

拓片元数据标准比较研究

张 靖

摘 要 在对拓片元数据标准总体情况梳理的基础上,采用比较研究的方法,对台湾中研院史语所《拓片与古文书数位典藏计划暨辽金元拓片数位典藏计划拓片 Metadata 需求规格书》、《北京大学古籍数字图书馆拓片元数据标准(草案)》以及国家科技部重大基础课题“我国数字图书馆标准规范建设”子项目“专门元数据规范”的拓片元数据规范研究成果等三家代表性拓片元数据标准进行比较分析,总结了三部标准规范的各自特点,最后提出拓片元数据信息的提取与拓片元数据标准的设计同等重要。参考文献 5。

关键词 拓片数字化 书目控制 拓片元数据 标准

分类号 G254.36

ABSTRACT The digitization of rubbings has attracted the attention of related institutions. The bibliographic control of rubbings is one of the important steps in digitization. Since the design of metadata is the basis of bibliographic control, it has been the focus of research. Treating the status quo as a whole, this paper makes a comparative study of three representative metadata established by the Institute of History & Philology, Academia Sinica of Taiwan, Peking University and a major project of the Ministry of Science and Technology, People's Republic of China. Finally, it points out that the extraction of metadata information of rubbings is as important as the design of bibliographic control metadata. 5 refs.

KEY WORDS Digitization of rubbings. Bibliographic control. Metadata of rubbings. Standard.

CLASS NUMBER G254.36

1 前言

随着计算机技术被越来越普遍地运用于古籍整理,数字化成为古籍整理和保存的热门课题,拓片作为一种特殊形式的历史文献,其数字化工作为国内外各主要相关机构所重视。在拓片数字化工作上,国际各机构间形成既合作又竞争的态势。目前致力于拓片数字化研究、拓片数据库建设的国内外学术研究机构主要包括中国国家图书馆、北京大学图书馆、上海图书馆、台湾中研院史语所、美国加州大学伯克利分校东亚图书馆以及日本京都大学图书馆。其中,书目控制作为拓片数字化的重要组成部分以及图书馆这一类型拓片收藏机构的传统业务,受到了较为普遍的重视(上述机构中,除美国加州大学伯克利分校东亚图书馆和日本京都大学图书馆外,均各有专门的拓片元数据研究)。

书目控制研究自术语产生、概念确立到理论变革,统一书目记录格式的探讨始终是研究的一个重要方面。拓片书目控制研究是在元数据的概念及理论研究发展较为成熟之后才开始展开的,因此拓片元数据标准(或元数据方案)的设计和整合作为拓片书目控制的基础,是目前研究最为集中的部分。

2 拓片元数据标准研究概述

据“我国数字图书馆标准与规范建设”项目阶段性研究成果《国内专门数字资源对象元数据实例列表分析报告》所述,截止到2003年3月中国(包括台湾)进行拓片元数据研究和维护的机构有中国国家图书馆、北京大学图书馆、上海图书馆和台湾中研院史语所四家。其中,上海图书馆拓片元数据设计简单,仅由八个元素组成(号码、名称、年代、类别、书者、版本、款式、备注);中国国家图书馆中文石刻资源库元数据是

在继承 CNMARC 格式的基础上(其中 191 编码数据字段是专门为拓片这一文献类型所设置的特色元素),参考 DC 元数据,并结合中文石刻拓片的特点所设计的描述性元数据(包括拓片题名、责任者、年代、地点、拓片原物状况、附刻、拓片版本、拓片特征<拓片描述、书体与行款、形款、装裱、获得方式、附注>、数目与尺寸、题跋印记、文种、拓片内含书目索引、关联关系、拓片录文、馆藏信息等)。CNMARC 格式几乎未能反映出拓片这一类型文献的特殊性,DC 元数据针对拓片的特点设置了特色元素,然而元素的设置体现出单纯的描述性,不足以实现拓片的数字化保存和管理;除此之外,元数据元素之间的层级结构关系也不清晰。

