

ICS 01.140.20

A 14

C A D A L 项 目 标 准

CADAL 20802—2012

数字对象交换协议规范

Digital Object Exchange Agreement Standard

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	57
引言	58
1 范围	59
2 规范性引用文件	59
3 术语和定义	59
3.1 开放电子书	59
3.2 电子发行本	59
3.3 便携文件格式	59
3.4 都柏林核心元数据	60
3.5 标签图像文件格式	60
3.6 DjVu 图像文件	60
4 数字对象结构及存储格式规范	60
4.1 数字对象组织结构	60
4.2 数字对象存储格式	61
参考文献	63
图 1 书的存放目录结构	62
表 1 <dc-metadata>元素	61
表 2 <x-metadata>元素	61

前 言

《数字资源服务协议规范和接口标准集》包括以下 4 个方面的内容：

- 第 1 部分：元数据互操作协议规范；
- 第 2 部分：数字对象交换协议规范；
- 第 3 部分：资源检索协议规范；
- 第 4 部分：服务集成标准规范。

本标准是其中的第 2 部分。

本部分的制定依据了标准化工作导则第 1 部分(GB/T 1.1—2009)。

本部分由数字图书馆教育部工程研究中心提出。

本标准由 CADAL 项目管理中心归口。

本标准起草单位：数字图书馆教育部工程研究中心。

本标准主要起草人：张鹏、张寅。

引 言

随着信息化社会的快速发展,很多传统的文本资源因为传播、阅读、保存等因素的限制往往不能适应现代社会需求,因此逐渐被更易传播、阅读、存储的数字化资源取代。为了保证数字资源易于传播和交换,需要制定一套数字对象交换协议规范。然而,目前我国的数字图书馆建设还未形成一个广泛的数字资源交换协议规范。本规范是基于 CADAL 数字图书馆对于数字资源交换的需求而编制的。

本标准的编制目的是为了给数字图书馆的数字对象交换提供一个参照标准,从而提出数字对象组织、存储格式规范。本标准参照了业界流行且成熟的相关开发规范和标准,有利于迅速地推广 CADAL 数字图书馆的建设成果。

数字对象交换协议规范

1 范围

本标准确定了数字图书馆数字对象交换协议规范。
本标准适用的数字资源对象包括印刷文献的数字化衍生物。
本规范规定了各类数字对象交换协议规范。

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。

Dublin Core Metadata Element Set
ISO 32000
GB/T 23286.1—2009

3 术语和定义

3.1 开放电子书 Open eBook 缩写: OEB

开放电子书是用于格式化和包装电子书的一种行业标准。OEB 基于 XML, 定义了电子出版物的文本如何被标记, 以及一本 eBook 的各部分(封面、目录、正文、说明、索引等)应如何包裹在一起。

3.2 电子发行本 Electronic Publication 缩写: EPUB

电子发行本是一种电子图书标准, 由国际数位出版论坛(International Digital Publishing Forum, IDPF)提出; 其中包括 3 种文件格式标准, 文件的存档名为 .epub, 这个格式已取代了早期的 Open eBook 开放电子书标准。EPUB 格式的内部结构域 OEB 基本一致, 可以方便地相互转换。

3.3 便携文件格式 Portable Document Format 缩写: PDF

便携文件格式是一种电子文件格式, 这种文件格式与操作系统平台无关, 主要便于打印和跨平台使用。

3.4 都柏林核心元数据 Dublin Core 缩写: 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。

3.5 标签图像文件格式 Tagged Image File Format 缩写: TIFF

标签图像文件格式是一种主要用来存储包括照片和艺术图在内的图像的文件格式。它最初由 Aldus 公司与微软公司一起为 PostScript 打印开发。

TIFF 文件格式适用于在应用程序之间和计算机平台之间的交换文件,它的出现使得图像数据交换变得简单。

3.6 DjVu 图像文件

DjVu 图像文件是主要用于存储扫描的文档。这种格式的特色包括图像分层、渐进载入、算术编码、对二进制图像进行有损压缩,从而以较小的空间,存放高质量的可读图像。渐进载入使得 DjVu 适合应用于因特网。DjVu 对于大部分的扫描文档,表现都优于 PDF,故被作为 PDF 的替代品来进行推广。这种格式已经在文件共享网络中,被广泛使用于分发数学书籍。

4 数字对象结构及存储格式规范

4.1 数字对象组织结构

目前,国际上流行的数字对象组织结构有 OEB、EPUB、PDF 等。其中,OEB 和 EPUB 是针对电子书制定的标准,EPUB 是 OEB 的替代版,两者可以方便地转换。

本标准规定使用 OEB 结构组织数字资源。OEB 基于 XML,定义了电子出版物的文本如何被标记,以及电子书的各部分(封面、目录、正文、说明、索引等)应如何包裹在一起。OEB 实际上是一个综合体,可以说是 HTML 4.0、XML 1.0、Dublin Core 与 CSS1.0 的子集,主要是让编辑与系统商花最少的力气加入 eBook 行列,因为以上语言都是它们所熟悉的。OEB 2.0 以 XML 为主干,严谨地把编排格式与资料内容分开。

