



中国大学数字博物馆丛书

6

博物馆数字资源 的管理与展示

齐越 沈旭昆 著



上海科学技术出版社

中国大学数字博物馆丛书

博物馆数字资源的管理与展示

齐越 沈旭昆 著

上海科学技术出版社

内 容 提 要

《中国大学数字博物馆丛书》是国家教育部重大科研项目——大学数字博物馆共享平台与规范标准研究项目成果的重要组成部分。本书是该丛书中的一本,共计6章,主要从博物馆数字资源的管理与展示的角度,展示了大学数字博物馆共享平台与规范标准研究的主要成果,它将阐释如何对作为大学数字博物馆基础的数字资源进行科学规范管理和统一协同却富有个性的展示。

本书可供博物馆和文物保护科研机构专业工作者、大专院校博物馆专业师生阅读,也可供关注数字博物馆建设的社会各界人士参考阅读。

图书在版编目(CIP)数据

博物馆数字资源的管理与展示 / 齐越, 沈旭昆著.
上海: 上海科学技术出版社, 2008.6
(中国大学数字博物馆丛书)
ISBN 978-7-5323-9201-8

I. 博... II. ①齐...②沈... III. 数字技术 - 应用
- 博物馆 - 研究 IV. G26-39

中国版本图书馆 CIP 数据核字 (2007) 第 180155 号

上海世纪出版股份有限公司 出版、发行
上海科学技术出版社
(上海钦州南路71号 邮政编码200235)
新华书店上海发行所经销
常熟市兴达印刷有限公司印刷
开本 787×1092 1/16 印张 9.375
字数:196千字
2008年6月第1版 2008年6月第1次印刷
印数:1-3000
定价:30.00元

本书如有缺页、错装或坏损等严重质量问题,
请向工厂联系调换

《中国大学数字博物馆丛书》序

现代意义上的博物馆是收藏、保护、研究、展示人类活动和自然环境的见证物，并为公众提供知识、教育和欣赏的非营利性社会服务机构，同时也是一个国家和民族的物质与精神文化遗存宝库。在不同主题博物馆中保存和展示着过去的文明、技术、生命和自然现象。当一种文明（历史考古类博物馆）、一种技术（工程技术类博物馆）、一种生命（生命科学类博物馆）、一种自然现象（地球科学类博物馆）随着时间的流逝而成为过去，当人们从一种全新的角度去看待它们的时候，一些原本不为人所知的价值就会显露出来，启迪着人们新的思想和灵魂。正是由于这样的原因，博物馆的建设有了更深的科学内涵和人文意义，它担负着知识载体收集、科研解读、教育传播以及审美休闲的重任，总能使人联想到美好的生活品质而心向往之。

目前全国高校系统的博物馆约有100座，大致可分为地球科学类、生物类、技术类和人文（历史、考古、艺术）类四种不同类型。这些博物馆大多为配合教学、科研工作而建，很少对社会公众开放，其知识传播和影响仅局限于很小的范围。由于人力、场馆或经费方面的因素，大学博物馆的资源利用率也较低，不仅藏品难于补充，很多珍贵藏品难以同公众见面，也无法对现有藏品所承载的自然和历史文化内涵进行更深层次的挖掘，从而不能完全发挥出博物馆应有的作用。因此，教育部《面向21世纪教育振兴行动计划》“现代远程教育工程”于2001年专门设立了“现代远程教育网上公共资源建设——大学数字博物馆建设工程”项目，以中国大学实体博物馆丰富的馆藏资源为基础，以资源系统整合与共享为主线，以信息化技术为手段，坚持以人为本，搭建一个具有系统性、先进性、可持续性的中国大学数字博物馆，

为大学教学、科研以及全民素质教育提供强有力的支撑。

目前全国已有超过18家高等院校建设了各具特色的大学数字博物馆。为深化中国大学数字博物馆工程项目，教育部还特意拨专款为其关键技术研究立项，支持由六所重点大学的多学科教授组成的核心专家组成员总结经验、开发工具，为该项建设事业提供相关理论和技术支持，本套丛书就是该研究项目成果之一。作为教师身份的 authors 还期待达到另一个目的——为已经在大学开展的数字博物馆人才培养工作提供教材。

