



团 体 标 准

T/TAC 6—2021

多语种国际传播大数据服务 基础元数据

Big data service for multilingual international communications—
Primary metadata

2022-04-01 发布

2022-05-01 实施



中国翻译协会 发布

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国翻译研究院提出。

本文件由中国翻译协会归口。

本文件起草单位：中国标准化研究院、当代中国与世界研究院、太极计算机股份有限公司、中央广播电视总台、中国翻译协会、对外经济贸易大学、传神语联网网络科技股份有限公司、中国质量标准出版传媒有限公司、华南师范大学。

本文件主要起草人：王海涛、袁林、黄振、王恩平、曹馨宇、刘嘉、梅剑平、刘强、陈玉忠、魏洁、吴晓蕊、闫栗丽、崔启亮、刘晓东、郝天永。

多语种国际传播大数据服务 基础元数据

1 范围

本文件规定了多语种国际传播大数据服务中基础元数据的分类框架、元数据属性、基础元数据及其定义描述、元数据扩展规则。

本文件适用于多语种国际传播大数据资源加工处理、资源管理、应用服务、交换共享、平台设计开发等。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；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 4880.1—2005 语种名称代码 第1部分：2字母代码

GB/T 7156—2003 文献保密等级代码与标识

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 20093—2013 中文新闻信息分类与代码

RFC 2396 统一资源标识符(URI):通用语法[Uniform Resource Identifiers (URI): Generic Syntax]

3 术语和定义

下列术语和定义适用于本文件。

3.1

元数据 metadata

定义和描述其他数据的数据

[来源:GB/T 18391.3—2009,3.2.18]

3.2

基础元数据 primary metadata

多语种国际传播大数据服务中,描述基础、通用数据资源的元数据

4 缩略语

下列缩略语适用于本文件。

CIP:图书在版编目(Cataloguing In Publication)

DOI:数字对象唯一标识(Digital Object Identifier)

ID:身份标识号(Identity Document)

ISBN:国际标准书号(International Standard Book Number)

6.4 英文名称

元数据的英文名称采用小写形式。英文名称由两个及两个以上单词组成时,单词之间以连接符“_”连接。

6.5 数据类型

对元数据的有效值域的规定。

示例: 整型、浮点型、布尔型、字符串型、时间型等。

6.6 值域

用于说明元数据的允许取值范围。

6.7 最大出现次数

元数据在实际使用时允许重复出现的最大次数。只出现一次的用“1”表示;重复出现多次的用“N”表示。

6.8 是否必选

表示元数据的选取属性。该属性值表示为:

- a) 必选:该元数据必须选用;
- b) 可选:根据实际具体情况决定元数据是否选用,可以选用也可以不选用。

7 基础元数据

7.1 资源建设相关基础元数据

资源建设相关基础元数据可分为事件相关基础元数据、言论相关基础元数据、数据资源相关基础元数据、地理区域相关基础元数据、机构相关基础元数据、人物相关基础元数据等类型。资源建设相关基础元数据及其属性说明见表 1。

表 1 资源建设相关基础元数据及属性说明

序号	分类	元数据中文名称	说明	英文名称	数据类型	值域	最大出现次数	是否必选
1	事件相关基础元数据	事件 ID	系统赋予的、用于唯一标识事件的编码	event_id	字符串型	自由文本	1	必选
2		事件地点	事件发生的地点	event_place	字符串型	自由文本	1	必选
3		事件摘要	事件的概要描述	event_summary	字符串型	自由文本	1	可选
4		事件发生日期	事件发生起始日期	event_start_date	日期型	格式符合 GB/T 7408—2005 规定	1	必选

表 1 资源建设相关基础元数据及属性说明 (续)

