分图的文件命名: 当该图既是夹页又需分图扫描, 其扫描文件名为 Txxxxx_yyzz.tif, 其中 xxxxx 为数字, 表示夹页的前一页顺序号; yy 为数字, 表示夹页, 按原书顺序依次排序; zz 为数字, 表示该图的多画幅顺序号。发布格式需做拼图处理, 拼接后该图的扫描文件名为 Txxxxx_yy.pdf, 其中 xxxxx 为数字, 表示夹页的前一页顺序号; yy 为数字, 表示夹页, 按原书顺序依次排序。

(5) 后附页部分是由正文之后的无页号或单独排页号的部分组成, 可能是后记、附录、附表、附图等, 直接扫描在正文之后。文件名为 Yxxxxx_00.tif, 其中 xxxxx 为数字, 按原书顺序依次排序。若后附页中含版权页, 应在该存储光盘的 readme.txt 文件中对版权页的文件名予以说明。

(6) 封底扫描文件名为 Zxxxxx_00.tif, 其中 xxxxx 为数字, 按原书顺序依次排序。

从附表和以上的例子可得出, 国家图书馆自建数字资源, 已有详尽细致的唯一标识符命名规则, 能覆盖了各种类型的数字资源, 也能标识包括各种揭示深度的资源。但目前的唯一标识符命名规则缺乏一致性, 各种资源的命名方式自成体系; 不同深度的数字资源内容之间缺乏关联关系; 整个命名规则处于零散各自割裂的工作状态, 也没有解析系统负责对各种命名的资源进行解析。

13.3 国家图书馆应用系统唯一标识符的应用情况

13.3.1 自动化系统

目前, 国家图书馆的数字资源不能通过国家图书馆的自动化系统 ALPH500

的 OPAC 进行检索，OPAC 的书目数据中只反映印本资源的书目信息，不能和数字资源实现关联。这是由于国家图书馆采用的两套元数据：印本资源书目数据和数字资源元数据，在印本资源的书目数据中没有反映相对应的数字资源的可获取地址。

国家图书馆建立数字资源门户，能嵌入到 OPAC 中，提供“保存到我的收藏夹、保存/邮寄选中记录、文献索取”的服务。在原则上，通过数字资源的门户，即 SFX 的连接，应能够获取数字资源的全文。但国家图书馆在数字资源门户的获取全文中数据库中没有配置相应的全文数据库，目前获取全文的数据库指向方正 Apabi 电子书，因而从 OPAC 检索印本资源，通过数字资源门户关联数字资源也不能实现。

13.3.2 信息门户

国家图书馆数字资源门户旨在有机地整合国家图书馆收藏的多文种、多学科、多载体、多类型，且分布存在的印刷型和数字化的信息资源，面向社会公众提供方便快捷的一站式检索和信息获取服务。它将国家图书馆购买、自建及通过其他方式获取的 70 余种中文数据库，120 余种外文数据库，2 万种余中外电子期刊，博硕士学位论文、学术会议、企业名录等以及国图 OPAC 公共目录检索系统进行有机的整合，实现了这些资源之间的无缝互连。

国家图书馆的个性化服务，也是通过国家图书馆数字资源门户提供的，在“我的图书馆”中提供“电子书架”、“我的数据库”、“我的电子期刊”、“检索历史”、“使用偏好”5 项个性化服务。

国家图书馆数字资源门户是基于 ExLibris 公司推出的 Metalib & SFX 系统建立，其系统的核心是 OpenURL 标准，是对 HTTP 协议中 POST 和 GET 数据交互方式的一种扩展，SFX 将数据库的 metadata 定义为标准的 OpenURL 格式，提供不同数据资源的互操作功能。

目前国家图书馆数字资源门户集成的国图自建资源数据库由 5 个，即碑帖菁华、民国期刊、西夏碎金、数字方志和博士学位论文。通过一站式的跨库检索，通过 OpenURL 标准，调用自建数据库系统的检索结果界面。

如 在 国 家 图 书 馆 数 字 资 源 门 户
(<http://dportal.nlc.gov.cn:8331/portal/metasearch.htm>) 中检索《儒家思想与中国传统战略文化》，能在“馆藏博士论文与博士后研究报告数字化资源库”和“馆藏学位论文数据库”中找到，通过数字资源门户的链接能直接跳转到所属的数据库的检索结果界面
(<http://res1.nlc.gov.cn/doctor/servlet/msearch?pos=-1&clause=&pagec=&fld=D01&>

vlU=%E5%84%92%E5%AE%B6%E6%80%9D%E6%83%B3%E4%B8%8E%E4%B8%AD%E5%9B%BD%E4%BC%A0%E7%BB%9F%E6%88%98%E7%95%A5%E6%96%87%E5%8C%96)

13.3.3 自建系统

国家图书馆自建的数字资源，如碑帖菁华、敦煌遗珍、西夏碎金、数字方志、博士论文、民国期刊、年画撷英、中国学、在线讲座（视频库）、甲骨世界、民国法律和民国图书，都有其专门的元数据，是独立于物理形态资源的书目数据。国图自建系统对唯一标识符的调用，采取的是：数据库名/对象数据的唯一标识号/文件名 ID 号，全文检索时通常调用首页的文件名 ID。

如检索民国图书“怎样读书”，其书目数据地址为：
<http://res3.nlc.gov.cn/rocbook/detail.jsp?bookid=01jh000271>

访问其全文的 URL 为：

http://res3.nlc.gov.cn/rocbook/book.jsp?bookid=01jh000271&volid=001&filename=A00001_00.PDF&booktitle=%D4%F5%D1%F9%B6%C1%CA%E9

13.3.4 网络搜索引擎

为了方便读者从国家图书馆的数字图书馆中查找资料，国家图书馆在 2005 年 12 月开始与 Google scholar 合作。国家图书馆的读者在利用 Google scholar 检索时，只需在使用偏好中预先限定图书馆链接为国家图书馆。读者在检索到感兴趣的摘要或引用时，就能在搜索结果中根据特殊链接，直接访问电子版全文。目前能够通过网络搜索引擎检索，检索和链接的全文仅限于国家图书馆购买的部分数字资源，而不包括国家图书馆自建的数字资源。

13.4 小结

从以上分析可看出，国家图书馆现在的唯一标识符主要有两种：一部分来自外购资源，唯一标识符是由其数据库生产厂商根据一定标准所应用。这部分由于不涉及本地保存的问题，属于直接调用外部购买资源，数据库生产厂商有约定的命名方法、解析、调用方法；另一部分是国家图书馆自建数字资源的唯一标识符，虽然目前国家图书馆在数字加工时规定数字资源命名规则，它已能覆盖各种类型、各种揭示深度的数字资源，但它并不是唯一标识符标准规范，要作为唯一标识符标准存在以下几方面的问题：按照资料类型规定的唯一标识符命名方式，反映的是载体资源，即简单数字对象，但不能反映复杂数字对象，也不能对复杂数字对象和所属简单熟悉对象的关系进行说明。国家图书馆自建数字资源元数据

来自于印本资源的书目数据，在此基础上又继续补充了描述元数据、结构元数据和技术元数据。要解决数字资源和传统文献的元数据关联关系属于元数据关联问题，需要在元数据规范中约定。现在国图应用系统对数字资源的调用，主要是采用 URL 对因特网上数字资源进行标识和链接的方式，不能适应分布式动态环境的要求。

因而必须建立唯一标识符的框架体系来标识国家图书馆的基于不同应用系统和数据库的自建数字资源，要包括名称空间、唯一标识符、命名机构、命名登记系统和解析系统。

14 国家图书馆唯一标识符实施指南

14.1 概述

14.1.1 目的和作用

唯一标识符规范规定了唯一标识符的语法规则、名称分配规则、解析规则、管理规则等方面的内容。规范的制定与规范的实施应用是相辅相成，缺一不可的。国家图书馆唯一标识符实施指南提出了唯一标识符规范应用和实施的细则。

本章从标识符与唯一标识符的概念出发，阐明了 CDOI 可以标识的对象。同时本章面向中国唯一标识符服务体系的需求，描述了全球唯一标识符服务系统框架，提出了未来国家图书馆唯一标识符服务系统框架。

本章也提出了国家图书馆唯一标识符后缀的分配规则，详细阐述了各种复杂的数字对象的唯一标识符分配的原理和方法以及唯一标识符分配与维护管理中需要关注的问题。

唯一标识符的生成、注册与服务是一个系统过程，唯一标识符系统与国家数字图书馆其他系统相互关联，共同构成有机的数字图书馆服务体系。所以本章阐述的唯一标识符实施指南也比较详细地分析唯一标识符系统与国家数字图书馆其他系统的关系，说明唯一标识符服务系统实现方法和实现过程。在本指南中，详细描述了唯一标识符的解析过程；系统描述了唯一标识符的生成与注册过程，包括用户角色的界定与生成示例；提出了唯一标识符规范应用对相关各系统的要求，为国家图书馆应用唯一标识符标准规范提供流程机制、系统需求方面的建设性参考意见。

14.1.2 CDOI 与唯一标识符

标识符是用于标识实体的明确的字符串或标签，通过正式的标准、行业的

规定或者统一的语法规则来标识和区别不同的实体。比如 ISSN, ISBN 等都是标识某一专门对象的标识符。

唯一标识符是因特网通用名称机制的具体体现。唯一标识符作为应用系统对数字对象命名机制的组成部分,通过在对象的内部标识号前加全球唯一的命名授权号的方式来维持对象标识的唯一性。网络环境对唯一标识符提出了更高要求,它不但要求唯一标识符能够唯一标识对象,更需要持久、可操作地标识对象,以满足不同系统间互操作的需要。

因此唯一标识符体系并不仅仅是一套命名机制,更需要在命名机制的基础上实现对标识对象的解析、发现与调用。例如 IETF 工作组针对 URN (Uniform Resource Name) 定义了一套语法规则、名称解析机制和具体资源标识符的名称空间注册过程。对于每个经过注册的资源对象,都被赋予了一个唯一的名称空间标识 ID,任何解析系统在解析这些名称空间时,同时需要在 URN 规定的一个 RDS (资源发现系统) 中注册。

CDOI 的全称是中国数字对象唯一标识符 (Chinese Digital Object Identifier,CDOI)。CDOI 是一种基于因特网的唯一标识符的命名、管理和解析的唯一标识符系统,CDOI 目标是实现网络化分层的唯一标识符管理和解析服务,与其他唯一标识符系统,如 DOI,一起构成对因特网数字对象的持久稳定标识和解析服务。

14.1.3 CDOI 标识的对象

CDOI 可以标识各种类型、各种格式、各种粒度的资源:

- (1) 各种格式的资源,如文本、图像、视频、音频等格式的资源;
- (2) 各种形态的资源,如数据库、软件、图书、期刊、古籍、拓片、年画、手稿、舆图等;
- (3) 物理对象和抽象对象。为了服务系统达到某种能力,有时需要唯一标识符能标识抽象对象,比如标识抽象的作品、概念、主题等;
- (4) 各种粒度的资源,如可以标识单册图书和一套丛书、多媒体资源集合、图书的章节、期刊、期刊中的一期、一期中的一篇论文、一篇论文中的某一页、单页中的一幅图片、公式、一首乐曲和乐曲中的一段音乐等。