OEB 格式的所有描述信息存放于一份后缀名为 .opf 的文件中。OPF 文档是 EPUB 电子书的核心文件,且是一个标准的 XML 文件;依据 OPF 规范,主要由五个部分组成:

- `<metadata>`: 元数据信息,由 `<dc-metadata>` 和 `<x-metadata>` 两个子元素组成。
 - `<dc-metadata>` 元素构成采用 DC 的 15 项核心元素,见表 1;
 - `<x-metadata>` 是扩展元素,如果有些信息在上述元素中无法描述,则在此元素中进行扩展,见表 2。
- `<manifest>`: 文件列表,由于列出 OEBPS 文档及相关的文档,由一个子元素构成:
 - `<item id="" href="" media-type="">`,该元素由三个属性构成: id 表示文件的 ID 号,href 表示文件的相对路径,media-type 表示文件的媒体类型。

- `< spine >`: 书脊, 其主要功能是提供书籍的线性阅读次序。由一个子元素构成: `< itemref idref="" >`, 由一个属性构成: `idref` 是参照 `manifest` 列出的 ID。
- `< guide >`: 指南, 依次列出电子书的特定页面, 例如封面、目录、序言等, 属性值指向文件保存地址。
- `< tour >`: 导读。可以根据不同的读者水平或者阅读目的, 按一定次序, 选择电子书中的部分页面组成导读。

OPF 文件的实例可参见附录中 A. 1 部分。

表 1 `< dc-metadata >` 元素

元素名称	意义
title	题名
creator	主要责任者
subject	主题词或关键词
description	内容
contributor	贡献者或其他次要责任者
date	日期
type	类型
format	格式
identifier	标识符
source	来源
language	语种
relation	相关
coverage	覆盖范围
rights	权限描述

表 2 `< x-metadata >` 元素

元素名称	意义
edition	版本信息
marc	MARC 记录
degree	学位

4.2 数字对象存储格式

本标准规定对于从纸质书中扫描的书页, 原始格式采用 TIFF 格式。标签图像文件格式(tag image file format, TIFF)是一种主要用来存储包括照片和艺术图在内的图像的文件格式。TIFF 文件格式适用于在应用程序之间和计算机平台之间的交换文件, 它的出现使图像数据交换变得简单。

本标准规定使用 DjVu 格式存储扫描的电子书文档，并提供阅览服务。这种格式的特色包括图像分层、渐进载入、算术编码、对二进制图像进行有损压缩，从而以较小的空间，存放高质量的可读图像。此外，DjVu 支持多种格式文件的相互转换，例如 PDF、DWG、JPG、TIFF、BMP 和 DOC。

对于数字对象而言，首先使用 TIFF 图片格式存储原始的扫描文件，之后转换成 DjVu 图片格式对外提供阅览服务。此外，对于某些需要高清显示的资源，也可以直接使用 TIFF 提供阅览服务。尽管 TIFF 格式具有更好的图像质量，但是占用存储空间大、解析复杂，尽量少使用。另外，需要把 TIFF 图片文件单独存放至外部存储设备以长期保存。

对于一本电子书，每一页纸张经过扫描后存放为一份 DjVu 格式的文档，另外还需要人工生成元数据文件。opf。一本完整的电子书的目录格式如图 1 所示。

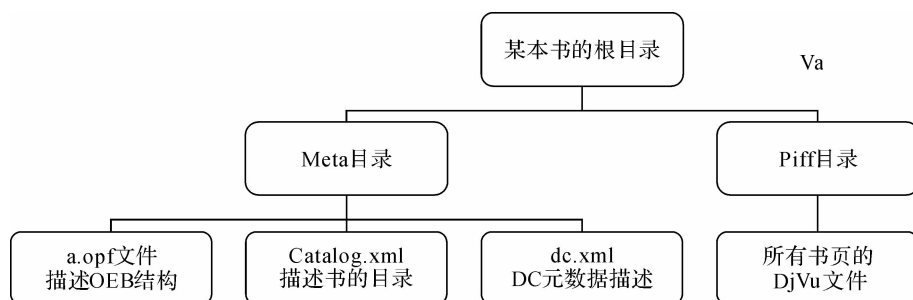


图 1 书的存放目录结构

参 考 文 献

[1] Dublin Core Metadata Element Set 都柏林核心元数据元素集[S/OL]. <http://dublincore.org/documents/dces/> 2012.

[2] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会. GB/T 1.1—2009 中华人民共和国国家标准[S]. 北京:中国标准出版社,2009.

[3] 侯瑞芳. OEBC(开放式电子图书)格式标准与安全性研究[J/OL]. 新世纪图书馆. [2003-03]. <http://www.cnki.com.cn/Article/CJFDTotol-JSTS200303005.htm>.

[4] EPUB electronic publication 数位版权管理[S/OL]. <http://idpf.org/epub/30>. 2007.