

图像数据库元数据元素集调查研究和改进建议*

王荣群 李婷 魏蕊
(河北大学管理学院, 保定 071000)

摘要: 数字图像是数字信息资源的重要组成部分,但又不同于一般的信息资源。数字图像的组织需要专门的元数据标准。本文首先通过文献调研归纳并总结国内外数字图像元数据的研究现状;然后选择不同类型的图像数据库调查分析其图像的元数据元素,并将数据库内图像元素与数字图像相关的元数据标准进行对照分析;最后在理论分析和实践调查的基础上提出图像数据库元数据元素集的改进建议,具体包括17个元数据元素和11个元素修饰词,以期进一步完善图像数据库元数据在实践中的应用。

关键词: 图像; 图像数据库; 元数据

中图分类号: G254 DOI: 10.3772/j.issn.1673-2286.2022.03.002

引文格式: 王荣群, 李婷, 魏蕊. 图像数据库元数据元素集调查研究和改进建议[J]. 数字图书馆论坛, 2022(3): 15-21.

随着数字技术的发展和智能手机时代的到来,图像类信息资源大规模增加。图书馆等机构开始将数字图像类数据库纳入馆藏建设的范围。数字图像作为数字信息资源的重要组成部分,它不同于一般的信息资源,其内容特征涉及非文字性的图形、色彩、色调、纹理、内容对象、物理制作要素等信息^[1]。元数据是信息资源描述、组织与管理的重要工具,利用元数据对图像进行描述、组织和规范处理,可以提高检索的准确性和快捷性,从而增强其可发现性和可用性。从数字图像的获得方式来看,包括直接拍摄的数字图像和从纸质文献资源中以扫描方式提取的数字图像两大类。直接拍摄的数字图像主要包括实物展示类图像和新闻纪实类图像两类。因此,本文选择实物展示类图像、新闻纪实类图像和数字化提取类图像这三类图像数据库作为调查样本,调查分析其图像的元数据元素,然后与数字资源元数据标准和图像类资源元数据标准进行对照分析,结合实践调查和理论分析提出图像数据库元数据元素集的改进建议。

1 数字图像元数据的研究现状

以中国知网CNKI为数据源,使用“图像”“图像数据库”“元数据标准”分别进行题名、主题的组合检索。检索不限“年份”,经过人工筛选后,得到检索结果75条。经过分析可以看出,近年来国内图像元数据的研究主要集中在两个方面。一是对国际和行业上相关的数字图像元数据项目和规范进行介绍。例如:袁莉等^[1]对国外数字图像的元数据格式MOA2 metadata、VRA Core等做了简单介绍和比较,将元数据分为描述性元数据、结构性元数据、管理性元数据;黄崑等^[2]将可用于描述和著录图像的元数据标准归纳为专门的图像元数据标准、文化资源相关的元数据标准和通用元数据标准并分别介绍。二是对我国大量古籍文献和特殊地方文化等相关元数据标准的制定进行探究。例如:北京大学图书馆提出古籍元数据标准、拓片元数据标准;郭瑞芳等^[3]根据国家图书馆图像元数据元素及修饰词著录细则设计了一套描述端砚图像的元数据方案,同时

* 本研究得到2021年河北大学研究生创新资助项目“数字图像元数据元素集研究与设计”(编号:HBU2021ss011)资助。

结合端视作品自身特点给出完整描述；王若宸等^[4]基于Panofsky和Shatford的图像学理论模型构建专门面向非遗数字图像的语义描述框架，提出一种专门面向非遗图像语义信息的描述方法。

20世纪90年代，国外开始逐渐出现图像元数据的研究项目，相关研究以介绍元数据在各行业领域的实践应用为主，侧重于元数据的实践技术等方面研究。关于元数据的具体应用实践，范围很广，比较有代表性的是数字图书馆中数字图像元数据规范项目的建设、将图像元数据的标准应用于构建地标图像的AI训练数据集^[5]，以及从实验日志和文本文件中提取元数据，处理图像并将其上传到服务器，从而创建完全注释的多维数据集^[6]。Kim等^[7]针对多媒体图像的侵犯版权和非法使用等问题，提出以JPEG格式的数据保护特性作为元数据扩展到远程访问控制中，通过将访问控制数据的位置信息作为元数据保存在图像文件中，图像所有者可以允许或拒绝其他人的数据消耗，起到隐私保护作用。在图书馆、档案馆、博物馆等方面，Crowe等^[8]以丹佛大学档案馆馆藏为例，分析了高校档案图像元数据最佳实践。Saleh^[9]探讨了网络文化遗产数字馆藏中使用图像嵌入元数据的可用性。