14.1.4 元数据

14.1.4.1 各种元数据及其关系

在国家数字图书馆中,在多个系统、多个层次上都包含有元数据。在本指南中,涉及到几个层次的元数据,特说明如下:

(1) 国家图书馆数字资源服务系统中的元数据，或书目数据库系统中的元数据。比如特色资源库（如民国期刊、敦煌文献等）中既有数字对象，又有对象的详细的元数据信息。

(2) 将来国家数字图书馆的集中元数据仓库中的元数据，按照国家数字图书馆建设规划方案的设计，集中元数据仓库将存放国家图书馆所有数字对象的元数据。

(3) CDOI 服务系统中，将注册已分配了唯一标识符的对象的核心的元数据。

上述三个层次的元数据的关系基本是这样的：

第（1）中的元数据是最详细最完整的元数据；

第（2）中的元数据可以是基本元数据集合，内容可以小于或等于（1）中的元数据；

第（3）中的元数据是最小、最核心的元数据，数据元素见《国家图书馆唯一标识符规范》的附录 A。

本指南中提及的国家图书馆元数据系统泛指（1）和（2）中的元数据。

14.1.4.2 元数据与唯一标识符的关系

一个资源如果需要作为一个独立对象提供服务，并被用户引用或第三方系统调用，就需要给这个资源对象分配一个唯一标识符，同时需要有一条元数据记录对该资源的有关信息进行描述和揭示。用于资源描述的元数据记录一方面可以在国家图书馆元数据系统中提供用户对该资源的查询服务，另一方面其核心的元数据会注册到唯一标识符服务系统中，提供唯一标识符解析过程中的元数据服务（参见第 3 章）。

唯一标识符系统中的元数据与所标识对象的元数据一一对应，唯一标识符服务系统中的元数据与国家图书馆元数据系统中的元数据也必须一一对应。

在国家图书馆元数据系统中的元数据除了提供查询服务、提供用户对资源的识别和确认服务外，从唯一标识符应用的角度还有一个重要作用是发布该资源对象的唯一标识符信息，用户在查询到该资源的元数据的同时，可获取资源的唯一标识符，使得用户可以方便的采用唯一标识符对资源进行引用、链接或调用。

14.2 唯一标识符服务系统技术框架

唯一标识符服务系统实现唯一标识符命名授权的管理、唯一标识符的注册服务和唯一标识符的解析服务。这是唯一标识符服务系统最基本的功能。唯一标识符服务系统的扩展服务还可以包括不同来源的元数据方案的存储，元数据的映射与互操作，情景敏感的多重解析服务的自动选择等等。

国家图书馆构建唯一标识符服务系统需要实现唯一标识符生成、注册和解析三个最基本功能。

14.2.1 Handle System 体系架构

Handle System 是实现唯一标识符注册与解析服务的全球通用的技术体系。国际主要的唯一标识符系统，如 DOI 系统，都是 Handle System 技术体系上的一个应用。

Handle System 定义了一套分层的 service 模型。顶层 service 由一个全球 service 构成，称为全球 Handle 注册中心——Global Handle Registry(GHR)。其下由其它的 handle service 构成，通常被称为区域 Handle service——Local Handle Services (LHS)。LHS 处于 GHR 的下层，它负责某个或某些名称空间下所有的 handle service，而不是指只在某个机构内部起作用的 handle service。

GHR 可以管理所有的 handle 名称空间。它与其它的 handle service 的功能基本相同，只是 GHR 还能够提供用于管理命名授权的服务，命名授权是作为 handle 来管理的，也就是说，命名授权 (NA) 同样由 handle 来标识并管理。NA handle 为客户端访问和利用 LHS 提供必要的信息。

一般来说，LHS 由对特定的命名授权有管理权限的组织来设立并维护。一个 LHS 可以对任意数量的 handle 名称空间负责，每个名称空间由一个唯一的命名授权来标识。LHS 和它所负责的 handle 名称空间集必须在 GHR 下注册。

Handle System 的一个重要特点就是它采用的是分布式架构。Handle System 整体是由许多 handle service 来构成的。

图 14-1 给出了 Handle System 分布式体系架构的示意图。最上层的 GHR 管理着 n 个区域 handle service (LHS)。

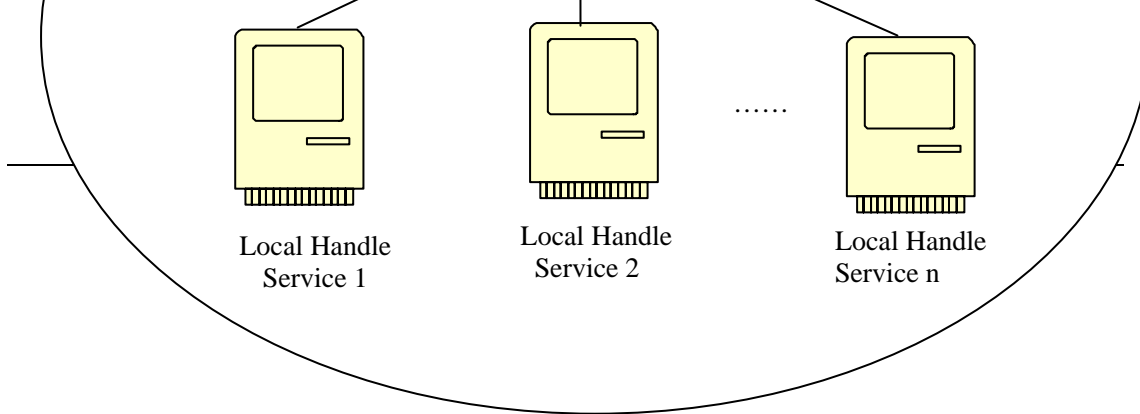


图 14—1 Handle 系统体系架构

每个 handle 服务管理着 Handle System 下的一个子名称空间 (sub-namespace)，不同 handle 服务下的名称空间互不重叠。子名称空间通常由一些命名授权下的 handle 集组成，负责这些命名授权的 handle 服务称为主服务 (home service)，并且是唯一一个为这些命名授权下的 handle 提供解析和管理服务的 handle 服务。在开始解析一个 handle 前，客户端需要判定所请求解析的 handle 的主服务在什么位置。每个 handle 的主服务也就是该 handle 的命名授权的主服务，并且该服务必须在 GHR 注册。客户端通过向 GHR 查询命名授权 handle (naming authority handle, 简称 NA handle) 来找到每个 handle 对应的主服务。

GHR 维护着所有的 NA handle。每个 NA handle 则维护着描述该 NA handle 所指示的命名授权的主服务相关的服务信息。这些服务信息列出了提供 handle 服务的服务站点，以及与每个站点内的每台 handle 服务器通讯的接口信息。当需要查找某个 handle 的主服务时，客户端可以查询 GHR，从对应的 NA handle 处获得相关的服务信息，该服务信息为客户端与主服务通讯提供了必要的信息。

图 14—2 给出了一个典型的 handle 解析的过程。

在该例中，主服务为 Local Handle Service。客户端试图解析 handle "cnri.dlib/july95-arms"，它必须先从 GHR 找到该 handle 对应的主服务。为此，客户端向 GHR 发出一个查询请求，要求得到与 "cnri.dlib" 对应的相关的 NA handle 信息。GHR 返回对命名授权 "cnri.dlib" 下的 handle 负责的 LHS 的服务信息，该服务信息使得客户端可以和 LHS 通讯以便对 handle "cnri.dlib/july95-arms" 进行解析。

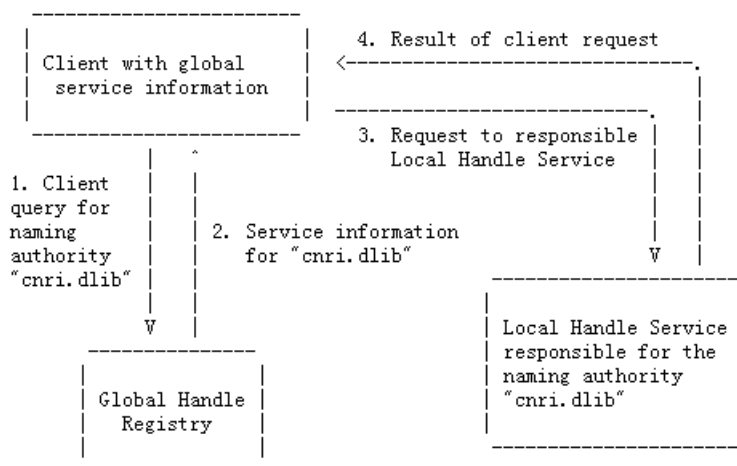


图 14-2 Handle 解析过程

14.2.2 国家图书馆唯一标识符服务系统体系框架

建议国家图书馆唯一标识符服务系统成为全球 Handle 服务下的一个区域服务 (LHS)，管理 CDOI 子命名授权下的所有唯一标识符的注册和解析服务。其中一级命名授权需要在全局 Handle 注册中心注册。这样，国家图书馆的唯一标识符都可以通过全球 Handle 分布式服务体系进行解析服务。

为实现国家图书馆唯一标识符的生成、注册与解析服务，国家图书馆唯一标识符服务系统应由三个子系统组成：唯一标识符生成子系统、唯一标识符注册子系统与唯一标识符解析子系统。

唯一标识符生成子系统负责国家图书馆资源对象的唯一标识符的生成，包括通过元数据批量生成唯一标识符，或单个资源对象唯一标识符生成。

唯一标识符注册子系统负责将唯一标识符和标识对象的元数据注册到国家图书馆本地 CDOI 服务中。

唯一标识符解析子系统负责国家图书馆唯一标识符名称的解析服务。

14.3 唯一标识符解析服务

14.3.1 什么是解析

唯一标识符是网络环境下一个实体的标识码。解析是通过向唯一标识符系统的网络服务器提交一个实体的唯一标识符，从而获取被标识实体当前的一个或者多个信息，这些信息可能是标识具体的实体对象地址，关于实体对象的服务、或者是与实体相关的元数据。

例如，在 DNS 系统中，解析的过程就是从域名，如 www.doi.org，到能用

于网络主机识别的 IP 地址的过程，如 132.151.1.146。对于国家图书馆的唯一标识符，解析的过程是从 CDOI 的名称，如 cdoi.011001/123456，解析到一个或多个 URL 地址：如具体的数字对象或者数字对象的元数据。

使用多重解析，唯一标识符能在网络环境下解析到任意的不同的值：多个 URL 地址、其他的唯一标识符名称或者代表实体的元数据。解析需要反馈所有当前信息的属性，或者一个数据类型的所有属性；这些属性需要进一步被其他的客户端软件所应用。简单地说，用户能获得一个选择的表单，或是通过一个复杂的自动过程，自动选择合适的值以进一步处理。

