



中华人民共和国文物保护行业标准

WW/T XXXXX—XXXX

文物数字对象唯一标识符

Digital Object Identifier for Cultural Heritage

(征求意见稿)

(本稿完成日期: 2017-03-16)

XXXX - XX - XX 发布

XXXX - XX - XX 实施

中华人民共和国国家文物局

发布

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语与定义	1
4 CHDOI 体系框架结构	2
4.1 语法规则	2
4.2 命名规则	3
4.3 元数据规则	4
4.4 管理规则	4
4.5 解析规则	4
4.6 兼容性规则	5
附录 A（规范性附录） 可移动文物收藏单位编码	6
附录 B（规范性附录） 文物保护单位档案全宗号的编制规则	8
附录 C（规范性附录） 数字资源类型代码及用途	10
参考文献	12

前 言

本标准依据国际上通用的唯一标识符标准规范制定。

本标准规定了文物数字资源唯一标识符的体系框架结构等,是文物数字资源唯一标识符标准的指导性文件。

主要包括:制定文物数字资源唯一标识符的语法规则、命名规则、解析规则、管理规则等。

本标准由中华人民共和国国家文物局提出。

本标准由全国文物保护标准化技术委员会(SAC/TC 289)归口。

本标准起草单位:北京大学。

本标准起草人:王燕。

本标准是首次发布。

文物数字资源唯一标识符规范

1 范围

数字资源唯一标识符也称为数字对象唯一标识符，是实体或对象持久标识的基础设施。CHDOI是文物数字资源唯一标识符（Cultural Heritage Digital Object Identifier）的英文首字母缩写。本规范的主要目的是建立文物数字资源唯一标识符系统。

CHDOI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。

本规范规定了CHDOI系统的体系框架、语法规则、命名规则、解析规则、管理规则等。

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。

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

3 术语与定义

下列术语和定义适用于本规范。

3.1 CHDOI名称 CHDOI name

在CHDOI系统内为数字资源定义一个字符串，以唯一标识该数字资源。CHDOI名称可以由遵循CHDOI语法规则的字母、数字、字符组成。

3.2 CHDOI解析 CHDOI resolution

向CHDOI系统提交一个CHDOI名称的请求，可返回一条或多条与CHDOI相关的信息的过程。解析结果可能返回标识对象的实例，也可能不返回。

3.3 CHDOI语法 CHDOI syntax

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

3.4 CHDOI系统 CHDOI system

用于管理、分配、解析CHDOI等应用的系统。

3.5 持久性 Persistent

持久性指唯一标识符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于各种服务中。

3.6 CHDOI元数据 CHDOI Metadata

CHDOI元数据用最小集合来描述该数字对象，有助于发现、确认、获取、管理、评价和保存对象。

4 CHDOI 体系框架结构

CHDOI标准由唯一标识符的语法规则、命名规则、元数据规则、解析规则、管理规则和兼容性规则组成。使用CHDOI可以标识实体文物或文物数字资源。

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。

解析规则规定从应用系统中的CHDOI出发，通过CHDOI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。

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。

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，提供用户通过对对象的元数据的检索，获得唯一标识符。

兼容规则说明唯一标识符前缀的扩展原则，及加入其它唯一标识符体系的可能性。

4.1 语法规则

4.1.1 基本规则

一个CHDOI标识唯一的一个数字资源，CHDOI必须是唯一的，其名称应遵循URI语法规则。为了使唯一标识符能被浏览器处理，CHDOI名称中不宜采用URI中的保留字符。CHDOI字符串的长度最长不应超过255个字符。CHDOI中的英文字母不区分大小写，但是规范的写法是小写，它可以是除ISO/IEC 10646字符集保留字符外的任何字符。

