



文献元数据 设计指南和实践



沈仲祺 张建勇 ◎ 编著



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

文献元数据设计指南和实践

沈仲祺 张建勇 编著



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

· 北京 ·

图书在版编目 (CIP) 数据

文献元数据设计指南和实践/沈仲祺, 张建勇编著. —北京: 科学技术文献出版社, 2017. 6

ISBN 978-7-5189-2860-6

I. ①文… II. ①沈… ②张… III. ①元数据—研究 IV. ①G254.364

中国版本图书馆 CIP 数据核字 (2017) 第 137934 号

文献元数据设计指南和实践

策划编辑: 周国臻 责任编辑: 王瑞瑞 白建刚 责任校对: 张叫噪 责任出版: 张志平

出 版 者 科学技术文献出版社
地 址 北京市复兴路 15 号 邮编 100038
编 务 部 (010) 58882938, 58882087 (传真)
发 行 部 (010) 58882868, 58882874 (传真)
邮 购 部 (010) 58882873
官 方 网 址 www.stdp.com.cn
发 行 者 科学技术文献出版社发行 全国各地新华书店经销
印 刷 者 虎彩印艺股份有限公司
版 次 2017 年 6 月第 1 版 2017 年 6 月第 1 次印刷
开 本 787 × 1092 1/16
字 数 263 千
印 张 13.5
书 号 ISBN 978-7-5189-2860-6
定 价 58.00 元



版权所有 违法必究

购买本社图书, 凡字迹不清、缺页、倒页、脱页者, 本社发行部负责调换

编 委 会

主 编 委 会	编	沈仲祺	张建勇		
		彭以祺	吴波尔	张晓林	曾建勋
		孙 坦	宋 文	姜爱蓉	王继民
		孟连生	梁 芳	胡铁军	梁 冰
		赵瑞雪	赵 艳		
撰写工作组		张建勇	于倩倩	黄永文	董智鹏
		邓彩虹	周 毅	李 芳	袁国华
		刘筱敏	鲜国建	王 星	葛红梅
		王 昉	寇晶晶		

前 言

在新的信息环境下,数字文献资源逐渐成为主流资源形态,开放出版迅猛发展,用户信息需求和使用行为发生了重大变化,图书馆的资源组织和服务面临重大挑战。图书馆的文献管理和服务系统不仅要揭示文献、发现文献线索,而且要进行数据挖掘和知识化服务。文献元数据在揭示和发现文献方面起着重要的作用,为文献数据的汇聚融合、统一管理、集成揭示和内容挖掘创造了有利条件。多年来,各个领域和各个行业都制定了相关的元数据标准,支持自身业务的发展,其中都柏林核心元数据倡议(DCMI)已历经20多年,在各个领域充分应用,而图书馆、档案馆、教育机构、出版机构、数字资源仓储机构等纷纷制定了多样化的元数据标准。2003年,科技部支持建设的我国数字图书馆标准规范建设项目全面梳理了数字图书馆领域的标准规范,形成了广泛的影响。此后国家图书馆也制定了数字图书馆元数据行业标准。美国国家医学图书馆提出的期刊文档标签集(JATS 1.1版本)用标签的方式对期刊论文全文进行了结构化描述,该标准被出版社和长期保存机构广泛采用。Web of Science、Scopus等数据服务系统也逐步公开了自己的元数据标准。

图书馆用户的需求也从一般的文献获取向数据分析、结构化全文获取方向发展,相应地图书馆的文献描述也从一般的文献编目向论文元数据描述和全文结构化及语义描述方向发展。现代图书馆的服务不仅要能提供文献的馆藏,同时也能提供数据分析和知识发现服务。在广泛调研国内外相关系统和标准基础上,为适应新的信息环境,满足

用户的需求,本书撰写工作组基于前瞻性、协同化、最小粒度、模块化和兼容国际标准等原则,参照DC元数据应用纲要(DCAP)的基本框架,撰写了文献元数据设计指南,确定了文献元数据制定的基本流程,即功能需求分析、领域模型构建、设计元数据记录、编制使用指南、计算机语言形式化描述;详细说明了流程中每个步骤的内容、要求、方法,为制定文献元数据标准确定了基本的方法体系。

