



地方志数字化加工 规范应用指南

白鸿叶 著



學苑出版社

中国地方志数字化关键技术研究
与演示平台设计

地方志数字化加工 规范应用指南

白鸿叶 著

學苑出版社

图书在版编目 (CIP) 数据

地方志数字化加工规范应用指南 / 白鸿叶著 .

—北京 : 学苑出版社 , 2017.12

ISBN 978-7-5077-5399-8

I . ①地… II . ①白… III . ①地方志—古籍整理—数据处理—规范—中国—指南 IV . ① K290-65

中国版本图书馆 CIP 数据核字 (2017) 第 319150 号

出 版 人 : 孟 白

责任编辑 : 战葆红

出版发行 : 学苑出版社

社 址 : 北京市丰台区南方庄 2 号院 1 号楼

邮政编码 : 100079

网 址 : www.book001.com

电子信箱 : xueyuanpress@163.com

联系电话 : 010-67601101 (营销部) 010-67603091 (总编室)

经 销 : 新华书店

印 刷 厂 : 北京京华虎彩印刷有限公司

开本尺寸 : 787 × 1092 1/16

印 张 : 15.25

版 次 : 2017 年 12 月第 1 版

印 次 : 2017 年 12 月第 1 次印刷

定 价 : 98.00 元

前 言

中国地方志是一种珍贵的文献资源，是地方的百科全书。地方志同时具备了地域性、时代性、系统性、资料性、科学性和连续性，既包含丰富的内容信息，又适合与现代技术相结合。2015年国家图书馆成功申请到科技部，国家科技支撑计划“中国地方志数字化关键技术与演示平台设计”项目，本项目包含三个课题：“地方志资源调查与数字化加工规范研究”“地方志数字化与知识抽取技术研究”“地方志可视化技术与演示平台实现”。其中课题一“地方志资源调查与数字化加工规范研究”主要研究目标是通过点面结合的调查方式，研究清康熙时期纂修方志的收藏情况和文献特性，重点考察山西介休历代方志的纂修与收藏情况；利用国家图书馆数字方志项目与本课题相关的地方志图像44000筒子页，补充6000筒子页清康熙时期纂修方志、介休县志等资料，作为本课题的数据基础；研制8个地方志数字化加工规范和4个相关软件等。

地方志数字化加工规范研究是本项目和本课题的重要组成部分之一，是方志数字化、知识化和可视化的基础。通过遵循同一数据加工规范，保证数据的形式统一，提高数据的通用性和可用性；保证软件系统处理数据的有效性，降低演示平台建设的难度。依据项目的实际需求和地方志的内容特性，本课题拟研制5类8个地方志数字化加工规范，包括：元数据规范2个：地方志专门元数据规范和地方志卷目数据标引规范；对象数据规范2个：地方志图像数据规范和地方志文本数据规范；文字处理规范2个：汉字集外字描述规范和文字认同描述规范；语料数据规范1个：地方志语料数据规范；知识库数据规范1个：古今地名数据描述规范。

地方志专门元数据规范用于整体描述地方志图像数据和文本数据，内容包括书

目信息、数字化参数、版权信息、用户使用提示等。地方志卷目数据规范用于描述地方志图像数据的卷次结构,内容包括卷目信息、层次结构、分类、起始页码等。地方志图像数据规范用于描述地方志图像及规范图像采集方法,内容包括图像类型、画幅内必备要素、图像参数、数字化设备参数、图像压缩、色彩管理等。地方志文本数据规范描述地方志文本数据的基本结构和全文文本描述方式,内容包括文本数据格式、文字描述、版式描述、图形图像描述等。汉字集外字描述规范用于描述地方志文本数据中的集外字,内容包括集外字定位、集外字图像(或图形)、IDS描述等。文字认同描述规范用于描述地方志文本数据中的文字认同情况,内容包括认同字定位、认同方式、认同类型等。地方志语料数据规范描述地方志语料数据的基本结构和标引方式,内容包括文本、文字描述、版式描述、图形图像描述等。古今地名数据描述规范是基于古今地名词典和本项目所调研的地方志,将古地名的名称进行规范,与今地名进行匹配。

本指南的编写,则是为了更好地推广和利用以上8个地方志数字化加工规范。

编者

目 录

前 言	1
第一章 地方志专门元数据规范应用指南	1
1.1 说明	1
1.2 引用及参考的标准和规范	1
1.3 地方志元数据的著录原则	2
1.3.1 著录对象	2
1.3.2 著录单位	2
1.3.3 著录用文字和符号	2
1.3.4 著录信息源	2
1.3.5 著录项目	3
1.3.6 著录规则的内容结构	5
1.4 著录细则	6
1.4.1 题名	6
1.4.2 创建者	9
1.4.3 其他责任者	10
1.4.4 出版者	11
1.4.5 主题	13
1.4.6 描述	14
1.4.7 日期	17
1.4.8 类型	19
1.4.9 格式	19
1.4.10 标识符	21
1.4.11 来源	21
1.4.12 语种	22
1.4.13 关联	22
1.4.14 时空范围	24
1.4.15 权限	26

