

博物馆元数据标准构建研究： 以盖蒂研究所元数据标准为例^{*}

□张俊娥 王亚林

摘要 元数据在博物馆数字化中起着至关重要的作用,要有效地组织和揭示博物馆文物信息,需要各类元数据标准作支撑。通过对盖蒂研究所发布的描述元数据标准(CDWA)、数据内容标准(CCO)、数据值标准(AAT、ULAN、TGN、CONA)和数据交换标准(LIDO)等元数据标准的研究,分析不同类型的元数据标准在数字博物馆建设中所起的作用,提出在我国的数字化或智慧化博物馆建设中不仅需要首先构建适合于我国文物数字化建设的描述元数据标准和著录规范,而且也应该逐步建立各种数据值标准和数据交换标准,为未来我国文物元数据的统一管理和数据共享打下基础。

关键词 元数据 数据结构标准 数据内容标准 数据值标准 数据交换标准

分类号 G264.1

DOI 10.16603/j.issn1002-1027.2018.06.009

元数据是数据的数据,是对信息资源的结构化描述,用来组织、描述、发现、检索、索引、集成、浏览、保存和管理信息资源。元数据标准为各种形态的数字化信息单元和资源集合提供规范、普遍的描述方法。为了特定的目标如数据结构的设计、数据值的规范、数据内容和数据交换的有效和一致性,需要设计和制定不同的元数据标准。

参考威斯康星大学密尔沃基分校(UWM)斯蒂文·米勒(Steven Miller)的“元数据标准类型”(Typology of Metadata Standards)和曾蕾、秦健合著元数据(Metadata)一书对元数据标准的论述,元数据标准类型有数据结构标准(Data Structure Standards)、数据内容标准(Data Content Standards)、数据值标准(Data Value Standards)、数据交换(和收割)标准(Data Format / Technical Interchange Standards)和数据呈现标准(Data Presentation Standards)^{[1][2]}。

博物馆的元数据标准在国外发展较早,常用的元数据标准见表1。

盖蒂研究所隶属于美国盖蒂艺术中心,致力于视觉艺术领域文化遗产的保存、普及与相关研究。上述

博物馆元数据标准中,盖蒂研究所参与研究并推荐使用的描述元数据结构标准 CDWA、数据内容标准 CCO、数据值标准 AAT、ULAN、TGN、CONA 和数据交换标准 LIDO 是博物馆元数据标准的一个套完整解决方案。下面对这些标准进行详细介绍和分析。

1 数据结构标准 CDWA^[3]

CDWA 的全称是“艺术作品描述类目”(Categories for the Description of Works of Art,以下简称 CDWA),是描述博物馆馆藏的元数据方案。它诞生于 1990 年代的美国,由美国国家人文基金会、保罗·盖蒂信托基金(J. Paul Getty Trust)及学院艺术协会(College Art Association)联合资助的艺术信息工作组 AITF (the Art Information Task Force) 制定,用于描述艺术品、建筑以及艺术品的视觉或文字替代品。2000 年 9 月推出 2.0 版,2014 年 3 月的版本是最新的修订版,在保罗·盖蒂信托基金的官网发布,发布网址是 http://www.getty.edu/research/publications/electronic_publications/cdwa/index.html。

* 本文系科技部国家科技支撑计划课题“文物数字化保护元数据标准研究”(项目编号:2014BAK07B02)的研究成果之一。
通讯作者:张俊娥,ORCID:0000-0002-3430-6230,邮箱:zhangje@lib.pku.edu.cn。

表 1 国外博物馆常用元数据标准

元数据类型	英文简称	英文全称	说明
数据结构标准	DCMES	Dublin Core Metadata Element Set	都柏林核心元数据元素集
	CDWA	Categories for the Description of Works of Art	艺术作品描述类目
	VRA Core	Visual Resources Association Core Categories	视觉资源协会核心类目
	MODS	Metadata Object Description Schema	元数据对象描述标准
	EAD	Encoded Archival Description	编码档案著录标准
	Reach	Records Export for Art and Culture Heritage Project	艺术品与文化遗产数据传输项目
	AMS	The ARCO Metadata Element Set	欧盟博物馆藏品数字化和建立网上虚拟展示的元数据数据集
	MIDAS	Monument Inventory Data Standard	英国皇家历史古迹委员会推出的历史遗迹元数据标准
	Object ID	Object ID	物品标识
数据内容标准	CCO	Cataloging Cultural Objects	文物编目指南
	DACS	Describing Archives Content Standard	档案描述内容标准
数据值标准	AAT	Art & Architecture Thesaurus	盖蒂艺术和建筑词表
	ULAN	The Union List of Artist Names	盖蒂联盟艺术家名称表
	TGN	theGetty Thesaurus of Geographic Names	盖蒂地理名称表
	CONA	the Cultural Objects Name Authority	文物名称规范
数据交换 (收割)标准	CDWA LITE	Categories for the Description of Works of Art	CDWA 的 xml 交换标准
	LIDO	Lightweight Information Describing Objects	轻量信息描述对象
	CIMI	Consortium for the Computer Interchange of Museum Information	美国博物馆信息交换联盟