北京大学图书馆中文元数据标准研究项目组于 2001 年 3 月发布了《北京大学古籍数字图书馆拓片元数据标准(草案)》(以下简称北大图《标准》)^[1]。台湾中研院史语所于 2004 年 6 月发布了《拓片与古文书数位典藏计划暨辽金元拓片数位典藏计划拓片 Metadata 需求规格书(2.0 版)》(以下简称史语所《规格》)^[2]。与国图及上图的拓片元数据相比,北大图和台湾中研院史语所的拓片元数据的设计更为合理。2004 年 5 月,科技部重大基础课题“我国数字图书馆标准与规范建设”之子项目“专门数字对象描述元数据规范”完成,该规范基本上是北大图《标准》的一部分,可视为北大图的节选本(北大图《标准》将拓片元数据分为描述性元数据元素、管理性元数据元素以及 GIS 元数据元素三个部分,并分别进行结构及元素定义说明;《拓片描述元数据规范(标准草案)》节选了其中的描述性元数据部分,并有所修订)。2007 年 4 月基于上述科技部课题子项目研究成果的《中文元数据概论与实例》出版之前,史语所《规格》和北大图《标准》是可以公开获得的较为成熟的拓片元数据方案。

2005 年年底,国家科技部重大基础课题“我国数字图书馆标准规范建设”子项目“专门元数据规范”完成。内容包括拓片在内的 11 种中文元数据规范、相关著录规则及中文元数据基本理论的研制^[3]。在该研究成果的基础上编写成

《中文元数据概论与实例》第五章,其中提供了一套拓片元数据规范(以下简称“科技部规范”)。该规范与上述北大图《标准》以及《拓片描述元数据规范(标准草案)》关系密切,可视为后者的修订本。

3 史语所《规格》、北大图《标准》与科技部规范比较分析

台湾中研院史语所在拓片的整理和研究工作上积累了丰富的经验,在拓片元数据(台湾地区将 metadata 译作“后设资料”,其所谓“拓片后设资料”即指拓片元数据)的设计上起步较早,2001 年度“金石拓片与其他媒材古文书数位典藏计划”已执行“青铜器全形拓”、“墓志铭”、“碑刻”、“杂刻”、“佛教石刻造象拓本”等项目的拓片元数据分析,且不断地有项目资金注入。史语所《规格》在较准确把握拓片这一类型文献特点的基础上,参考国际元数据标准 DC (Dublin Core,都柏林核心元素集)和 CDWA (Categories for the Description of Works of Art,艺术作品描述类目)进行设计,并藉之与国际接轨。施汝瑛总结中研院史语所的拓片元数据具有层级性、特殊性、扩充性、结构性、整合性、互通性和国际性等特点^[4]。笔者认为该拓片元数据方案在以下几个方面值得借鉴:

第一,拓片信息被划分成拓片资料、原件资料 and 影像资料三个部分,每部分又包括若干相应的元数据元素;拓片元数据的层级结构相当分明,这一方面有利于拓片书目控制数据的形成,另一方面有利于拓片数字化工作流程的顺畅进行;再者,拓片本身、拓片原刻、拓片影像的信息得以专门反映又彼此联系,全面地呈现了拓片书目信息。

第二,以元数据单元群组关系描述丛拓关系(丛拓关系包括复本、残件复原、同物件拓片、同组件或建筑拓片、成套拓片、摹刻本以及附件),以一组丛拓群组为著录单位形成“丛拓资料”,使得相互关联却分别著录的拓片通过丛拓记录识别码在数据库中建立了自动的联系,既利于馆藏的管理,又方便从事相关研究的用户