我国的数字博物馆是随着新世纪的到来而出现的新生事物，既有单体建设如数字故宫，也有群体建设如中国数字科技馆、中国大学数字博物馆，更有大量准备建设的单位存在。对于那些打算而尚未启动数字博物馆建设的单位来说，应该汲取已有相关经验以保障人财物力投入的有效性。本套丛书的出版填补了我国数字博物馆系统性专著的空白，但毕竟是根据一时的实践及范围有限的大学数字博物馆特点而作，难免带有不足或不到之处，敬请读者不吝批评指正！

《中国大学数字博物馆丛书》编委会

2006年秋

《中国大学数字博物馆丛书》

编辑委员会

主 编

赵沁平

编辑委员

(以姓氏拼音为序)

陈红京 陈 源 董少春 李鸣光 娄 晶 齐 越
沈旭昆 武港山 谢焕忠 徐士进 杨雪琴 袁 伟 查红彬



目录

第 1 章 综述	1
1.1 数字博物馆概述	1
1.2 数字资源	3
1.3 博物馆数字资源的管理	4
1.4 博物馆数字资源的展示	4
第 2 章 博物馆数字资源的组织形式与管理方法	6
2.1 博物馆数字资源管理的信息模型	6
2.2 元数据	8
2.3 数字藏品的概念模型	10
2.4 数字藏品仓储	14
2.5 博物馆数字资源的提交	18
2.6 博物馆数字资源库的访问授权	29
第 3 章 博物馆数字资源的分布式管理系统	30
3.1 博物馆数字资源管理系统的整体框架	30
3.2 数字资源的定位	33
3.3 地方站点到中心站点的注册	44
3.4 各地方数字资源库与数据中心的信息交互	46
第 4 章 博物馆数字资源的展示系统	65
4.1 博物馆数字资源展示存在的主要问题	65
4.2 数字博物馆展示系统	66
4.3 检索功能	79
第 5 章 博物馆数字资源的交互展示技术	82
5.1 多媒体交互展示	82
5.2 三维模型的构造	85
5.3 三维交互浏览	109
第 6 章 博物馆数字资源保护技术	129
6.1 数字水印技术	129
6.2 远程渲染技术	137

第1章 综 述

1.1 数字博物馆概述

大学博物馆是收集、保护、展示各种重要文物、标本的场所,是高等学校文化基础设施的重要组成部分,是实施素质教育、提高国民文化素质的重要教育基地,大学数字博物馆的建设对于保护和合理利用国家珍贵标本、文物,建设数字典藏系统,开展现代化教学、科研及科普教育,推动我国高等院校网上公共资源建设,对于构建我国的远程教育基础设施具有重要意义。

大学博物馆大多是为了配合教学和科研工作而建立的,很少对社会公众开放,其知识传播和影响也仅局限于很小的范围。此外,因为人力、场馆或经费方面的原因,这些大学博物馆的资源利用率也较低,一方面藏品难以补充,很多珍贵藏品难以与公众见面,另一方面也无法对现有藏品所承载的自然和历史文化内涵进行更深层次的挖掘,从而不能完全发挥出博物馆应有的作用。

正因为如此,教育部《面向 21 世纪教育振兴行动计划》“现代远程教育工程”于 2001 年专门设立了“现代远程教育网上公共资源建设——大学数字博物馆建设工程”项目,由教育部科技司主持,18 所高等院校承担,力图将各高校已有的博物馆内容、计算机技术和专业优势整合起来,建立一个覆盖中国大部分地区的综合性分布式数字博物馆。2003 年 7 月一期建设结束,初步建成了地球科学、生命科学、人文艺术、科学技术四类大学数字博物馆。大学数字博物馆一期工程的建设成果不仅能够直接为高等院校教学科研服务,而且充分注重了素质教育在网上的实现,力求体现科学与人文精神的结合,阐明基本理论的同时还特别叙述了科学发展史上重要发现的创新历程,以及科学思想的诞生经过,体现自然科学的演进方式和科学家的人文意识,为学生远程自主学习建立了不可多得的第二课堂。