2 图像数据库元数据元素调查分析

为了保证所调研数字图像的全面性，本文选择实物展示类图像数据库、新闻纪实类图像数据库和数字化提取类图像数据库为研究对象，对不同类型数字图像数据库内的元数据元素进行调查。图像类目是将图像按不同特征进行归类处理，以满足用户检索、获取所需图像的需求。图像元数据是描述图像文件的一些信息标签，具体包括图像拍摄时自动标注的图像尺寸、像素等信息，以及图像名称、作者、地址等后期对图像进行人工描述的元数据信息。本节主要通过网络调查法对数据库内图像类目和图像元数据进行总结分析，以期改进图像数据库元数据元素提供一定的实践基础。

2.1 实物展示类图像数据库

实物展示类图像就是将各类实物直接地予以反映，通过图像拍摄真实还原其本质美感。实物展示类图像数据库保存的是相机等拍摄工具直接捕捉的图像或者计算机内绘图工具等创作出的图像。这类数据库既

有图像资源种类丰富的综合性图像数据库，如图虫图库^[10]，它是字节跳动旗下的优质摄影图片整合图像库，由摄影爱好者上传作品，已有超过4.6亿张正版素材，包括风景、动物等摄影图片；也有将某一类型的作品收集在一起，为用户提供某一专业领域的图像检索及利用服务的专类图像数据库，如中药材图像数据库^[11]，它由香港浸会大学中医药学院建立，图像涉及常用中药材420余种，以高清图像展示药材的外观特征，并以文字记载药材的来源、主产地、性味功效等信息。作为实物展示类图像数据库，这两类数据库中的图像类目多种多样，且都对不同特征、不同类别的图像进行整合区分。

经调查，图虫图库主要图像类目包括“人像”“自然风光”“节日”等，为更加细致地划分不同类别的图像，图虫图库对每一大类下的类目进行细分，例如，“人像”类目下可以细分为“外国人”“青年”“女性”“肖像”等，“自然风光”类目下可以分为“山川”“草原”“海洋”等，以便更好地帮助用户根据图像的不同特征精确检索图像；中药材图像数据库对常用类型中药材图像的药材类别、味觉特性、品质特性等进行划分，又对每一个类目进行细分，如“味觉特性”可以细分为“苦”“甘”“辛”等。通过对数据库内图像元数据调查发现，根据图像类型的不同，所描述的信息侧重点也有所区别。图虫图库作为摄影图像数据库，更注重对图像的摄像标准系数进行描述，如白平衡、色彩模式、图像宽高度、分辨率等，同时也对图像标题和作者等主要信息进行描述；中药材图像数据库内的图像主要为高清拍摄中药材图像，除了对图像尺寸、像素等信息进行描述外，更注重体现中药材图像的属性特征，对药材科名、中外文名称、药材来源、药材产地等图像元数据进行描述。

2.2 新闻纪实类图像数据库

新闻纪实图像的主要功能为向用户传递故事和信息，在新闻纪实类图像中需要展示人物、时间、地点、事件、原因、发生过程六要素，在呈现图像的同时搭配文字介绍，更加完善地向外界呈现一个新闻信息。中国新闻图片网^[12]属于新闻纪实类图像数据库，将每日新闻图片制作成电子文本，通过互联网整合成可视化的图片分享平台。新闻的不同类型决定图像也有不同的类目。中国新闻图片网按反映社会生活的内容将新闻图像分为政治、经济、文化、体育、教育等类目，并对不同类目下

的具体新闻图像进行细分,例如“体育”大类下具体分为“奥运”“足球”“NBA”等,以保证新闻覆盖面的广泛性。关于新闻图像的元数据,通过调查可知,新闻图像具有新闻的特点,其图像元数据也是围绕新闻六要素,即人物、时间、地点、事件、原因、发生过程进行描述。还需要注意的是,中国新闻图片网通过“图像文件名”“图片作者”“作者单位”和“拍摄地址”等元数据对图像信息进行描述,以保证新闻的真实性与准确性。

