

ICS 01.140.20

A 14

C A D A L 项 目 标 准

CADAL 10303—2012

复合数字对象描述和管理规范

Description and Management of Compound Digital Object Specification

第一稿

2012-05-08 发布

2012-05-09 实施

CADAL 项目管理中心 发 布

目 次

前言	181
引言	182
1 范围	183
2 规范性引用文件	183
3 术语和定义	183
3.1 元数据	183
3.2 数字对象	183
3.3 复合数字对象	184
3.4 元数据编码与传输规范	184
3.5 开放电子出版物结构规范	184
3.6 都柏林核心元数据	184
3.7 唯一标识符	184
4 复合数字对象描述和管理规范——元数据编码与传输规范(METS)	184
4.1 METS 文档及其基本结构	184
4.2 METS 各节的基本结构	185
5 复合数字对象描述和管理规范——开放电子出版物结构规范(OeBPS)	189
5.1 OeBPS 文件及其组成	189
5.2 OeBPS 包文件各节说明	190
参考文献	192
图 1 METS 基本结构图	185
图 2 METS 头结构图	185
图 3 描述型元数据结构图	186
图 4 管理型元数据结构图	187
图 5 文件节结构图	187
图 6 结构图的结构图	188
图 7 结构链接结构图	188
图 8 行为节结构图	188
图 9 OeBPS 包文件结构图	189

前 言

《CADAL 项目数字对象标准》分 5 个部分,由 5 个标准组成。

——第 1 部分:CADAL 10301—2012 数字对象唯一标识符规范。

——第 2 部分:CADAL 10302—2012 数字对象标识与命名规范。

——第 3 部分:CADAL 10303—2012 复合数字对象描述和管理规范。

——第 4 部分:CADAL 10304—2012 数字对象阅读规范。

——第 5 部分:CADAL 10305—2012 CADAL 数字对象唯一标识符解析系统应用规范。

本标准为第 3 部分。

本标准由以下几部分组成:

——第 1 部分:元数据编码与传输规范(Metadata Encoding and Transmission Standard, METS)。

——第 2 部分:开放电子出版物结构规范(OeBPS)。

本标准的制定依据美国数字图书馆联盟(Digital Library Federation)发布的《元数据编码与传输规范》(*Metadata Encoding and Transmission Standard, METS*),同时参照了美国国会图书馆开放电子图书论坛(Open eBook Forum, OeBF)2002 年发布的 OEBPS 1.2 规范。

本标准的某些内容可能涉及专利,本标准的发布机构不承担识别这些专利的责任。

本标准由浙江大学图书馆制定。

本标准由大学数字图书馆国际合作计划(CADAL)项目管理中心提出并归口。

本标准起草单位:浙江大学图书馆。

本标准起草人:叶杭庆、周云平、李懿、陈振英。

引 言

《复合数字对象描述和管理规范》是大学数字图书馆国际合作计划(China Academic Digital Associative Library, CADAL)标准规范集《数字对象标识标准规范集》的第 3 部分。根据 CADAL 标准规范招标指南、研制需求书研制。

本规范基于 CADAL 项目一期及二期数字图书馆建设的实际情况,参考国内外主要的数字对象的封装和管理标准,为 CADAL 项目数字图书馆建设提出合适的复合数字对象描述及管理的解决方案。

复合数字对象描述和管理规范

1 范围

本部分规定了大学数字图书馆国际合作计划(China Academic Digital Associative Library, CADAL)的数字图书馆复合数字对象的描述及管理方法。

本部分所指的数字对象,包括 CADAL 建设中加工制作生成的文本、图片、音频、视频等各种类型的数字对象。

本部分的开放电子出版结构规范(OeBPS 规范)适用于 CADAL 项目中电子图书文件(djvu 格式、PDF 格式等类型文件)的封装及描述。

本部分的元数据编码与传输规范(METS)适用于 CADAL 项目中由文本、图片、音频、视频等多种类型数字对象所构成的复合数字对象的描述及管理,以及采用 OeBPS 进行封装后的数字对象的管理。

2 规范性引用文件

下列文件对于本规范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规范。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本规范。

ISO 15836: 2009 *Dublin Core Metadata Element Set Version 1.1*
METS Schema & Documentation Version 1.9.1
Open eBook Publication Structure Version 1.2
Extensible Markup Language (XML) 1.1 (Second Edition)