14.3.2 CDOI 的解析过程

唯一标识符生成注册后，通过唯一标识符的解析服务来实现对数字资源的定位。唯一标识符的解析是由两层的的服务实现的，全球唯一标识符解析服务和国家图书馆唯一标识符解析服务。解析服务可以是正向解析服务，也可以是反向解析服务。

14.3.2.1 正向解析过程

当唯一标识符服务系统收到唯一标识符的解析请求时，其过程基本如下：

(1) 客户端引导

当客户端遇到一个 CDOI 名称时，通过网络向全球 Handle 注册中心服务器提交请求，全球 Handle 注册中心根据请求，返回该唯一标识符的相应的主服务的服务器信息，即国家图书馆 CDOI 系统解析子系统。

(2) 查询操作

客户端发送查询请求给国家图书馆的 CDOI 解析子系统，国家图书馆 CDOI 解析系统对该 CDOI 名称进行解析，反馈客户端对象的 URL 地址或对象的相关描述性信息。

正向解析过程又可以分为简单解析和多重解析。

(1) 简单解析

简单解析是单一的指向性解析。每个唯一标识符只有一个可解析到的 URL 地址。简单解析过程见图 14-3。

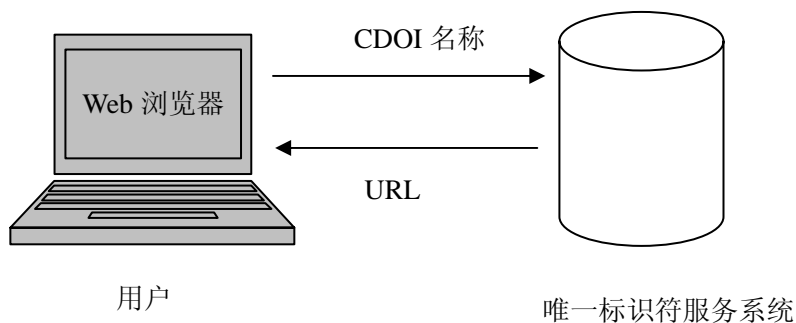


图 14-3 唯一标识符解析到单一的 URL 地

(2) 多重解析

多重解析允许一个实体被解析到该实体的多个 URL 地址、多个数据类型/值对, 多个其他相关实体, 如父子关系和其他各种关系。一个唯一标识符解析到多个结果, 用户手工可选择一个解析结果。多重解析示意图见图 14-4。

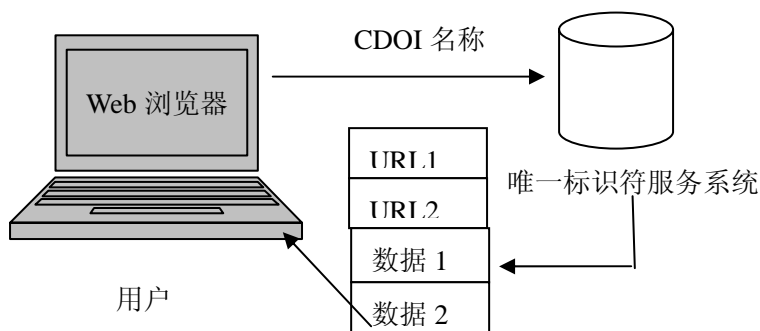


图 14-4 唯一标识符解析多重解析示意图

14.3.2.2 反向解析过程

除根据唯一标识符进行解析之外, 唯一标识符系统也可根据元数据查询和解析唯一标识符。

客户端向国家图书馆 CDOI 系统的解析子系统提出通过标识对象的元数据查询唯一标识符的请求, 国家图书馆唯一标识符服务系统解析子系统根据请求, 查询返回相对应的 CDOI 名称或者返回系统判断信息。

14.4 唯一标识符生成与注册

14.4.1 总体流程

在整个国家图书馆数字资源加工与服务体系中, 唯一标识符生成与注册发生在哪个阶段, 与数字资源加工、元数据加工、数字资源服务系统等等的关系是怎样的呢?

唯一标识符的生成与注册过程是在完成数字资源加工、数字资源元数据描述、数字资源上载到应用服务系统后开始。唯一标识符是数字对象的一个逻辑标识符, 与数字资源加工无关, 与数字资源加工时对象的物理文件划分、数字资源的存放地址、数字资源的文件名都没有关系。

唯一标识符是在国家图书馆元数据系统基础上, 由唯一标识符生成系统自动生成。唯一标识符生成与注册过程将与国家图书馆元数据系统、唯一标识符生

成子系统、唯一标识符注册子系统发生关联。国家图书馆数字资源加工系统与唯一标识符的生成与应用不直接发生关系。唯一标识符生成与注册过程以及相关系统的关系见图 14—5。

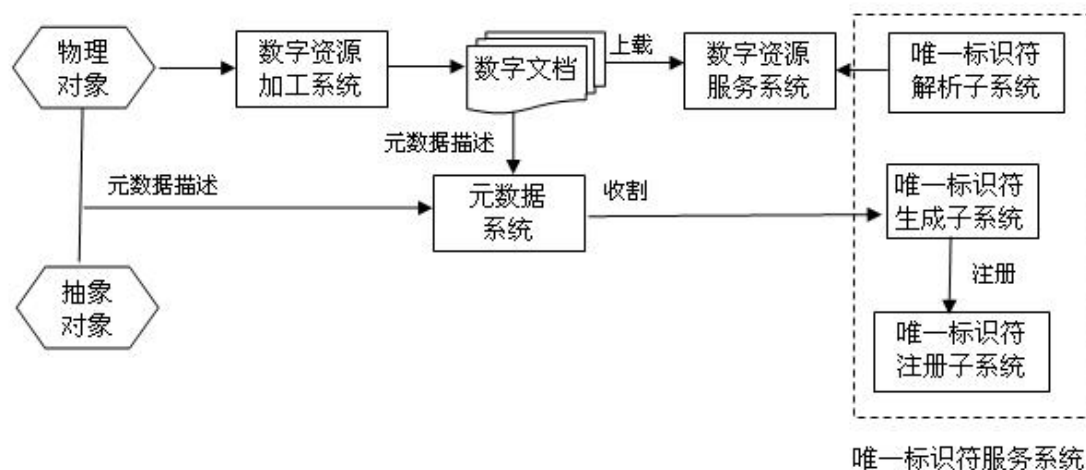


图 14—5 唯一标识符生成、注册过程中相关系统的关系

在上图中，物理对象、抽象对象、数字文档、元数据涵盖了国家图书馆所有的资源类型。

物理对象可以包括中文图书、西文图书、中文期刊、西文期刊等等。

抽象对象可以包括作品、表达，也包括概念、主题等的抽象概念。

数字文档包括由国家图书馆馆藏物理对象经数字化加工生成的数字对象，也包括购买的数字对象、从网络获取用于长期保存的数字对象等等。

国家图书馆元数据系统中的元数据包括了国家图书馆物理对象的元数据，比如图书的书目数据是一种元数据，抽象对象的元数据，数字对象的元数据，还包括购买的元数据资源。

在上图中，元数据系统泛指第 1.6 节中的（1）和（2）中的元数据。元数据系统中的元数据包括了所有物理对象、抽象对象与数字对象的元数据信息，同时元数据描述也揭示各对象之间的关系。

所以，国家图书馆唯一标识符的生成过程的设计是以国家图书馆元数据系统为基础，采取从元数据系统抽取元数据，通过唯一标识符生成子系统自动生成各种资源对象的唯一标识符的批处理方法。在特殊情况下，也可以根据国家图书馆的元数系统中的元数据信息，手工向唯一标识符生成子系统输入元数据信息生成唯一标识符。

唯一标识符生成子系统从国家图书馆元数据系统采集对象的元数据，并按唯一标识符分配规则自动给每个对象分配唯一标识符，然后通过唯一标识符注册子系统进行注册登记，最后才能提供唯一标识符的名称解析服务。

14.4.2 唯一标识符生成与注册过程中的用户角色

唯一标识符生成与注册过程中涉及注册管理员和注册用户两种类型的用户。

(1) 注册管理员

注册管理员是在唯一标识符服务系统中进行唯一标识符和元数据注册登记的人员，他负责：

- 用批次或个别的方式进行唯一标识符的注册登记；
- 在解析系统中对唯一标识符和相关的元数据及其数字对象的 URL 地址进行增、删、改操作；
- 将登记错误报告发送给各应用系统管理员；
- 负责唯一标识符服务系统的用户权限管理。

(2) 注册用户

注册用户按资源类型进行划分。每个资源类型的注册用户负责：

- 用自动收割方式或手工批次输出方式，从国家图书馆元数据系统获取该类型资源的元数据及其 URL 地址，将元数据载入唯一标识符生成系统，进行唯一标识符生成操作；
- 将生成的唯一标识符及元数据提交唯一标识符系统管理员进行注册；
- 接受唯一标识符注册管理员反馈的注册错误，并及时进行错误处理；
- 执行成功注册的唯一标识符回填入国家图书馆元数据系统的操作；
- 负责该资源类型后缀规则的制定，保证在同一前缀命名授权下后缀的唯一性，从而保证 CDOI 名称的唯一性；
- 保证一个对象分配一个 CDOI 名称；
- 保证每个注册 CDOI 的对象有一对应的元数据记录。

14.4.3 唯一标识符生成与注册过程

唯一标识符生成与注册过程涉及国家图书馆元数据系统、唯一标识符生成系统和唯一标识符注册系统三个系统，需要注册用户和注册管理员两种角色。

在唯一标识符服务系统成功搭建后，国家图书馆首先需要在全球 Handle 注册中心注册国家图书馆唯一标识符命名授权前缀，只有保证国家图书馆命名授权前缀唯一，才能保证在该命名授权下分配的唯一标识符唯一。