教育部大学数字博物馆工程自 2001 年启动第一期以来,已经在因特网(Internet)上建成 18 个大学数字博物馆。这些数字博物馆已经拥有了大量的数字化藏品资源,如果能将分布于各地博物馆的数字资源结合起来,统一地管理和展示,这样,用户就不需要知道他们访问的资源存储在哪一个站点上,便可以获得相应数字资源的信息。因此,从 2004 年起,开始了大学数字博物馆二期工程,目标是将 18 个博物馆所有数字资源进行无缝整合,所有的数字资源按主题统一管理,无需考虑其物理位置。本书主要介绍大学数字博物馆系统管理和展示数字资源的方法。

数字博物馆(Digital Museum)是以数字形式对实体藏品的各方面信息进行采集和管理,实现藏品的永久保存,并可以通过互联网为用户提供数字化的展示、教育和研究等多种服务的信息系统。它摆脱了实体博物馆所必需的建筑、陈列、参观时间等条件的束缚,突破了时间与空间的限制,使任何人在任何时间、任何地点都能够

获取所需的知识信息。

数字博物馆采用计算机网络技术、数据库技术、多媒体技术、虚拟现实技术、人工智能技术、人机交互技术,构造出分布式的数字信息资源系统,为文化遗产提供永久的数字化保存、管理和传播手段。利用数字博物馆,可以解决许多当前困扰我国文博事业发展的问題。

(1) 扩大服务对象的数量

普通博物馆提供的是与观众面对面的直接服务,因此服务对象的数量十分有限。数字博物馆的观众是虚拟的,是看不见的,数量也是巨大的。数字博物馆通过网络间接地为用户服务,根本地改变了普通博物馆的服务模式。

(2) 提高文化遗产的展出率

普通博物馆的可移动文物主要存放、展示在库房和展厅中,基本陈列和临时展览需要什么,就从库房调出什么。对于不可移动文物,开放哪些文物场所,观众就参观哪些场所。库房的藏品不给观众看,不开放的场所不让观众参观,这已是普通博物馆从未改变的传统做法。所以“展什么、看什么”是普通博物馆的真实写照。数字博物馆将库房、展厅的可移动文物,古建筑、古遗迹、历史遗存等不可移动文物,以及与之相关的无形信息全部数字化,为用户提供丰富的、可以自由选择的内容。利用现代信息检索技术,让观众做到“看什么、有什么”,增强观众的主动参与性,使服务由被动转变成主动。

(3) 提高服务质量

普通博物馆为了提高服务质量,尽管采取延长开放时间、节假日不休息的办法,但这仍然是一种面对观众“开馆能看,闭馆不能看”的方式。数字博物馆将展示内容放到因特网上,超越了时间和空间的限制,观众可以借助计算机,“愿意什么时候看,就什么时候看”,即查即用。

(4) 改善展示效果

运用数码摄像、三维建模以及扫描技术,对文物进行数字化,为用户提供三维的、动画的、影像的以及文字的全方位展示。特别是运用虚拟现实技术,可以根据考古研究的数据和文献记载,生动形象地再现已挖掘或已经湮灭了遗址、遗存。总之,通过这些技术,可以将所有的文化遗产资源统一整合起来,全面地向社会传播,达到教育、研究和欣赏的目的,而丝毫不会影响到文化遗产本身的安全。

数字博物馆的特点主要表现在以下几个方面。

(1) 信息全面、规模巨大

长期以来,实体博物馆作为向公众展示人类各个方面文化遗产的主要途径,积累了种类繁多、数量庞大的藏品资源,其自身的职能就决定了对实体藏品进行数字化后的数字资源信息必然是全方位的,规模必然是巨大的。而各种数字资源之间的跨度也十分巨大,这些因素都导致数字博物馆的数据量是海量的。