2.3 数字化提取类图像数据库

数字化提取类图像数据库主要指通过扫描仪等识别图像的机器,将文献中的图像传输到计算机中,通过图像处理软件最大限度地还原文献图像,整理保存形成数据库。插图是插附于书刊文字间的特殊图画形式,将文献中的纸质插图进行数字化提取和整合,有助于更加直观地理解文本内容。

古籍插图图像数据库^[13]是由首都图书馆创办的一个包含古籍插图数据的线上查询系统,目前收录了1

万多张插图,每张插图都有其所属类别以及具体的图像信息。古籍插图图像数据库中每条图像数据包括插图影像和内容描述两部分。根据插图主题内容的不同分为小说、戏曲、宗教、风景等几大类目。在对插图进行数字化提取过程中,主要通过“插图题名”“绘图者”“图像类别”“绘制年代”等元数据对插图图像进行描述,以揭示图像内蕴含的信息。同时还通过“文献题目”“责任者”对提取插图的古籍文献信息进行描述。因此在构建图像元数据时可注重数字化提取类图像与文献本身的内容联系,保障数字化提取类图像元数据描述的完整性。

2.4 三类图像数据库的对比分析

通过对上述三类图像数据库的调查分析发现,不同类别图像数据库中的图像呈现方式和信息描述不同,图像类目和图像元数据描述侧重点有所差异,结果见表1所示。

表1 三类图像数据库的对比分析

图像类别	数据库名称	图像类目	图像元数据
实物展示类图像	图虫图库	人像、自然风光、动物、美食、节日、商务、科技、教育、文化	场景类型、白平衡、数据色彩、色彩模式、色彩降噪、色彩空间、图像宽高度、分辨率、景深、视角
	中药材图像数据库	药材类别、缩图、性状特征、品质特性、性味功效	拉丁学名、科名、中文名称、英文名称、来源、主产地
新闻纪实类图像	中国新闻图片网	政治、经济、军事、体育、教育、民族与宗教、法制	人物、时间、地点、事件、原因、发生过程
数字化提取类图像	古籍插图图像数据库	人物、小说、戏曲、军事、宗教、动植物、风景、建筑、历史故事	成图方式、图像类别、绘制年代、母体文献题名、母体文献索取号

实物展示类图像主要是对于事物本身的展示,向外界传递图像自身所携带的信息。因此在构建图像元数据元素集时多以描述图像本身的元素为主。根据用途不同,在提取元数据元素时有各自不同的依据,从图虫图库中提取的图像类目都是很直观的图像元素,如“自然风光”“美食”;而中药材图像这种某一具体类别的图像则更多以图像背后的信息元素作为提取的元数据,如味甘、味辛是不能在图像中直观表示的。因此要根据图像的功能定位,以及用户对于图像的需求等具体情况构建实物展示类图像的元数据。新闻纪实类图像更加注重事件的表达和信息的呈现,在进行图像

描述时,主要以围绕新闻的六要素,即人物、时间、地点、事件、原因、发生过程构建元数据,此外还需明确新闻图像的所属类别,注重图像所反映的新闻主题。数字化提取类图像与上述两类数字图像不同,在构建图像元数据元素集时,除了描述图像本身的元素之外,还要对提取图像的文献资料进行描述,插图所附文献的题名、故事、人物等也要作为图像元数据描述加以呈现。综上所述,构建数字图像元数据元素集时,要根据数据库的类型和定位,考虑用户对于不同图像的信息需求,综合考量数字图像的基本信息、所属范畴、主要功能、用户定位等因素。

3 图像元数据元素与相关元数据标准的对照

图像数据库元数据元素集构建的最终目的是方便数字图像的收集与管理,因此需在构建之初将数字图像相关标准与实际应用中的图像元素进行比较研究。本文选取数字资源元数据标准《都柏林核心元素集》

(Dublin Core Element Set, DC)和图像类资源元数据标准《图像元数据规范》(WH/T 51—2012)、《中文新闻图片内容描述元数据规范》(GB/T 35311—2017),对比元数据标准与上述4个图像数据库的具体实践情况,分析其图像的元数据元素。为缩小元素的含义范围,使其更具有专指性,所选取的《都柏林核心元素集》和《图像元数据规范》都给定相应的元素修饰词,因此本节针对元素修饰词在具体图像数据库内的适应性也进行一定的分析介绍,从而为图像数据库元数据元素集的改进提供参考。