1.1 CDWA 的结构

CDWA 元数据标准是针对描述艺术品的需求而设计的,不仅要兼顾艺术品的物理形态和数字化的图像,而且要描述时空、人物、历史文化等方面的上下文关系;并且强调某些专门元素中受控词汇表的使用,如人名/团体名称规范、地名规范、通用概念规范、主题规范等。如下图是 CDWA 的“实体关系图”。

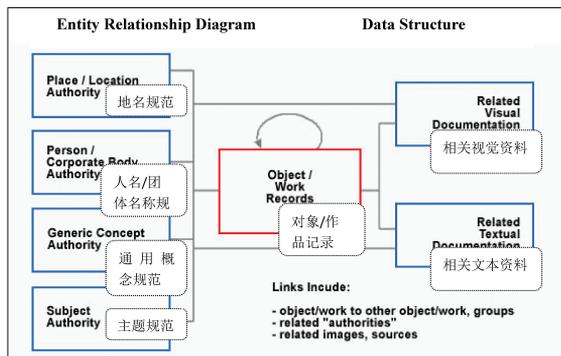


图 1 CDWA 的实体关系图

(图 1 改编自 http://www.getty.edu/research/publications/electronic_publications/cdwa/entity.html)

从 CDWA 的实体关系模型可以看出,一个“对象/作品记录”实体与其他相关对象或作品有关联,与规范档实体(个人/团体名称规范档、地点/位置规范档、类属概念规范档、主题规范档)相关联,与相关的视觉资料和相关文献资料相关联。这种关联关系的广泛存在使得描述的元数据记录高度系统化,对深度揭示一个对象或作品奠定了数据基础,最终方便了研究人员对某一对象实体的利用和研究。

1.2 CDWA 的类目组成

CDWA 通过表达艺术作品、建筑,其他物质文化作品或收藏及相关图像等描述信息的概念框架,来描述艺术数据库的内容。CDWA 定义了 31 个主类目、501 个子类目,表 2 为主类目。

1.3 CDWA 的特点分析

CDWA 元数据标准是针对描述艺术品的需求而设计的,艺术品首先是具有物理形态的,也可能同时具有数字化了的图像,此外时空、人物、历史文化等方面的上下文关系在描述一件艺术品时也是非常

表 2 CDWA 的主类目

1	对象/作品 (Object/Work)	17	背景 (CONTEXT)
2	分类 (Classification)	18	描述附注 (Descriptive Note)
3	名称 (Titles or Names)	19	批评性响应 (Critical Responses)
4	创作 (Creation)	20	相关作品 (Related Works)
5	风格/时期/流派/运动 (Styles/Periods/Groups/Movements)	21	现藏地点 (Current Location)
6	计量 (Measurements)	22	版权/限制 (Copyright/Restrictions)
7	材质与技法 (Materials and Techniques)	23	所有权/收藏历史 (Ownership/Collecting History)
8	题识/标记 (Inscriptions/Marks)	24	展览/借展史 (Exhibition/Loan History)
9	状态 (State)	25	编目史 (Cataloging History)
10	版本 (Edition)	26	相关视觉文献 (Related Visual Documentation)
11	制作手法 (Facture)	27	相关文本参考数据 (Related Textual References)
12	方位/布置 (Orientation/Arrangement)	28	个人/团体名称规范 (Person/Corporate Body Authority)
13	外观描述 (Physical Description)	29	地名/位置规范 (Place/Location Authority)
14	现况/鉴定历史 (Condition/Examination History)	30	类属概念规范 (Generic Concept Authority)
15	保存/处理历史 (Conservation/Treatment History)	31	主题规范 (Subject Authority)
16	主题 (Subject Matter)		