利用。

第三,描述性元数据的设计全面。拓片资料描述性元数据元素主要包括:类型、登录号、题名、高广、装潢、版本、拓印范围、拓工、技法、铭刻<包括作者、书体、行款、文向、位置、语文、楷定、释文>、图像/纹饰、印记、著录、保存状况、丛拓资料等;原件资料描述性元数据元素主要包括:品名、年代、材质、出土地点、刻立地点、保存状况、收藏机构/现存地点等;影像资料描述性元数据元素主要包括:图档档名、影像类型、来源属性、影像范围、影像出处等,从而使拓片书目信息得到较有深度的揭示。

第四,管理性元数据及保存性元数据的设置(管理性元数据元素主要包括:使用限制、展览、版权所有等;保存性元数据元素主要包括:典藏资讯等)简化并规范化了拓片收藏机构对于实物形式和数字化形式拓片的管理和保存工作,也在一定程度上提高了拓片数据库使用的安全性及合法性。另外,著录拓片相关参考文献形成“书目资料”,也使拓片数据库的研究色彩更为突出。

中研院史语所典藏金石拓片的数字化工作,从2001年开始分“青铜器全形拓”、“墓志铭”、“碑刻”、“杂刻”、“佛教石刻造象拓片”等项目分别进行拓片元数据分析,至2003年始加入“青铜器铭文拓片”与“佛教石刻画像拓本”;2004年又加入“甲骨文拓片”及“辽金元拓片数位典藏计划”,最终统一于“拓片与古文书数位典藏计划暨辽金元拓片数位典藏计划”,形成史语所《规格》。因此,该元数据方案在客观上需解决主题分散的前期拓片元数据的整合问题。体现在元数据设计上则是先通过三个层级的基本元素反映不同主题拓片的相同属性,再分类设置各主题拓片的特殊需求元素,从而既实现拓片元数据的整合,又保留和利用了前期的工作成果。但就国内拓片收藏机构而言,一则所藏拓片种类较为单一,二则一般没有前期的数字化工作积累,因此在拓片元数据的设计上无须考虑整合问题,即使有意在拓片数据库中实

现分主题展示,亦可通过考虑周全的元数据方案统一完成。

北大图《标准》认为原器物、拓片与数字图像之间的关系紧密,在元数据结构上,关于三者的记录无法绝对分离,与史语所《规格》的拓片信息三级结构意见相反。但两个元数据方案在对著录对象之间关系的理解和定位上则基本一致。北大图《标准》将拓片元数据分为三个部分,包括19个描述性元数据元素(题名、责任者、主题/关键词、内容及注释、金石刻制时间、金石类型、资源形式、拓片标识、语种、相关资源、时空范围、馆藏信息<以上为12个核心元素>,版刻/版本、外观特征<以上为2个本馆核心元素>,收藏历史、书法特征、金石刻立/出土地点、金石材质的、原器物标识<以上为5个拓片个别元素>、4个管理性元数据元素(对象创建说明、实例说明、用户评论/评注、文档结构项)以及2个GIS元数据元素(空间项、时间项),并分别对三个部分的元数据进行内容结构及元素定义说明。其优于史语所《规格》之处主要在于管理性元数据的专门突出与使用,另一值得注意的地方是北大图《标准》在按《北京大学数字图书馆中文元数据标准框架方案》^①将描述性元素分为核心元素、本馆核心元素及拓片个别元素之外,同时将描述性元素分别归入主要信息、内容信息、外观信息以及馆藏与标识信息四类,与史语所《规格》的拓片资料、原件资料和影像资料分层方式异曲同工,使得元数据元素的层级结构清晰明朗。

“科技部规范”属于描述性元数据,由11个核心元素(题名、责任者、金石年代、资源类型、语种、附注、关键词、时空范围、相关资源、标识符、权限)、4个古文献系列核心元素(版刻与版本、载体形态、收藏历史、馆藏信息)以及4个拓片个别元素(金石所在地、金石材质的、书法特征、金石类型)组成,共19个元素,其中部分元素扩展了若干元素修饰词及编码体系修饰词^{[3]90-91}。基本对应北大图《标准》的19个描述性元数据

