

人物档案多级著录应用研究

——以上海交通大学钱学森图书馆特藏“629 袋”为例

孙 逊^{1, 2} 于英香¹ 孙 安^{1, 3}

(1 上海大学图书情报档案 上海 200444; 2 上海交通大学钱学森图书馆 上海 200030; 3 河南科技大学图书馆 洛阳 471023)

摘 要 人物档案因其收藏场所各异,具备档案、文献、文物的多重属性。同时,它具有多级的内部结构。文章在梳理档案多级著录研究现状的基础上,分析人物档案著录存在的主要问题,指出人物档案多级著录的必要性。重点以上海交通大学钱学森图书馆特藏“629 袋”为例,分析其多级结构特征,试图运用档案多级著录的思想与规则对其进行分级整理,实现多级著录的应用。

关键词 人物档案 多级著录 来源原则 元数据描述 EAD(编码档案著录规则)

DOI: 10.16065/j.cnki.issn1002-1620.2019.04.010

The Application Research on Private Archives Multiple–Levels Description: A Case of “629” from QLM

SUN Xun^{1,2}, YU Yingxiang¹, SUN An^{1,3}

(1 School of Library, Information and Archives, Shanghai University, Shanghai 200444; 2 Qianxuesen Library and Museum, Shanghai Jiaotong University, Shanghai 200030; 3 Library, Henan University of Science and Technology, Luoyang 471023)

Abstract: Private archives have multiple properties of archives, literature and relics depending on their different collecting institutions. Meanwhile, it has a multi-level inner structure. Basing on describing the research situation of archive multi-level description, the paper analyzes major problems existing in private archives description and stresses the importance of multi-level description. Taking “629” specially stored in QLM of SJTU as an example, this paper analyzes its multi-level structure and tries to use the idea and rules of multi-level description of archives to sort them out in different levels so as to realize the application of multi-level description in private archives.

Key words: private archives; multi-level description; principle of provenance; metadata; EAD

1 问题的提出

一直以来,有关人物档案的探讨主要集中在其管理方面。具体地讲,包括如何征集、建档等。随着档案信息化建设的发展,人们对于档案信息需求的增长,如何对人物档案进行科学整理和信息组织是

人物档案整理者、保管者面临的迫切问题。

2 人物档案多级著录研究现状

2.1 档案著录研究现状

有关人物档案著录的探讨与思考,应当回归到

“档案著录”这一大的理论研究与工作实践的环境下加以审视。

首先,国外著录规则的引介。国外档案著录针对档案的多重结构体系进行档案著录规则的制定,如国际档案著录规则(总则)、EAD等。国内积极引入和推广EAD来推动档案著录的标准化建设,促进档案信息资源的共建共享。只有实现档案信息资源著录的规范化、标准化,编制出高效的档案检索工具,才能满足网络环境下档案资源开发利用的要求,才能为档案信息资源共享、数字档案馆建设以及档案科学管理打好基础^[1]。

其次,国内外著录规则的对比研究。尚珊^[2]对已有著录准则进行过对比,提出EAD主要适用于档案和手稿资源,包括文本和电子文档、可视材料和声音记录,其实用范围非常广。根据国际档案理事会的EAD官方网站的统计,参与EAD项目的机构大多数为高校和国家单位,并且多数机构位于发达国家,尤其是美国,所占比例超过80%。ISDA(G)在著录方法上提出了以尊重全宗理论为基础的多级著录。其遵循原则如下:从总体到部分、与著录级别相关、著录条目连接、无冗余著录^[3]。ISDA(G)在特性说明项中规定了著录项层级,包括全宗级、子全宗级、类别级、子类别级、案卷级、文件级层次的著录。而我国档案著录规则(DA/T 18—1999)只是对文件级、文件组合级别、案卷级、案卷组合级进行描述,只能反映文件一案卷间的关系,没有“全宗级与类别级的著录,其著录结果是只见树木(案卷与文件)而不见森林(全宗与类别),这样就违背档案尊重全宗的基本原则”^[4]。

最后,对国内著录规则的反思。我国于1985年5月颁发了国家标准《档案著录规则》第一版,在此基础上,适当吸取了国际标准《档案著录规则(总则)》的思路和决定,于1999年以行业标准DA/T 18—1999的形式再次发布。赵芳^[5]曾多次撰文分析我国档案著录规则的应用局限性。

综上,我们不难得出结论,尽管ISDA和EAD的重合度较高,但这两种著录标准都非常注重对著录层次性和结构性的把握,并对所要著录的项目进行了明确的规定和细化描述,从而为网络时代用户从不同著录级别检索到所需信息提供了更多的可能^[6]。