重要的,CDWA 中有很多元素都反映了艺术品这些特点。有代表性的有:方位/布置 (Orientation/Arrangement)、题识/标记 (Inscriptions/Marks)、背景信息 (Context)、(Style/Periods/Groups/Movements 风格/时期/流派/运动)、评论 (Critical opinions)。物理形态方面有:计量 (Measurement)、材质与技法 (Materials and Techniques)、外观描述 (Physical description)。

在对艺术品的保存、管理方面的内容也用一些很有特色的元素来描述,如:现况/鉴定历史 (Condition/Examination History)、保存/处理历史 (Conservation/Treatment History)、所有权/收藏历史 (Ownership/Collecting History)、展览/借展史 (Exhibition/Loan History)。

此外,CDWA 比较强调专门元素中受控词汇表的使用。在新版的类目定义中,增加了人名/团体规范、地点/方位规范、通用概念规范、主题规范这四个规范类目。

CDWA 用于满足学术研究者等人员的记录、保存、检索艺术信息的需求。它确定了能够兼容各种信息系统的内容描述框架,有利于保持数据的完整性,也便于长期保存和迁移数据。同时 CDWA 还制定了与其他元数据标准 (Dublin Core、MARC、VRA 等) 的映射表,为实现数据的交换和共享奠定了基础。

2 数据内容标准 CCO^[4]

CCO 的全称是“文物编目指南”(Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images, 以下简称 CCO)。CCO 是一个对文化作品及它们的可视替代品进行描述的编目标准。CCO 主要针对艺术和建筑,包括但不限于绘画、雕塑、印刷品、手稿、照片、建筑作品、设施以及其他可视媒介。CCO 同样覆盖了一些其他类型的文化作品,包括考古遗址、史前古器物以及物质文化范畴的有用物品。CCO 项目由 VRA 成员的编辑团队,以及来自博物馆、图书馆和档案馆的专家顾问委员会构想和开发。在 2001 年 12 月正式启动。主要资金由数字图书馆联盟和盖蒂资助计划提供,目标是为描述和检索文物及其图像的信息、提供内容标准即编目规则,CCO 的官网地址是 <http://cco.vrafoundation.org/>。

2.1 CCO 的内容组成

CCO 包括三部分内容、“一般准则”、“元素及编目规则”和“规范文档”。

第一部分“一般准则”说明了 CCO 描述的对象:作品的实体对象和数字对象、实体对象和数字对象的定义及其之间的关系、最简记录、记录类型、相关作品、数据库设计及关系、规范和受控词汇表等,以

及在应用 CCO 时的十大主要原则。

第二部分“元素及编目规则”,列举了编目描述的 9 大类信息及每一类所包含元素的描述和定义,说明元素的编目规则,包括大写、缩写、语法、术语,针对不同编目对象的应用规则以及数据的展现形式,包括显示、索引以及例子。其中有 11 个元素强烈推荐为必备元素(Required),14 个元素值为受控词汇,取值来自受控词汇表,或者来自简单的取值表。

表 3 CCO 编目必备元素

9 大类信息	必备元素	是否受控
对象命名信息 (Object Naming)	文物类别(Work Type)	受控
	名称(Title)	
责任者信息 (Creator Information)	显示的责任者(Creator Display)	
	责任者(Creator)	受控
	责任方式(Role)	受控
物理特征 (Physical Characteristics)	计量(Measurements)	
	材质和技法(Materials and Techniques)	
风格和年代(Stylistic, Cultural, and Chronological Information)	年代(Date)	
所在位置 (Location and Geography)	当前位置(Current Location Display)	
主题(Subject)		
级别(Class)		
描述(Description)		
显示信息(View Information)	显示的描述信息(View Description)	
	显示类型(View Type)	受控

第三部分“规范文档”包括个人和团体名称规范、地理位置规范、概念规范以及主题规范。CCO 所采用的规范主要来自盖蒂词汇表的艺术与建筑词典(AAT)、艺术家名字联合目录(ULAN)、盖蒂地名词典(TGN)及美国国会图书馆规范词表等。

2.2 CCO 的十大原则

在使用 CCO 作为编目著录规则时,需要遵守以下十大原则:

(1)无论是单个物件、一个由几个部分组成的作品,还是一个组合件或作品集合,都要建立每个作品记录的逻辑重点,明确区分是实体记录还是图像

记录。

(2)要包含所有 CCO 的必备元素。

(3)在对 CCO 进行本地化,制定本地规则时,要遵循 CCO 规则,以便有效地检索,重新使用和交换信息。

(4)要使用受控词汇,如盖蒂词汇和国会图书馆规范档。

(5)要使用标准发布的受控词汇以及用本地术语和名称建立的本地规范档。尽可能地将本地规范作为叙词表使用。

(6)编目元数据标准使用已建立的元数据标准,如 CDWA、VRA、DC 等。

(7)要了解编目、分类、索引和显示是不同但相关的功能。

(8)在作品和它的影像之间、作品集合和作品之间、一个作品集合的不同作品之间以及同一作品的不同影像之间建立关系。

(9)遵守大小写、标点符号和语法的规定。避免使用缩写词,但必要时可使用缩写的标准代码和列表(例如 ISO 标准中的国家缩写)。

(10)对于英文信息系统和用户,尽可能使用英文数据值。

2.3 CCO 的特点分析

CCO 是为博物馆和可视资源收藏、档案、以艺术和建筑收藏为主的图书馆设计的。CCO 专注于描述性编目的数据内容标准。CCO 的重点是描述性元数据和规范控制——它倾向于描述一个文化作品,数据被用来为作品和作品的图片生成目录。不包含管理性元数据和技术元数据。

CCO 是一个数据内容标准。它与 21 世纪图书馆采用的编目标准“资源描述和检索”(Resource Description and Access, RDA)一样:

(1)都是编目标准,与采用格式无关;

(2)与其他元数据格式兼容;

(3)重编目员的判断而不是盲目坚持规则;

(4)从英文语言发展而来,但致力于适应于世界范围;

(5)重视文物实体之间的关系;

(6)重视规范的重要性;

(7)CCO 与 RDA 的不同点在于两者对“作品”的定义不同,CCO 的描述对象是唯一的物品。

3 数据值标准——Getty 词表^[5]

艺术和建筑类词表(The Art & Architecture Thesaurus,以下简称 AAT)、盖蒂地理名称表(The Getty Thesaurus of Geographic Names,以下简称 TGN)、盖蒂联盟艺术家名称表(The Union List of Artist Names, ULAN,以下简称 ULAN)和文物名称规范(The Cultural Objects Name Authority,以下简称 CONA)提供有关文物名称、艺术家、概念和在艺术、建筑和物质文化领域具有重要性的地方的名称和相关信息,是用来提高获取艺术、建筑和物质文化信息的结构化的词汇表。它们可以被用作文物编目的数据值标准,编目员和编制索引者可把它们作为一个控制词表或规范来使用。还可以作为一个数据库检索系统的检索助手,因为这些词汇表本身是一个包括不同概念链接和路径的语义网络,所以这些关联能够使检索更有效率。也因为提供丰富的关联知识,它们也是一个有价值的研究工具。

这 4 个规范词表总称盖蒂词汇(Getty Vocabulary),始建于 80 年代中期,是为了满足艺术博物馆、视觉资源、档案和艺术图书馆等单位的需求而建立的,是对艺术、建筑和物质文化等编制索引和相关资源检索的规范词表。盖蒂词汇由美国盖蒂研究所(Getty Research Institute,GRI)的盖蒂词汇项目组(Getty Vocabulary Program)编制、维护和发布。词汇的增长依赖于用户群体,由有资格的艺术类图书馆、博物馆和专家批量增加或单条编制在线提交。盖蒂词汇的官网地址是 <http://www.getty.edu/research/tools/vocabularies/>。

3.1 构建原则

Getty 词表在构建之初就确立了以下原则:

(1)遵循 ISO 和美国国家标准(NISO)的叙词表结构标准而构建;

(2)词表结构支持多语言,所有术语词汇都采用 Unicode 编码,变音代码与 Unicode 做映射;

(3)收录词汇包括编目需要的术语和有关视觉艺术和建筑的检索信息;

(4)各个词表的术语中,有一个被标记为首选的规范词。可能有多个规范词,用不同的文种表达。

(5)各词表由专门的学术顾问团队来审批和更新,顾问团队由艺术和建筑史学家、建筑师、策展人员、视觉资源编目者、档案馆、博物馆人员和叙词表构建专家组成。

(6)收录的词汇取自从古到今的全球范围;