CHDOI语法规则规定了组成CHDOI字符串的形式和顺序。

CHDOI由三个部分组成：

- 前缀
- 前后缀之间的分隔符
- 后缀

4.1.2 CHDOI 前缀

CHDOI前缀是命名授权段，命名授权由多个非空的命名授权段来构成，每个段由字符“.”（半角点号）分隔。

一般情况下，CHDOI的前缀由两个命名授权段组成：

- a) 第一个命名授权段是国家文物局命名授权。
- b) 第二个命名授权段是各文博机构命名授权，即国家文物局为各文博机构制定的文物收藏或保护单位编码，由14位数字组成，具体编码规则取自“中华人民共和国文物保护行业标准：馆藏文物登录规范 WW/T 0017-2013”附录C，并作为本规范的附录A。

示例：设北京市西城区某博物馆的收藏单位代码为11010221800001，则其数字资源唯一标识符的前缀为chdoi.11010221800001。

4.1.1 CHDOI 后缀

CHDOI后缀跟在前缀之后，与前缀之间用斜线“/”分隔。在同一个前缀元素下的每个后缀都应是唯一的。

- a) 利用CHDOI标识实体文物

实体文物的CHDOI后缀包含对应文物的唯一标识符信息。

数字资源对应文物的唯一标识符：建议采用文物收藏或保护单位使用的唯一标识符，如可移动文物可采用总登记号或可视为总登记号的其他编号，不可移动文物则可包含文物保护单位档案全宗号，由9

位数字组成，具体编码规则取自国家文物局《文物保护单位记录档案档号编制规则》，并作为本规范的附录B。

示例1：设北京市西城区某博物馆的收藏单位代码为 11010221800001，该博物馆的一件铜器文物的唯一标识（总登记号）为：60.31.1234，则该文物的CHDOI为：chdoi.11010221800001/60.31.1234。

示例2：设陕西省半坡博物馆的单位代码为 61011121800001，半坡博物馆的半坡遗址全宗号为 611100001，半坡遗址中一座半地穴房屋的唯一标识为 611100001.0123，则半坡遗址的唯一标识为 chdoi:61011121800001/611100001，该半地穴房屋的CHDOI为：chdoi:61011121800001/611100001.0123。

b) 利用CHDOI标识文物数字资源

文物数字资源的CHDOI后缀包含对应文物的唯一标识符、数字资源类型、数字资源用途、序号四部分信息，采用“对应文物的唯一标识符-数字资源类型-数字资源用途-数字资源序号”形式编码。

——数字资源类型代码：见“附录C：数字资源类型代码及用途”。

——数字资源的用途代码：见“附录C：数字资源类型代码及用途”。

——数字资源序号：可以是一个顺序号。

以上各项之间用英文半角“-”符号分开。

示例1：设北京市西城区某博物馆的收藏单位代码（地域和行业编码）为 11010221800001，该博物馆的一件铜器文物的总登记号为：60.31.1234。则一幅该文物存档级图像的唯一标识符为：chdoi.11010221800001/60.31.1234-CG-A-001（CG代表图形/图像，A代表用于长期保存，001为顺序号）。

示例2：设陕西省半坡博物馆的单位代码为：61011121800001，半坡遗址隶属半坡博物馆，其全宗号为 611100001，半坡遗址中一座半地穴房屋的唯一标识为 611100001.0123，则该房屋用于网上发布的三维模型的唯一标识符为：chdoi:61011121800001/611100001.0123-CG-LV-A01（CG代表图形/图像，LV代表用于网上发布的矢量图形，A01为顺序号）。

4.1.2 CHDOI的字符集与编码规定

CHDOI可以是除ISO/IEC 10646字符集保留字符外的任何字符，规定CHDOI字符串使用UTF-8规则编码。

CHDOI名称通过Web浏览器解析时，要求CHDOI名称应遵循IETF的URI表示指南标准。由于唯一标识符经常需要通过浏览器来处理，当不可避免需要使用保留字符时，必须用转义的方法来表示。