3 术语和定义

下列术语和定义适用于本文件。

3.1 元数据 Metadata

元数据是关于数据的数据。它描述一个具体的资源对象并能对这个对象进行定位、管理,且有助于它的发现与获取。

3.2 数字对象 Digital Object

数字对象是数字资源管理系统中存储、管理和发布信息的基本实体,通常由标识符、数据流和元数据三要素组成。

3.3 复合数字对象 Compound Digital Object

复合数字对象是指封装了数据、元数据及相关服务的一个具有逻辑关系的数字集合。这里的数据是指各种类型数据文件(如文本数据、图像数据、音频数据、视频数据等)或各种格式的文件(如 Word 文档、PDF 文档、JPEG 图像、WAV 等)。

3.4 元数据编码与传输规范 Metadata Encoding and Transmission Standard, METS

元数据编码与传输规范由美国数字图书馆联盟(Digital Library Federation)发布,METS 是用于将数字对象的元数据编码和传输的标准,采用 W3C(World Wide Web Consortium)的 XML Schema 语言表达。

3.5 开放电子出版物结构规范 Open ebook Publication Structure, OeBPS

开放电子出版物结构规范由开放电子图书论坛(Open eBook Forum, OeBF)发布。OeBF 是一个居于电子书产业领导地位的国际性商业与标准组织,为规范电子图书的制作格式,实现电子图书阅读系统的互操作性,制定了开放电子出版物结构规范,以作为电子书内容描述的标准。

3.6 都柏林核心元数据 Dublin Core, DC

都柏林核心元数据是数字图书馆中使用的一组简单的包括 15 个“核心元素”的元数据元素集合,主要用于描述数字对象、馆藏管理和元数据交换。这 15 项元数据不仅适用于电子文献目录,而且适用于各类电子化的公务文档目录,产品、商品、藏品目录。

3.7 唯一标识符 Unique Identifier

唯一标识符是在特定命名域内按一定规则给予数字对象的唯一和永久名字。

4 复合数字对象描述和管理规范——元数据编码与传输规范(METS)

4.1 METS 文档及其基本结构

一个 METS 文档由描述复合数字对象的元数据、复合数字对象数据及其结构关系组合而成。

METS 文档采用 XML Schema 语言表达。

METS 的基本框架由 7 个节组成,7 个节名称如下:

- METS 头<metsHdr>;
- 描述型元数据<dmdSec>;
- 管理型元数据<amdSec>;
- 文件节<fileSec>;
- 结构图<structMap>;
- 结构链接<structLink>;
- 行为节<behaviorSec>。

上述 7 个节中,结构图部分为必备项,其他部分为可选项。

一个 METS 文档对应一个数字对象,CADAL 项目中,该数字对象可包含多种不同格式文件,也可包含一个或多个同一格式的文件。

该数字对象的相关元数据,包括描述型元数据和管理型元数据,都需编码到一个文档中。各类型元数据分别在描述型元数据<dmdSec>和管理型元数据<amdSec>节中予以记录。

文件节<fileSec>记录该数字对象所有数据文件的相关信息。

结构图<structMap>描述数字对象的完整框架,包括该结构中元素与相应的内容文件和元数据之间的逻辑关系。

METS 的基本结构如图 1 所示。

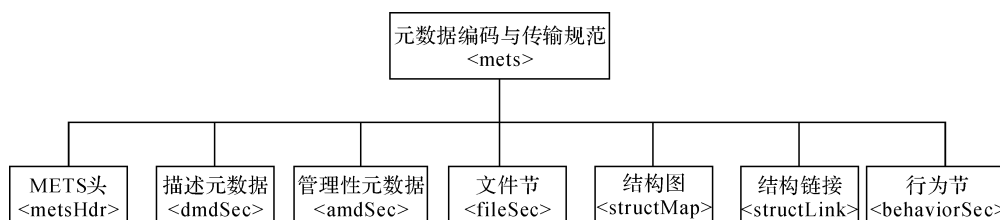


图 1 METS 基本结构图

4.2 METS 各节的基本结构

4.2.1 METS 头<metsHdr>的基本结构

METS 头记录关于 METS 文档本身的元数据。这些元数据包括 METS 文档的创建日期、上次修改日期以及该 METS 文档的状态。

METS 头采用元素<agent>记录一个或多个承担过与该 METS 文档有关的某种任务的机构,注明他们扮演的角色,再为其活动附上小说明之类。METS 头采用元素<metsHdr>的 OBJID 属性记录 METS 文档的主要标识符。