2.2 人物档案多级著录研究现状

2.2.1 有关人物档案的界定及其研究

人物档案的称法不一。在图书馆,有人将其称作

“名人手稿”;在博物馆,则将其视为名人遗物或者文物藏品。早在20世纪90年代,徐锦^[7]曾撰文对“名人档案”“名人全宗”“名人档案资料”“人物档案”等概念作了分析。实则,至今尚未有正式而统一的称谓和定义。2012年,王利伟^[8]对“人物档案基本问题”进行了思考与总结,指出人物档案真实记录了一个人从出生、成长乃至成才、成功的全过程,因而包含了丰富的知识信息,具有较高的历史与经济价值。对于人物档案收集途径方面,提到人物多媒体资料库建设。蒋友钧^[9]梳理了人物档案整理的步骤和方法,提出采用“人名一问题”方法进行分类、组卷。不难发现,作为一类专门档案,人物档案的研究主要集中于收集、整理、管理等方面。然而,随着档案管理信息化水平的提高及人们对档案信息利用需求的不断增长,对于人物档案信息的利用也变得更为迫切。特别是一些人物纪念馆,其承担着展现某一重要人物的辉煌一生及留存给世人宝贵的精神和物质财富,并弘扬这一人物崇高价值和精神品质的重大责任。例如,美国肯尼迪总统图书馆(<https://www.jfklibrary.org/>),实则兼顾档案馆、博物馆的属性,其所藏资源非常丰富,更主要的是它们被详尽而系统地呈现在网络上,供全世界民众访问并使用,这实际与其著录工作的发展水平不无关系。

2.2.2 人物档案著录存在的问题

人物档案是某个人在其一生的社会活动中所形成的原始记录。不管如何称谓,其表现形式和内部结构应当保持原貌。这一原则正是类似于档案学中的“来源原则”。对人物档案的研究大多集中于如何征集、管理方面。但其实,人物档案整理是人物档案资源开发、利用(资料库检索、展览展示)及价值发现的基础。整理的过程中,信息著录是对人物档案进行规范控制的方式。2004年,上海图书馆刘炜^[10]等人曾对该馆名人手稿的元数据方案进行过深入探索。2016年,上海交通大学图书馆张洁^[11]等人曾撰文分析科学手稿如何组织的问题。两者的共同点在于,他们均对“资源”的类型及定义做了较好的界定,并且对元数据定义模型等做了探讨,其成果给业界带来了较大的参考价值。然而不足的是,尚未系统地揭示有关人物的资源整体,未能揭示所谓“森林”的模。这种现象其实在整个国内档案界也较为普遍。那么,目前人物档案著录中到底存在哪些棘手的问题?经过分析与研究,笔者认为,大致有如下几点:

首先,人物档案的分类界限模糊,存在交叉。论述人物档案分类的文章不少,主要问题在于,将内容

分类和形式分类进行交叉、混合分类的现象较为常见。以上海图书馆名人手稿馆为例,他们在其研究中将资源进行了严格分类。当然,这大致是建立在上图所收集到的资源自身情况基础之上的,对于每个机构而言,仍需要结合自身的馆藏实际进行区分。由此我们应当借鉴的是,针对人物档案的分类问题,我们必须要按照某一标准严格划分,不能有所交叉。

其次,人物档案查全率和查准率低。目前,国内档案检索的普遍现象是基于案卷级和文件级的检索并呈现,不能详细检索到其他层级的文件。甚至因为信息揭示的不足,导致查询结果严重缺失的情况屡见不鲜。当然,这与档案开放、获取的规则不无关系。

最后,人物档案描述层级简单。通俗地讲,查询的目标对象“只见树木,不见森林”。人物档案的检索本身具有现实局限性,很多门户上人物档案的查询和展示呈现单一化、一维性的特点。我们无法全面、系统地知道该资源在整个体系中的位置及与其他资源间的关系。当然,有图书情报工作者采用了“资源集合”的概念,甚至会跟关联数据(Linked data)相联系。然而,作为档案本身,其具有自身的结构,应在层级上进行科学而客观的剖析。国际档案著录标准理事会认为,著录信息应能反映档案整理的各个层次,精确表述档案信息的收集、鉴定、分析和组织,以便于揭示档案材料所形成的活动全貌,对其进行信息控制和提供利用。

人物档案的整理类似于普通档案整理,又不完全相同。人物档案通常以全宗形式固定。通常以大类进行区分。具体的细分类不尽相同。这些划分的标准因馆各异,仁者见仁、智者见智。从管理层级上看,通常仅包括两层,即大多数局限在案卷级和文件级的著录。

2.2.3 人物档案多级著录的必要性

多级著录的概念和实践均来源于国外。多级著录思想最早由Oliver W. Holmes^[12]提出,他提出了多级著录的5个层级,分别是存储处(depository)、文件组(records group)、系列(series)、案卷(file unit)、文档(document),依据从上至下的先后顺序进行著录,Oliver的分级整理思想为档案多级著录理论奠定了基础。随着信息技术的发展和理论研究的深入,多级著录逐渐由思想发展为理论。在现代档案著录工作中,要达到对档案控制的规范、可靠及有效,就要遵循档案著录的基本原则,即来源原则、尊重全宗原则、反映管理级次原则,这三个方面的内容又被