CHDOI不宜使用的保留字符包含以下字符，如果需要使用这些保留字符，需要转义。

表1 需转义的保留字符

字符	空格	:	/	?	#	[]	@	;	{	}
转义编码	%20	%3A	%2F	%3F	%23	%5B	%5D	%40	%3B	%7B	%7D
字符	!	\$	&	'	()	*	+	,	“	%
转义编码	%21	%24	%26	%27	%28	%29	%2A	%2B	%2C	%22	%25
字符	\	^	~	`	<	>		=	#		
转义编码	%5C	%5E	%7E	%60	%3C	%3E	%7C	%3D	%23		

4.2 命名规则

4.2.1 范围

任何时候当需要将一个数字资源作为一个被识别的实体时，可给这个数字资源分配一个CHDOI名称。

CHDOI可以分配给任何类型、格式、载体的数字资源。CHDOI名称分配的主要对象包括但不限于：文字资源、图像资源、视频资源、音频资源、其它多媒体资源等。

4.2.2 粒度

CHDOI名称能在数字资源的任意粒度层次上进行分配。如一件文物由若干附件组成时，如果这些附件有相应的数字资源，则可以对整件文物及其附件的数字资源分别分配CHDOI。

4.2.3 CHDOI 的唯一性

一个完整的CHDOI在系统内是唯一的，在编制CHDOI时不得编制重码，可以将一个以上的CHDOI赋予同一文物。

4.2.4 CHDOI 的持久性

持久性指唯一标识符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于各种服务中。

4.3 元数据规则

CHDOI注册时，注册者应同时注册CHDOI所标识的数字资源的元数据信息。CHDOI元数据应具有与国内外相关元数据标准的互操作能力，能在任何粒度层次上标识一个实体。

由于文物资源的特殊性，CHDOI元数据设计的原则是用最小的元素集合唯一地标识文物数字资源，同时体现其对应文物的最基本、最核心的信息。

表2 CHDOI 的元数据

元素名称	说明	必备性	可重复性
CHDOI	文物数字资源的CHDOI	必备	可重复
名称	CHDOI对应的文物数字资源的名称	必备	可重复
现藏机构/位置	实体文物现藏机构或所属地理位置	有则必备	可重复
URI	文物数字资源的URI地址	必备	可重复
文物类型	数字资源对应的文物所属类型，取自《中华人民共和国文物保护行业标准 WW/T 0017-2013:馆藏文物登录规范》和《第三次全国文物普查不可移动文物分类标准》	必备	可重复
数字资源类型	文物数字资源的类型说明，如数字资源是视频、图像等。资源类型代码见附录C。	有则必备	不可重复
标识符	实体文物的标识符，如总登记号等	有则必备	可重复
关联	实体文物与其他文物的关系，如与其它文物的关系或同一文物的不同部件	有则必备	可重复
权限	文物数字资源的访问及使用权限	有则必备	可重复

4.4 管理规则

CHDOI由国家文物局唯一标识符管理中心进行管理及注册登记。

4.5 解析规则

解析是向CHDOI系统提交一个CHDOI的检索请求时，CHDOI系统反馈的该CHDOI所标识的数字对象相关数据的过程。解析可返回数字资源本身，也可以返回数字资源的元数据信息。

CHDOI解析记录可包含多个数字资源的存放地址（URL）及其元数据信息，如数字资源名称、数字资源类型、文物类型、其它标识符、现藏机构/位置、与其它文物的关系或同一文物的不同部件等。CHDOI系统可实现简单解析、多重解析、反向解析。

- a) 简单解析：简单解析是单一的指向性解析。每个 CHDOI 只有一个可解析到的 URL 地址，解析结果将返回该 CHDOI 对应的数字对象实例。
- b) 多重解析：多重解析允许一个 CHDOI 被解析到多个 URL 地址。一个 CHDOI 解析到多个结果时，用户可选择一个解析结果。
- c) 反向解析：反向解析是从 CHDOI 的元数据解析到 CHDOI 的过程。CHDOI 服务系统应能注册文物数字资源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 CHDOI，供用户选择。