METS 头的组成元素包括元素<metsHdr>、元素<agent>、元素<altRecordID>、元素<name>和元素<note>。METS 头的结构如图 2 所示。

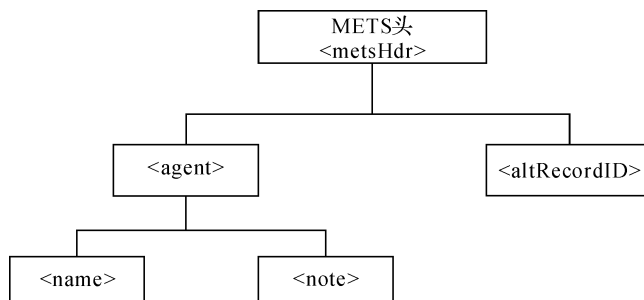


图 2 METS 头结构图

4.2.2 描述型元数据<dmdSec>的基本结构

描述型元数据是指向该 METS 文档外部的描述性元数据,也可以在该 METS 文档内部嵌入描述性元数据,或二者兼有。

描述型元数据由一个或多个<dmdSec>元素组成。每个<dmdSec>元素可以包含一个指向外部元数据的指针(元素<mdRef>),也可以包含内嵌元数据(嵌在元素<mdWrap>内),或二者兼备。描述性元素必须要有 ID 属性。此属性为每个<dmdSec>元素提供了唯一的内部名称,它可以应用于结构图中,把文档层次结构中指定的 div 与指定的<dmdSec>元素联系起来。

描述型元数据的组成元素包括元素<dmdSec>、元素<mdRef>、元素<mdWrap>、元素<xmlData>和元素<binData>。描述型元数据的结构如图 3 所示。

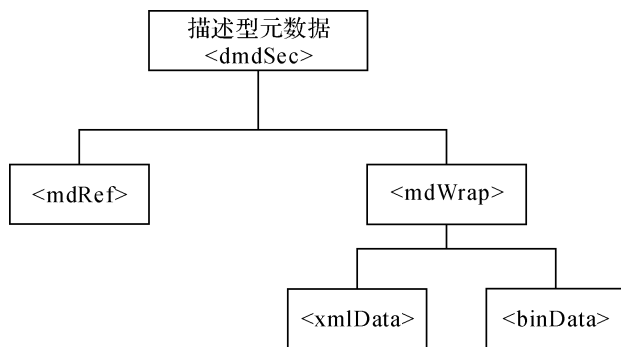


图 3 描述型元数据结构图

4.2.3 管理型元数据<amdSec>的基本结构

管理型元数据描述组成数字对象的文件及生成该对象的原始素材。

METS 文档中的管理型元数据有四种:①技术元数据(关于文件的创建、格式和使用特征的信息);②知识产权元数据(版权和许可信息);③来源元数据(关于该数字对象之模拟来源的描述型元数据和管理型元数据);④数字起源元数据(关于作品的原始数字化形态与作为数字对象的当前形态之间的关系信息)。

管理型元数据的组成元素包括元素<amdSec>、元素<techMD>、元素<rightsMD>、元素<sourceMD>、元素<digiprovMD>、元素<mdRef>、元素<mdWrap>、元素<xmlData>和元素<binData>。管理型元数据结构如图 4 所示。