同时,数字博物馆的资源不仅包含有形部分,如藏品的外形、装饰等;还包含无

形的文字记载信息,如制造工艺、化学物理成分,以及其生产、使用、流传等相关的多种信息。这就意味着,博物馆数字资源必须通过有效的管理整合信息,才能成为更好地为公众提供服务的有组织的统一体。

(2) 数字博物馆技术缺乏积累

从事数字博物馆研究的专业人士经过长期的积累,获得了大量经过实践证实的经验,这对解决实际问题是行之有效的,但是这些经验目前还没有形成一套严谨的理论体系,导致很多方法难以得到准确的描述和精确的分析。因此,需要建立一个在数字博物馆领域对问题进行准确描述的术语体系,形成一套精确表达知识的科学方法,同时形成一整套针对具体问题选择相应知识的逻辑性的策略和途径,从而使数字博物馆的诸多实际问题转化为计算机科学领域的可计算性问题。

(3) 标准化问题

博物馆学界的操作处理长期以来缺乏清晰统一的程序。因此在建立大学数字博物馆的过程中,需要加大力度,完善标准化及规范化水平。

(4) 表现形式

数字博物馆庞杂海量的数据资源决定数字博物馆的表现形式必然是多种多样的,这与以文献记载内容为主体的数字图书馆信息对象截然不同。后者主要着眼于文字和图像。而数字博物馆中有形数字资源对象注重资源的三维信息,无形数字资源注重音、视频以及图像信息。因此,数字博物馆的数字采集需要更加全面、复杂的数字化采集工具。

另一方面,数字博物馆对视觉上的要求,远远高于数字化图书馆。公众使用数字图书馆的时候,其信息获取方式以浏览文献资源为主,通常是逐字逐句的阅读,这无形中降低了对数字图书馆带宽的要求。然而,在数字博物馆中,公众更重视视觉上的体验和参与性,他们希望数字博物馆成为他们与各种知识交互的窗口,从而需要连续地、全方位地观赏,因此数字博物馆的数据流量巨大,网络带宽要求很高。数字博物馆的信息处理对象复杂,对系统的容量、结构、流量、管理等方面都比数字化图书馆的技术实施更有挑战性。

综上所述,需要一套科学合理的体系,针对博物馆数字资源自身的特点进行管理和展示。

1.2 数字资源

数字资源是指以数字化的形式存在于计算机中的资源。它的获得渠道有两种,一种是将传统的资源进行数字化,如将图书扫描,得到图片;另一种是通过信息技术直接获得,如用数码相机拍摄下来的图片。数字资源有很多属性,其中主要有两个方面:一方面,与存储和管理数字资源相关的属性,包括资源类型和格式、元数据资源的网络分布等;另一方面,与提供服务相关的属性,也即资源所提供的服务。

数字资源有各种各样的分类方法,按照资源的媒体类型来分,可以分为以下四种。

(1) 文本

文本资源是最普遍的数字资源,在数字资源领域内占有很重要的地位。表示文本的常用格式有 Txt、Doc、Pdf、Html、Xml 等。

(2) 图片

图片也是很普遍的一种数字资源,其常用格式有 Bmp、Gif、Jpeg、Png、Tiff 等。

(3) 音频

数字音频资源是由数字化的声音信息构成的,包括经过数字化处理的音乐、语音、自然音响等。音频常用的表示格式有 Wav、Mid、Mp3、Rm 等。

(4) 视频

随着多媒体技术的迅速发展和日益普及,数字视频资源占信息资源的比重越来越大。它的主要信息来源是电影、电视、录像和动画等动态图像信息。其表示格式有 Avi、Mpg、Mpeg、Divx、Rm、Asf、Wma、Mov 等。

1.3 博物馆数字资源的管理

博物馆数字资源建设的目标是通过有效的管理和合理的展示,形成一个共享的知识网络,使其成为一个全面发挥博物馆功能的平台。在博物馆数字资源管理方面,大学数字博物馆项目采用大规模的分布式系统,包括南北两个中心站点,连接各地方站点,是一个涵盖 4 个学科分类的巨大虚拟博物馆群。用户访问时,将不会察觉出藏品的物理存储位置不同。此外,大学数字博物馆基于美国惠普公司和麻省理工学院图书馆联合研制开发的开源系统 DSpace(见 <http://www.dspace.org>),整合博物馆数字资源,实现结构化、系统化管理。