在本书第二部分,针对国家科技图书文献中心(NSTL)的建设目标和资源状况,根据文献元数据设计指南的流程和方法,设计了满足NSTL文献管理和应用需求的统一文献元数据标准,适用的文献对象涵盖NSTL购买、交换、赠予等方式获取的科技类资源,包括期刊、会议录、科技丛书、科技专著、文集汇编、工具书、科技报告、期刊论文、会议论文、学位论文、开放课程、开放课件等;可统一描述文献的印刷版本、数字版本,统一描述文献对象各个层次的信息,满足NSTL数字业务流程中文献数据管理和应用服务的需求。

本书提出的文献元数据设计指南和统一文献元数据标准的视角侧重于元数据的应用,重视用户使用需求,采用用户需求驱动的方式规划设计文献元数据标准。在撰写过程中,撰写工作组广泛调研了现有相关元数据标准,调研各类用户的需求,召开了多次专家研讨会,征求专家的意见,几易其稿,并在实际应用环境中测试标准的可用性。

在本书成书过程中,沈仲祺主编确立了设计指南和统一文献元数据标准编制的目标和原则,指明标准研究的框架和方向,对最终版本进行了修订;张建勇主编确定了编制元数据设计指南的流程和方法体系,主持了设计指南和统一文献元数据标准具体撰写过程,并撰写了导言、功能需求章节,与黄永文合作撰写了领域模型章节,最后审定了全稿;于倩倩撰写了指南部分的元数据记录设计章节,具体组织协调统一文献元数据标准的撰写、修订工作,完成4个数据对象的撰写工作,并进行统稿、审校、修订和完善;董智鹏、黄永文、邓彩虹、

周毅、李芳分别负责 1~2 个数据对象的撰写、修订工作，并进行内容审校；董智鹏撰写了形式化描述方法和形式化描述章节。此外，刘筱敏、鲜国建、王星、葛红梅、王昉、袁国华、寇晶晶参与了标准内容修改和审校工作。

在本书撰写过程中，多位专家对本书进行了评审、论证和指导，他们是彭以祺、吴波尔、张晓林、曾建勋、孙坦、宋文、姜爱蓉、王继民、孟连生、梁芳、胡铁军、梁冰、赵瑞雪、赵艳。其中，曾建勋研究员对书中引用的标准逐一进行了审定，并对全书审校；孟连生研究员对全书内容进行了审校。NSTL 资源工作组和数据库建设工作组相关成员参与了讨论，并提出了重要建议。在此，对他们的贡献一并表示感谢。

希望本书对图书馆和文献领域机构在文献的描述、组织和系统建设方面的工作具有一定的指导作用。

沈仲祺 张建勇

2017 年 3 月 20 日

目 录

第一部分 文献元数据设计指南	1
1 导言	1
1.1 元数据的描述对象	1
1.2 元数据的通用技术性要求	1
1.3 元数据技术体系	2
1.4 设计元数据的人员要求	2
2 元数据应用框架	2
3 功能需求分析	3
3.1 功能需求的必要性	3
3.2 如何创建合适的功能需求	3
3.3 功能需求创建过程及方法	4
3.4 功能需求创建实例	5
4 领域模型构建	7
4.1 领域模型的界定	7
4.2 领域模型建立的步骤	7
4.3 领域模型的检验	8
4.4 领域模型实例分析	8
4.5 详细的数据模型	9
5 元素选取和定义方法	10
5.1 元素选择方法	10
5.2 元素定义方法	11
5.3 元素选取实例分析	13
6 设计元数据记录	13
7 使用指南	14
8 元数据形式化描述	15
参考文献	15