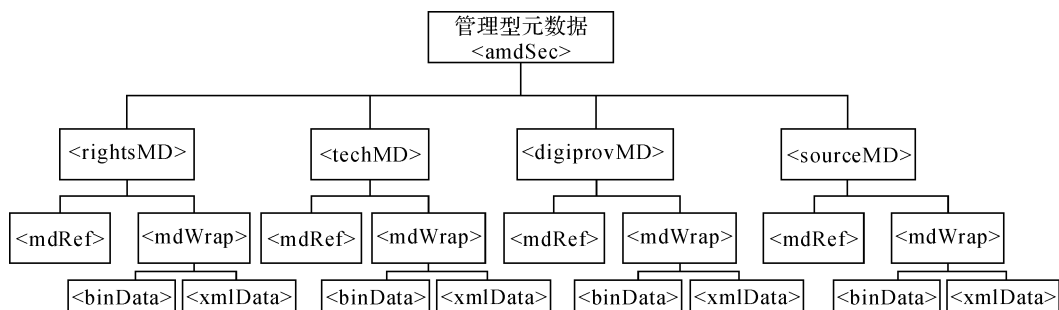


图 4 管理型元数据结构图

4.2.4 文件节<fileSec>的基本结构

文件节记录 CADAL 数字对象中所有数据文件的相关信息。数字对象是由若干内容文件组成的,这些文件要全部列在文件节中。

CADAL 采用 fileGrp 来表示数字对象,采用 file 来表示数字对象中的内容文件,一个 METS 文档中可以包含一个或多个 fileGrp,而每个 fileGrp 中可以包含一个至多个 file。在整个 fileSec 段中,只是列出了数字对象中的内容文件,而这些文件的组织结构是在 METS 的结构图节 structMap 中完成的。

文件节的组成元素包括元素<fileSec>、元素<fileGrp>、元素<file>、元素<FLocat>、元素<FContent>、元素<stream>、元素<transformFile>、元素<xmlData>和元素<binData>。文件节结构如图 5 所示。

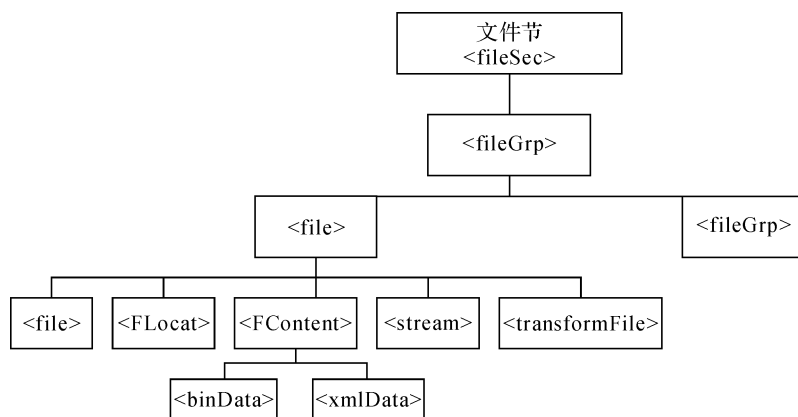


图 5 文件节结构图

4.2.5 结构图<structMap>的基本结构

结构图定义了数字对象的内部层次结构,并将该结构中的元素与相应的内容文件和元数据关联起来。

CADAL 数字对象的层次结构是通过子元素“div”组成的树状结构实现的。任何一个“div”通过 mptr 元素可以指向其他的 METS 文档,或通过 fptr 元素以及其他附属元素指向

一个单个的文件或一个文件组或一个单个文件的一部分。

结构图的组成元素包括元素<structMap>、元素<div>、元素<fptr>、元素<mptr>、元素<par>、元素<seq>、元素<area>、元素<seq>和元素<area>。结构图的结构如图 6 所示。

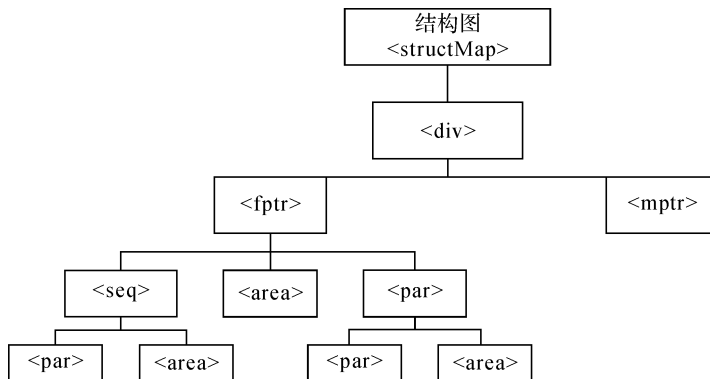


图 6 结构图的结构图

4.2.6 结构链接<structLink>基本结构

结构链接是记录 METS 结构图内文件之间存在的超链接。

结构链接的组成元素包括元素<structLink>和元素<smLink>。结构链接具体结构如图 7 所示。

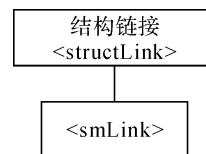


图 7 结构链接结构图

4.2.7 行为节<behaviorSec>的基本结构

行为节用于把可执行的行为与 METS 文档的内容联系起来。

行为节包含一个或多个<behavior>元素，每一个<behavior>元素都有接口定义元素，而接口定义元素表示了一系列行为的抽象定义，这些定义由一个特定的行为节表示。<behavior>元素还包含<mechanism>元素，它指向一个可执行代码模块，运行在接口定义中定义的行为。