1.4.16 版本	27
1.4.17 载体形态	28
1.4.18 收藏历史	30
1.4.19 馆藏信息	31
1.4.20 体例	32
1.4.21 志书类型	33
1.5 参考文献	34
1.6 著录样例	35
1.6.1 样例 1	35
1.6.2 样例 2	38
1.6.3 样例 3	40
第二章 地方志卷目数据标引规范应用指南	44
2.1 说明	44
2.2 地方志卷目数据标引	44
2.2.1 正文卷目标引	46
2.2.2 正文外卷目标引	66
2.2.3 特殊情况标引	104
2.3 地方志卷目数据格式	110
第三章 地方志图像数据规范应用指南	114
3.1 地方志图像数据加工流程	114
3.2 地方志图像数据加工原则	115
3.3 地方志图像采集	115
3.4 地方志图像后期处理	116
3.5 地方志图像保存	118
第四章 地方志文本数据规范应用指南	119
4.1 规范对象	119
4.2 版式规范	120
4.2.1 版式选择	120
4.2.2 边框类型规范	122
4.2.3 栏线规范	126
4.2.4 页面标签规范	127
4.3 字体设置	130

4.4 文本规范	131
4.4.1 文本位置	131
4.4.2 文字样式	132
4.4.3 大字和小字	134
4.5 图像规范	135
4.5.1 图像文件	135
4.5.2 图像坐标	135
4.6 集外字规范	135
4.7 不规则版框规范	136
4.8 附录	137
附录 1 书 XML 示例	137
附录 2 卷册 XML 示例	138
附录 3 版式 XML 示例	139
附录 4 书叶 XML 示例	141
第五章 汉字集外字描述规范应用指南	142
5.1 说明	142
5.2 汉字集外字描述的拆分原则	143
5.2.1 遵循原貌	143
5.2.2 描述最简	143
5.2.3 最少拆分	143
5.2.4 优先关注字形	143
5.2.5 尽量避免扩展	143
5.3 使用表意文字描述序列 (IDS) 描述汉字集外字的推荐流程	143
5.4 汉字集外字描述细则	146
5.4.1 一般细则	147
5.4.2 特殊细则	149
5.5 IDS 的扩展	151
5.6 汉字集外字描述元数据	151
5.6.1 集外字描述说明	152
5.6.2 集外字描述范围	152
5.6.3 集外字描述依据	152
5.6.4 集外字描述规则说明	152

5.6.5 基础字符集说明	152
5.7 汉字集外字数据	153
5.8 附录：应用示例	154
附录 1 汉字集外字描述元数据著录示例	154
附录 2 汉字集外字数据著录示例	155
第六章 文字认同描述规范应用指南	163
6.1 说明	163
6.2 引用及参考的标准和规范	163
6.3 文字认同描述数据的基本原则	163
6.3.1 客观性原则	163
6.3.2 一致性原则	164
6.3.3 实用性原则	164
6.3.4 可扩展性原则	164
6.4 文字认同描述数据著录细则	164
6.4.1 文字认同描述元数据	164
6.4.2 文字认同规则描述	167
6.4.3 项目文字认同数据	172
6.5 著录示例	175
6.5.1 著录示例 1	175
6.5.2 著录示例 2	179
6.5.3 著录示例 3	181
附录	183
(资料性附录) 文字认同描述数据 XML Schema	183
第七章 地方志语料数据规范应用指南	190
7.1 规范对象	190
7.2 文本规范	190
7.2.1 条目规范	190
7.2.2 文章规范	193
7.3 图像规范	195
7.4 附录	197
附录 1 地方志碎片化结果 1	197
附录 2 地方志碎片化结果 2	202

附录 3 地方志碎片化结果 3	205
附录 4 地方志碎片化结果 4	221
第八章 古今地名数据描述规范应用指南	226
8.1 说明	226
8.2 范围	226
8.3 词条规范	226
8.3.1 地名标识 (toponymid)	226
8.3.2 地名名称 (name)	227
8.3.3 其他名称 (othername)	227
8.3.4 时间类别 (timetype)	227
8.3.5 属性类别说明 (classdesp)	227
8.3.6 属性类别代码 (classification)	228
8.3.7 地名创建时间 (createdtime)	228
8.3.8 地名废止时间 (endtime)	228
8.3.9 地理位置 (location)	229
8.3.10 地理坐标 (coordinate)	229
8.3.11 现今地名 (modernname)	229
8.3.12 现今地名标识 (modernid)	230
8.3.13 地名简介 (description)	230
8.4 附录	231
附录 1: 地名类别及代码	231
附录 2: 数据示例	233