附录 1 推荐的编码体系标准	16
附录 2 形式化描述语言	17
第二部分 NSTL 统一文献元数据标准	34
1 导言	34
2 功能需求分析	34
2.1 愿景	34
2.2 基本原则	35
2.2.1 模块化设计原则	35
2.2.2 最小粒度原则	35
2.2.3 协同化原则	35
2.2.4 与国际相关标准兼容原则	35
2.3 需求分析	36
2.4 用例分析	36
3 规范性引用文件	37
4 术语和定义	37
4.1 文献 document	37
4.2 期刊 journal	37
4.3 会议录 conference proceedings	37
4.4 科技丛书 scientific and technical series	37
4.5 科技专著 scientific and technical book	38
4.6 文集汇编 collection book	38
4.7 工具书 reference book	38
4.8 科技报告 scientific and technical report	38
4.9 期刊论文 journal article	38
4.10 会议论文 conference paper	38
4.11 学位论文 thesis	38
4.12 开放课程 open course	38
4.13 开放课件 open courseware	38
4.14 元数据 metadata	38
4.15 单篇文献 article	39
4.16 来源 source	39
4.17 数据标识符 data identifier	39
4.18 数据唯一标识符 data unique identifier	39
4.19 元素集 element set	39

4.20	元素 element	39
4.21	属性 attribute	39
5	领域模型构建	39
5.1	确定实体对象, 并对实体对象命名	39
5.2	确定实体对象之间的相互关系, 定义实体对象之间的关联和约束	40
5.3	领域模型图示	41
5.4	NSTL 统一文献元数据领域模型的验证	41
6	元数据结构	44
6.1	元素选取和定义	44
6.1.1	元素选取原则	44
6.1.2	字母及符号定义	44
6.1.3	元素定义	45
6.2	元数据框架和结构	45
6.2.1	来源元素集结构	46
6.2.2	单篇文献元素集结构	48
6.2.3	主题/分类/关键词元素集结构	48
6.2.4	贡献者/机构元素集结构	49
6.2.5	会议元素集结构	49
6.2.6	基金元素集结构	49
6.2.7	操作信息元素集结构	49
6.2.8	获取管理元素集结构	49
6.2.9	全文文件元素集结构	49
6.2.10	图元素集结构	49
6.2.11	表元素集结构	49
6.2.12	附加资料元素集结构	49
6.2.13	参考文献元素集结构	49
6.2.14	引用关系	50
6.2.15	归一关系	54
6.2.16	规范关系	55
6.2.17	沿革关系	55
7	描述性元素	56
7.1	描述性元素集简表	56
7.1.1	来源元素集简表	56
7.1.2	单篇文献元素集简表	57
7.1.3	主题/分类/关键词元素集简表	58

7.1.4	贡献者/机构元素集简表	58
7.1.5	会议元素集简表	59
7.1.6	基金元素集简表	60
7.1.7	操作信息元素集简表	61
7.1.8	获取管理元素集简表	61
7.1.9	全文文件元素集简表	62
7.1.10	图元素集简表	62
7.1.11	表元素集简表	62
7.1.12	附加资料元素集简表	62
7.1.13	参考文献元素集简表	63
7.2	描述性元素定义	64
7.2.1	abbrev-source-title 来源题名缩写	64
7.2.2	abstract 摘要	64
7.2.3	addr-line 地址信息描述	65
7.2.4	alt-text 替代性文本描述	66
7.2.5	alt-title 交替题名	66
7.2.6	article-id 单篇文献唯一标识符	67
7.2.7	article-title 题名	67
7.2.8	article-type 单篇文献类型	68
7.2.9	award-acronym 基金项目名称缩写	68
7.2.10	award-amount 资助金额	69
7.2.11	award-date 基金项目日期	69
7.2.12	award-id 基金项目标识符	70
7.2.13	award-name 项目名称	71
7.2.14	baseurl 网站基地址	71
7.2.15	bio 个人简介	72
7.2.16	caption 文字说明	72
7.2.17	city 城市	73
7.2.18	classification 分类号	73
7.2.19	collab 团体作者	74
7.2.20	conf-acronym 会议名称缩写	74
7.2.21	conf-date 会议日期	75
7.2.22	conf-id 会议标识符	75
7.2.23	conf-name 会议名称	76
7.2.24	conf-num 会议届次	77

