

数字视频资源元数据及描述

段明莲¹, 李燕²

(1.北京大学信息管理系, 北京 100871; 2.冶金工业信息标准研究院, 北京 100730)

摘要: 本文详细阐述视频资源的定义、类型、特征, 介绍国内有代表性的数字视频资源元数据及编目规则, 指出其现存问题及解决问题的对策。文本认为数字视频资源元数据及编目规则是创建视频资源数据库的基础, 编目标准化是共建共享数字视频资源数据库的必由之路。

关键词: 数字视频资源; 元数据; 编目规则

中图分类号: G254.364

DOI: 10.3772/j.issn.1673-2286.2016.12.003

1 视频资源的定义与类型

视频资源 (video resource) 是指用于存储活动或静止图像的媒体, 所记录的视频需用特定的回放设备 (如视频播放器、DVD播放机等) 播放, 包括用数字信号和模拟信号存储的视频资源, 但不包括静态图像资源。

根据不同的划分标准, 视频资源可分为不同的类型。按视频信号划分, 视频资源类型包括用模拟信号存储的视频资源和用数字信号存储的数字视频资源; 按视频资源的存储载体划分, 视频资源类型包括实体视频资源 (如DVD、VCD、录像带等) 和网络视频资源; 按内容划分, 视频资源类型包括新闻、讲座、访谈、电影、电视剧、舞蹈、动漫、相声、小品、杂技、戏剧、戏曲、音乐会、综艺节目、体育节目等。

2 视频资源的特征

2.1 结构的特殊性

视频资源是集语言、音响、图像或字幕于一体的信息资源, 其结构具有特殊性。无论电影还是录像均是按顺序承载一系列图像, 在快速按顺序投影时会产生运动错觉的原理来呈现画面。视频资源物理结构的最小单位

是帧, 由帧构成镜头, 由镜头构成场景, 由场景构成片段, 由片段构成节目, 即一部完整的录像资料或电影。

2.2 特征的独特性

视频资源是集图、文、声、像于一体的信息资源, 物理特征复杂。视频资源通常包括时长、入点、出点、画面宽高比、色彩、分辨率、声道格式、制式、视频数据码率、视频编码格式、视频取样格式、音频编码格式、音频数据码率、音频采样频率、音频位深度等物理特征, 数字视频资源尤为如此。

2.3 题名的层次性

视频资源题名层次多, 尤其是影视节目题名的层次。如电视连续剧《三国演义》由五部分84集组成, 不仅有系列题名《三国演义》还有分部总题名 (如第一部《群雄逐鹿》、第二部《赤壁逐鹿》等), 甚至在各部下还载有分集题名 (如在第一部《群雄逐鹿》中包括第1集《桃园三结义》、第2集《十常侍乱政》等)。

2.4 责任方式的复杂性

视频资源责任者及责任方式复杂多样。如主持、受

访、表演、演唱、演奏、演讲、编剧、导演、导播、策划、作词、作曲、音乐总监等,其中导演又可细分为总导演、合成导演、导演组、电视导演、外拍导演、执行导演、录音导演、配音导演、舞台导演、原排导演、译制导演。

2.5 新闻的时效性

新闻资讯类视频资源包括消息、信息推介、深度报道三种节目形态。这类视频资源具有信息量大、涉及领域广泛、时效性强、内容简明、语言简洁的特点。随着时间的推移,有些内容虽会老化,但历史性视频信息仍有再利用的价值。

3 数字视频资源的收藏与制作

目前,我国数字视频资源的收藏与制作具有两个特点。

3.1 收藏机构的广泛性

2005年之前,我国视频资源的收藏机构主要是电视台音像资料馆、电影资料馆、电化教育馆,而且收藏视音频资源的图书馆为数不多(如中国国家图书馆、北京大学图书馆、广东省立中山图书馆等)。

随着人们对视频资源需求的增长,视频资源的收藏机构已从音像资料馆、电影资料馆扩展至省市公共图书馆、大学图书馆、文化部全国公共文化发展中心(分中心)以及科学技术协会等单位。