3.2 结构

Getty 词表都是层级化的术语表,包含等级关系、等同关系和相关关系。表 4 是 AAT 的分面和层级^{[6][7]}。

表 4 AAT 的分面和层级

1	面(Facets)	层级(Hierarchy)
2	相关概念分面(Associated Concepts Facet)	相关概念(Associated Concepts)
3	物理属性分面(Physical Attributes Facet)	属性和特性(Attributes and Properties),条件和效果(Conditions and Effects)、设计元素(Design Elements)、颜色(Color)
4	风格和时代分面(Styles and Periods Facet)	风格和时代(Styles and Periods)
5	代理分面(Agents Facet)	人(People)、机构(Organizations)和团体(Living Organisms)
6	活动分面(Activities Facet)	学科(Disciplines)、功能(Functions),事件(Events)、身体和心理活动(Physical and Mental Activities)、过程和技术(Processes and Techniques)
7	材料分面(Materials Facet)	材料(Materials)
8	对象分面(Objects Facet)	对象分类和系统(Object Groupings and Systems)、对象类型(Object Genres)、组件(Components)
9	品牌名称分面(Brand Names Facet)	品牌名称(Brand Names)

ULAN 目前有 5 个分面(Facets)名称:个人艺术家(Persons, Artists)、法人团体(Corporate Bodies)、非艺术家(Non-Artists)、身份不明的人(Unidentified Named People)、不知名的人(Unknown People by Culture)。除了法人团体,其他通常都没有分支元素。

TGN 目前有两个分面(Facets)的地理名称:世界地名和行星地名(World and Extraterrestrial Places)。在世界地名分面,有层级关系,地名在层次结构中的排列一般反映了当前的政治上的和实际上的地理名称,也包括历史上曾经存在的国家和帝国。

CONA 最上层的主要分面(facets)有建筑物(

Built Work)和可移动作品(Movable Work),另外还有概念性作品(Conceptual Works)、视觉作品(Visual Surrogates)、未名作品(Unidentified Named Works)。CONA可能存在多个更广泛的上下文关系,使得CONA多层次化。除了等级关系(例如,一个印刷品和它的卷期之间)之外,CONA具有等同关系(例如在等效标题/名称之间)和关联关系(例如,草图和最终作品之间)。

3.3 特点分析

(1)多层次的结构化把不同级别、多种形式的术语关联起来。

GETTY词表虽然显示为一个词表,实际上是一个关联的数据库。每个记录有一个唯一的数字标识,通过这个标识把与每一个记录相关的所有术语词汇、相关及上一级词汇、数据源和附注都关联起来。任何术语都可以包括复数形式、单数形式、自然语序、倒装语序、拼音形式、科学写法和普通写法等各种形式的说法、各种语源的同义词。在这些术语中,有一个被标记为首选的规范词。可能会有多个规范词,用不同的文种表达。

(2)提供了内容全面的规范档。

AAT已被完全翻译为西班牙语和荷兰语。德语和葡萄牙语的翻译也正在进行中,而且在增加法语词汇和意大利语词汇。中国台湾的“台湾数字学习和数字档案项目”(Taiwan E-Learning and Digital Archives Program)已经把约8000多个术语翻译为汉语,推出了AAT-Taiwan,目的是将台湾的数字典藏与数字学习科技计划多年来累积的数量庞大的数字藏品与国际分享^[6]。

ULAN词汇提供了艺术家规范术语,专注于其中的文化信息。包括艺术家专有名称或非常用称谓、传记、相关的个人或团体。适用于公司、工作室、博物馆、图书馆特藏部、赞助人、捐助者、安保部门和其他群体共同参与创作、收集、维护有关的信息。艺术家可能是单独的个人或一起工作的团体。ULAN的艺术家通常是视觉艺术的创作者、概念产生者或生产者,但ULAN并不侧重于演员、舞者或其他的表演艺术家^[8]。

TGN词汇提供了地名规范术语,专注于其中的文化信息。适用于公司、工作室、博物馆、图书馆特藏部、赞助人、捐助者、安保部门和其他群体共同参与创作、收集、维护有关的信息。TGN包括名称和

与地名相关的信息,地名包括行政实体(例如城市、国家)和地理特征地名(例如山脉、河流),当前和历史的地名也包含在内。TGN发展的领域包括更新行政区划的层次结构;添加新发现考古遗址、新增世界遗产名称和其他历史遗迹,特别专注于亚洲、前哥伦布时期、中东和其他地方;建立历史上的国家和帝国的层次结构^[9]。