7.2.25	conf-theme 会议主题	77
7.2.26	contrib-id 贡献者标识符	78
7.2.27	copyright-holder 版权所有 者	78
7.2.28	copyright-statement 版权声明	79
7.2.29	copyright-year 版权年	79
7.2.30	count 总数	80
7.2.31	country 国家	80
7.2.32	database-name 数据库名称	81
7.2.33	date 日期	81
7.2.34	day 日	82
7.2.35	degrees 学位	82
7.2.36	elocation-id 电子位置标识符	83
7.2.37	email 电子邮箱	83
7.2.38	ext-link 外部链接	83
7.2.39	fpage 起页	84
7.2.40	full-name 全名	85
7.2.41	funding-statement 资助说明	85
7.2.42	given-names 名	86
7.2.43	graphic 图像	86
7.2.44	holding-number 馆藏号	87
7.2.45	institution 机构	87
7.2.46	institution-id 机构标识符	88
7.2.47	isbn 国际标准书号	89
7.2.48	issn 国际标准连续出版物编号	89
7.2.49	issn-l 连接 ISSN	89
7.2.50	issue 期	90
7.2.51	issue-part 分期	90
7.2.52	issue-total 总期	91
7.2.53	kwd 关键词	91
7.2.54	level 加工深度	92
7.2.55	license-p 使用许可描述	92
7.2.56	lpage 止页	92
7.2.57	major 专业	93
7.2.58	mixed-citation 参考文献原始信息	93
7.2.59	mode 加工方式	94

7.2.60	month 月份	94
7.2.61	notes 注释	95
7.2.62	object-id 对象标识符	95
7.2.63	open-access 开放获取说明	96
7.2.64	page-range 页码范围	96
7.2.65	phone 电话号码	97
7.2.66	postal-code 邮政编码	98
7.2.67	prefix 姓名前缀	98
7.2.68	process-date 操作日期	98
7.2.69	pub-date 出版日期	99
7.2.70	pub-id 出版物标识符	100
7.2.71	research-subject 研究方向	100
7.2.72	role 职称职务	100
7.2.73	season 季度	101
7.2.74	series 丛书题名	101
7.2.75	size 大小	102
7.2.76	source-id 来源唯一标识符	102
7.2.77	source-subtitle 来源副题名	103
7.2.78	source-title 来源题名	103
7.2.79	source-type 来源类型	104
7.2.80	state 州或省	105
7.2.81	string-conf 会议信息描述	105
7.2.82	string-issue 期信息描述	106
7.2.83	subject 主题词	106
7.2.84	subtitle 副题名	106
7.2.85	suffix 姓名后缀	107
7.2.86	supplement 增期	108
7.2.87	surname 姓	108
7.2.88	td 标准单元格	109
7.2.89	th 表头单元格	109
7.2.90	title 章节题名	110
7.2.91	trans-abstract 其他语种摘要	110
7.2.92	trans-source 来源题名译名	111
7.2.93	trans-subtitle 副题名译名	111
7.2.94	trans-title 题名译名	112

7.2.95	volume 卷	113
7.2.96	volume-series 丛卷	113
7.2.97	year 年份	113
8	辅助性元素	114
8.1	access-group 获取管理	114
8.2	address 地址	115
8.3	alternatives 数据对象其他形式信息	115
8.4	article-meta 单篇文献元数据	116
8.5	award-group 基金项目组	118
8.6	cited-by 引用关系	118
8.7	citings 施引信息	119
8.8	class-group 分类	120
8.9	conference 会议	121
8.10	conf-loc 会议地点	122
8.11	conf-sponsor 会议举办者	123
8.12	contrib 贡献者	124
8.13	contrib-group 贡献者组	125
8.14	counts 计数	126
8.15	database 数据库	127
8.16	fig 图	127
8.17	fig-group 图组	128
8.18	fulltext-file 全文文件	129
8.19	funding-group 资助信息组	129
8.20	funding-source 基金项目资助者	130
8.21	has-appellation 规范关系	131
8.22	history 历史信息	132
8.23	holding 馆藏信息	132
8.24	institution-wrap 机构信息	133
8.25	kwd-group 关键词信息描述	134
8.26	license 使用许可	135
8.27	name 姓名	135
8.28	permissions 使用权限	136
8.29	process-group 操作信息	137
8.30	publisher 出版者	137
8.31	record 文献	138