3.2 视频资源制作者的多样性

随着互联网和数字技术的飞速发展,视频资源的制作者已不局限于电视台、出版社及网站,而是扩展至图书馆乃至个人。目前,我国许多图书馆既收藏、组织、揭示数字视频资源,还自行策划、出品、拍摄、制作视频资源。现在不少图书馆成为视频资源的产出机构。如首都图书馆策划出品的《京味儿趣玩》系列等儿童动画片、辽宁图书馆制作的《辽宁战争往事》、上海图书馆制作的《螺旋藻人工养殖和加工》农业专题片等;部分图书馆还设立视频拍摄室和视频制作室,为图书馆和个人用户拍摄和制作视频创造良好条件。

简而言之,图书馆已不仅是收藏、揭示、组织视频

资源的机构,还成为自行策划、出品、拍摄、制作视频资源的产出机构,部分图书馆甚至创建了许多专题视频资源的特色数据库。

4 我国数字视频资源元数据及描述现状

在我国,数字视频资源迅猛发展,收藏机构广泛,新创建的数字视频资源元数据规范层出不穷。下面介绍四部我国参照《都柏林核心元数据元素集》(*Dublin Core metadata element set, DC*)创建的数字视频资源元数据规范或编目规则。

4.1 《广播电视音像资料编目规范 第1部分:电视资料》

2004年,国家广播电影电视总局发布的GY/T 202.1—2004《广播电视音像资料编目规范 第1部分:电视资料》行业标准在节目层复用15个DC元素,并设置97个修饰词。如在“题名”元素中设置正题名、并列正题名、副题名、题名说明、系列题名、分集总数、分集次、并列系列题名8个修饰词;在“格式”元素中设置实长、入点、色彩、字幕形式、声道格式、声音质量、画面质量、画面宽高比、载体类型、制式、音频数据码率、音频编码格式、音频采样频率、音频位深度、视频数据码率、视频编码格式、视频取样格式及文件格式18个修饰词。

《广播电视音像资料编目规范 第1部分:电视资料》行业标准的特点:明确规定按节目层、片段层、场景层、镜头层四个层次揭示视频资源,并规定各层次元素与修饰词。深度揭示视频资源的内容,以满足专业用户的需求。该标准存在两点不足:第一,图书馆学方面的名词术语,其定义和英文名称未与相关标准保持一致,缺乏规范性;第二,设置一些无实际意义的修饰词,如“系列”“分类”等。

4.2 《中央电视台音像资料编目细则》

2005年中央电视台音像资料馆落成,参考《都柏林核心元数据元素集》和《广播电视音像资料编目规范 第2部分:广播资料》制定《中央电视台音像资料编目细则》。2008年经修订形成《中央电视台音像资料编目细则(3.0版)》,2015年再次修订。其元数据集由题名、主题、描述、责任者、出版者、版权、语种、日期、类型、格

式、时空覆盖范围、来源以及关联12个元素构成,并设置67个修饰词。《中央电视台音像资料编目细则》包括1个通则和有关新闻类节目、专题节目、综艺节目、影视剧节目、体育赛事节目、纪录片及节目素材的7个分则。

与广播电视行业GY/T 202.1—2004标准的区别在于。

第一,元数据集未复用DC的标识符、创建者、其他责任者这3个元素。

第二,根据视频资源责任者的特点,将DC元数据的“创建者”和“其他责任者”两个元素合并为“责任者”;根据中央电视台电视节目版权管理的实际需要,将“权限”改为“版权”。

第三,根据中央电视台专业用户的特殊需求,在GY/T 202.1—2004行业标准的基础上增设17个修饰词。如在“描述”元素中增设现场同期声、拍摄地点、拍摄方式、景别、拍摄角度、限用、馆方分类、磁带条形码、索带号以及表现主体类型10个修饰词;删除或修改广播电视行业标准中的一些修饰词,如“分类号”是“广电分类法”修饰词中的著录内容,故将GY/T 202.1—2004行业标准“主题”元素中的“分类法”和“分类号”合并为一,并将其修饰词名称改为“广电分类法”。

第四,明确规定节目层、片段层、场景层、镜头层各层记录的元素、修饰词及分层方法。截至2016年7月,中央电视台数字视频资源的编目量达96万小时,特别是视频资源书目记录与数字视频资源实现超链接(见图1)。