唯一标识符生成与注册过程见图 14-6 所示。唯一标识符的生成和注册过程如下：

(1) 注册用户从国家图书馆元数据系统获取资源对象的元数据和资源对象的 URL 地址，获取的方式可以通过标准接口协议，由唯一标识符生成系统从

元数据系统自动收割，或元数据系统批次输出；

(2) 注册用户使用唯一标识符生成系统生成数字对象唯一标识符；

(3) 将生成的唯一标识符和资源对象元数据列表按规定格式提交给唯一标识符注册管理员；

(4) 唯一标识符注册管理员负责在唯一标识符服务系统进行唯一标识符注册和元数据登记，并将注册错误信息反馈给注册用户；

(5) 注册用户对注册错误结果进行处理，重新生成唯一标识符，并提供注册管理员继续注册，如此反复，直到全部注册成功；

(6) 注册用户启动国家图书馆元数据系统的唯一标识符获取和回填功能，将注册成功的唯一标识符回填到元数据系统中。这需要通过唯一标识符服务系统提供的标准协议接口。

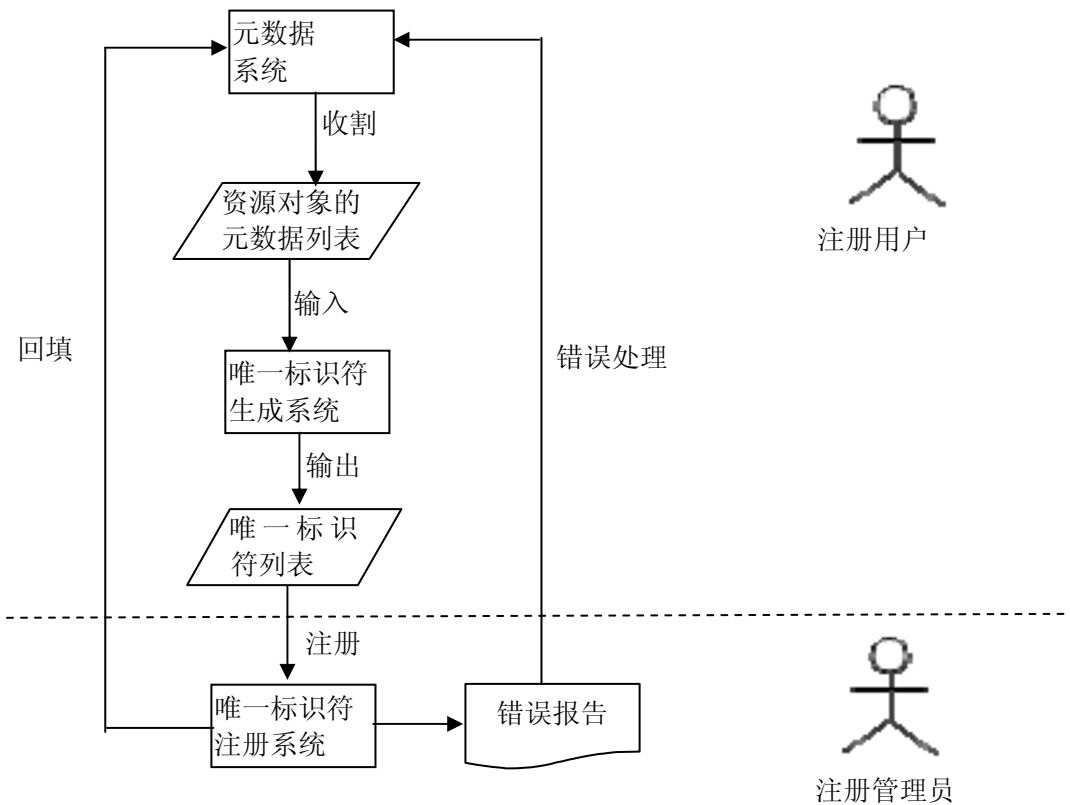


图 14-6 唯一标识符生成与注册流程图

14.4.4 唯一标识符生成的例子

14.4.4.1 博士论文的唯一标识符生成

博士论文作为国家图书馆收藏的物理资源，按照唯一标识符规范，可以为其分配唯一标识符。博士论文唯一标识符分配过程如下：

(1) 博士论文到馆后, 需要对其进行元数据描述, 元数据进入国家图书馆书目数据库系统, 如果国家图书馆有集中元数据仓库, 博士论文的元数据也可以同时进入到集中元数据仓库揭示;

(2) 负责博士论文唯一标识符注册的注册用户定期(如一周或一个月)从书目数据库或集中元数据库中抽取新编目的博士论文元数据记录。抽取的方式可以是元数据系统批次输出或唯一标识符生成子系统自动收割, 这取决于国家图书馆将来的系统设计;

(3) 注册用户启动唯一标识符生成子系统, 唯一标识符生成子系统输入博士论文的元数据记录, 输出唯一标识符和对应的元数据列表;

(4) 博士论文注册用户将生成的唯一标识符和元数据列表交注册管理员, 注册管理员在唯一标识符注册子系统中进行唯一标识符的注册;

(5) 注册管理员将注册成功或未成功的结果反馈博士论文注册用户;

(6) 注册用户对错误结果进行修改, 重新进行生成和注册过程, 直至全部待注册博士论文完成注册;

(7) 注册用户启动国家图书馆元数据系统的收割唯一标识符功能, 从唯一标识符系统收割唯一标识符回填到元数据系统中。

14.4.4.2 西夏文献的唯一标识符生成

西夏文献是国家图书馆的特色资源, 国家图书馆对西夏文献进行数字化扫描, 在西夏碎金特色资源库中提供服务。西夏文献的数字资源按照唯一标识符规范可以为其分配唯一标识符, 西夏文献的数字化加工和唯一标识符分配过程如下:

(1) 对西夏文献进行数字化扫描, 将扫描生成的数字文件按国家图书馆数字资源加工过程的要求保存;

(2) 为该西夏文献进行元数据加工;

(3) 将西夏文献的数字版本和元数据按西夏碎金特色资源库的格式上载到西夏碎金特色资源库中。如果国家图书馆有集中元数据仓库, 西夏文献的元数据记录也同时转入集中元数据仓库;

(4) 负责西夏文献唯一标识符注册的注册用户定期(如一周或一个月)从西夏碎金特色资源库或集中元数据库中抽取新的西夏文献的元数据记录。抽取的方式可以是元数据系统批次输出或唯一标识符生成子系统自动收割, 这取决于国家图书馆将来的系统设计;

(5) 注册用户启动唯一标识符生成子系统, 唯一标识符生成子系统输入西夏文献的元数据记录, 输出唯一标识符和对应的元数据列表;

(6) 西夏文献注册用户将生成的唯一标识符和元数据列表交注册管理

员，注册管理员在唯一标识符注册子系统中进行唯一标识符的注册；

(7) 注册管理员将注册成功或未成功的结果反馈西夏文献注册用户；

(8) 注册用户对错误结果进行修改，重新进行生成和注册过程，直至全部待注册西夏文献完成注册。

(9) 注册用户启动国家图书馆元数据系统的收割唯一标识符功能，从唯一标识符系统收割唯一标识符回填到元数据系统中。

14.5 国家图书馆唯一标识符名称分配规则

14.5.1 唯一标识符分配的基本原则

是否需要给一个对象分配唯一标识符，取决于这个对象是否有一个独立的元数据记录。如果在元数据系统中，该对象有一个独立的元数据记录，那么可以给这个对象分配唯一标识符，但是如果两个对象在元数据描述中是作为一个对象的，比如两个数字复本，用一条元数据记录进行描述，那么唯一标识符分配时，给这两个对象分配相同的唯一标识符。

那么，什么情况下两个或多个对象应作为一个对象进行元数据描述，又在什么情况下，两个或多个对象作为多个对象分别进行元数据描述呢，答案是需要依据国家图书馆元数据标准和元数据著录规则的规定。

14.5.2 各种特殊对象的唯一标识符分配

14.5.2.1 多个数字复本

一个数字对象可能存放在多个资源库中，有多个数字复本，有多个访问地址。比如：分别来自超星与方正的同一版本图书的数字资源，那么需要对来自不同来源的同一数字对象进行元数据描述，在同一条元数据记录中记录多个来源的 URL 地址，而不是多条元数据记录。当解析系统解析唯一标识符时，通过多重解析机制，可以有选择地访问存放在不同地址中的对象。

14.5.2.2 相同内容、不同数字格式的资源

一种资源被数字化为多种数字格式，如 tif 格式、pdf 格式等，是作为同一种对象，还是作为不同的数字对象，这取决于国家图书馆元数据标准和元数据著录规则的规定。

如果作为同一种资源对象，用一条元数据记录对其进行描述，那么就给其分配一个唯一标识符，那么需要在元数据记录中记录不同格式对象的 URL 地址，并且要求元数据描述信息中能区分每个 URL 地址所对应的格式。唯一标识符解析系统解析该唯一标识符时，提供用户多个数字格式的 URL 地址，用户可以根

据权限和需要选择不同的格式访问资源。

如果作为多种对象对待，每种格式对象对应有一条元数据记录，那么就分别对不同的数字格式对象分配唯一标识符，用户引用不同格式的对象唯一标识符，解析系统解析该唯一标识符时，通过记录在对象元数据中的对象的 URL 地址直接调用该数字格式的对象。

14.5.2.3 物理资源

按照唯一标识符规范，物理资源也可以分配唯一标识符。数字资源通过唯一标识符解析系统，可以直接调用数字对象。与数字资源不同的是，物理资源通过唯一标识符解析系统无法直接调用到物理对象，通过解析系统调用到该物理资源的详细元数据信息，比如书目数据，通过索书号来获取物理资源。

离线数字资源，如光盘、磁带等，实质上是不能通过唯一标识符直接调用的，从唯一标识符的分配和服务上采取与物理资源相同的处理办法。

14.5.2.4 不同粒度层次的资源

按照唯一标识符规范，可以对任何粒度层次的资源分配唯一标识符。比如可以对一种期刊从整体上分配唯一标识符，也可以对期刊中的某一期分配唯一标识符，也可以对某一期中的每一篇论文分配唯一标识符。再比如可以对一首乐曲分配唯一标识符，也可以对这首乐曲中的一段音乐分配唯一标识符。在各种粒度层次上分配唯一标识符都是可以的，重要的是只要一个对象分配了唯一标识符，就需要有对应的元数据对其进行描述，在元数据中揭示该对象的唯一标识符，也通过元数据揭示不同粒度层次对象之间的关系。

14.5.2.5 简单对象与复合对象

如果多个简单对象被重新组合为一个复合对象，简单对象的元数据记录和已分配的唯一标识符继续保留，同时为复合对象重新编制一条元数据记录，并分配唯一标识符。在元数据记录中可以描述复合对象与其所包含的简单对象的关系。

14.5.3 唯一标识符后缀分配规则

按照第 4 章的设计考虑，国家图书馆唯一标识符的分配是通过唯一标识符生成系统自动生成唯一标识符。为保证自动生成的唯一标识符的唯一性，也需要根据国家图书馆的实际状况，制定相应的唯一标识符命名细则。

14.5.3.1 有集中元数据仓库下的名称分配

在国家数字图书馆系统建设集中元数据仓库的情况下，所有的物理对象、

数字对象、虚拟对象的元数据都在集中元数据仓库存储，提供快速的集中检索服务。这种条件下，唯一标识符的生成系统可以直接从集中元数据仓库收割元数据并分配唯一标识符，这时唯一标识符的后缀建议直接采用元数据记录在集中元数据仓库中的记录标识号。

例如：“西游记年画”的元数据记录在集中元数据仓库中的记录号假如是 H08718，那么分配给“西游记年画”的唯一标识号是：cdoi.011001/H08718

14.5.3.2 无集中元数据仓库下的名称分配

如果国家图书馆无集中元数据仓库，或者集中元数据仓库不是所有印本文献、数字文献等的全部资源的元数据大集中，这时的唯一标识符名称构成采用分段式后缀方法。

后缀采用两段式标识方法，两者之间用圆点（‘.’）分隔。

第一段为元数据系统的标识号，直接采取数字编码方式，如 001，002。

第二段为元数据记录在该元数据系统中的记录号。

例如，给印本文献的书目数据库赋予唯一标识号：001，《胡适评传》在书目数据库中的 ID 号是：003582409，那么胡适评传的唯一标识号是：cdoi.011001/001.003582409。