第一章 地方志专门元数据规范应用指南

1.1 说明

本指南旨在对“地方志元数据规范”(以下简称“规范”)做进一步的阐释和说明,为规范的正确、有效应用提供通用性的指导。

本指南由说明、引用及参考的标准和规范、地方志元数据的著录原则、地方志元数据著录细则、参考文献五部分组成。

使用者可根据实际需求,在不违背本指南一般应用原则的前提下,制定专门的应用规则。

1.2 引用及参考的标准和规范

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 2659—2000 世界各国和地区名称代码

GB/T 3469—2013 信息资源的内容形式和媒体类型标识

GB/T 4880.2—2000 语种名称代码 第2部分:3字母代码

GB/T 7408—2005 数据元和交换格式 信息交换 日期和时间表示法

GB/T 18391.3—2009 信息技术 元数据注册系统(MDR) 第3部分:注册系统元模型与基本属性

GB/T 25100—2010 信息与文献——都柏林核心元数据元素集

GB/T 33286—2016 中国机读书目格式

WH/T 48—2012 数字对象唯一标识符规范

WH/T66—2014 古籍元数据规范（中华人民共和国文化行业标准）

DCMI Namespace Policy [EB/OL].[2017-08-01].<http://dublincore.org/documents/dcmi-namespace/>

DCMI DCSV [EB/OL].[2017-08-01]. <http://dublincore.org/documents/dcmi-dcsv/>

1.3 地方志元数据的著录原则

1.3.1 著录对象

本规则所指的地方志资源是以拍照、扫描等数字化方式形成的派生性地方志数字资源。

1.3.2 著录单位

本规则的著录单位根据地方志资源实际情况可分为以下几种：

- a) 地方志的一部藏本原物所转换的数字资源。
- b) 地方志的一部藏本原物的组成部分所转换的数字资源。
- c) 由不同地方志原物转换的一种数字资源。
- d) 由不同地方志原物的组成部分转换的一种数字资源。

1.3.3 著录用文字和符号

著录用文字一般根据地方志资源本身的文字客观著录。著录人员描述或标引的信息则使用简体中文。

题名元素中的卷数，编码体系修饰词中的年号纪年应依据原书使用汉文数字著录，其他的数字一般用阿拉伯数字著录。

元素 / 元素修饰词可重复时，优先采用重复元素 / 元素修饰词的方式著录。

著录用符号建议采用《国际标准书目著录》（International Standard Bibliographic Description，简称 ISBD）规定的著录标识符。如果 ISBD 中的规定不适用或元素修饰词不可重复时，推荐采用 DCSV（Dublin Core Structured Values）规范。DCSV 规定使用分号作为并列数据值的分隔，如出现多个关键词、多个语种，使用分号作为分隔符。

1.3.4 著录信息源

著录信息源是地方志资源本身。地方志资源本身信息不足时，可参考其他信息源或其他有关文献资源。自拟的著录信息置于方括号“[]”内，取自规定信息源以

外的信息，建议在著录时根据情况在附注元素中说明著录来源。

各元素规定信息源如表 1 所示。

表 1 元素规定信息源

元素	规定信息源
题名	正文首卷卷端
版本	全书的版本类型特征及有关文字记述
出版者	版权页、题名页、牌记、版心、序跋、正文首卷卷端、其他各卷卷端及全书其他部分有关写刻印刷的文字记述
载体形态	全书的完整本信息
关联	题名页、版心、序跋、凡例、正文首卷卷端、其他各卷卷端及全书其他部分有关与本文献相关文献的文字记述
描述	任何信息源

1.3.5 著录项目

本规则著录内容包括 21 个元素，详见表 2。

表 2 地方志元数据著录项目列表

元素	元素修饰词	编码体系修饰词
题名	版心题名	
	内封题名	
	书衣题名	
	书根题名	
	卷端题名	
	其他题名	
创建者		
其他责任者		
出版者	出版地	
	制作者	
	制作地	
	印刷者	
	印刷地	