图1 中央电视台数字视频资源检索结果界面

4.3 《国家图书馆视频资源元数据规范》

《国家图书馆视频资源元数据规范》是“国家图书

馆专门元数据”项目的组成部分,是“国家数字图书馆工程专门元数据标准与著录规范——视音频”项目组于2009年10月—2011年2月制作而成。项目组主要成员来自北京大学信息管理系和北京大学现代教育技术中心。

《国家图书馆视频资源元数据规范》的特点主要体现在四方面。

第一,元数据规范既适用于数字视频资源、网络视频资源,也适用于实体视频资源。

第二,元数据规范是建立在对国内外视频资源元数据调研、国家图书馆视频资源分析以及对视频资源元数据标准设计研究的基础上,并参考DC元数据和GY/T 202.1—2004行业标准制定而成。元数据集在DC15个元素的基础上,增设“版本”“受众”“馆藏信息”及“源载体”四个元素。

第三,元数据规范将19个元素分为核心元素、资源类型核心元素及个别元素三部分。核心元素指“使用频率高的、共性的、可用于不同类型的信息资源描述的元数据元素”^[15],其收录DC15个元素作为该元数据规范的核心元素;资源类型核心元素指“在制定不同类型和不同资源的元数据规范时,根据资源对象的特点设计出的、相似资源共同所需的元素及修饰词”^[15],其将版本、受众、馆藏资源视为资源类型核心元素;个别元素指“为某一特定的资源对象设计的、仅适用于这类对象的元素,不用于交换”^[15],其中源载体属于个别元素。

第四,视频资源著录层次包括集合层、个体层及分析层。以单个视频文件或无层次的实体视频资源为单位编制个体层记录。为满足用户深度检索和利用视频资源的需求,适当地以个体层视频中析出的片段、场景、镜头为单位编制分析层记录。此外,有选择地为系列视频资源编制集合层记录,以便全面揭示一种视频资源的全貌。

4.4 文化部全国文化发展中心的《数字资源元数据规范》

2013年6月—2014年6月,北京大学信息管理系主持制定的“数字资源元数据标准规范、交换标准规范及著录规则”是“国家公共文化数字支撑平台数字资源标准规范”项目的子课题之一,其研究成果由《数字资源元数据规范》和《数字资源编目规则》两部分构成。《数字资源元数据规范》参考GB/T 25100—2010《信息与文献 都柏林核心元数据元素集》和《国家图书馆视频资

源元数据规范》，设置21个元素和105个修饰词，其中复用12个DC元素（即标识符、题名、日期、描述、主题、类型、格式、语种、来源、关联、权限、时空范围，复用率达80%）。根据数字资源特点，将DC元数据的“创建者”与“其他责任者”合并为“责任者”。由于DC元数据的“出版者”元素无法涵盖“出版地”“出品地”以及“出品者”而设立“出版发行”元素，故将DC“出版者”元素调整为修饰词。根据文化部全国文化发展中心数据管理的实际需要新增6个元素，即受众、源载体、馆藏信息、资源加工、资源服务、资源验收。

《数字资源元数据规范》的特点表现在四个方面。

第一，适用范围不局限于视频资源，还包括音频资源、图像资源以及视音频数据库。

第二，根据文化部全国文化发展中心（分中心）的实际需求和资源特点，《数字资源元数据规范》明确规定按集合层、个体层及分析层三个层次设计元数据方案，以揭示数字资源（含数字视频资源）。该元数据规范还明确规定以单个具有独立标识的数字文化资源文件为单位编制个体层记录。必要时可为成套或成系列的数字文化资源编制集合层记录，或为从个体层中析出的视频信息编制分析层记录。

第三，物理记录与逻辑记录相结合。以文件为单位创建的个体层记录是物理记录，而集合层记录和分析层记录是逻辑记录。个体层记录的元素及修饰词设置详尽，而集合层记录和分析层记录的元素及修饰词设置则相对简化。

第四，存在从属关系、引用关系、派生关系及并列关系的资源间，通过“关联”元素中的“原版本、其他版本、包含、包含于、参考、被参考”修饰词进行链接。如系列数字视频资源与其所含的具有独立标识的单个数字视频资源文件之间，或单个数字视频资源文件与其析出的数字视频信息之间存在从属关系，可用“包含”或“包含于”关联方式予以链接，具体见例1和例2。