CONA所包含的是文物名称的规范档,文物包括建筑和可移动文物,如绘画、雕塑、版画、素描、手稿、照片、纺织品、陶瓷、家具、其他视觉媒体(如壁画和建筑的雕塑,行为艺术和考古文物),及各种由博物馆收藏和物质文化领域的文物部件。CONA主要针对已经编目的艺术品,如在学术文献中的艺术品、博物馆收藏中的艺术品、其他视觉资源收藏中的艺术品、档案馆和图书馆收藏的艺术品以及在艺术、建筑或考古的编制索引项目中的艺术品。每个CONA记录对应一件艺术品或一座建筑物^[10]。

(3)使用灵活,用户广泛。

盖蒂词汇的主要用户包括博物馆、艺术图书馆、档案馆,视觉资源编目员、收藏家、考古项目、有关艺术的编目、艺术和艺术史领域的研究者,以及处理这些用户需求的信息专家。此外,还有相当多的学生用户和公众用户。

编目员和索引编制者通常以两种方式使用盖蒂词汇:一种方式是通过数据许可文件把他们集成到一个收藏管理系统中,另外一种是通过盖蒂网站使用在线数据库,网站上的数据库可以支持有限的研究和编目工作。广泛使用盖蒂词汇的机构应向盖蒂词汇项目组获取许可,采用集成方式来使用。

4 数据交换标准——LIDO^{[11][12]}

轻量信息描述对象(Lightweight Information Describing Objects,以下简称LIDO)是在国际博物馆界的合作努力下,为向门户和其他聚合资源知识库贡献文化遗产内容而提出的共同解决方案。LIDO是用来作为聚合博物馆数据的格式,可利用这种格式转换和发送各博物馆的文物数据信息到一个整合文物元数据的机构。LIDO的官方网址是<http://www.lido-schema.org>。

4.1 LIDO的设计原则

LIDO XML架构(和规范)的设计使其可以:

(1)对文物进行适当描述;

(2)允许一个机构决定其向特定门户提供的元数据记录的丰富程度(即在不同场景中可能不同);

(3)允许一个机构在传递文物实体的元数据时链接到相关的数字对象;

(4)在 LIDO 的记录中要包括回连到实体记录的链接;

(5)允许传送包含用于正确显示和检索单个文物的所有信息的元数据记录;

(6)通过提供引用受控词汇和规范文件,使所提供的实体(即谁、什么、何时和在哪里)被识别;

(7)提供用于检索和显示的元数据,即具有显示和索引元素;

(8)支持多种语言记录,无论是结构元素级别还是单个文本元素级别(或两者);

(9)能够区分文物实体标识符和描述文物的网页标识符及它的在线数字对象的标识符。

4.2 LIDO 的结构

LIDO 是由一套封装的元数据组成的结构数据。此元数据设计的重点是从国际文献工作委员会的概念参考模型(Conceptual Reference Model)引入了“事件”概念。例如,对象的创建、收藏和使用被定义为“事件”,相关的实体包括时间、地点、人物。这样,这些“事件”就可以用一致的方式来表现。

为了满足不同门户的需要,LIDO 用“数据集”来组织信息。LIDO 定义了 14 组信息,其中只有 3 组是必备的。这利于最大限度地保存不同来源信息的完整性。可以由机构决定哪些数据提供给门户,或者在网上发布。

LIDO 允许一个机构提供为检索和获取而优化的索引信息和为门户的用户在线提供信息而优化的显示信息。LIDO 的结构化元素包含“数据元素”,可以容纳服务环境下已被收割的信息和发送给用户的信息。LIDO 允许记录关于数据的来源信息(例如来自某书)和受控术语(例如术语在某词典中的识别码)。

LIDO 记录的信息,从概念上可以分成 7 个领域,其中 4 组为描述性信息,3 组为管理性信息。见表 5。

4.3 LIDO 的特点分析

LIDO 作为专门为博物馆设计的一个国际收割标准,有如下特点:

(1)LIDO 是一个 XML 规范,用于把一个机构

的文物对象和其他资源的信息提交到门户,聚合资源库和其他类似的数据库;

(2)LIDO 是作为一个元数据传输机制被使用,包括一个机构自己的在线收藏数据库,以及汇集来自许多机构外部门户网站的资源信息;但 LIDO 不是一个完整开发的数据交换格式,即不应该用于机构之间的点到点的信息交换,特别是不用于从一个收集管理系统迁移到另一个收集管理系统。