8.32	ref 参考文献	139
8.33	ref-list 参考文献列表	141
8.34	relation 沿革关系	142
8.35	relation-type 沿革关系类型	142
8.36	same-as 归一关系	143
8.37	section 章节	144
8.38	source-meta 来源元数据	144
8.39	source-title-group 来源题名组	145
8.40	subj-class-kwd 主题/分类/关键词	146
8.41	subj-group 主题词信息描述	147
8.42	supplementary-material 附加资料	147
8.43	table 表格	148
8.44	table-group 表格组	149
8.45	tbody 表格主体	150
8.46	thead 表头	151
8.47	title-group 题名组	151
8.48	toc 目录	152
8.49	tr 行	153
8.50	trans-title-group 题名译名组	154
8.51	variants 文献其他形式信息	154
8.52	volume-issue-group 卷期组	155
8.53	xref 交叉引用	155
9	属性	156
9.1	元素集属性简表	156
9.1.1	来源元素集属性简表	156
9.1.2	单篇文献元素集属性简表	157
9.1.3	主题/分类/关键词元素集属性简表	157
9.1.4	贡献者/机构元素集属性简表	157
9.1.5	会议元素集属性简表	158
9.1.6	基金元素集属性简表	158
9.1.7	操作信息元素集属性简表	159
9.1.8	获取管理元素集属性简表	159
9.1.9	全文文件元素集属性简表	160
9.1.10	图元素集属性简表	160
9.1.11	表元素集属性简表	160

9.1.12	附加资料元素集属性简表	161
9.1.13	参考文献元素集属性简表	161
9.2	属性定义	162
9.2.1	abbrev-type 缩写类型	162
9.2.2	abstract-type 摘要类型	162
9.2.3	access-type 获取方式	163
9.2.4	award-id-type 基金项目标识符类型	163
9.2.5	award-type 基金项目类型	164
9.2.6	calendar 日历类型	165
9.2.7	class-group-type 分类法	166
9.2.8	collab-type 团体作者类型	166
9.2.9	conf-id-type 会议标识符类型	167
9.2.10	conf-name-type 会议名称类型	167
9.2.11	conf-num-type 会议届次类型	168
9.2.12	content-type 内容类型	168
9.2.13	contrib-id-type 贡献者标识符类型	169
9.2.14	contrib-type 贡献者类型	170
9.2.15	corresp 通信作者	170
9.2.16	count 数量	171
9.2.17	country 国家	172
9.2.18	count-type 计数对象类型	172
9.2.19	currency 货币类型	173
9.2.20	data-source 数据来源	173
9.2.21	date-type 日期类型	174
9.2.22	ext-link-type 外部链接类型	175
9.2.23	facet-type 描述类型	176
9.2.24	gbt-7408-date GB/T 7408 格式日期	177
9.2.25	initials 姓名首字母	178
9.2.26	institution-id-type 机构标识符类型	179
9.2.27	kwd-group-type 关键词类型	179
9.2.28	license-type 使用许可类型	180
9.2.29	mime-subtype MIME 子类型	181
9.2.30	mimetype MIME 类型	181
9.2.31	name-style 姓名类型	181
9.2.32	notes-type 注释类型	182

9.2.33	pub-id-type 出版物标识符类型	182
9.2.34	publication-format 出版物格式	184
9.2.35	publication-type 参考文献类型	184
9.2.36	ref-type 引用目标对象类型	186
9.2.37	relation-type 沿革关系类型	186
9.2.38	rsno 内部序号引用	187
9.2.39	sec-type 章节类型	187
9.2.40	seq 序号	187
9.2.41	sno 内部序号	188
9.2.42	source-id-type 来源唯一标识符类型	188
9.2.43	specific-use 具体应用	189
9.2.44	state 数据状态	190
9.2.45	subj-group-type 主题词表	191
9.2.46	units 测量单位	192
9.2.47	version 标准版本	192
9.2.48	xlink: href 超链接	192
9.2.49	xml: lang 语种	193
10	形式化描述	194
10.1	元数据 Schema	194
10.2	元数据 DTD	194
附录 A	NSTL 统一文献元数据的数据唯一标识符规则	194
附录 B	特殊字符处理方法	196