大学数字博物馆资源管理系统主要有以下功能。

- (1) 将数字化对象及其元数据作为一个整体进行保存;
- (2) 允许藏品有机地结构化组织起来;
- (3) 提供必需的访问控制;
- (4) 索引、搜索和提取;
- (5) 管理存储;
- (6) 实现数字资源的长期保存。

1.4 博物馆数字资源的展示

博物馆数字资源的展示技术主要包括两个方面:展示系统和数字资源保护。

对于展示系统,从艺术设计角度讲,数字博物馆的外观设计对于能否吸引观众,达到寓教于乐的目的是十分重要的;从技术角度讲,通过集成现有技术、研发先进技术,可以更好地展示数字博物馆的数字资源。在分布式博物馆的架构下,通过统一

的框架结构,将不同主题不同风格的博物馆通过同一套数字资源管理系统搭建起来,从而降低博物馆展示的设计周期,为整体集成提供方便。

对于数字资源保护,基于互联网的数字博物馆,将大量的数字资源通过网络展现给观众,服务于社会。然而,如何保护这些数字资源的版权是一个重要的问题。目前数字资源保护的主要技术有数字水印技术和远程渲染技术。

第2章 博物馆数字资源的组织形式与管理方法

2.1 博物馆数字资源管理的信息模型

2.1.1 概述

在大学数字博物馆中,对博物馆数字资源采用了分层结构的信息模型进行管理。整个博物馆首先划分为若干个顶层结构,即资源群落(Community),然后划分藏品结构,把不同藏品划分到不同的资源群落中。若资源群落的规模和定义仍然过大,则可以将其再次细分为若干个子资源群落(Sub-Community),子资源群落还可以再继续划分,各个子资源群落又对藏品的信息进行了更进一步的界定和说明。另外,资源群落和子资源群落可划分为若干个资源集合(Collection)。与资源群落不同,资源集合是数字博物馆的最小存储单位,它通常是一系列拥有相同属性的藏品集合,数字藏品将根据其特征存放在某个特定的资源集合之下。需要说明的是,由于划分标准并不是绝对的,因此某一个资源集合可能出现在多个资源群落中;同理,某一个数字藏品也可能出现在多个资源群落中。但资源群落、资源集合或数字藏品,都拥有唯一独立的标识符。

这种信息模型解决了不同分馆馆藏结构不同的问题。例如,在人文科学中,可能存在4个子群落,每个子群落下面可能各有3层细化的子群落;而对于生命科学领域,可以存在6个子群落,每个子群落则可由4层再细化的子群落组成。

2.1.2 资源群落

资源群落是整个馆藏结构中的最上层,它将整个博物馆划分成几个部分,从结构上看可以将资源群落看作是当前博物馆的分馆。每个资源群落都含有一些属性描述当前资源群落的性质,如Name表示当前资源群落的名称;Description是对当前资源群落的描述;Handle是系统为资源群落分配的唯一标识符,用来区别不同的数字资源。一个博物馆必须至少包含一个资源群落(也就是不划分子馆的情况)。另外,一个资源群落可以根据需要划分为多个子资源群落,这主要是由于某些资源群落在划分时的范围过于宽泛,不利于藏品的存储、管理和展示,故将其继续划分为概念更加清晰的几个部分。例如,名为“飞行器”的资源群落,如果不作划分,则所有可看作“飞行器”的藏品都会集中在该资源群落之下,这将导致存储和展示都很混乱。但如果将其进一步划分为:“飞机”、“火箭”、“飞艇”、“风筝”,则这几类有明显区别的藏品将分类存放,博物馆的结构会更加清晰。