3.1 数字资源元数据标准

DC是由联机计算机图书馆中心(Online Computer Library Center, OCLC)与美国国家超级计算机应用中心(National Center for Supercomputer Applications, NCSA)联合发起,目的是建立一套描述网络电子文献的方法,以便实现网上信息检索。DC元数据元素集可以归纳为内容属性、知识产权属性、形式属性3种属性,共包括15个核心元素和24个元素修饰词。对照DC元数据元素集,将图虫图库、中药材图像数据库、中国新闻图片网和古籍插图图像数据库内图像元数据元素的利用情况进行对照分析(见表2)。

上述4个数字图像数据库都包含DC元数据中的“题名”“主题”“描述”“创建者”和“日期”,表明这些元素都是图像数据库中最基本的信息。DC是描述电子文献的元数据标准,而数字图像具有强烈的数字资源属性,因此数字图像数据库内图像元素与DC的核心元素适配性较高。例如:“主题”,古籍插图图像数据库中的有“个人主题”和“地名主题”,这都是对于“主题”元素的扩展应用;“语种”,中药材图像数据库内图像的每一个信息元素都是以中文名、拉丁名、英文名3种语言表示;“权限”,图虫图库内图像标注“限时免费”“授权书”“第三方权利说明”等与使用下载相关的权限内容。在对比元素时发现,一些DC元素修饰词

表2 DC元数据元素在图像数据库中的应用

元数据属性	元数据元素	图虫图库	中药材图像数据库	中国新闻图片网	古籍插图图像数据库
内容属性	题名	√	√	√	√
	主题	√	√	√	√
	描述	√	√	√	√
	来源	√	×	√	√
	覆盖范围	×	×	×	×
	语种	√	√	√	×
	关联	×	×	×	×
知识产权属性	创建者	√	√	√	√
	权限	×	×	√	×
	出版者	×	×	√	√
	其他责任者	×	×	√	√
形式属性	日期	√	√	√	√
	类型	×	√	√	×
	格式	√	×	√	×
	标识符	√	×	×	×

注:元素在数据库中存在应用情况用“√”表示,无则用“×”表示

可以更加精确地描述图像信息,如“日期”下的元素修饰词在图虫图库中有所体现,“创建日期”可表示作者拍摄图像的时间,“可获得日期”表示用户获取利用图像资源的时间区间,“发布日期”表示图像资源正式对外公开发布的日期。值得注意的是,这种对比只是在对照DC元数据标准的基础上分析图像数据库元素集在具体项目应用中的体现,对于图像数据库元数据元素集的构建还应该参考各类相关的元数据标准规范。

3.2 图像类资源元数据标准

数字图像资源不同于一般的信息资源,对其内容特征元素描述存在特定类型图像元数据标准,这类标准往往对于描述实物本质特征、派生的数字化资源以及新闻类图像资源内容更具有针对性。本文选取《图像元数据规范》和《中文新闻图片内容描述元数据规范》两个图像元数据标准对所调研的图像数据库内的元素进行一一对照分析。

3.2.1 《图像元数据规范》

《图像元数据规范》是国家文化部发布的一项文化行业标准,由上海交通大学图书馆牵头,联合国家图书馆、浙江省图书馆和山西省图书馆牵头制定。标准规

定了图像资源(包括所有原生和派生的图像资源)的内容和外观描述,给出图像资源定位与管理的一般性方法,适用于描述数字形态的图像资源,包括22个元素和24个元素修饰词,其中15个元素对应DC核心元素,另外7个其他元素分别为出版地、版本、受众、收藏机构、背景、源载体和收藏历史。由于15个对应的DC核心元素已在表2进行对比,这一部分就不再重复对比,只对4个数字图像数据库应用《图像元数据规范》中另外7个元数据元素情况进行对照分析(见表3)。

表3 《图像元数据规范》元数据元素在图像数据库中的应用

元数据元素	图虫图库	中药材 图像数据库	中国新闻 图片网	古籍插图 图像数据库
出版地	√	×	√	×
版本	√	×	√	×
受众	√	√	√	√
收藏机构	×	√	√	√
背景	√	√	√	√
源载体	√	√	√	√
收藏历史	×	√	√	√

注:元素在数据库中存在应用情况用“√”表示,无则用“×”表示

这7个元素均不同程度地体现在4个图像数据库内:

“背景”,作为图像资源的特有元素,主要体现为图像资源有关的责任者和其他相关人物的资料,这一元素适用于古籍插图图像数据库中的“出处”,适用于中国新闻图片网内的“拍摄地址”;“受众”,作为图像资源的各类实体,包括创建者、出版者等,不同类型图像数据库的受众也会有所不同,如图虫图库的受众更多为专业摄影师和摄影爱好者。

需要注意的是,《图像元数据规范》是在DC的基础上进行设计,其元素修饰词相较DC有所补充,主要针对图像资源的内容特征进行描述,一些核心元素的修饰词对于图像类资源的描述更加精确。例如“格式”的元素修饰词“篇幅”复用DC“大小”,如中国新闻图片网内的图像“篇幅”不仅指图像的尺寸大小,还包括图像数量,明确表明每组图有几张图像,以“组图”的形式出现,“格式”下还新增元素修饰词“技术细节”描述数字图像,表现图片拍摄制作成电子文本所需的技术细节处理;“描述”的元素修饰词“目次”复用DC“目录”,作为资源内容的子单元列表,在描述时,几乎适用于上述提到的所有数据库内的图像,如图虫图库内“冬奥之约”板块下的“单板滑雪”“双板滑雪”“跳台滑

雪”等,“描述”下还新增元素修饰词“风格”,在4个图像数据库中均有体现,如古籍插图图像数据库内风景、建筑等图像资源的内容具有强烈鲜明的风格特点。

3.2.2 《中文新闻图片内容描述元数据规范》

《中文新闻图片内容描述元数据规范》是2017年12月底正式发布的国家标准,是世界范围内第一个以图片视觉内容和语义特征描述为主的图片元数据规范,专业性强、结构简明,具有较高的应用价值。标准的主要内容是对图片元数据元素做出准确定义,尤其突出对图片视觉内容特征和语义信息的描述。标准以编辑类图片和创意类图片为主要对象,定义了24个元数据元素,分别归属到公共元数据、新闻编辑类图片元数据和新闻创意类图片元数据3个部分。经调研发现,新闻编辑类图片元数据和新闻创意类图片元数据侧重描述新闻图像,对图片类型的指向性更强,强调新闻的属性。而图虫图库、中药材图像数据库和古籍插图图像数据库中不包含新闻图像,所以只对比分析公共元数据在4个图像数据库内的具体应用情况(见表4)。

表4 公共元数据元素在图像数据库中的应用

元数据	图虫图库	中药材 图像数据库	中国新闻 图片网	古籍插图 图像数据库
图片标识	√	√	√	√
篇幅	√	√	√	√
色彩	√	√	√	√
拍摄角度	√	√	√	√
被摄体方向	√	√	√	√
景别	√	√	√	√
自定义分类	√	√	√	√
地点	√	√	√	√
主题	√	√	√	√
情景	√	×	√	√

注:元素在数据库中存在应用情况用“√”表示,无则用“×”表示

由于公共元数据的“公共”特征,因此图片标识、篇幅、色彩、拍摄角度、被摄体方向等元素均体现在4个数字图像数据库中。从表4可见,“景别”作为描述图像的特写、近景、全景及人物身体部位的元素,在上述4个数字图像数据库中均有体现,如中药材图像数据库中的“山楂”药材图像对其果肉、果核、果梗和外皮进行局部特写描述,图虫图库内对人物的“身体部位”进

行描述,如脸部、手、腰部。因此建议“景别”作为数字图像元数据元素集的一项元素扩充其中;“情景”描述画面中主要元素呈现出的情形和活动信息,偏向于展示图像的动态信息,中药材图像数据库主要是对于药材进行展示,图像不涉及动态的相关内容,因此“情景”未在中药材图像数据库进行体现,而其他3个数字图像数据库内的图像都不同程度地展现了图像的画面信息,如图虫图库内的“动物捕食”图像、古籍插图图像数据库的“祭祀”图像和中国新闻图片网内的“运动会”图像都呈现了图像内主要元素的活动情形。

