ICS 01.140.20
A 14

C A D A L 项 目 标 准

CADAL 20506—2012

数字资源压缩和索引规范

Standard Specification for Digital Resource Compression and Index

第一稿

2012-05-08

2012-05-08 发布　　2012-05-09 实施

CADAL 项目管理中心　　发　布

金旭照排　五校样　开本:787×1092　成品:185mm×260mm　版芯:39 字×39 行　行距:＊2　正文字体号:5SS　2014.12.4

目　　次

金旭照排　五校样　开本:787×1092　成品:185mm×260mm　版芯:39 字×39 行　行距:＊2　正文字体号:5SS　2014.12.4

前　言

《CADAL 项目数字对象存储标准》分为 6 个部分，由 6 个标准组成。

——第 1 部分：CADAL 20501—2012 通用数字资源存储标准。

——第 2 部分：CADAL 20502—2012 数字对象文本类型存储标准。

——第 3 部分：CADAL 20503—2012 数字对象图像类型存储标准。

——第 4 部分：CADAL 20504—2012 数字对象音频类型存储标准。

——第 5 部分：CADAL 20505—2012 数字对象视频类型存储标准。

——第 6 部分：CADAL 20506—2012 数字资源压缩和索引规范。

本部分为《CADAL 项目数字对象存储标准》的第 6 部分。

本标准制定了通用数字资源的存储要求。

本标准是大学数字图书馆国际合作计划(CADAL)项目二期研制成果之一。

本标准制定 CADAL 项目通用数字资源存储标准，CADAL 项目的各种类型数字资源存储标准应至少满足本标准提出的要求。

本标准由大学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。

本标准的起草单位：浙江理工大学图书馆。

本标准的主要起草人：刘翔、黄志强、施干卫。

金旭照排　五校样　开本:787×1092　成品:185mm×260mm　版芯:39 字×39 行　行距:*2　正文字体号:5SS　2014.12.4

引　言

本标准针对 CADAL 项目的实际情况，以制定 CADAL 项目数字资源存储通用标准为目标，考虑了 CADAL 项目“数字对象加工标准规范集”、“数字对象元数据标准规范集”、“数字对象标识标准规范集”等子项目相关成果的关系，根据 CADAL 项目“数字资源存储标准集”的要求制定。

本标准在制定过程中参考了全球网络存储工业协会(Storage Network Industry Association，SNIA)、国家数字图书馆工程等国内外同类型项目与单位相关公开文档，重点借鉴了美国国会图书馆的 METS 标准，同时还参考了 MooseFS 分布式文件系统相关文档。

本标准针对数字资源压缩和索引的共性进行规范。本标准基于 CADAL 项目一期及二期当前存储情况及今后发展的需要，优先考虑 CADAL 项目现有的资源基础，有选择地借鉴国内外数字资源存储的经验。

数字资源压缩和索引是数字图书馆中的资源保存的重要手段之一，目前尚未有统一存储标准。因此本标准主要在 CADAL 数字对象视频类型存储标准、CADAL 数字对象图像类型存储标准、CADAL 数字对象文本类型存储标准、CADAL 数字对象音频类型存储标准和 CADAL 通用数字资源存储标准的基础上，提出适合 CADAL 项目的数字资源压缩和索引标准，从而为建设 CADAL 项目提供一个科学、合理、可行的数据格式标准规范。

金旭照排　五校样　开本:787×1092　成品:185mm×260mm　版芯:39 字×39 行　行距:*2　正文字体号:5SS　2014.12.4

数字资源压缩和索引规范

1 范围

本部分规定了数字资源的压缩和索引规范，包括通用数字资源与专门数字对象(分为文本类型数字对象、图像类型数字对象、音频类型数字对象以及视频类型数字对象)的压缩和索引细则，供 CADAL 项目数字资源存储使用。

2 数字资源保存级别

为了满足不同来源数字资源在生命周期的不同阶段的使用需要，把数字资源分为长期保持级、复制加工级和发布服务级。

长期保存级：作为档案保存及出版用的数字资源，不用于发布服务，可作为格式转换和复制的母本。

复制加工级：有较高质量的数字资源，是加工复制各种精度、大小的数字资源的母本文件，供专家、合作伙伴及专门组织成员通过网络有限权限的访问。

发布服务级：供普通读者通过网络访问，可进行浏览、下载、复制的数字资源。

3 数字资源压缩

3.1 长期保存级

对于长期保存级别的数字资源由于需要具有格式开放透明、持续可解释；不包含加密协议，也不包含加密选项；可转换，支持其他格式转换为长期保存格式，因此，对长期保存级别的数字资源不采取压缩格式。