序号	分类	元数据 中文名称	说明	英文名称	数据类型	值域	最大出现次数	是否 必选
22	数据资源相关基础元数据	资源 ID	系统赋予的、可唯一标识资源的编码	resource_id	字符串型	自由文本	1	必选
23		标题	资源的标题名称	document_title	字符串型	自由文本	1	必选
24		目录	报告、图书、期刊等资源的目录	table_of_contents	字符串型	自由文本	1	可选
25		主题词	根据特定主题词表,指明资源主题的词	subject_term	字符串型	自由文本	N	可选
26		关键词	由资源作者给出的表示资源内容主题的中英文词汇	key_word	字符串型	自由文本	N	可选
27		简介	对资源主要内容的概括	description	字符串型	自由文本	N	可选
28		全文	报告、论文或图书等资源的全部文字	full_text	字符串型	自由文本	1	必选
29		页数	报告、期刊、论文、图书等资源的页数	count_of_pages	整型	整数	1	可选
30		作者	创建资源知识内容的主要个人或团体	creator	字符串型	自由文本	N	必选
31		最后修改者	对资源进行最后一次修改的人员	final_modifier	字符串型	自由文本	N	可选
32		资源语种代码	描述资源知识内容所使用的语言的语种代码	language_code_of_resource	字符串型	来源 GB/T 4880.1—2005	N	必选
33		密级	资源的保密等级名称	confidentiality_level	字符串型	来源 GB/T 7156—2003	1	可选
34		版本	资源的版本信息	edition	字符串型	自由文本	1	可选
35	发布日期	资源正式发布(例如出版)的日期	issue_date	日期型	格式符合 GB/T 7408—2005 规定	1	可选	
36	发布时间	资源正式发布(例如出版)的时间	issue_time	时间型	格式符合 GB/T 7408—2005 规定	1	可选	

表 1 资源建设相关基础元数据及属性说明（续）

序号	分类	元数据中文名称	说明	英文名称	数据类型	值域	最大出现次数	是否必选
53	数据资源相关基础元数据	分辨率	图像、视频的分辨率	resolution	字符串型	自由文本	1	可选
54		画面宽高比	视频或图像信号所显示画面的宽与高的比率	screen_aspect_ratio	字符串型	自由文本	1	可选
55		量化精度	度量采样后离散信号幅度的分级数	quantitative_level	字符串型	自由文本	1	可选
56		拍摄地	视频或图像拍摄时所处的地理区域名称	creation_location_	字符串型	自由文本	1	可选
57		拍摄日期	图片或视频的拍摄日期	creation_date	日期型	格式符合 GB/T 7408—2005 规定	1	可选
58		拍摄时间	图片或视频的拍摄时间	creation_time	时间型	格式符合 GB/T 7408—2005 规定	1	可选
59		片幅	图片的片幅大小	frame	字符串型	自由文本	1	可选
60		色彩	图像或视频的颜色模式	color	字符串型	自由文本	1	可选
61		色彩空间	视频或图像的色彩范围	color_space	字符串型	自由文本	1	可选
62		时长	音乐作品或视频等资源的时间长度	duration	整型	以秒为单位的整数	1	可选
63		视频序列	由时空联系的多个视频图像的排列顺序	video_sequence	字符串型	自由文本	1	可选
64		帧率	单位时间内传输的视频画面数量	frame_rate	整型	整数	1	可选
65		位深度图像	存储每个像素所用的位数	bit_depth_picture	字符串型	自由文本	1	可选
66		音频采样	视频数据中音频信号的采样方式	audio_sampling	字符串型	自由文本	1	可选
67		压缩标准	图像、视频等数据进行压缩时采用的标准	compression_standard	字符串型	自由文本	1	可选
68	压缩率	文件压缩后的大小与压缩前的大小之比	compression_ratio	浮点型	实数	1	可选	

表 1 资源建设相关基础元数据及属性说明 (续)

序号	分类	元数据 中文名称	说明	英文名称	数据类型	值域	最大出现次数	是否 必选
85	机构相关基础元数据	成立时间	机构成立的时间	establishment_time	时间型	格式符合 GB/T 7408—2005 规定	1	必选
86		组织机构	机构的人员组织结构	organization_structure	字符串型	自由文本	1	可选
87		相关人员	机构内的主要人员	relevant_person	字符串型	自由文本	N	可选
88		城市	机构所处的城市名称	city	字符串型	自由文本	1	必选
89		国家/ 区域代码	该机构所属国家或区域的代码	country_code	字符串型	来源 GB/T 2659—2000	1	必选
90		简介	对机构基本信息的介绍	description	字符串型	自由文本	1	可选
91		重要事件	机构发展中的重点事件	import_events	字符串型	自由文本	N	可选
92		专注领域	机构主要关注、熟悉的领域	field	字符串型	自由文本	N	可选
93		机构规模	机构的规模大小	size	字符串型	小规模、中规模、大规模	1	可选
94		影响级别	机构的影响级别	influence_level	字符串型	自由文本	1	可选
95		关联机构	与机构存在重大影响关系的机构名称	related_organization_name	字符串型	自由文本	N	可选
96		电话号码	机构的联系电话	phone_number	字符串型	自由文本	N	可选
97		联系地址	有效实际的地址,包括地区名、街道名及门牌号码	contact_address	字符串型	自由文本	1	必选
98	邮箱	机构的电子邮箱	email	字符串型	自由文本	N	可选	
99	官方主页	机构的官方网址	official_homepage	字符串型	自由文本	1	可选	