14.6 唯一标识符的维护

在唯一标识符的分配和管理中，重点需要关注如下几个方面的问题：

14.6.1 重复控制

一个唯一标识符不能分配给多个对象，这通过唯一标识符生成子系统和唯一标识符注册子系统来进行控制。

一个对象有时可能会被分配了多于一个的唯一标识符，虽然这种情况不至于引起唯一标识符服务系统的瘫痪，但如果出现这种重复分配将给用户造成混乱和迷惑，影响系统的服务质量。

控制一个对象只分配一个唯一标识符要从控制对象的元数据描述是唯一的入手。在集中元数据仓库中，要保证一个对象只有一条元数据描述记录，这样才能保证一个对象被分配一个唯一标识符。

14.6.2 对象移动

如果一个已分配了唯一标识符的数字对象移动存放位置，需要修改国家图书馆元数据系统中元数据记录对象的 URL 地址，同时也要对唯一标识符服务系统中的元数据记录进行维护与修改，保证解析到准确的对象地址。

14.6.3 对象删除

如果一个已分配了唯一标识符的对象被删除，这个唯一标识符将永不再分配给其他对象。在唯一标识符系统的元数据记录中需要注明对象被删除信息，当系统解析这个唯一标识符时，返回客户端错误信息。

如果由于重复的元数据记录引起重复分配唯一标识符，这时可以对元数据系统中的两条元数据记录进行合并。但是对象已经被分配了两个或多个唯一标识符，提供服务的唯一标识符是不能删除的，都应该定位到一个对象，这时唯一标识符服务系统中的元数据记录是应该继续保持两条元数据记录，都定位到同一个对象，还是将两条元数据记录合并，用一条元数据记录对应两个唯一标识符，采用何种处理方式应与系统开发商详细讨论。

14.6.4 唯一标识符维护

唯一标识符生成后，需要对唯一标识符服务系统中的唯一标识符和相关的元数据进行维护，其任务包括：

(1) 保证每个唯一标识符相对应的元数据是准确的、简化的，与元数据系统的元数据描述相一致。

(2) 当唯一标识符标识的数字资源在国家图书馆元数据系统中的元数据发生变更后，如元数据合并、删除、对象的存放地址变更、元数据信息修改等，应及时维护和更新国家图书馆唯一标识符服务系统中的唯一标识符及对应的元数据。

14.7 唯一标识符对相关系统的要求

14.7.1 对唯一标识符服务系统的要求

14.7.1.1 总体要求

(1) 因特网兼容性。CDOI 经由全球信息系统传输，唯一标识符服务系统应与全球唯一的地址空间和相关的通讯协议保持兼容；

(2) 开发标准。在系统的开发过程中，相关的工作严格遵循软件开发标准和规范，或采用软件业普遍遵循的事实标准，以保障开发工作的质量；

(3) 系统功能实现采用的技术标准推荐：

- CDOI 注册、管理和解析：Handle System
- 元数据封装与互操作协议：OAI-PMH、METS
- 资源链接：OpenURL

- 开放式应用交互和服务描述：Web Services 系列规范 (SOAP\UDDI\WSDL)

(4) 支持跨平台。系统应支持在 Windows、UNIX、Linux 等平台上部署和运行；

(5) 系统健壮性。系统应充分考虑到运行过程中可能出现的异常情况 & 问题，并在系统运行中对这些情况 & 问题进行及时判断 & 处理，从而保证系统在各种合理的异常情况中长期、稳定地运行。对于一些特殊的异常情况，应有预案，通过人工或其他手段保证系统的正常运行；

(6) 可维护性。系统结构应合理、清晰，代码标准、规范，注释与文档齐全，可读性强，并对可预见的技术发展预留接口，以易于今后系统的修改 & 各类维护性工作；

(7) 系统响应能力。在面向客户端 CDOI 解析 & 查询请求的功能模块，系统应支持 20 个用户的并发操作，延迟时间不超过 30ms；面向管理员 CDOI 注册、管理 & 维护功能模块，系统应支持 20 个用户的并发操作，延迟时间不超过 50ms；

(8) 数据处理 & 容灾备份能力。系统设计与实施应考虑 & 保证唯一标识符服务系统在今后五年内对数据增长 & 用户增长保持充足的处理能力，并能保证安全、顺畅的并发处理能力。建立安全、可靠的数据备份、恢复机制；

(9) 系统安全性。应用系统设计应采用基于公钥/私钥的安全处理协议，系统安全包括数据传输安全、数据存储安全、数据管理安全等方面，在操作系统层次，推荐使用 Unix 或 Linux 系统。应用系统设计不应留有漏洞。系统设计完成后应通过安全测试。

14.7.1.2 对生成子系统的要求

唯一标识符注册子系统提供数字对象的唯一标识符 & 相关元数据（包括 URL）的登记 & 注册服务。在唯一标识符进行登记 & 注册之前，应该准备好数字对象的唯一标识符、数字对象存放的 URL 和数字对象的元数据。而唯一标识符生成子系统的主要功能就是从国家图书馆元数据系统获取资源对象的核心元数据（即 CDOI 的唯一标识符元数据方案中规定的元数据）、资源对象的 URL，并自动给每个资源对象分配一个唯一标识符。

唯一标识符生成系统应实现如下功能：

(1) 提供手工单条方式 & 批处理两种模式的唯一标识符生成；

(2) 单条方式的元数据生成要求唯一标识符生成系统提供图形化编辑界面，供注册用户输入对象的元数据信息，系统自动生成对象的唯一标识符；

(3) 批处理生成模式要求唯一标识符生成系统能够自动获取国家图书馆元数据系统中数字对象的元数据。应支持 OAI-PMH、Z39.50、SRU/W 等接口协议，通过元数据自动收割的方式从国家图书馆元数据系统中获取元数据。批处理模式也需要能接受国家图书馆元数据系统输出的标准格式（XML、ISO 2709 等）的元数据；

(4) 支持按照唯一标识符前缀、后缀命名规则，通过注册用户的参数配置自动生成唯一标识符。每个数字对象的唯一标识符由前缀和后缀组成，两者之间由“/”分隔。其中，前缀由命名授权组成，国家图书馆所有的资源都有相同的前缀“cdoi.011001”。后缀是元数据记录在国家图书馆集中元数据仓库中的唯一标识号。

唯一标识符生成系统工作流程如图 14-7 所示。数字对象元数据可以通过单条记录输入到唯一标识符生成系统，可以通过批量导入到唯一标识符生成系统，或者是从国家图书馆元数据系统自动收割。输入、导入或者收割的数字对象元数据，根据制定的后缀分配规则，生成该数字对象的后缀，同时和已经分配的前缀共同生成唯一标识符。将生成的唯一标识符和 URL 注册到唯一标识符服务系统，完成注册后，唯一标识符服务系统返回注册结果列表，如果成功，则表示唯一标识符的生成和注册成功；如果失败，则重新返回到唯一标识符生成系统，重新生成唯一标识符并再次进行注册。

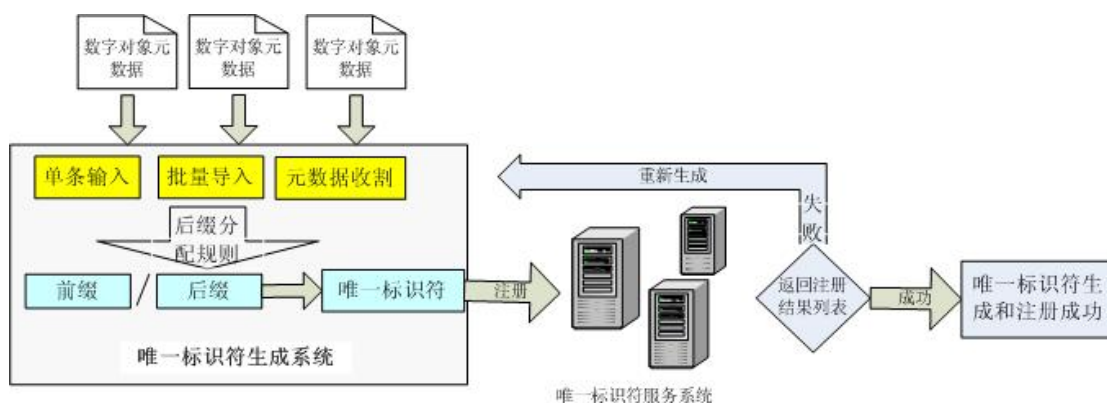


图 14-7 唯一标识符生成系统

14.7.1.3 对注册子系统的要求

唯一标识符注册子系统的要求是：

- (1) 支持唯一标识符注册，支持唯一标识符标识对象的元数据注册；
- (2) 支持两种方式的唯一标识符及其元数据登记注册：
 - 注册管理员通过 Web 界面手工登记；
 - 注册管理员批次登记。
- (3) 注册与结果反馈。系统具有唯一标识符及元数据注册过程中的格式

校验与查重等能力，并反馈成功注册结果和注册未成功记录；

(4) 唯一标识符及元数据的管理。支持对唯一标识符名称、元数据的管理功能，包括增加、删除、修改唯一标识符名称和修改唯一标识符对应的元数据值。在完成管理请求之前，服务器应先进行用户身份验证；

(5) 命名授权管理。将命名授权作为一个唯一标识符进行管理。客户端可以通过向全球 Handle 注册中心服务器发送请求来获取所有命名授权列表，以及命名授权下的子命名授权。对某个命名授权有管理权限的人也就是获得了该命名授权下的唯一标识符的管理权限；

(6) 提供唯一标识符获取的接口。系统应提供公开调用接口，供国家图书馆元数据系统获取已登记的唯一标识符并回写到其系统中；

14.7.1.4 对解析子系统的要求

(1) 支持按规范第 9 章规定的解析规则提供唯一标识符名称的简单解析服务；

(2) 支持多重解析。解析请求应能够返回对象的所有相关信息、个别值或一个数据类型的所有的值，如多个 URL 地址；

(3) 客户端认证。客户端认证包括客户端和服务器之间交换信息，如服务器到客户端对客户端的认证，客户端对来自服务器端认证的服务器端的认证，密钥认证时对请求和返回结果的认证请求等。在成功的客户端认证的基础上，服务器根据管理员是否对请求进行认证来返回结果，如果管理员有足够的权限，则对请求进行处理并返回结果，否则服务器将返回错误信息。

(4) 服务参照。当发送到 CDOI 名称服务器的查询请求的 CDOI 是通过其他 CDOI 名称服务器或者服务管理时，服务器可以返回一个该 CDOI 所属的名称服务的参照信息，或者是简单地返回一个错误信息。

(5) 提供元数据和唯一标识符查询。唯一标识符服务系统应提供按数字对象的核心元数据查询 CDOI 以及按 CDOI 查询数字对象的元数据的功能。查询的方式可以是向查询用户提供 Web 界面的检索入口，或接受用户客户端发送的查询请求。查询所需的参数和返回结果的方式需预先定义、公布，供查询者在查询时指定；