新闻编辑类图片元数据和新闻创意类图片元数据强调新闻图片的特有属性,10个新闻编辑类图片元数据“标题”“图说”“拍摄时间”“拍摄地”“人物姓名”“事件”“专题”“新闻分类”“民族”“宗教”和4个新闻创意类图片元数据“人物信息”“画面主体”“特定时间”“摄影技术”均在中国新闻图片网内有所体现。例如:“民族”和“宗教”两项可以描述中国新闻图片网站内的“民族与宗教”板块内的新闻图片;“图说”描述新闻图片的创作背景,新闻事件的时间、地点和人物等,如中国新闻图片网中的“河南增派中医医疗队驰援上海”这一新闻组图的文字说明——“4月12日,河南省增派75名队员组成的中医医疗队驰援上海。医疗队队员分别来自河南省中医院、河南中医药大学第三附属医院、河南省中医药研究院附属医院3家医疗机构,均为业务能力强的中青年骨干中医医师”,其中对图像涉及背景、时间和地点等元素均有介绍。

通过对3个元数据标准和4个图像数据库进行对比可知,DC作为描述网络资源的元数据标准在4个数据库内均有体现,图像类资源元数据标准则根据图像资源的不同特征在具体应用过程中有所差异,《图像元数据规范》适用于描述实物展示类图像、新闻纪实类图像和数字化提取类图像,而《中文新闻图片内容描述元数据规范》更具专指性,侧重对新闻图像描述,更适用于对新闻纪实类图像描述。因此本文结合具体图像类别的不同,参考借鉴不同类型的图像元数据标准,在元数据标准和图像数据库具体实践对比分析的基础上,提出图像数据库元数据元素集的改进建议。

4 图像数据库元数据元素集改进建议

通过调查实物展示类图像数据库、新闻纪实类图像数据库和数字化提取类图像数据库中的元数据元

素,将数据库中图像的元数据元素与数字资源元数据标准《都柏林核心元素集》和图像类资源元数据标准《图像元数据规范》《中文新闻图片内容描述元数据规范》进行对照分析,再根据图像数据库的图像特点,参考3个元数据标准,共精炼出17个元数据元素和11个元素修饰词。其中复用《都柏林核心元素集》的16个元数据元素和“摘要”“目次”“创建日期”“发布日期”“可获日期”“篇幅”“载体”7个元素修饰词;复用《图像元数据规范》的“风格”“技术细节”作为元素修饰词;复用《中文新闻图片内容描述元数据规范》的“景别”作为元数据元素,“情景”“特定时间”作为元素修饰词。

根据元数据在图像数据库中的具体应用情况,将改进的17个元数据元素划分为3个基本检索点、5个有价值的检索点和9个辅助检索点。“题名”“创建者”“主题”可作为基本检索点进行描述,用户可以根据“题名”和“主题”进行简单检索,也可通过作者及拍摄者进行检索。有价值的检索点包括:①“描述”,其元素修饰词包括“摘要”和“风格”,其中“摘要”是对图像具体内容的概括和总结,“风格”包括中国风、表现主义和写实风格等,例如中国风就是图像中较有特色的一个风格,包含强烈鲜明的中国元素事物,可以对中国风的服饰、陶瓷、图画等实物资源进行描述;②“日期”,其中元素修饰词“创建日期”和“发布日期”是图像资源的必备元素,尤其新闻图片的发布日期对新闻报道尤为重要,而“特定时间”(例如节气、节日等)也是画面内容所表现的时间信息,且都强调新闻图片的时效性和故事性,可作为图片重要检索点;③“类型”,图像资源的特征或类型,按照不同图像主题检索;④“格式”,包括图像资源的媒体类型和大小,如尺寸、像素等;⑤“覆盖范围”,图像资源的覆盖范围包括时间范围和空间范围,时间范围是图像拍摄的时间间隔和日期,空间范围可以涉及图像资源的地名或地理坐标。一些检索点对于用户检索需求较少,可作为图像资源的辅助元素,包括:①“出版者”,使资源可以获得和利用的责任实体,主要指图像资源的出版者或颁布者,包括个人与团体,由个人或团体的名称来表示;②“其他责任者”,对资源做出贡献的其他责任实体;③“语种”,表达图像资源内容的语言;④“标志符”,给予图像资源的明确标志;⑤“来源”,与当前图像来源有关的资源;⑥“权限”,指与图像资源相关的各种产权说明,记录图像资源的权限管理说明等;⑦“关联”,与图像资