2.1.3 资源集合

资源集合是比资源群落更低一级的层次概念,资源集合是博物馆层次结构中的最低层级,它通常都与一个具有相同特性的藏品集合相对应。如仍然以“飞行器”为

例,在子资源群落“飞机”中,划分出“军用飞机”和“民用飞机”两个资源集合,则所有属于民用飞机类型的藏品(如波音 747、737 等)都存放在资源群落“飞行器”的子资源群落“飞机”下的资源集合“民用飞机”中,这也正是博物馆层次结构的体现。

2.1.4 数字藏品

数字资源也即藏品(Item),是整个数字博物馆的核心内容,也是数字博物馆存放的实际内容。

藏品摄入(Ingest)系统后应该包含两部分内容:数字化的博物馆藏品和元数据信息。每个藏品都分配有一个称为 Handle 的唯一标识符。

因此,一个藏品对应的数字资源有:Handle(系统为该藏品分配的全球唯一的标识符)和藏品元数据文件(描述这个藏品的属性信息,如名称、作者、日期等)。

一个数字藏品文件由多个“数字包”(Bundle)组成,每个数字包又由若干个“数字流”(Bitstream)组成。“数字流”是一些比特流文件,通常的形式是计算机常见的文件格式,如 Txt 文件或 Jpeg 文件等。把那些相互关联的数据流文件放在一起就组成了一个数字包。例如一个网页文件通常包含文字、图片、Flash 动画等,文字、图片、Flash 动画都称为数字流,这些数字流一同构成了一个数字包,若干个这种数字包的集合构成一个藏品。一个藏品必须至少包含一个数字包,且每个数字包必须至少包含一个数字流。数字包和数字流分别具有一些属性,如名字(Name),数字包还要指出其包含的处于首位的数字流,数字流则还要提供校验和等属性。每个数字流的结构都是相同或至少十分相似的,因为它们都遵循相同的结构定义。

2.1.5 群落、集合和藏品之间的关系

群落、集合都是博物馆的层次结构概念,它们既有联系又有区别。

资源集合从属于资源群落,任何一个资源集合都包含在某一个(或多个)资源群落中。可以说,资源集合所规定的藏品属性集合是资源群落规定的属性的子集,任何一个属于某资源集合的藏品也必然同时属于该集合的父群落。

资源群落的划分通常比较宽泛,它只是对博物馆的藏品做了最基本的分类,如在航空馆中“航空史”同“飞行器”两类完全没有交集的数字资源就可以划分为两个资源群落,但数字集合的划分非常细致,通常包含对数字藏品相关性质的明确限制,根据这些限制可以迅速的判断出一件藏品究竟是否属于该资源集合,从而避免了在藏品录入时模棱两可的局面。

需要注意的是,资源群落和资源集合的划分并不存在什么必须要遵守的规则或约束,因为这都是结构上的概念,引入这些概念的最终目的是使得整个博物馆层次结构清晰,藏品便于存放管理和展示。因此只要是合理且可行的划分方案,都是可以接受的。例如,一个概念在一个博物馆的方案中可能作为资源集合,但在其他情况下它也可能是一个子资源群落,如何选择需视实际情况而定。如上面的例子中,“民用飞机”也可以看作是又一个子资源群落,其下可以再按某种标准划分为若干个

资源集合,如按年代划分、按生产国划分或是按用途划分等。这样存放在各个资源集合中的藏品的限制就更多,划分也就更细致。资源群落和资源集合的划分层次越深,对藏品的分类就越细致,系统的层次结构就越深,而且从理论上说,整个博物馆的结构也会越清晰。但并不是越细致就越好,因为博物馆结构的划分最终必须为实际的需求服务,如当前的展示只对作为藏品的飞机用途(军用或民用)感兴趣,而不关注藏品的其他特性,那么再费力地去将藏品按年代或国别或其他的什么标准划分就是完全多余的,而且过于冗杂的层次结构也会为藏品的管理带来许多不必要的麻烦。因此在建设数字博物馆的时候,采用适当的层次结构是十分重要的。

正是由于划分上的灵活性,资源集合与资源群落之间并不存在一一对应的或是必然的从属关系,藏品也是如此,一个资源集合或藏品可能在当前情况下属于某个资源群落,但条件变化时又属于另外的资源群落。一个资源集合可能同时存在于多个资源群落中。如“德国生产的战斗机”这个资源集合可能既存在于子群落“战斗机”中,又存在于子群落“德国产飞机”中。同理,藏品也可能同时存在于不同的资源集合中。