(6) 按数字对象的核心元数据查询 CDOI，查询者可能提供该数字对象完整的元数据（如一条完整的期刊论文书目记录），或只能提供少数必备元数据（如论文题名和作者名）；允许命中多个结果，并指定是否将检索结果定向到数字对象的 URI 或显示 CDOI；

(7) 按 CDOI 查询，返回的结果形式可能是数字对象、数字对象的元数据信息；

(8) 服务系统可根据 CDOI 和元数据使用权益的需求, 决定元数据查询和 CDOI 查询是否需要经过用户认证或匿名访问的不同的用户使用权限。

14.7.2 对国家图书馆元数据规范和集中元数据系统的要求

唯一标识符生成系统需要从国家图书馆元数据系统提取数字对象元数据, 根据元数据的属性自动生成唯一标识符。对国家图书馆元数据规范和元数据系统有如下要求:

(1) 资源对象的唯一标识符分配粒度是以元数据的揭示粒度为基础的。元数据规范应能揭示任意粒度层次的数字对象, 能揭示任意格式和载体的对象, 能揭示物理对象和抽象对象;

(2) 元数据系统数据元素必须包含数字对象的 URL, 以便唯一标识符解析服务器能直接定位到数字对象。数字对象的 URL 也可以采用基于元数据动态生成的方式, 在生成唯一标识符的同时动态生成 URL, 这时元数据系统必须提供各种不同的资源对象的 URL 生成规则;

(3) 元数据规范应有一基本元数据描述集合, 用这个基本集合可以描述所有资源类型的共同特征, 唯一标识符生成系统需要从元数据系统抽取资源对象的基本元数据描述字段进行唯一标识符核心元数据的注册;

(4) 元数据系统应支持 Z39.50、OAI-PMH、SRW/U 等协议, 并提供相应的接口, 以便唯一标识符生成子系统可以自动收割元数据;

(5) 元数据系统应能从唯一标识符注册和解析系统中收割已成功注册的唯一标识符并回填到元数据系统中;

(6) 元数据系统中的数据应与唯一标识符服务系统中的核心元数据相对应, 当元数据系统中记录进行增、删、改操作后, 应能提供方法和手段动态更新唯一标识符服务系统中相应的核心元数据记录信息;

(7) 根据唯一标识符应用的需要, 有时唯一标识符解析响应页面不是直接到数字资源服务系统的数字资源页面, 而是到元数据系统的响应页。如对于物理对象, 这时元数据系统应该能够通过 URL, 提供唯一标识符的响应页。

14.7.3 对国家图书馆数字资源服务系统的要求

国家图书馆数字资源服务系统提供国家图书馆数字全文文献服务。数字资源服务系统、唯一标识符服务系统、国家图书馆元数据系统是唯一标识符服务体系的有机组成部分。第三方系统的用户通过唯一标识符系统的解析到达数字资源服务系统中的目标对象。

所以实现唯一标识符的解析和目标对象获取, 国家图书馆数字资源服务系

统应达到如下基本要求：

(1) 提供数字对象可获取的地址。数字资源的服务系统应能够向“唯一标识符系统”提供可直接获取数字对象的网络地址(URL)。

(2) 数字资源服务系统中的数字对象都需要有一对应的元数据描述，并且该元数据记录必须在国家图书馆元数据系统中反映，以便唯一标识符生成系统提取元数据信息生成唯一标识符。

14.8 国家图书馆唯一标识符系统运行管理

14.8.1 国家图书馆唯一标识符管理体系

国家图书馆唯一标识符技术框架为两层结构，上层为全球唯一标识符注册中心，在全球唯一标识符注册中心下面建立国家图书馆的区域唯一标识符服务，参见第2章。

在这样一个技术框架下，国家图书馆唯一标识符运行管理体系应包括一个国家图书馆唯一标识符管理中心和多个国家图书馆唯一标识符注册用户。注册用户可按资源类型分组，也可以按部门分组。图 14-8 是一个按资源类型分组的管理体系框架示意图。



图 14-8 国家图书馆唯一标识符管理体系

14.8.2 国家图书馆唯一标识符管理中心

国家图书馆唯一标识符管理中心全面负责国家图书馆唯一标识符的有关政策的制定和唯一标识符服务系统日常运行，具体应履行如下职责：

(1) 向全球唯一标识符注册管理中心注册申请中国唯一标识符一级命名授权，这个申请下来的一级命名授权可以是“cdoi”，也可以是数字代码，如“101”等；

(2) 负责一级命名授权段下的二级命名授权段的分配，同时负责与全球

唯一标识符注册中心的协调，负责已分配的命名授权的注册和区域服务有关参数在全球注册中心的配置等。比如，负责分配国家唯一标识符二级命名授权，按唯一标识符规范，国家图书馆二级命名授权是“011001”，在分配二级命名授权后，负责将国家图书馆唯一标识符前缀在 GHR 进行注册，将国家图书馆区域唯一标识符服务器的相关技术参数在 GHR 中进行设置等；

(3) 负责国家图书馆唯一标识符运行管理的有关政策、规则的制定，负责唯一标识符命名和分配的实施细则的制定，比如可以根据国家图书馆需求的发展，制定唯一标识符的扩展规则；

(4) 负责接收国家图书馆唯一标识符注册用户的注册申请，并实施唯一标识符的注册；

(5) 负责国家图书馆唯一标识符服务系统的稳定可靠和持续运行；

(6) 代表国家图书馆与国内外有关的唯一标识符系统及相关机构合作，促进国家图书馆唯一标识符与其他系统的合作与互操作，促进唯一标识符中国的标准化、规范化发展。

14.8.3 国家图书馆唯一标识符注册用户职责

按照唯一标识符生成过程（参见第 4 章），唯一标识符的生成是在完成资源的元数据加工、数字资源上载到服务系统提供服务之时开始。国家图书馆唯一标识符的注册用户应该是对国家图书馆的资源对象的元数据进行加工与管理的人员或机构，各类资源对象的元数据管理员通过管理元数据实施对资源的管理，他们全面了解资源对象的元数据状况和资源对象的服务状况，是最合适的注册用户。注册用户在唯一标识符管理运行体系中的角色和任务是：

(1) 负责某一或几类资源的元数据加工、元数据管理和维护更新；

(2) 负责按唯一标识符规范生成对象的唯一标识符；

(3) 向国家图书馆唯一标识符管理中心提交唯一标识符注册申请，申请注册唯一标识符时必须提供唯一标识符标识对象的元数据描述信息和唯一标识符；

(4) 在资源的生命周期中，负责资源对象的描述性元数据的维护更新，负责向唯一标识符管理中心提交由元数据维护更新或其他变化（如数字资源移

动、删除等)引起的唯一标识符注册系统中注册元数据和唯一标识符的维护申请,并提供相关维护信息,由唯一标识符管理中心的注册管理员统一对注册元数据和唯一标识符进行维护更新。

14.8.4 国家图书馆唯一标识符注册用户划分方案

由于国家图书馆唯一标识符管理体系采用了唯一标识符管理中心和注册用户两级管理体系,通过上面的职责和任务的划分,可以看到注册用户是贯穿于数字对象的元数据加工、数字对象服务和唯一标识符分配的关键角色,注册用户在数字图书馆的运行中起着至关重要的作用。注册用户划分需要考虑国家数字图书馆建设的需要、国家图书馆业务组织结构、国家图书馆资源对象的类型等各种复杂因素,下面提供两个注册用户划分方案,供参考:

(1) 按部门和业务组划分

按照国家图书馆目前的业务组织结构进行注册用户分组。目前国家图书馆有中文文献采编部、西文文献采编部、古籍部、讲座展览部等等。各部门负责各自所负责的资源唯一标识符分配。

例如:以中文文献采编部为基础,设置中文文献注册用户,负责中文文献的数字化加工、元数据加工,唯一标识符分配;以西文文献采编部为基础,设置西文文献注册用户,负责西文文献的数字化加工、元数据加工,唯一标识符分配;以古籍部为基础,设置古籍文献注册用户,负责古籍文献的数字化加工、元数据加工,唯一标识符分配等等。

(2) 按资源类型划分

按国家图书馆资源类型进行划分,例如,按照唯一标识符规范附录 C,将国家图书馆资源类型划分为古籍、拓片、舆图、地方文献、家谱、普通图书、连续出版物、学位论文、报告、专利文献、视频资料、音频资料等等,设置各种资源类型的注册用户,负责该类资源的唯一标识符生成。比如设置拓片注册用户,负责拓片资源的数字化加工、元数据加工,唯一标识符分配;设置学位论文注册用户,负责学位论文的数字化加工、元数据加工,唯一标识符分配;设置视频资料注册用户,负责视频资料的数字化加工、元数据加工,唯一标识符分配等等。

总之,如何设置和划分注册用户,需要在综合规划国家图书馆数字资源加工流程、元数据加工管理体系、集中元数据仓库管理机制、数字资源服务与管理机制下,以最合理、高效、保证唯一标识符分配的不交叉重复或遗漏为基本原则。

附录 A（规范性附录）：CDOI 元数据

由于国家图书馆资源类型的复杂性，CDOI 元数据设计的原则是用最小的元素集合唯一的标识对象。CDOI 元数据设计参考了 DOI 核心元数据方案、DC 元数据、中国数字图书馆标准规范建设项目基本元数据规范。CDOI 元数据规范包括对象的描述型元数据、管理型元数据。

A1 CDOI 描述型元数据

表 A-1 CDOI 描述型元数据

元素	出现频次	说明
CDOI 名称	1	分配给一个对象的唯一的 CDOI 名称。
名称	1	对象最通常被使用的名称，如题名。
创建者	0-n	对对象的创建和出版负有主要责任的实体，如作者。
URL	1-n	对象的 URL 地址。
标识符	0-n	对象的其他标识符，如 ISBN，ISSN 等。
结构类型	1	所描述的对象是物理对象或抽象对象的规范说明（如对象可以是一个物理对象，也可以是数字对象，可以是作品内容的表达也可以是抽象作品等）。结构类型词汇见附录 B。
资源类型	1	对象的资源类型说明，如对象是音频文档，期刊，乐谱，数据集、期刊论文等。资源类型代码见附录 C。
关联	0-n	所描述的对象与其他对象的关系。

A2 CDOI 管理型元数据

表 A-2 CDOI 管理型元数据

元素	说明
注册代理	发布 CDOI 名称的注册中心，如国家图书馆。
发布日期	CDOI 名称注册到 CDOI 服务系统的日期。

附录 B（规范性附录）：结构类型词汇

结构类型用于标识 CDOI 元数据中的结构类型元素。结构类型从资源整体

角度出发，按三个类型进行分类。一个结构类型可以包含在另一个结构类型中。例如，以一张 CD 盘（物理的）可以包含文件（数字的），文件又包含一首歌（抽象的）的演唱。如果对 CD 盘整体赋予唯一标识符，元数据描述中结构类型选择物理的。如果对 CD 盘中的歌赋予唯一标识符，元数据中的结构类型选择抽象的。结构类型受控词汇见下表