4.6 兼容性规则

CHDOI系统可适应其他国际上流行的标识符方案，未来不管加入哪一个网络资源组织，其注册资源代码都是要由注册机关自定，只要再加上注册机关代码即可为国际间唯一的识别码，因此可以快速简单地转换成该组织命名方式，使其符合系统扩充性及未来性。

附录 A
(规范性附录)
可移动文物收藏单位编码

取自“中华人民共和国文物保护行业标准 WW/T 0017-2013”《馆藏文物登录规范》附录C。

A.1 概述

收藏单位代码即地域和行业编码，用14位阿拉伯数字表示。

为区分不同的单位，将编号设计为四部分：行政区划代码+单位性质代码+行业分类代码+单位顺序号。

A.2 编号的命名规则

命名内容：行政区划代码+单位性质代码+行业分类代码+单位顺序号。

命名长度： 6位 1位 2位 5位

	行政区划代码						单位 性质 代码	行业 分类 代码		单位顺序号				
长度	1	2	3	4	5	6	1	1	2	1	2	3	4	5

各项的顺序要固定，名称中不应存在任何空格。无论任何原因，名称中均不应存在中文字符。

A.3 行政区划代码

馆藏文物收藏单位所在地的行政区划号码，用6位阿拉伯数字表示。以国家统计局公布的最新《全国县及县以上行政区划代码表》为准。

A.4 单位性质代码

按表A.1所示内容选择后填写，用1位阿拉伯数字表示。

表A.1 单位性质代码表

单位性质代码	单位性质
1	国家机关
2	事业单位
3	国有企业
4	人民解放军、武警部队

A.5 行业性质代码

以馆藏文物收藏单位所属的行业分类为准，用2位阿拉伯数字表示。见表A.2所示。

表A.2 行业性质代码表

代码	与《国民经济行业分类与代码》 中“门类”的对应关系	行业名称
01	A	农、林、牧、渔业
02	B	采矿业
03	C	制造业
04	D	电力、热力、燃气及水生产和供应业
05	E	建筑业
06	F	批发和零售业
07	G	交通运输、仓储和邮政业
08	H	住宿和餐饮业
09	I	信息传输、软件和信息技术服务业
10	J	金融业
11	K	房地产业
12	L	租赁和商务服务业
13	M	科学研究和技术服务业
14	N	水利、环境和公共设施管理业
15	O	居民服务、修理和其他服务业
16	P	教育
17	Q	卫生和社会工作
18	R	文化、体育和娱乐业
19	S	公共管理、社会保障和社会组织
20	T	国际组织

A.6 单位顺序号

由国家统一编制，用5位阿拉伯数字表示。从个位开始编号，不足5位的前面用“0”补齐。

示例：北京市西城区某博物馆

110102	2	18	00001
北京市西城区	事业单位	文化、体育和娱乐业	西城区的第1家

附 录 B
(规范性附录)
文物保护单位档案全宗号的编制规则

取自国家文物局《文物保护单位记录档案档号编制规则》（2014年补充修订稿）。

B.1 文物保护单位档案全宗号，依据文物保护单位所在省份、文物保护单位的级别和类别，及排列顺序进行标识。

B.2 文物保护单位档案全宗号第一、二位代码，引用国家标准GB/T 2260 的省份代码，来表示要标识的文物保护单位所在地域，参见表 1。对于跨省份的文物保护单位，这两位代码以“99”表示。

表B.1

代码	名 称	代码	名 称
11	北京市	43	湖南省
12	天津市	44	广东省
13	河北省	45	广西壮族自治区
14	山西省	46	海南省
15	内蒙古自治区	50	重庆市
21	辽宁省	51	四川省
22	吉林省	52	贵州省
23	黑龙江省	53	云南省
31	上海市	54	西藏自治区
32	江苏省	61	陕西省
33	浙江省	62	甘肃省
34	安徽省	63	青海省
35	福建省	64	宁夏回族自治区
36	江西省	65	新疆维吾尔自治区
37	山东省	71	台湾省
41	河南省	81	香港特别行政区
42	湖北省	82	澳门特别行政区