例1：视频个体层记录

题名：艺文中国

【并列题名】：Art China

【其他题名信息】：本期人物展望

责任者【个人名称】：翁菱

【责任方式】：策划

【个人名称】：卢敏捷

【责任方式】：策划

【个人名称】：王平

【责任方式】：策划

【个人名称】：王水泊

【责任方式】：导演

【个人名称】：叶小虎

【责任方式】：摄像

【个人名称】：王赫泽

【责任方式】：摄像

【个人名称】：刘泽民

【责任方式】：摄像

【个人名称】：高原

【责任方式】：摄像

格式【数量】：168.20M

【时长】：00:39:52

例2：视频分析层记录

题名：不锈钢假山石系列

责任者【个人名称】：翁菱

【责任方式】：策划

【个人名称】：卢敏捷

【责任方式】：策划

【个人名称】：王平

【责任方式】：策划

描述【摘要】：1995年，展望从古代园林假山石中获取灵感，开始创作“不锈钢假山石系列”。这一系列成为他最具代表性最为世人所关注的作品。

格式【时长】：00:00:24

【入点】：00:01:37

关联【包含于】：王水泊. 艺术中国：展望. 2010

5 我国元数据规范分析与现存的问题

5.1 元数据规范大同小异

DC元数据是一种简单结构的元数据，适合组织网络环境下的数字资源。因此，我国无论是广播电视行业制定的GY/T 202.1—2004《广播电视音像资料编目规范 第1部分：电视资料》《中央电视台音像资料编目细则》，还是GC-HD090190《国家图书馆视频资源元数据规范》、文化部全国公共文化发展中心的《数字资源元数据规范》均参考DC元数据扩展而成。但是，在各元数据规范间存在差异。如部分元数据规范复用DC元

数据的“创建者”和“其他责任者”，部分元数据规范则将其合并为“责任者”。DC的创建者指“创建资源的主要责任者”，其他责任者指“对资源做出贡献的其他责任实体”^[2]。在实现计算机检索的今天，书目记录无须区分主要责任者和次要责任者，尤其是视音频资源很难选取负主要责任的创建者，故“创建者”和“其他责任者”元素失去了存在的必要性。

5.2 名词术语欠规范

目前，我国基于《都柏林核心元数据元素集》创建的视频资源元数据规范，对于同一事物却使用不同名称术语。如GC-HD090190《国家图书馆视频资源元数据规范》和文化部全国公共文化发展中心的《数字资源元数据规范》分别将“has part”和“is part of”译为“包含”和“包含于”，而GY/T 202.1—2004《广播电视音像资料编目规范 第1部分：电视资料》和《中央电视台音像资料编目细则》则分别译为“部分为”和“组成部分”；又如，有的元数据规范将赋予资源的名称称为“题名”，有的则称为“名称”。

5.3 修饰词设置方法不一

主题词表和分类法是信息资源主题标引和分类标引的依据。在修饰词设置方面，图书馆界通常参照DC元数据修饰词的设置方法。如通常用主题词表题名（《汉语主题词表》《中国分类主题词表》等）或分类法题名（《中国图书馆分类法》《广播电视节目资料分类法》等）作为“主题”元素的修饰词，而GY/T 202.1—2004《广播电视音像资料编目规范 第1部分：电视资料》则是在“主题”元素下设“分类”修饰词，在“分类”修饰词下设置“分类法”和“分类号”两个子修饰词。“分类”不仅是无实际意义的修饰词，而且所设置的“分类法”和“分类号”两个子修饰词也缺乏科学性。

5.4 编目规则各行其是

《数字资源编目规则》依据GB/T25100—2010《信息与文献 都柏林核心元数据元素集》国家标准将“题名”定义为“赋予资源的名称”^[3]。《国家图书馆视频资源元数据著录规则》和GY/T 202.1—2004《广播电视

音像资料编目规范 第1部分：电视资料》则分别将其定义为“赋予视频资源的名称”^{[1]15}“创作者或出版者赋予一个电视音像资料的正式名称信息”^[4]。“题名”是各类资源共同的特征，在视频资源元数据规范或编目规则中不宜过于强调视频资源题名的个性，否则不利于各类数字资源记录横向间的协调与统一。

