

ICS 01.140.20

A 14

C A D A L 项 目 标 准

CADAL 10101—2013

数字对象采集规范

Specification of Digital Resource Collection

第二稿

2013-10-15

2013-10-15 发布

2013-10-16 实施

CADAL 项目管理中心 发 布

目 次

前言	7
引言	8
1 范围	9
2 规范性引用文件	9
3 术语与定义	9
3.1 数字对象	9
3.2 版本	9
3.3 原生电子数据	9
3.4 OCR	9
3.5 标识符	10
3.6 原始图像	10
3.7 原始音频	10
3.8 原始视频	10
3.9 DC元数据	10
3.10 目录结构	10
3.11 资源结构	10
4 总则	10
4.1 数字对象采集基本要求	10
4.2 数字对象采集管理	10
4.3 数字对象采集版本选择	10
4.4 数字对象采集内容	11
5 数字对象采集要素	11
5.1 图书与期刊	11
5.2 报纸	11
5.3 文档资料	11
5.4 图片	11
5.5 古籍	12
5.6 音频	12
5.7 视频	12
参考文献	13

前 言

《CADAL 数字对象加工规范》分成 4 个部分，由 13 个标准组成。

——第 1 部分：CADAL 10101—2013 数字对象采集规范。

——第 2 部分：CADAL 10102—2013 数字对象制作基本流程规范，这部分根据加工对象的不同又分成 8 个子规范。

- 第 1 子规范：CADAL 10103.1—2013 图书期刊数字对象制作规范。
CADAL 10103.2—2013 Book Digitalization Specification。
- 第 2 子规范：CADAL 10104—2013 报纸数字对象制作规范。
- 第 3 子规范：CADAL 10105—2013 文档数字对象制作规范。
- 第 4 子规范：CADAL 10106—2013 图片数字对象制作规范。
- 第 5 子规范：CADAL 10107—2013 古籍数字对象制作规范。
- 第 6 子规范：CADAL 10109—2013 视频数字对象制作规范。
- 第 7 子规范：CADAL 10110—2012 音频数据加工标准与操作规范。
- 第 8 子规范：CADAL 10227—2012 缩微胶片数字化加工标准与操作规范。

——第 3 部分：CADAL 10111—2013 数字内容编码与内容标记规范。

——第 4 部分：CADAL 10112—2013 数字对象加工与应用等级标准。

本标准为第 1 部分。

《CADAL 数字对象加工规范》代替 CADAL 项目一期制定的《数字化文本加工规范草案》。

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：深圳市点通数据有限公司、杭州中元数据科技有限公司、浙江大学图书馆。

本标准主要起草人：郑传双、周小芳、薛霏。

引 言

数字对象加工规范是数字图书馆资源建设的基础,制定数字对象加工规范的目的是让数字图书馆资源建设单位,在数字对象采集、加工、封装、存储等环节中有统一的规格和操作方法,保持数字资源的格式与内容形式的一致性。

《CADAL 数字对象加工规范》是 CADAL(China Academic Digital Associative Library)项目关于数字对象加工的规范集,是 CADAL 项目数字对象加工必须遵从的基础性企业标准。

《数字对象采集规范》的基本目的是保证 CADAL 项目资源采集质量,主要解决:

- (1) 界定资源采集范围;
- (2) 明确采集内容;
- (3) 明确资源版本管理的基本方案。

数字对象采集规范

1 范围

本部分规定了 CADAL 项目认可的数字对象采集范围、方式、版本甄选等。
本部分适用于 CADAL 项目可加工资源的收集与甄选过程。

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 3469	文献类型和文献载体代码
GB/T 3792.1	文献著录 第1部分 总则
CADAL 10301—2012	数字对象唯一标识符规范
CADAL 10302—2012	数字对象标识与命名规范