表 5 LIDO 的描述性信息

信息领域(Information Area)	说明
对象的类别信息 (Object Classification)	关于对象类型的信息
对象的识别信息 (Object Identification)	关于对象的基本信息,尤其指物理特征等
事件(Event)	对象参与的事件,包括创造、发现、修改、使用等
对象的相关信息 (Object Relation)	主题信息以及其他相关的对象等

表 6 LIDO 的管理性信息

信息领域(Information Area)	说明
权限信息(Rights Work)	与实体资源与它的数字对象相关的使用权限信息
记录信息(Record)	关于编目记录的基本信息
资源信息(Resource)	提供给服务环境的数字资源的信息

(3)LIDO 旨在囊括关于文物(如艺术、历史和科技)的全面描述性信息。但不是用于完整编目的格式。

(4)LIDO 支持多语种门户环境。它为每个元素提供了语言属性,可用于多语种资源,支持多语言环境。

LIDO 作为一个对象描述元数据的收割集,允许所在机构将他们的元数据以标准的方式提供给他们其他门户。所以目前已经在世界范围内得到广泛应用,例如 Athena Plus 是欧盟竞争力和创新框架计划的一个项目,它始于 2013 年并于 2015 年结束。该联盟由来自 21 个欧盟国家的 40 个成员组成。Athena Plus 提供平台将成员单位提交的元数据转化为 LIDO 中间标准,然后再传输给欧洲的数字图书

馆。LIDO 在欧洲、美国的文物项目和博物馆中应用较为广泛。

5 我国博物馆的元数据标准现状

相比国外博物馆以艺术品和建筑雕塑为主的藏品,我国博物馆的藏品更具复杂多样性。我国博物馆元数据的设计始于对国外主流文物元数据标准的研究,但大都进行了本地化,或简化,或扩展,或改造,因此基本上每个文博机构采用的元数据都有区别。业内公认的当属国家文物局第一次和第三次文物普查推出的相关规范文件。

从公开发表的论文及调研来看,我国尚未形成完整的博物馆元数据内容、结构、取值、交换、发布体系。现有的博物馆元数据标准多为内容标准,个别数据结构标准、专门领域有数据取值标准。数据交换(和收割)和数据呈现,虽然在一些项目中有所应用,但未见形成通用标准。

(1) 数据内容标准。

2013 年国家文物局编撰的《第一次全国可移动文物普查工作手册》的第三章“可移动文物普查登录内容”,规定了馆藏可移动文物登录的四部分内容,包括馆藏文物的基本信息、管理信息、影像信息、收藏单位情况等,并给出各项著录的详细规则。它可以称为博物馆元数据的一个数据内容标准。各文博单位内部使用的元数据标准多为内容标准^[13]。

(2) 数据结构标准。

2001 年国家文物局发布的《博物馆藏品信息指标体系规范》可以称得上是博物馆的一个数据结构标准。该规范将博物馆元数据分为“藏品信息”“藏品管理工作”“藏品文档信息、研究论著信息与声像资料信息”三个指标群,其下又分 33 个指标群、139 个指标项,每个指标群或项都有一个独立的指标代码^[14]。

(3) 数据值标准。

1992 年湖南省文物考古研究所李科威主持编写的《中国考古主题词表》,收词量 12449 条,是我国文物界较早的一个数据取值标准。但更多的数据取值是在元数据内容标准中规定,尚未形成专门的取值标准。有些则直接参考其他国家标准,比如《中华人民共和国行政区划代码》《中国各民族名称代码》等^[15]。

(4) 数据交换(和收割)标准和数据呈现标准。

在文物普查工作中,数据收割采用统一的数据格式,但并未形成标准,所以这两类标准在我国文物界都未形成通用标准。

6 结语

我国在博物馆元数据的标准体系制定及利用方面,同国外发达国家存在较大差距。虽然近几年在数字博物馆建设方面取得长足进步,积累了可观的元数据及数字资源,但标准不统一,大都处于分散的、无组织状态,文物机构间缺乏数据交换与数字化资源共享。在缺乏大数据检索需求的前提下,目前已在使用的博物馆元数据的设计颗粒度不够精细,不利于检索利用和数据挖掘。同时,数据取值标准的重要性尚未得到充分重视,文物元数据标准的统一与推广已经成为文物数字资源共享和信息交换的瓶颈。