续表

元素	元素修饰词	编码体系修饰词
主题	关键词	
	主题词	中国分类主题词表
	分类法	四部分类法
		其他分类法
元素	元素修饰词	编码体系修饰词
描述	卷次篇目名称	
	责任者附注	
	摘要	
	一般性附注	
		年号纪年
		公元纪年
		干支纪年
	纂修日期	
	刻版日期	
	印刷日期	
	重印、增补日期	
	获取日期	
类型		DCMI 类型词表
	内容形式	GB/T 3469-2013
	资源载体	
	文件大小	
	页码	
	技术细节	
标识符		统一资源标识符 (URI)
		数字对象唯一标识符 (CDOI)
		统一资源标识符 (URI)
		数字对象唯一标识符 (CDOI)
语种		GB/T 4880.2-2000

续表

元素	元素修饰词	编码体系修饰词
关联		统一资源标识符 (URI)
		数字对象唯一标识符 (CDOI)
	其他版本	
	包含于	
	包含	
	合刻书名	
	被参照	
时空范围	空间范围	DCMIPoint
		GB/T 2659-2000
	时间范围	DCMIPeriod
		年号纪年
		公元纪年
权限	版权拥有者	
	使用权限	
元素	元素修饰词	编码体系修饰词
版本	版印说明	
载体形态	装帧方式	
	数量	
	尺寸	
	版式	
收藏历史	获得方式	
	题跋印记	
馆藏信息	典藏址	
	典藏号	
	其他编号	
体例		
方志类型		

1.3.6 著录规则的内容结构

在著录细则中以元素及元素修饰词为主线撰写著录规则。每个元素及元素修饰

词的定义与描述做了统一的规定，详见表 3。

表 3 地方志元数据元素和元素修饰词的定义与内容

序号	项目	项目定义与内容	必备性
1	名称	赋予术语的唯一标记	必备
2	标签	描述术语的可读标签	必备
3	定义	对术语概念与内涵的说明	必备
4	注释	关于术语或其应用的其他说明，如特殊的用法等	有则必备
5	著录内容	所著录的资源对象，在著录规则中可以有具体、细化的说明	有则必备
6	元素修饰词	是对元素的语义进行修饰，提高元素的专指性和精确性的词	有则必备
7	编码体系修饰词及其用法	元素或元素修饰词取值依据的各种受控词表和规范标记，或者其形式遵循的特定解析规则	有则必备
8	规范档	说明著录元素或元素修饰词内容时依据的各种规范。其取值可能来自各种受控词表和规范。它可以和编码体系一致，也可以是适应具体需要而做出的相关规则	有则必备
9	必备性	说明元素或元素修饰词是否必须著录。取值有：必备（M）、有则必备（MA）、可选（O）	必备
10	可重复性	说明元素或元素修饰词是否可以重复著录。取值有：可重复、不可重复	必备

在每个元素及元素修饰词叙述之后，以“示例”描述典型实例。

1.4 著录细则

1.4.1 题名

名称：title

标签：题名

定义：地方志的名称。

注释：正题名以及题名说明文字如卷数等著录于此。

著录内容：

（1）题名信息一般依据正文首卷卷端所题著录。正文首卷卷端所题题名有误，应照原样著录，并在描述元素中说明。

（2）正文首卷卷端题名不能代表全书或未题名，可选择版心、内封、书衣、书根、其他卷端题名，并在相应的元素修饰词中著录，除此之外取自文献中的其他部分，置于其他题名，并在描述元素中说明来源。

（3）规定信息源以及文献中其他部分均未提供适当的题名信息，由其他资料查考得出，依查考出的题名著录，外加方括号“[]”，并在描述元素中说明来源。

(4) 正文首卷卷端书名有“新修”、“重修”、“续修”、“增修”等字样时,一律保留。志书的题名前按纂修年代或记事所止时间建议增加年号,并置于方括号“[]”中。

(5) 卷数除主卷外,卷首、卷末、附录等卷数信息,均须著录。

(6) 残书依全书原卷数著录。

元素修饰词: 版心题名; 内封题名; 书衣题名; 书根题名; 卷端题名; 其他题名

必备性: 必备(M)

可重复性: 可重复

示例 1:

题名: 乾道臨安志:十五卷卷首一卷割記一卷

示例 2:

题名: [康熙]重修富陽縣志:十卷

1.4.1.1 版心题名

名称: running title

标签: 版心题名

定义: 在地方志版心上书写或印刷的题名。

必备性: 有则必备(MA)

可重复性: 可重复

示例:

题名: 瓊州府志

版心题名: 瓊郡志

1.4.1.2 内封题名

名称: title-page title

标签: 内封题名

定义: 在地方志内封上书写或印刷的题名。

注释: 内封相当于现代图书的题名页, 故又称题名页题名, 主要用来题写题名和版刻情况。

必备性: 有则必备(MA)

可重复性: 可重复

示例:

题名: 瓊州府志