表 1 资源建设相关基础元数据及属性说明 (续)

序号	分类	元数据中文名称	说明	英文名称	数据类型	值域	最大出现次数	是否必选
115	人物相关基础元数据	机构英文名	人物所属机构的英文名称	english_organization_name	字符串型	自由文本	1	可选
116		任职日期	人物加入任职机构的日期	employed_date	日期型	格式符合 GB/T 7408—2005 规定	1	可选
117		职务	人物职务名称	position	字符串型	自由文本	1	可选
118		职称	人物的技术职称	professional_title	字符串型	自由文本	1	可选
119		联系方式	用于与该人物联系的通讯方式,包括通讯地址、电话等	contact_information	字符串型	自由文本	N	必选
120		居住地	人物当前时间的居住地区	address	字符串型	自由文本	1	可选
121		简介	对人物基本信息介绍	introduction	字符串型	自由文本	1	可选
122		重点关注言论	该人物所发表的重要言论内容	marked_speech	字符串型	自由文本	N	可选
123		重点事件	与该人物相关的重要事件	key_event	字符串型	自由文本	N	可选
124		关联人物	与该人物具有联系的其他人物	connected_person	字符串型	自由文本	N	可选
125		关注领域	人物关注的领域,如方向、话题等	field	字符串型	自由文本	N	可选
126		人物标签	人物特点的标记	label	字符串型	自由文本	N	可选
127		社交媒体情况	人物所使用的社交媒体	social_media	字符串型	自由文本	N	可选
128		是否故去	该人物是否健在	alive	布尔型	是或否	1	必选
129		是否有图片	是否有关于该人物的图片	has_photo	布尔型	是或否	1	可选
130	图片地址	存储该人物图片、照片的链接地址	photo_url	字符串型	自由文本	N	可选	

表 2 资源采集存储相关基础元数据及属性说明 (续)

序号	元数据 中文名称	说明	英文名称	数据类型	值域	最大出现次数	是否必选
13	审核时间	对采集的资源进行审核的时间	verification_time	时间型	格式符合 GB/T 7408—2005 规定	1	可选
14	审核人员 ID	对采集的资源进行审核的人员编码	verifier_id	字符串型	自由文本	1	可选
15	存储位置	资源存储的位置	storage_location	字符串型	自由文本	1	必选

7.3 资源加工整合相关基础元数据

资源加工整合相关基础元数据及其属性说明见表 3。

表 3 资源加工整合相关基础元数据及属性说明

序号	元数据 中文名称	说明	英文名称	数据类型	值域	最大出现次数	是否必选
1	资源加工 ID	系统赋予的、用于唯一标识资源加工的编码	resource_processing_id	字符串型	自由文本	1	必选
2	角色	参与资源加工整合过程的人员的角色	role	字符串型	自由文本	N	可选
3	是否合并	类型、结构等相同的资源,考虑是否可以进行合并	to_be_merged	布尔型	是或否	1	可选
4	数据类型	需要加工的资源的数据类型	data_type	字符串型	自由文本	1	可选
5	加工领取日期	领取需要加工资源的日期	receive_date	时间型	格式符合 GB/T 7408—2005 规定	1	必选
6	加工领取时间	领取需要加工资源的时间	receive_time	时间型	格式符合 GB/T 7408—2005 规定	1	必选
7	加工领取用户	领取需要加工资源的用户	receive_user	字符串型	自由文本	1	必选
8	加工无效原因	资源加工无效的原因说明	processing_invalid_reason	字符串型	自由文本	1	可选