源存在某种关系的其他资源,例如原格式图像;⑧“受众”,指利用图像资源的各种实体,不同类型图像数据库的受众会有所区别;⑨“景别”,作为描述的近景、中景、全景及人物身体部位的元素。需要注意的是“受众”这一元素复用于DC的非核心元素,其不包括于表2中15个核心元素内,将其扩充进图像数据库元数据元素集可以较之前更全面地描述数字图像,从而完善图像数据库元数据元素集的构建。

5 结语

为推动数字图像类数据库的建设,加强图像资源的可发现性和可用性,需要一个更具有针对性,且与数字图像实践应用紧密结合的标准。由于数字图像元数据的组织和规范具有一定的复杂性,同时还受到各方面因素的制约,本文通过分析当前相关图像数据库的建设现状和数据特点,根据现有国家标准、行业标准,提出合理并且实用的元数据改进方案,这对于整理数量庞大的图像资源和提高图像的检索利用具有重要意义。需要注意的是,随着技术的发展,图像数据库元数据元素集标准的建设也会是一个不断完善的过程,当前对于数字图像元数据的描述需要进一步合理规范,以保证数字图像资源的保存和共享。

参考文献

- [1] 袁莉, 张晓林. 数字图像的元数据格式[J]. 大学图书馆学报, 2001(2): 27-30, 41.
- [2] 黄崑, 王珊珊, 耿赛. 国内图像元数据应用研究现状与分析[J]. 国家图书馆学报, 2015, 24(4): 60-66.
- [3] 郭瑞芳, 张会章. 基于端砚图像数字化建设中元数据的标准设计[J]. 图书馆论坛, 2012, 32(6): 36, 57-60.
- [4] 王若宸, 朱学芳. 面向非物质文化遗产的图像语义描述研究[J]. 信息资源管理学报, 2021, 11(2): 128-137.
- [5] KIM J. A Study on designing metadata standard for building AI training dataset of landmark images[J]. Korean Society for Library and Information Science, 2020, 54(2): 419-434.
- [6] HAY J, TROUP E, CLARK I, et al. PyOmeroUpload: a python toolkit for uploading images and metadata to OMERO [EB/OL]. [2022-01-01]. <https://wellcomeopenresearch.org/articles/5-96/v1>.
- [7] KIM J Y, SOUNDRA PANDIYAN R. Protecting metadata of access indicator and region of interests for image file [J/OL]. Security and Communication Networks, 2020 [2022-01-01]. <https://doi.org/10.1155/2020/4836109>.
- [8] CROWE K M, MEAGHER E S. Metadata best practices for university archives images [J]. Journal of Library Metadata, 2015, 15(2): 79-101.
- [9] SALEH E I. Image embedded metadata in cultural heritage digital collections on the web [J]. Library Hi Tech, 2018, 36(2): 339-357.
- [10] 图虫图库 [EB/OL]. [2022-02-12]. <https://tuchong.com/>.
- [11] 中药材图像数据库 [EB/OL]. [2022-02-12]. <https://library.hkbu.edu.hk/electronic/libdbs/mmd/index.html>.
- [12] 中国新闻图片网 [EB/OL]. [2022-02-12]. <http://www.cnsphoto.com/>.
- [13] 古籍插图图像数据库 [EB/OL]. [2022-02-12]. <http://query.cicn.net.cn/GJAndST/gjct1.htm>.

作者简介

王荣群, 女, 1997年生, 硕士研究生, 研究方向: 信息组织。

李婷, 女, 1997年生, 硕士研究生, 研究方向: 资源获取与利用。

魏蕊, 女, 1982年生, 博士, 副教授, 通信作者, 研究方向: 信息组织, E-mail: weirusxyq@foxmail.com。

Investigation and Research on Metadata Element Set of Image Database and Suggestions for Improvement

WANG RongQun LI Ting WEI Rui
(School of Management Hebei University, Baoding 071000, P. R. China)

Abstract: As an important part of digital information resources, digital image is different from general information resources. The organization of digital images requires a special metadata standard. Through literature research, this paper summarizes the research status of digital image metadata at home and abroad. Select different types of image databases, investigate and analyze the metadata elements of their images, and compare the image elements in the database with the metadata standards related to digital images. On the basis of theoretical analysis and practical investigation, this paper puts forward some suggestions to improve the metadata element set of image database, including 17 metadata elements and 11 element modifiers, so as to further improve the application of image database metadata in practice.

Keywords: Image; Image Database; Metadata

(收稿日期: 2022-02-12)