行为节的组成元素包括元素<behaviorSec>、元素<behavior>、元素<interfaceDef>和元素<mechanism>。行为节具体结构如图 8 所示。

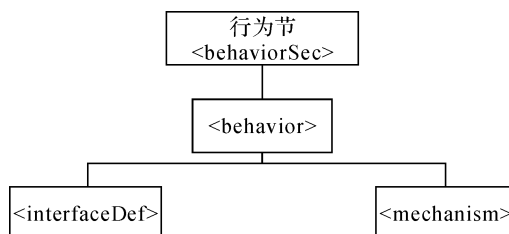


图 8 行为节结构图

5 复合数字对象描述和管理规范——开放电子出版物结构规范(OeBPS)

5.1 OeBPS 文件及其组成

一个 OeBPS 文件构成一本符合 OeBPS 规范的 CADAL 项目电子图书。

一个 OeBPS 文件由若干 OeBPS 文档、一个 OeBPS 包文件,以及其他文件(图像文件、CSS 文件等)组合而成。

OeBPS 文档采用 XHTML 编辑器制作。一本 CADAL 电子图书可以按章节制作形成多个章节的 XHTML 文档,也可按整本图书制作,具体视电子图书的内容而定。

OeBPS 包文件(OeBPS Package file)是不可或缺的,该包文件后缀名为“.opf”,语法为 XML,MIME 媒体类型为“text/xml”。OeBPS 包文件描述了电子图书的完整框架,包括 OeBPS 文档、图像文件和其他文件对象,以及它们之间的逻辑关系。

OeBPS 包文件的主要组成部分如下:

- 包标识符(package identity);
- 元数据(metadata);
- 文件列表(manifest);
- 阅览顺序(spine);
- 指南(guide);
- 导读(tour)。

OeBPS 包文件结构图如图 9 所示。

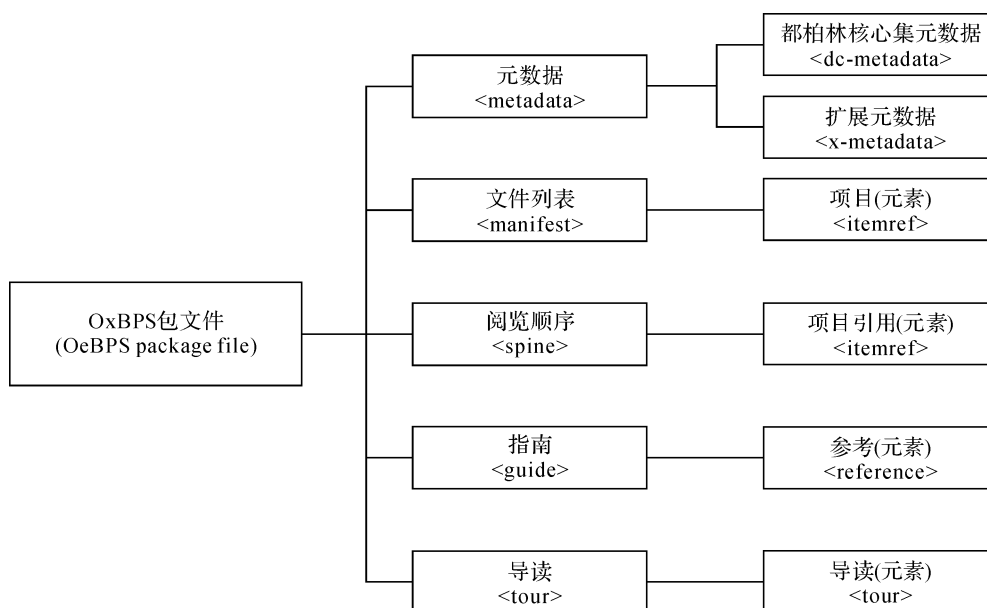


图 9 OeBPS 包文件结构图

5.2 OeBPS 包文件各节说明

5.2.1 包标识符 Package Identity

包标识符是电子图书作为一个整体的唯一识别符号。

包标识符采用 CADAL 数字对象唯一标识符(CADOI)。关于 CADOI 参见标准《CADAL 数字对象唯一标识符规范》，标准号:Q/CADAL 10301—2012。