B.3 文物保护单位档案全宗号第三位代码，表示要标识的文物保护单位的文物级别，根据表 2 的对应关系来确定。

表B.2

级别代码	表示对象
1	全国重点文物保护单位
2	省、自治区、直辖市级文物保护单位
3	县、市级文物保护单位
4	未定级的文物保护单位

B.4 文物保护单位档案全宗号第四位代码，表示要标识的文物保护单位的文物类别，根据表3的对应关系来确定。

表B.3

类别代码	文物保护单位类别	类别代码	文物保护单位类别
1	古遗址	4	石窟寺及石刻
2	古墓葬	5	近现代重要史迹及
3	古建筑	6	其它

B.5 文物保护单位档案全宗号第五至第九位代码，表示在第一、二位代码所标识的地域范围内，级别相同，且类别相同的文物保护单位的顺序号，在00001~99999的取值范围内升序编号。排序时应以公布的时间先后为依据；公布的时间相同和不明确时，按公布的文物保护单位年代为依据；公布的文物保护单位年代不明确时，按其名称的音序为依据。尚未核定公布为文物保护单位的不可移动文物的顺序号，在地域、时代的限定下，可按其名称音序排列。

附 录 C
(规范性附录)
数字资源类型代码及用途

图片、文字、音频、视频、动画，还有衍生的一些表现方式等。

数字资源类型	数字资源类型代码	表现形式或用途	用途代码	常见格式
图形/图像	CG	点阵图像	A——长期保存级 (Archives)	BMP、GIF、JPEG、TIFF、PSD、PNG、RAW 等
			P——复制加工级 (Process)	
			L——网上发布级 (Service) 精细级	
			M——网上发布级 (Service) 普通级	
			S——网上发布级 (Service) 袖珍级	
			MrSID——地图浏览	
图形/图像	CG	矢量图形	AV——长期保存级 (Archives)	SVG、DXF、WMF、EMF、CDR、AI、DWG、TGA、OSGB、OBJ、FBX、STL、3DS、WRL、MB 等包含三维建模渲染、三维动画、三维模型数据格式在内的矢量图形格式
			LV——网上发布级 (Service)	
文本	TXT		T——文本文件	TXT、DOC、DOCX、PDF、WPS、HTML、XML 等
视频资源	video	本地播放或保存	L——低 (相当于 VCD)	MP4、RMVB、AVI、RM、RMVB 等
			M1——中 1 (相当于 SVCD)	
			M2——中 2 (相当于 DVD)	
			H——高 (相当于 DVD)	
		流式播放	VM1——标清/流畅	
			VM2——高清	
			VM3——超清	
VH——原画				

音频资源	audio	本地播放或保存	H——高（特殊录音资料，用于保存或本地播放）	MP3、CDA、WAV、MIDI、WMA、RA、BWF 等
			M1——中（极高质量要求，用于保存或本地播放）	
			M2——中（CD 质量标准，用于保存或本地播放）	
			L——低（录音最低质量，用于保存或本地播放）	
		流式播放	D——低（用于流式播放）	
其它多媒体资源	Multimedia		F——动画等	FLA、SWF 等
其它资源	Others		0——其它资源	EXE、RAR、ZIP 等

参 考 文 献

- (1) 孙坦, 宋文, 贺燕主编. 国家图书馆数字资源唯一标识符规范和应用指南[M]. 北京: 国家图书馆出版社, 2010
- (2) 中华人民共和国文物保护行业标准 WW/T 0017-2013:馆藏文物登录规范
- (3) 第一次全国可移动文物普查工作手册
- (4) 第三次全国文物普查实施方案及相关标准、规范. 2007
- (5) 文物保护单位记录档案档号编制规则, 国家文物局, 2014