2.2 元数据

按狭义理解,元数据(Metadata)是用来描述数据的数据,其可以反映数据的结构、数据的更新情况、数据之间的依赖关系、数据的质量等。按广义理解,元数据不仅能实现对数据的描述,而且能实现数据的操作和管理。狭义元数据和广义元数据是从不同的层次上对操作数据进行描述。因此,可以这样理解:元数据是关于数据的数据,是操纵数据的过程和应用程序的结构与意义的描述信息,其主要目标是提供数据资源的全面指南。

国际档案理事会的《电子文件指南》中给出的定义是:元数据是关于数据的数据。但是,在定义完元数据这个概念之后,其又解释:元数据是关于文件的背景信息和结构的数据。《澳大利亚联邦机构电子文件管理元数据标准》中的定义是:元数据是关于电子文件背景信息的著录信息。英国国家档案馆《电子文件管理指南》中给出的定义是:元数据是单一电子文件和文件组合的背景及其相互关系的结构化著录数据。

根据《电子文件归档与管理规范》这一国家标准(以下简称规范),电子文件的元数据被定义为:描述电子文件数据属性的数据,包括文件的格式、编排结构、硬件和软件环境、文件处理软件、字处理软件和图形处理软件、字符集等数据。由此可以看出,元数据实际上就是对数据进行著录而得到的著录信息,这些著录信息专门用于电子文件的管理,以保证电子文件的真实性、可靠性,元数据就是为了提高电子文件的凭证性而提出来的。创建元数据的过程就是对文件特征进行描述的过程。元数据不仅包括对文件本身的描述,还包括对文件的管理信息的描述。

元数据直译为关于数据的数据,一般属于计算机领域中的术语。这可以用一个简单的例子来说明:有一本书《政治经济学》,对它的书名、作者、出版社等信息做一

个简单的摘要,那么这个摘要信息就可以称作元数据。同样的,关于物质世界的和初始事物的简单(相对于源)再描述所得到信息都可以称作元数据,这就是元数据的一般定义。元数据最基本的用途就是管理数据,从而实现查询、阅读、交换和共享。

元数据包括核心元数据和扩展元数据。核心元数据(Core Element)是被普遍认同的一组最重要的、简明的描述数字资源的元数据元素,主要描述所涉及的所有数字资源对象的共有属性;扩展元数据(Extended Element)是核心元数据之外的主要描述数字资源对象内容的元数据元素。根据不同种类的数字资源,要确立不同的扩展数据元素集。限定词(Qualifiers)是在数据集或元数据集中经常应用的和详细定义项的一个规定的词汇集合。应用限定词汇能提高检索的准确性。

元数据的应用目的有四个方面:确认和检索、著录描述、资源管理、资源保护与长期保存。确认和检索(Discovery and Identification)主要致力于如何帮助人们检索和确认所需要的资源,数据元素往往限于作者、标题、主题、位置等简单信息,Dublin Core 是其典型代表。著录描述(Cataloging)用于对数据单元进行详细、全面的著录描述,数据元素囊括内容、载体、位置与获取方式、制作与利用方法、甚至相关的数据单元方面等,数据元素的数量往往比较多,MARC、GILS 和 FGDC/CSDGM 是这类元数据的典型代表。资源管理(Resource Administration)支持资源的存储和使用管理,数据元素除了比较全面的著录描述信息外,还往往包括权利管理(Rights/Privacy Management)、电子签名(Digital Signature)、资源评鉴(Seal of Approval/Rating)、使用管理(Access Management)、支付审计(Payment and Accounting)等方面的信息。资源保护与长期保存(Preservation and Archiving)支持对资源进行长期保存,数据元素除了对资源进行描述和确认外,往往还包括详细的格式信息、制作信息、保护条件、转换方式(Migration Methods)、保存责任等内容。