5.2.2 元数据 Metadata

元数据定义的是电子图书各方面的信息。

元数据由都柏林核心集元数据(dc-metadata)和扩展元数据(x-metadata)两个子元数据集组成。

所使用的都柏林核心集十五项核心元素详见《大学数字图书馆国际合作计划基本元数据标准》中定义的都柏林核心集元素标准。

都柏林核心集元素命名格式为:以前缀“dc:”开头。“dc-metadata”必须包含“dc: Title”和“dc: Identifier”元素。都柏林核心集元数据的各项元素可以任何顺序排列。

扩展元数据“x-metadata”用于描述超出都柏林核心元素集范围的、自定义的元数据。

5.2.3 文件列表 Manifest

文件列表列举组成一本 CADAL 电子图书的所有文件。

文件列表可包含一个或多个“item”子元素。每个“item”元素描述一个文件、图像文件、样式表或者电子图书的其他组成部分。每个“item”元素必须有一个“id”属性、一个“href”属性及一个“media-type”属性。“id”属性表示所描述的文件 ID 号,“href”属性指示文件的相对路径,“media-type”属性指定文件的媒体类型。

5.2.4 阅览顺序 Spine

阅览顺序指定一本电子图书的线性阅读顺序。

阅览顺序由一个子元素“itemref”构成。

一本电子图书必须具备一个“Spine”元素。

5.2.5 导读 Tour

导读提供电子图书其他可选的阅读顺序,比如针对不同的阅读目的或不同的阅读群体等不同需求选择合适的阅读指引信息。

由于此导读功能与 CADAL 电子书的树状目录导航要求有差距,CADAL 电子书的包文件中不采用该元素。采用的解决方案是根据书本的目录结构录入目录信息,生成 catalog.xml 文件,该 xml 文件遵循 METS 标准。

5.2.6 指南 Guide

指南定义的是电子图书的结构参考信息。

指南可包含一个或多个“reference”子元素。“reference”元素列举电子图书的结构组成部分,包括电子图书的目次、前言、插图表单、序言和书目资料等。

每个“reference”元素必须有一个“href”属性及一个“type”属性。“href”属性引用一个文件列表中的电子图书文档,可以包含一个字段标识符,用于指向文件保存地址。“type”属性用于描述“href”属性引用到的电子图书组成部分。

参 考 文 献

- [1] 复合数字对象描述规范[S]. 中国高等教育文献保障系统管理中心,2011.
- [2] 郭迎九,林闯,尹浩. 复合数字对象及版权保护权利推理机制研究[J]. 计算机工程与应用,2008,44(31):25-27,32.
- [3] 何朝晖,王波,朱强. 国外复合数字对象管理研究摘要[J]. 2005,(5):6-14.
- [4] 李蓓. 数字化图书馆资源仓库的基础——METS[J]. 2004,22(11):1375-1387.
- [5] 李春旺,张晓林. 复合数字对象研究[J]. 情报学报. 2004,(4):444-451.
- [6] 马蕾. 元数据及其封装标准 METS 研究. 情报技术[J]. 2002,(2):56-57.
- [7] 数字图书馆联盟(Digital Library Federation). METS Schema & Documentation Version 1.9.1[S/OL]. <http://www.loc.gov/standards/mets/docs/mets.v1-9.htm>[Access 2012-3-21].
- [8] 唐光前. 开放式电子图书出版结构(OEBPS)规范的解析[J]. 图书情报工作,2002(5):58-62.
- [9] 薛崧,郑龔. 基于 OEB 规范电子书的结构及优化[J]. 现代图书情报技术,2004,117(11):10-13.
- [10] 张铮,李蓓. 元数据家族中的新成员——MODS 和 METS[J]. 2005,18(7):3.
- [11] 曾婷,张成昱,姜爱蓉. 电子图书内容格式 OEBPS 的研究[J]. 现代图书情报技术,2003,102(4):24-26.
- [12] Michael L. Nelson, Herbert Van de Sompel. International Journal on Digital Libraries[J]. 2006,6(2):113-114.