3 术语与定义

下列术语和定义适用于本部分。

3.1 数字对象 Digital Object

数字对象指一组通过数字化加工得到的，描述一个特定的实物资源的，可存储于计算机并可利用计算机技术进行再现的数据集合。

3.2 版本 Edition

版本指采用直接接触、照相、复制或其他方法，源于同一原始输入信息制作，并且由同一机构或同一组机构或个人发行的一种资源的全部复本。

3.3 原生电子数据 Original Electronic Copy

原生电子数据指在信息记录到文献载体前的、可编辑的计算机文件。

3.4 OCR Optical Character Recognize

OCR 指利用计算机软件，将栅格化字符点阵信息转换成计算机字符编码的过程。

3.5 标识符 Identifier

标识符指由管理机构按统一规范给定的编号, 简称为 ID。

3.6 原始图像 Original Image

原始图像指通过初始扫描、摄影、转换等手段直接获取的图像文件。

3.7 原始音频 Original Audio

原始音频指通过初始录制、转换等手段直接获取的音频文件。

3.8 原始视频 Original Video

原始视频指通过初始摄像、转换等手段直接获取的视频文件。

3.9 DC 元数据 DC Meta

DC 元数据指 Dublin Core 元数据。

3.10 目录结构 Catalog

目录结构指符合 XML 的 METS 规范的目录结构信息, 包括目录节点名称、链接指向的页面文件编号。

3.11 资源结构 Guide

资源结构指将资源各部分内容组合成一个整体的内部结构关系, 包括各资源片断间的并列、包含、从属、接续、引用关系等。

4 总则

4.1 数字对象采集基本要求

采集数字对象的目的是为了资源信息的保存和在线阅读, 所以要求:

——采集的数字对象应有实物资源支持, 或是由官方提供的原生电子数据;

——采集的数字对象应基本完整。

4.2 数字对象采集管理

同一版本的资源只进行一次数字对象采集。

数字对象应按统一的规范设定资源标识符, 标识符需要符合 Q/CADAL 10302—2012《数字对象标识与命名规范》和 Q/CADAL 10301—2012《数字对象唯一标识符规范》。

4.3 数字对象采集版本选择

优先选择来源权威、完整性好、版次新的资源进行数字对象加工。

对不同来源的原始资源，按以下原则进行版本认定与选择：

- 加印、影印、转拍、复印、转贴等直接复制的资源，应视为同一版本，应优先选择最新版本；
- 对重印、修订版、剪辑后的版本，应视为不同版本；
- 对流通、保有过程中附着了重要的、系统的收藏、批注类信息的，应视为不同版本；
- 对不同来源的手抄本应视为不同版本。

4.4 数字对象采集内容

数字对象采集内容包括：

- 原始图像文件、原始音频文件、原始视频文件；
- 数字对象的 DC 元数据信息；
- 数字对象的目录结构信息；
- 数字对象的资源结构信息。

5 数字对象采集要素

5.1 图书与期刊

- 原书页面原始图像文件；
- 用于检索和数字对象管理的 DC 元数据文件；
- 用于阅读导航的目录结构文件。

5.2 报纸

- 报纸原版面原始图像文件；
- 用于检索和数字对象管理的 DC 元数据文件；
- 用于阅读导航的目录结构文件。

5.3 文档资料

- 原文档页面原始图像文件；
- 文档内容清单；
- 用于检索和数字对象管理的 DC 元数据文件；
- 用于阅读导航的目录结构文件。

5.4 图片

- 原图片原始图像文件；
- 图片相关说明文件；
- 用于检索和数字对象管理的 DC 元数据文件；
- 用于阅读导航的目录结构文件。

CADAL 项目标准规范汇编(一)

CADAL 10101—2013

5.5 古籍

- 原书页面原始图像文件；
- 用于检索和数字对象管理的 DC 元数据文件；
- 用于阅读导航的目录结构文件。

5.6 音频

- 原始音频文件；
- 用于检索和数字对象管理的 DC 元数据文件；
- 用于阅读导航的目录结构文件。

5.7 视频

- 原始视频文件；
- 用于检索和数字对象管理的 DC 元数据文件；
- 用于阅读导航的目录结构文件。

参 考 文 献

[1] 国际图书馆协会和机构. 国际标准书目著录(统一版). 顾犇译. 北京: 北京图书馆出版社, 2008.

[2] The Library of Congress Standards. Metadata Encoding & Transmission Standard. 2012-03-22. [2013-10-15]. <http://www.loc.gov/standards/mets/mets-schemadocs.html>.

[3] LIZARD TECH, INC. DjVu Technology Primer. 2004-11. [2013-10-15]. http://djvu.org/docs/DjVu_Tech_Primer.djvu.

[4] LIZARD TECH, INC. Lizardtech DjVu Reference v3. 2005-11. [2013-10-15]. <http://djvu.org/docs/DjVu3Spec.djvu>.