表 B-1 结构类型词汇表

结构类型 词汇	词汇英文名 称	说 明
抽象的	Abstraction	抽象对象，作为概念存在，如作品（work），主题概念等。
数字的	digital	一个对象以数字形式存在。
物理的	physical	一个对象以物理形式存在。

附录 C（规范性附录）：资源类型代码

资源类型编码用于 CDOI 描述型元数据中资源类型标识。由于目前正在研制新的资源类型分类和编码的国际或国家标准，此资源类型编码表是根据国家图书馆的实际状况编制，因而不能包含所有资源类型和层次。如新的国际标准和国家标准制订后，建议国家图书馆唯一标识符服务系统采用国际或国家的有关标准。

如果一种资源类型属于下表中多个资源类型，宜采用更专指的类型编码。对资源类型进行分类和编码时，是根据数字资源的原生文献的资源类型的特征，而不是根据数字对象资源所属的数据库。

C1 标码规则

资源类型代码表采用英文字母与阿拉伯数字相结合的混合制号码。字母标志一个大类，在字母后用数字表示大类下类目的划分。

C2 扩展规则

资源类型编码表规定了一、二级类目，用户可结合具体情况，对二级类目进行适度扩展。例如在古籍下需增加新善本类目，可以 A3 为类型代码。如需要进一步对二级类目下的资源类型进行细分，可采用 1-9 的数字。例如，在善本类下要表示宋元善本，可采用 A21。但由于唯一标识符中的资源类型仅为表示资源类型的大类，不宜对资源进行过细分类，不建议采用三级类目。

国家图书馆在应用此标准时，可以根据数字资源建设的进展，对资源类型

代码进行不断的扩充与维护。

C3 资源类型代码主表

表 C-1 资源类型代码

序号	资源类型		类型代码	说明
1	古籍		A	如果资源库收录内容具有某一专门特征，可以归入更专指的类目，首选选择更专指类目归类。
		普通古籍	A1	
		善本	A2	包括中文善本和西文善本。即包括书、敦煌资料和佛经。
2	金石拓片		B	包括拓片、甲骨、刻石和玺印。
3	舆图		C	绘制、印刷于 1949 年以前的中外古旧地图。老照片、年画入此。
4	地方文献		D	记录某一地区的信息，反映疆域沿革、自然条件、经济物产、社会政治、文化教育、语言文字、风土人情、名胜古迹等记载。
		古代地方志	D1	1949 年以前，出版的地方文献
		现代地方文献	D2	1949 年以后出版
		其他	D3	
5	家谱		E	记载同一姓氏血缘家族的世系与事迹的史类文献，家谱的内容主要由序跋、凡例、目录、修谱名目、像赞、诰敕、源流、世系（图、录）、排行、仕宦录、传记、族规家训、祠堂祠产、坟墓、著述、五服图、余庆录、领谱字号诸类组成。
6	普通图书		F	以单册或有限的多册出版的专著，这里专指普通图书。如果资源库收录的图书内容具有某一专门特征，可以归入更专指的类目，首选选择更专指类目归类。
		旧平装图书	F1	1911 年-1949 年出版的普通图书，民国图书入此。
		中文图书	F2	1949 年以后出版的中文图书，包括海外出版的中文图书。
		外文图书	F3	
		少数民族语文	F4	少数民族语种的文献

序号	资源类型	类型代码	说明	
7	连续出版物	G	广义的期刊，等同于连续出版物，以连续分册形式出版。包括期刊、杂志、报纸、年报等。这里专指普通的连续出版物。如果资源库收录的连续出版物内容具有某一专门特征，可以归入更专指的类目，首选选择更专指类目归类。	
		中文期刊	G1	1949年10月迄今
		近现代杂志	G2	1949年10月前出版各类杂志，民国期刊入此。
		外文期刊	G3	
8	会议文献	J	在各种学术会议上宣读和交流的论文、报告及其他相关的文字或声像资料，以会议录、会议论文集、会议学术报告的形式出现。	
9	学位论文	K	研究生为取得学位，在导师指导下完成并向学位授予机构提交的学术性研究论文，表明作者从事科学研究取得创造性的成果或新的见解，作为提出申请授予相应学位时评审用的学术论文。	
		博士论文	K1	包括博士论文的书目数据、篇名数据、数字对象
10	报告	L	书面陈述，尤指科技报告，记录科学、技术研究结果或进展情况的文献。这里的科技报告也包括软课题研究报告、科研项目报告。科技报告可以包括正式出版的科技报告，也可以是非正式出版的科技报告或研究报告，如预印本。	
11	专利文献	M	各国专利局及国际性专利组织在审批专利过程中产生的官方文件及其出版物的总称。作为公开出版物的专利文献主要有：专利说明书、专利公报、专利文摘、专利索引和专利分类表等。	
12	标准文献	N	按照规定程序编制并经过一个公认的权威机构批准的，供在一定范围内广泛而多次使用，包括一整套在特定活动领域必须执行的规格、定额、规划、要求的技术文件。	
13	手稿	P	作家以传统书写方式书写的文字，包括文稿、诗稿、日记、书信、读书笔记、写作素材记录等。	
14	网络资料	Q	指由网络采集用于长期保存的信息资源。	

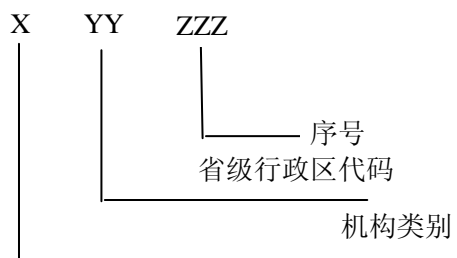
序号	资源类型		类型代码	说明
15	音频资料		R	主要以人的听觉器官来感知的信息资源。按照信息存储载体来分,有一般的录音载体资源(即以唱片、录音带、CD等为存储载体的音频资料)和数字化的网络资源。按信息内容来分,有以语言为主的语言类资源和以音乐为主的音乐类资源。这里主要是指数字化的网络音频资源,包括语言类与音乐类资源。
16	视频资料		S	以电视帧为基础存储的图像信息。
17	图片资料		T	用线条、色彩描绘的景物影像。
18	其他		U	不能归入上述类目的其他文献。

附录 D (规范性附录)：文献情报机构代码

本文献情报机构代码参照全国信息与文献标准化技术委员会制定的《文献情报机构代码》编制。本文献情报机构代码用于唯一标识符命名授权段中注册机构代码的分配。文献情报机构代码只是注册机构代码的一小部分,如果全国范围推行唯一标识符规范,需要国家唯一标识符管理中心编制相应的注册机构代码。

D1 编码规则

文献情报机构代码结构由三段六位代码组成:



机构类别代码: X 为第一段,由 1 位数字或英文字母组成,是文献情报机构所属类别的代码。文献情报机构所属类别代码见表 D-1。

表 D-1 文献情报机构类别代码表

代码范围	类别	系列名称说明	举例
0YYZZZ	公共图书馆	省、地、县独立的公共图书馆	国家图书馆、浙江省图书馆
1YYZZZ	情报所	省、地、县独立的综合性情报机构	中国科学技术信息研究所、广西壮族自治区情报所
2YYZZZ	高校图书馆	高等院校的文献收藏部门。部	

		队院校归国防系列	北京大学图书馆
3YYZZZ	中国科学院系统文献情报机构	中国科学院各单位文献收藏单位或部门	中国科学院文献情报中心、中国科学院物理所图书馆
4YYZZZ	中国社会科学院系统文献情报机构	中国社会科学院各单位文献收藏单位	中国社会科学院文献情报中心、中国社会科学院经济研究所图书馆
5YYZZZ	国务院各部委局所属	中央、国务院各部委局及所属情报文献收藏单位	公安部科技信息所；农业部植物检疫所图书馆
6YYZZZ	省、市、自治区所属	省、市、自治区所属各单位的独立文献部门	广西农科院情报所
7YYZZZ	企业所属	国有企业、私有企业等企业所属的文献情报机构	中国石油化工股份有限公司上海石油化工研究院图书馆

省级行政区划代码：YY 为第二段，由 2 位数字组成，是文献收藏单位所在地代码。采用 GB/T 2260 中华人民共和国行政区划代码。

序号：ZZZ 为第三段，由 3 位数字组成，是前 3 位编码下的文献机构的序号。

D2 部分文献情报机构代码

表 D-2 部分文献情报机构类别代码

代码	机构名称
011001	中国国家图书馆
011005	首都图书馆
031001	上海图书馆
111001	中国科学技术信息研究所
211010	北京大学图书馆
211020	中国人民大学图书馆
211030	清华大学图书馆
212010	南开大学图书馆
212020	天津大学图书馆
231010	复旦大学图书馆
231030	上海交通大学图书馆
232010	南京大学图书馆
233020	浙江大学图书馆
234030	中国科学技术大学图书馆
242010	武汉大学图书馆
244010	中山大学图书馆
261020	西安交通大学图书馆
311001	中国科学院文献情报中心
331001	中国科学院上海生命科学信息中心
411001	中国社会科学院文献信息中心
511001	农业部植物检疫实验所图书馆

511002	中国水产科学研究院图书馆
511080	航天科工集团二院 208 所图书馆
511081	中国航天科技集团公司第 1 研究院第 19 所图书馆
511084	中国空间技术研究院图书馆
511403	中国地质图书馆
511404	中国建筑图书馆
511408	机械工业信息研究院
511413	冶金工业信息标准研究院
511414	中国化工信息中心
511415	中国轻工业信息中心
511416	中国纺织科技信息研究所
511417	铁道科学研究院图书馆
511418	交通部科学研究院信息资源室
511421	中国农业科学院农业信息研究所
511422	中国林业科学研究院图书馆
511436	中国气象局图书馆
511437	中国地震局地震信息中心
511440	中国测绘科学研究院图书馆
511458	中国医学科学院图书馆
511460	中国原子能科学研究院科技信息中心
511501	国家知识产权局专利局文献部
731001	中国石油化工股份有限公司上海石油化工研究院图书馆

D3 机构变更

当文献情报机构代码分配以后，机构所属类别发生变化，如原属于国务院部委局所属机构，经体制改革转变为国有企业性质的机构，或者机构地址迁移，已分配给文献机构的机构代码不做修改，仍沿用原来的代码。

两个机构合并为一个机构，或一个机构拆分为两个机构，合并或拆分后机构是否需要重新分配机构代码，视该机构是否需独立注册唯一标识符而定。中国唯一标识符管理中心负责编制注册机构代码细则。

附录 E（资料性附录）：CDOI 名称中的保留字

CDOI 名称用 URL 的方式表示，经由 HTTP 协议传输将遵循 IETF 的 URI 表示指南标准。如果 CDOI 标识符中含有 URI 定义的保留字符、不安全字符，应用系统将进行转码，将 URI 中的保留字转换为十六进制编码表示。