针对当前文物数字化保护在元数据体系和规范方面的实际需要,从文物数字化保护的应用逻辑出发,应立足国内已有基础特别是国家相关科技计划项目的技术积累,在对国内外文物相关元数据标准及元数据管理的现状与应用情况进行分析研究的基础上,逐步建立我国的文物元数据标准体系,研制文物描述元数据标准、管理和保存元数据标准;以应用为导向,研究编制文物元数据标准的著录细则、实施指南;研究元数据应用规范、元数据封装与交换标准、访问协议等。

在建立文物的描述元数据标准的同时,要建立各种规范档,使文物描述能够在全国范围内进行统一规范化的描述。应用统一的元数据标准和取值规范来描述艺术作品及相关文献,可以使文物管理机构在统一标准的背景下利用数字技术更好地服务于艺术研究,并且通过特定字段将具有类似特征的作品联系起来。例如,具有同样印鉴的作品、具有同样风格的作品、同一主题的作品、特定尺寸的作品、被同一收藏家收藏过的作品等。

科技部国家科技支撑计划课题“文物数字化保护元数据标准研究”旨在建立我国文物元数据的各类标准规范,为文博单位开展文物元数据的制定和著录工作提供规范化、标准化的指导性文件,同时促进未来文物元数据的统一管理和数据共享。

参考文献

- 1 Steven J. Miller. Metadata and cataloging online resources: selected reference documents, web sites and articles. University of Wisconsin — Milwaukee School of Information Studies. Updated June 1, 2011. [2017-03-01]. <http://www.uwm.edu/~mll/resource.html>.
- 2 Marcia Lei Zeng and JianQin. Metadata .New York: Neal—Schuman, 2008.
- 3 Categories for the Description of Works of Art[EB/OL]. [2017-03-01]. http://www.getty.edu/research/publications/electronic_publications/cdwa/index.html.
- 4 Cataloging Cultural Objects A Guide to Describing Cultural Works and Their Images, Online Edition[EB/OL]. [2017-04-11]. http://cco.vrafoundation.org/index.php/toolkit/cco_pdf_version/.
- 5 Getty Vocabularies[EB/OL]. [2017-03-01]. <http://www.getty.edu/research/tools/vocabularies/>.
- 6 Art & Architecture Thesaurus[EB/OL]. [2017-03-01]. <http://www.getty.edu/research/tools/vocabularies/aat/index.html>.
- 7 张俊娥. 艺术和建筑词表 AAT 的中文文化研究[J]. 数字图书馆论坛, 2015(7):36-43.
- 8 Union List of Artist Names[EB/OL]. [2017-03-01]. <http://www.getty.edu/research/tools/vocabularies/ulan/index.html>.
- 9 Getty Thesaurus of Geographic Names[EB/OL]. [2017-03-01]. <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>.
- 10 Cultural Objects Name Authority[EB/OL]. [2017-04-01]. <http://www.getty.edu/research/tools/vocabularies/cona/index.html>.
- 11 LIDO's background [EB/OL]. [2017-04-11]. <http://network.icom.museum/cidoc/working-groups/lido/resources/>.
- 12 Implementing LIDO[EB/OL]. [2017-04-11]. <http://athe-naplus.eu/index.php?en/208/implementing-lido>.
- 13 国家文物局第一次全国可移动文物普查工作办公室. 第一次全国可移动文物普查工作手册. 北京: 文物出版社, 2013: 31-58.
- 14 国家文物局. 博物馆藏品信息指标体系规范, 2001.
- 15 李科威. 关于《中国考古主题词表》的编纂. 华夏考古, 1992(2):110.

作者单位: 北京大学图书馆, 北京, 100871

收稿日期: 2017年6月18日

Study on the Construction of Metadata Standards Recommended by Getty Research Institute

Zhang June Wang Yalin

Abstract: Metadata plays a vital role in the digitization of the museum, it is necessary to effectively organize and reveal the information of the museum artifacts, and it needs all kinds of metadata standards as the basic support. CDWA(Data Structure Standard), CCO(Data Content Standard)、AAT, ULAN, TGN, CONA(Data Value Standards) and LIDO(Data Format / Technical Interchange Standard), which are published by the Getty Research Institute, had been reviewed and the role of different types of metadata standards in the construction of digital museums had been analyzed. According to the current situation of cultural heritage Metadata in China, it is proposed that in the process of China's digital or intelligent museum construction, not only description metadata standards should be firstly built, but also Data Value Standards and data Interchange Standards should be established gradually as the foundation of unified metadata management and data sharing of cultural heritage in China in the future.

Keywords: Metadata; Data Structure Standard; Data Content Standard; Data Value Standard; Data Interchange Standard