7.4 资源计算分析相关基础元数据

资源计算分析相关基础元数据及其属性说明见表 4。

表 4 资源计算分析相关基础元数据及属性说明

序号	元数据 中文名称	说明	英文名称	数据类型	值域	最大出现次数	是否 必选
1	分析 ID	系统赋予的、用于标识资源分析的唯一编码	analysis_id	字符串型	自由文本	1	必选
2	资源标签	对资源主题、领域、类型等内容进行分析后,以关键词方式进行标记	resource_label	字符串型	自由文本	N	必选
3	外方国家名	新闻所涉及的除中国外其他国家名称	foreign_country_name	字符串型	自由文本	N	必选
4	外方国家代码	外方国家的国家代码	foreign_language_country_code	字符串型	来源 GB/T 2659—2000	N	必选
5	外方机构名	新闻所涉及的除中国机构之外的其他机构名称	foreign_organization_name	字符串型	自由文本	N	必选
6	外方人员名	新闻中所涉及的外方人员姓名	foreign_person_name	字符串型	自由文本	N	必选
7	外方人员职务	新闻中所涉及的外方人员的职务	foreign_person_position	字符串型	自由文本	N	可选
8	中方机构名	新闻所涉及的属于中国的机构的名称	chinese_organization_name	字符串型	自由文本	N	必选
9	中方人员姓名	新闻中所涉及的中国人姓名	chinese_person_name	字符串型	自由文本	N	必选
10	中方人员职务	新闻中所涉及中国人员的职务	chinese_person_position	字符串型	自由文本	N	必选
11	新闻标题	新闻的标题	news_title	字符串型	自由文本	1	可选
12	新闻分类代码	新闻所属分类的代码	news_classification_code	字符串型	来自 GB/T 20093—2013	1	可选

表 5 资源应用服务相关基础元数据及属性说明 (续)

序号	元数据中文名称	说明	英文名称	数据类型	值域	最大出现次数	是否必选
5	资源发布时间	资源提供方发布信息资源的时间	publication_time	时间型	格式符合 GB/T 7408—2005 规定	1	必选
6	资源最新修改日期	最近一次修改资源的日期	update_date	日期型	格式符合 GB/T 7408—2005 规定	1	可选
7	资源最新修改时间	最近一次修改资源的时间	update_time	时间型	格式符合 GB/T 7408—2005 规定	1	可选
8	资源链接地址	可以获取资源的网络地址	url	字符串型	自由文本, 格式符合 RFC 2396 规定	1	必选
9	资源版本号	资源的版本号	version	字符串型	自由文本	1	可选

7.6 系统管理相关基础元数据

系统管理相关基础元数据及其属性说明见表 6。

表 6 系统管理相关基础元数据及属性说明

序号	元数据中文名称	说明	英文名称	数据类型	值域	最大出现次数	是否必选
1	用户 ID	系统赋予的、用于唯一标识用户的编码	user_id	字符串型	自由文本	1	必选
2	用户名	用户登录系统的账号名	user_name	字符串型	自由文本	1	必选
3	用户角色	用户的角色, 例如系统管理员、管理员、用户、访客等	user_role	字符串型	自由文本	N	必选
4	用户密码	用户登录账号的密码	password	字符串型	自由文本	1	必选
5	真实姓名	用户的姓名	real_name	字符串型	自由文本	1	必选
6	性别	用户的性别	gender	字符串型	自由文本	1	可选

参 考 文 献

- [1] 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
-

中国翻译协会
团体标准
多语种国际传播大数据服务
基础元数据
T/TAC 6—2021

*

中国标准出版社出版发行
北京市朝阳区和平里西街甲2号(100029)
北京市西城区三里河北街16号(100045)

网址 www.spc.net.cn

总编室:(010)68533533 发行中心:(010)51780238
读者服务部:(010)68523946

中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

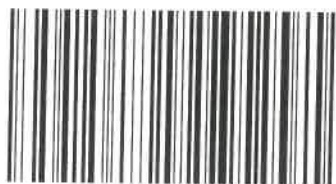
*

开本 880×1230 1/16 印张 1.5 字数 52 千字
2022年3月第一版 2022年3月第一次印刷

*

书号: 155066·5-4235 定价 42.00 元

如有印装差错 由本社发行中心调换
版权专有 侵权必究
举报电话:(010)68510107



T/TAC 6-2021



码上扫一扫 正版服务到