① 非专门的拓片元数据标准,故未予以专门讨论。

元素,部分元素的名称和位置有所调整:①原“本馆核心元素”调整为“古文献系列核心元素”②核心元素中,“主题/关键词”改称“关键词”,“内容及注释”改称“附注”,“金石刻制时间”改称“金石年代”,“资源形式”改称“资源类型”,“拓片标识”改称“标识符”,元素位置有所调整,并增加了新元素“权限”;③古文献系列核心元素中,“外观特征”改称“载体形态”,“收藏历史”由原拓片个别元素调入,“馆藏信息”由原核心元素调入;④拓片个别元素中,“金石刻立/出土地点”改称“金石所在地”,“原器物标识”删除,“金石类型”由原核心元素调入。调整后的元素名称和位置较原来更为合理和清晰。

作为重大基础课题子项目研究成果,“科技部规范”在拓片元数据结构上的安排需要与同属于项目研究范围的其他类型中文文献(除拓片外,还包括古籍、舆图、家谱、地方志、电子图书、网络资源、音频资料、学位论文、期刊论文、会议论文,共11种)的元数据结构保持一致。从我国数字图书馆建设这一宏观角度出发,统一各种类型中文文献元数据结构的处理方式是合理的考量。然而具体到专门的拓片书目控制,这种安排则在一定程度上限制了拓片信息的深入和全面揭示。

4 拓片元数据标准之后:拓片元数据信息提取

拓片元数据标准的设计是拓片书目控制及其研究的基础,有了合理的拓片元数据方案,加之提取和填入相应的信息,再通过有效的信息组织,才能形成拓片书目控制数据,实现拓片书目控制与数字化。由于这一部分属于实际操作的工作环节,通常不像拓片元数据的设计那样必然形成文本资料,而主要是工作人员的经验积累。目前尚未见有专门的文献对这些经验进

行总结,仅钟碧芬在介绍台湾中研院史语所“佛教石刻造像拓本数位典藏计划”时,论及拓片整理与内容校读。其指出,在进行整理佛教石刻拓片时,拓片的分类与之后的内容校读,面临许多问题,光是拓片的选择标准与分类方式就须经过反复思索与讨论;在内容校读方面,若要正确地校读碑铭,就必须在浩瀚庞杂的文献中,对各种版本的著录与相关的出土资料或研究报告,因而文献资料的准备,以及拓片读书会的定期举行,是拓片整理和内容校读的关键^[5]。拓片元数据信息的提取与拓片元数据标准的设计同样重要,相关理论研究和实践总结有待得到学界和业界的重视。

参考文献:

- [1] 北京大学图书馆中文元数据标准研究项目组. 北京大学古籍数字图书馆拓片元数据标准(草案)[R/OL]. [2007-03-02]. <http://www.idl.pku.edu.cn/1/page.htm>.
- [2] 台湾中研院史语所. 拓片与古文书数位典藏计划暨辽金元拓片数位典藏计划拓片 Metadata 需求规格书(2.0版)[R/OL]. [2007-03-02]. <http://www.ihp.sinica.edu.tw/%7Edahcr/>.
- [3] 肖珑,赵亮. 中文元数据概论与实例[M]. 北京: 北京图书馆出版社,2007.
- [4] 施汝瑛. “拓片资料库管理系统”现况[C/OL]. 2004年12月台湾中研院史语所“金石拓片数位典藏研讨会”会议论文. [2006-01-10]. <http://rub.ihp.sinica.edu.tw/download.htm>.
- [5] 钟碧芬. 佛教石刻造像拓本数位典藏计划[C/OL]. 2004年12月台湾中研院史语所“金石拓片数位典藏研讨会”会议论文. [2006-01-10]. <http://rub.ihp.sinica.edu.tw/download.htm>.

张靖 中山大学资讯管理系讲师,博士。通讯地址:广州新港西路135号。邮编510275。

(收稿日期:2009-04-12)