元数据的结构包括句法结构和语义结构。

(1) 句法结构

句法结构定义格式结构及其描述方式,例如,元素的分区分段组织、元素选取使用规则、元素描述方法(如 Dublin Core 采用 ISO/IEC 11179 标准)、元素结构描述方法(如 MARC 记录结构、SGML 结构、XML 结构)、结构语句描述语言(如 EBNF Notation)等。有时,句法结构需要指出元数据是否与所描述的数据对象捆绑在一起、或作为单独数据存在但以一定形式与数据对象链接。另外,句法结构还可描述与定义标准、DTD 结构和 Namespace 等的链接方式。

(2) 语义结构

语义结构定义元素的具体描述方法,例如,描述元素时所采用的标准、最佳实践(Best Practices)或自定义的描述要求(Instructions)。有些元数据格式本身定义了语义结构,而另外一些则由具体采用单位规定语义结构,例如 Dublin Core 建议日期元素采用 ISO 8601、资源类型采用 Dublin Core Types、数据格式可采用 MIME、标识符采用 URL 或 DOI 或 ISBN;又如 OhioLink 在使用 VRA Core 时要求主题元素

使用 A&AT、TGM 和 TGN,人名元素用 ULAN。

大学数字博物馆的元数据以 Dublin Core 为基础,提出了 9 个核心元数据,对部分元数据,加上了一个或数个限定词。9 个核心元数据包括: Title(题名)、Subject(主题)、Date(日期)、Format(格式)、Identifier(标识符)、Language(语种)、Relation(关联)、Copyright(版权)、Size(数据量)。

2.3 数字藏品的概念模型

同藏品与传统博物馆的关系一样,数字藏品是数字博物馆的核心,所有其他的一切都围绕数字藏品展开,可以说,没有数字藏品也就没有数字博物馆。在数字藏品中,其物质价值和文化价值已经融为一体,所体现的主要是其原生价值和再生价值。原生价值是数字藏品本身所蕴藏的,包括可能未被挖掘的价值,而再生价值则是数字藏品与公众的交流和交互过程中所产生的价值。

早期对数字藏品的评价使用实用主义标准,即如何更好地将藏品相关内容提供给用户,其中包括数字藏品的可用性、易访问性,另外还包括对预期用户需求的满足程度等。但是,数字藏品的建立仅仅给已知用户提供有效的信息服务是不够的,在当前的数字环境下,数字藏品要更好的被共享和利用,它的建设就要处于全球网络环境下而不应该仅仅在项目范围下进行考虑。基于这个思想,数字藏品的建立,需要考虑的因素有:可用性、重用性(重新利用和重新包装)、对服务的支持、对协同的支持和对知识产权控制的支持等。

在数字藏品的建立中包含了三个概念:数字藏品、数据文件和元数据。这里我们没有考虑数字博物馆的服务,但是有理由相信基于高质量的数字藏品、数据文件和元数据表示就可以建立各种高层次的数字博物馆服务。

数字藏品由数据文件(数字信息资源)组成,但数字藏品并不是数据文件的简单集合。数字藏品可以定义如下:它是数据文件有选择、有组织的集合,由元数据描述,并提供给用户相应的访问接口。其中数据文件可以分为两类:一类是各种不同格式的对象信息(如书籍、手稿、视频磁带、音频磁带)的数字化版本;另一类本身就是被机器所理解的数字信息(如数码照片、数据库信息等)。一个好的数据文件必须是对藏品的忠实体现,并尽可能的支持现有的和将来可能的各种应用。元数据被定义为关于数据的数据,或者关于信息的信息。其中有三种类型的元数据:有关内容的元数据、有关上下文的元数据和有关结构的元数据。

建立数字藏品需要遵循如下准则。

- (1) 需要制定一个明确的数字藏品开发规范,作为数字藏品建立的标准;
- (2) 数字藏品需要描述出藏品的主要特点(形式、访问控制、所有权等),还需要包含所有决定藏品真实性、整体性;
- (3) 数字藏品可以独立于项目经久存在和使用;
- (4) 数字藏品能被广泛的访问和利用;