名词术语的定义或编目规则的差异会导致编目结果各异，最终影响信息资源的检索与利用。如《国家图书馆视频资源元数据著录规则》（以下简称“国图著录规则”）将“并列题名”定义为“用另一种语言文字表示的题名”^{[1]16}，而GY/T 202.1—2004《广播电视音像资料编目规范 第1部分：电视资料》（以下简称“广电行业标准”）认为“正题名”是“用汉语文字表示的电视音像资源的正式名称”，并列题名是指“电视音像资源正题名的非汉语名称”^[4]。由于广电行业标准的错误导向，导致电视台误将“外语节目”称为“非汉语节目”，并试图制定不切实际的编目规则——《非汉语节目及素材编目规则》，甚至试图要求编目员将原版视频资源的外文题名译为中文，并作为正题名，而将外文节目的原文题名作为并列题名。错误的名词术语定义和编目规则势必导致错误编目结果，见例3和例4。

例3：原题：Cultural express 2006-12-30英文节目
【广电行业标准】正题名：文化报道 2006-12-30（英语）
并列题名：Cultural express 2006-12-30
【国图著录规则】正题名：Cultural express 2006-12-30
并列题名：文化报道 2006-12-30

例4：原题：Gone with the Wind（英语原声电影）
【广电行业标准】正题名：飘
并列题名：Gone with the wind
【国图著录规则】正题名：Gone with the wind

广电行业标准对“正题名”和“并列题名”的定义仅限于国内自产的汉语电视节目，未顾及我国及国外外语录制的电视节目，其名词术语的定义既缺乏科学性也缺乏实用性，尤其是与GB/T3792系列国家标准的有关规定格格不入。不规范、不科学的编目规则会影响信息资源描述的质量，更会阻碍数字视频资源数据库共建共享的进程。

6 建议

解决数字视频资源元数据规范和编目规则中现存的问题迫在眉睫。视频、音频、图像、文本等数字资源元数据规范的标准化是数字资源管理与组织的基础,是共建共享数字资源数据库的必由之路。

为解决我国数字视频资源元数据及编目规则现存的问题,建议全国信息与文献标准化技术委员会根据各类数字资源的特点和用户的需求,基于DC元数据制定数字资源元数据国家标准。数字资源元数据国家标准既要满足普通用户的需求,也要考虑专业用户的需求。力求统一图书馆界、广电行业、文化行业的数字资源元数据规范,使元数据规范的共享性和特殊性兼而有之。与此同时,数字资源元数据标准要充分揭示或链接各种描述对象间的关系,从而加强数字资源描述的

标准化,为我国共建共享跨行业、跨部门的数字资源数据库奠定坚实的基础。

参考文献

- [1] 段明莲,周晨,琚存华.国家图书馆视频资源元数据规范和著录规则[M].北京:国家图书馆出版社,2014.
- [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.信息与文献都柏林核心元数据元素集:GB/T 25100—2010[S].北京:中国标准出版社,2010:2.
- [3] 公共文化数字资源编目使用指南[M].北京:文化部全国公共文化发展中心,2015:19.
- [4] 全国广播电视标准化技术委员会.广播电视音像资料编目规范 第1部分:电视资料:GY/T 202.1—2004[S].北京:国家广播电影电视总局标准化规划研究所,2004:13.

作者简介

段明莲,女,1950年生,北京大学信息管理系教授,研究方向:图书馆学(含信息组织、元数据等),E-mail: duanml@pku.edu.cn。

李燕,女,1979年生,硕士研究生,冶金工业信息标准研究院馆员,研究方向:图书馆学(含信息组织、资源采集等),E-mail: liyan@cmisi.cn。

Digital Video Resource Metadata and Description

DUAN MingLian¹, LI Yan²

(1. Department of Library and Information Science, Peking University, Beijing 100871, China;

2. China Metallurgical Information and Standardization Institute, Beijing 100730, China)

Abstract: In this paper, the definition of video resource, types, and features are elaborated, several typical domestic digital video resource metadata and cataloging rules are introduced, also existing problems are pointed out and the solutions to these problems are provided. The author of this paper considers that metadata for digital video resource as well as relevant cataloging rules are the basis for video resource database construction, and cataloging standardization is the only way to co-construct and share digital video resource database.

Keywords: Digital Video Resource; Metadata; Cataloging Rules

(收稿日期: 2016-10-12)