示例：现有唯一标识符为：123"456"<abc>#xyz，如果使用专门的解析管理工具来操作该标识符，则使用 123"456"<abc>#xyz 作为标识符字符串。当该标识符需要在网页中以链接形式存在时，比如利用 http 代理解析该标识符的链接为：[http://cdoi.gov.cn/123"456"<abc>#xyz](http://cdoi.gov.cn/123)，则应对标识符部分进行十六进制编码，编码后的标识符为：123%22456%22%3cabc%3e%23xyz

作为一个 URI 字符串，它只是由有限的字符、数字和符号组成的，基本表述如下：uric = reserved | unreserved | escaped，即分为保留字（符）、非保留字符和逃逸字符。

CDOI 在 URL 环境部署时的字符集和编码规定如下：

- a) 可以安全使用的字符包括字母、数字和一些标记符号。CDOI 中能安全使用的标记符号如下：mark = "-" | "_" | "!" | "*" | "(" | ")"
- b) 当 CDOI 名称字符串中包含 URI 保留字、非 ASCII 码中的字符、ASCII 码中的非显示的控制字符（十六进制的 00-1F 和 7F）、和 URL 中不安全字符时，应对其进行十六进制编码。CDOI 名称中的保留字及其十六进制编码如下。

表 E-1 CDOI 名称中的保留字

字符	逃逸编码	字符	逃逸编码
;	%3b	{	%7b
/	%2f	}	%7d
?	%3f		%7c
:	%3a	\	%5c
@	%40	^	%5e
&	%26	~	%7e
=	%3d	[%5b
+	%2b]	%5d
\$	%24	`	%60
,	%2c	<	%3c
"	%22	>	%3e
%	%25	#	%23

空格	%20	'	%60
----	-----	---	-----



引用文献

- [1] R. Moats. URN Syntax. May 1997. <http://www.ietf.org/rfc/rfc2141.txt>
- [2] ANSI/NISO Z39.56-1996 Serial Item and Contribution Identifier. <http://sunsite.berkeley.edu/SICI/version2.html>
- [3] http://www.dtic.mil/cendi/presentations/ref_link_blixrud.ppt
- [4] Publisher Item Identifier. http://www.ch.ic.ac.uk/ectoc/ectoc_pii.html
- [5] DOI.URL. <http://www.doi.org/>
- [6] Crossref.org. <http://www.crossref.org/>
- [7] Keith Shafer, Stuart Weibel, Erik Jul, Jon Fausey. Introduction to Persistent Uniform Resource Locators. <http://purl.oclc.org/docs/inet96.html>
- [8] INFO URI. <http://info-uri.info/>
- [9] TagURI. <http://www.taguri.org/>
- [10] Sam X. Sun, Sean Reilly, Laurence Lannom. Handle System Overview. <http://www.handle.net/documentation.html>
- [11] 66th IFLA Council and General Conference. <http://www.ifla.org/IV/ifla66/papers/032-82e.htm>
- [12] OpenURL. http://www.niso.org/committees/committee_ax.html
- [13] Van de Sompel, Herbert and Patrick Hochstenbach. Reference Linking in a Hybrid Library Environment. Part 2: SFX, a Generic Linking Solution. D-Lib Magazine 1999. vol. 5, no. 4. http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt2.html
- [14] 唐世渭, 杨冬青, 王军. 数字图书馆的体系结构. http://web.tongji.edu.cn/~yangdy/computer/D_Lib/paper2.htm
- [15] Preservation Management of Digital Material Handbook <http://www.dpconline.org/graphics/intro/index.html>
- [16] 王燕. 我国数字图书馆数字资源唯一标识符应用现状研究. 我国数字图书馆标准规范建设项目报告.
- [17] CALIS 数字对象唯一标识符命名规范
- [18] CALIS 数字对象唯一标识符本地解析规范

[19] 毛军. 数字科研、远程教育等领域的唯一标识符应用现状研究. 我国数字图书馆标准规范建设项目报告.

[20] 乔晓东. 主要数字文献生产商唯一标识符应用现状研究. 我国数字图书馆标准规范建设项目报告.

[21] Security, authentication and Digital Rights Management (DRM)

<http://www.oclc.org/reports/escan/technology/security.htm>

[22] MPEG-21: 21 世纪多媒体技术框架.

http://www.chinavideoonline.com/mpeg4/mpeg4_049.htm

[23] cIDF.<http://www.cidf.org/>

[24] Universal Unique Identifier. <http://en.wikipedia.org/wiki/UUID>

[25] Globally Unique Identifier. http://en.wikipedia.org/wiki/Globally_Unique_Identifier

[26] 电子阅读引发图书产业加速向网络化转型

http://intro.apabi.com/news/news_2/20060519.htm

[27] 开放获取. <http://zh.wikipedia.org/wiki/开放获取>

[28] Handle System - Globus Toolkit Integration Project.

http://www-unix.globus.org/toolkit/projects/handle_system.html

[29] 中国科学院科学数据库标准规范

<http://www.sdb.ac.cn/list.jsp?boardid=0.projects.standard>

[30] <http://www.52ks.com/user0/407/show.asp?id=1719>

[31] 厦门大学学术典藏库. <http://dspace.xmu.edu.cn/dspace/>

[32] <http://oaps.lib.tsinghua.edu.cn:8080/dspace/index.jsp>

[33] 全国信息技术标准化技术委员会教育技术分技术委员会 <http://www.celtsc.edu.cn/>

[34] 郑晓惠. 开放源代码内容管理系统 Fedora.

<http://lib.utsz.edu.cn/adl2004/zhengxiaohui.ppt>

[35] <http://www.china.org.cn/chinese/zhuanti/356328.htm>

[36] 国家电子政务标准化指南.

<http://www.egs.org.cn/psdZoneDetail.jsp?id=1135733045406840>

[37] ANCC 系统的编码体系. <http://www.ancc.org.cn/knowledge/ancc-system/coding-system.asp>

[38] 条形码, 二维码, 龙贝码 (LP Code), QR Code, 3G. http://blog.terac.com/andy/e_407.html

[39] RFID. <http://www.systron.com.cn/rfid.htm>

[40] Robert Tansley. Building a Distributed, Standards-based Repository Federation.

<http://www.dlib.org/dlib/july06/tansley/07tansley.html>

[41] 借力知识创新工程 全力打造高层次数字化植物园.

<http://www.cas.cn/html/Dir/2003/03/15/9027.htm>

[42] 漫谈数字遗产的保护. http://www.chinainfo.gov.cn/data/200212/1_20021230_50964.htm

参考文献

- [1] ISO TC 46/SC 9/WG 7. ISO/CD 26324: "Information and documentation—Digital Object Identifier (DOI)". 2008
- [2] Dr Norman Paskin. The DOI® Handbook: Edition 4.4.1.[EB/OL]. [2008-4-28]. http://www.doi.org/handbook_2000/DOIHandbook-v4-4.pdf
- [3] 科技部科技基础条件平台专项资金项目. 数字资源唯一标识符解析系统应用规范（倪金松）[R]. 2005
- [4] 科技部科技基础条件平台专项资金项目. 我国数字资源唯一标识符应用模式与管理机制[R]. 2005
- [5] 科技部科技基础条件平台专项资金项目. 数字资源唯一标识符的现状与发展[R]. 2005
- [6] 倪金松. cdlis-唯一标识符-02-数字资源唯一标识符解析系统应用规范，2005 年

后 记

从 2002 年开始,我国先后从数字图书馆建设、电子出版、标准化研制的角度开展了数字资源唯一标识符标准规范的研究。2002 年 10 月,中国科学院国家科学图书馆承担科技部“我国数字图书馆标准规范建设”项目,对数字对象唯一标识符系统的相关标准规范进行了研究;中国高等教育文献保障系统(China Academic Library & Information System,简称 CALIS)研制出台了 CALIS 数字对象唯一标识符规范,数字出版公司针对其出版的电子资源对数字对象唯一标识也进行了规范。

国家图书馆数字资源唯一标识符规范项目,是国家数字图书馆工程标准规范项目之一。该项目的主要目标是:分析我国数字图书馆、出版发行、网络信息服务、远程教学等方面的需要,分析唯一标识符生成过程,并在调研国家图书馆已有数字资源、已有资源标识和国家图书馆潜在数字资源的基础上,提出适合国家图书馆数字资源利用的唯一标识符框架体系、语法规则、命名规则、解析规则、扩展规则、管理规则及其他相关规则,确立合适的数字资源唯一标识符规范,用于分配永久的、可互用的、优化组合的、可解析的标识符,用于系统互访和协同作业,并给出已有数字资源的唯一标识符的转换机制。为此,国家图书馆于 2006 年 4 月成立了数字资源唯一标识符子项目组,2007 年 10 月完成技术需求书,2007 年 12 月 4 日委托采购中心进行了竞争性谈判,中国科学院国家科学图书馆凭借着良好的研究基础,通过竞争成为项目研制单位。2008 年 1 月 15 日签订合同。

在项目研制过程中,国家图书馆项目组协助项目研制单位对国内外数字资源唯一标识符的应用情况、国家图书馆数字资源和唯一标识符应用情况进行了调研;项目研制单位借着全国图书馆标准化技术委员会数字对象唯一标识研讨会,与学界、业界、中文 DOI 的注册机构万方数据进行了交流;在完成标准规范的初稿后,又以专家座谈、访谈等多种形式听取了各方专家的意见,于 2008 年 6 月 17 日研制单位第一次提交研制成果,之后国家图书馆项目组与项目承担单位又经过多次沟通对成果进行了修改,先后几易其稿,最终形成了国家图书馆数字资源唯一标识符标准规范。2008 年 9 月通过子项目组验收,2008 年 9 月 25 日通过国家图书馆馆内专家验收,2008 年 10 月 25-11 月 24 日完成网站公开质询,2008 年 12 月 23 日通过国内业界专家验收,至此该项目全部完成。

此次结集成书正是项目的研究成果,本书主要由项目承担单位——中国科学院国家科学图书馆撰写,国家图书馆项目组提出修改意见并参与修改,孙坦、贺燕、宋文拟定大纲,孙坦、贺燕、宋文对全稿进行了审校。本书第一部分标准规范的撰写人为宋文、王昉、周静怡、毛军、刘峥、李欣,第二部分应用指南的撰写人为孙坦、宋文、刘峥、王昉、周静怡、梁娜、毛军、刘细文。

在标准的研制过程中,得到了国家图书馆汪东波研究馆员、申晓娟副研究馆员、苏品红研究馆员、富平研究馆员、孙一钢研究馆员、索传军研究馆员、王志庚副研究馆员等专家的大力支持;也得到来自北京大学图书馆朱强馆长和聂华副馆长、北京大学信息管理系李广健教授、中国出版科学研究所张书卿主任助理、清华大学图书馆姜爱蓉副馆长、上海图书馆刘炜研究馆员和赵亮副研究馆员、万方数据研究院张旭研究员等的多方帮助,在此一并表示感谢。