3.2 复制加工级

——文本数据转换为 PDF 格式，适度压缩。

——图像数据的压缩见表 1。

金旭照排　五校样　开本：787×1092　成品：185mm×260mm　版芯：39 字×39 行　行距：＊2　正文字体号：5SS　2014.12.4

表 1　复制加工级图像数据的压缩

文献类型	图像分辨率(dpi)	色彩位深	允许的编辑加工	文件格式压缩算法
普通图书类	300	黑白 8 位 24 位	锐化、裁切、纠偏、去噪，色彩管理	JPEG2000 或 PDF
古籍类	300～600	8 位 24 位 更高	锐化、裁切、纠偏、去噪，色彩管理	JPEG2000 无损压缩
手稿类	300～600	8 位 24 位 更高	锐化、裁切纠偏、去噪；成比例扩展，最低限度地调整彩色和色调	JPEG2000 无损压缩

——音频数据的压缩见表 2。

表 2　复制加工级音频数据的压缩

比特率/采样率	量化级	通道数	推荐压缩格式
64k～320kbps 44.1kHz	16bit	双声道	mp3 AAC WMA

——视频数据的压缩见表 3。

表 3　复制加工级视频数据的压缩

视频速率/kbps	音频速率/K	帧速/fps	音频采样	建议编码
2000	384	25/30	立体声 48kHz	MPEG 4 编码的 WMV、FLV 或 RM

3.3　发布服务级

——文本数据转换为 PDF 格式，高度压缩。

——图像数据的压缩见表 4。

表 4　发布服务级图像数据的压缩

文献类型	图像分辨率/dpi	色彩位深	允许的编辑加工	文件格式压缩算法
普通图书类	72～96	黑白 8 位 24 位	锐化、裁切、纠偏、去噪，色彩管理	JPEG2000 或 PDF
古籍类	72～96	8 位 24 位 更高	锐化、裁切、纠偏、去噪，色彩管理	JPEG2000 无损压缩
手稿类	72～96	8 位 24 位 更高	锐化、裁切纠偏、去噪；成比例扩展，最低限度地调整彩色和色调	JPEG2000 无损压缩

金旭照排　五校样　开本:787×1092　成品:185mm×260mm　版芯:39 字×39 行　行距:*2　正文字体号:5SS　2014.12.4

——音频数据的压缩见表 5。

表 5　发布服务级音频数据的压缩

比特率/采样率	量化级	通道数	推荐压缩格式
(16～46)kbps 22.05kHz	16bit	双声道/单声道	mp3 AAC WMA

视频数据的压缩见表 6。

表 6　发布服务级视频数据的压缩

视频速率/kbps	音频速率/K	帧速/fps	音频采样	建议编码
46	16	15	立体声 44.1kHz	MPEG 4 编码的 WMV、FLV 或 RM

4　元数据索引

在数字资源存储标准中采用 XML 存储描述数字资源对象的元数据，使用文件系统存储 XML 文件是一种长期保存的方法，但文件系统在访问并发性和查询效率上低下，需要对元数据进行索引提高访问效率。使用 Native 数据库方式(见表 7)来存储 XML 元数据文档，建立元数据索引提高资源查询获取效率。

表 7　Native 数据库方式索引

数据库类型	索引类型	查询方法	数据库系统
Native XML	标准索引 文本索引 结构索引	XPath XQuery	OrientX Tamino Timber

金旭照排　五校样　开本:787×1092　成品:185mm×260mm　版芯:39 字×39 行　行距:＊2　正文字体号:5SS　2014.12.4

参 考 文 献

[1] ADAL 基本元数据标准与扩展集标准(草). http: //www. cadal. cn.

[2] CADAL 音频数字对象制作规范(草). http: //www. cadal. cn.

[3] CADAL 音频资料数字化的元数据标准规范(草). http: //www. cadal. cn.

[4] Dublin Core. http: //www. dublincore. org/.

[5] Metadata Object Description Schema (MODS). http: //www. loc. gov/standards/mods/.

[6] MARCXML MARC 21 Schema (MARCXML). http: //www. loc. gov/standards/marcxml/.

[7] METS. http: //www. loc. gov/mets/.

[8] MIX. http: //www. loc. gov/mix/.

[9] PREMIS. http: //www. loc. gov/standards/premis.

[10] TextMD. http: //www. loc. gov/standards/textMD/.

[11] AudioMD/VideoMD. http: //www. loc. gov/standards/amdvmd/index. html.

[12] SNIA. http: //www. snia. org/.

金旭照排 五校样 开本:787×1092 成品:185mm×260mm 版芯:39 字×39 行 行距:*2 正文字体号:5SS 2014.12.4