称为档案著录的“马德里原则”。

档案多级著录概念的提出建立在现代档案著录概念的基础之上。就社会利用环境角度而言,在数字化条件下,现代档案利用需求出现了由传统手工向计算机、网络利用扩展,由本地利用向远程利用扩展,由单级、单方面利用向多级、多层面的梯度利用过渡与转变的新特点和趋势^[13]。档案多级著录是按照“全宗整理级别结构模式”实行的,它反映了档案全宗是一个不可分散体系的档案的本质与特点,因而档案的多级著录是对全宗理论的具体体现^[14]。实现人物档案的多级著录,其必要性有如下三点:

(1)实现人物档案的多级著录,可以提高人物档案的检索效率。传统的单级著录,不能体现由全宗到类别,再到案卷,最后到文件的逻辑关系。实行人物档案的多级著录,就可以建立起各著录级别之间的逻辑联系,这样只通过一次检索,就能获得由文件到全宗的全部信息,从而提高了检索速度,节约了检索成本。

(2)实现人物档案的多级著录,是保留人物档案原貌的途径。人物档案不同于一般档案,很多档案的形成来自于个人,他们留存下来的原始档案有其自身的逻辑与思考。通常具有明显的多层级关系。对其进行层级分析及多级著录,有助于揭示人物档案本来的结构和信息。

(3)实现人物档案的多级著录,是人物档案资源开发与利用的基础。人物档案资源开发与利用是在详细描述人物档案信息的基础上进行的深度组织和挖掘。

综上,本文认为,人物档案多级著录可以多维、立体、完整地体现人物档案的特点,不仅能揭示其形式特征和内容特征,还包括人物档案的背景信息及与之相关的关联信息,丰富著录对象的信息内容,为用户直观判断检索结果是否符合检索需求提供依据,同时为用户对其他关联信息的发现创造条件,为进一步扩检、缩减、改检提供了可能。下文将以钱学森的“629袋”私人档案作为案例,探讨多级著录思想与方法的应用。

3 实证分析:多级著录在特色馆藏“629袋”中的应用

在上海交通大学钱学森图书馆(以下简称“钱馆”)的馆藏中,有一批十分珍贵的馆藏文献——“629袋”。“629袋”是经由钱学森本人整理、归类

而成,经征集流程后,成为钱馆最为重要、最具特色的一类馆藏。“629袋”的形成,是钱学森在一生坚持读书看报、不断思考、与他人学术交流等活动中将原始材料留存下来,并经本人亲自分类、归档的结果,这与钱学森自身强烈的“档案意识”不无关系。1990年,杨恂南^[15]撰文指出,“钱学森早已关心和支持档案事业,对于档案和档案事业有过高度评价和深刻分析”。其中,钱老提出了“档案是‘一个巨大的事业’;档案是‘认识客观世界的锐利工具’,是‘人感觉器官的外延’;‘档案是人类的记忆工具’等”经典论述,是全社会必须重视档案和档案事业的理论根据之一,对于档案和档案事业的建设与发展具有重大的现实意义。

3.1 “629袋”结构分析及资源类型

从内容上看,“629袋”中既有钱学森亲手裁剪并粘贴的20000余份剪报,部分标有钱学森阅读后的批注,也有从全国各地甚至世界各地寄来并希望得到钱学森指导的材料;既有钱学森参加全国力学会议、科学会议、人大会议、政协会议所形成的报告、文稿、会议材料等,也有钱学森撰写的论文提纲、初稿、修改稿等手稿资料;还有钱学森与近千名专家、学者及各类人员的来往书信(从钱学森回国至20世纪80年代早期,书信寄出前,由钱学森本人及秘书对原书信重新抄写。复印机出现后,则由秘书将书信原件复印存档,因此保留了大量的书信手稿),等等。此外,还有大量藏书、期刊及其他捐赠^[16]。

维护其原貌,保留其内部原有的顺序是整理“629袋”的第一准则。仔细审视“629袋”的内部结构,不难看出,其中存在着多级结构,如图1所示。如何科学整理与著录“629袋”,全面而系统地揭示其中所含的信息,成为我们亟须考虑并有待解决的问题。

冯惠玲、张辑哲^[17]指出,档案分类涉及三个层面:第一,对档案的认识性分类(又称概念分类),即对档案的种类认识;第二,档案实体的管理分类(又称实体分类);第三,档案内容信息的管理性分类(又称信息分类或检索分类)。其中,实体分类是人物档案分类中采用的最为常见的分类方法。从档案实体性管理的分类角度,我们曾将资源分为论著、藏书、信函、贺卡、手稿、参阅资料、实物、声像等类型。照此分类(见图2),不难发现,存在诸多类型的交叉,不利于后续资源定义及其元数据描述。正如前文所述,我们可借助于图书馆对资源类型界定的经验,对“629袋”中资料的类型进行形式上的严格

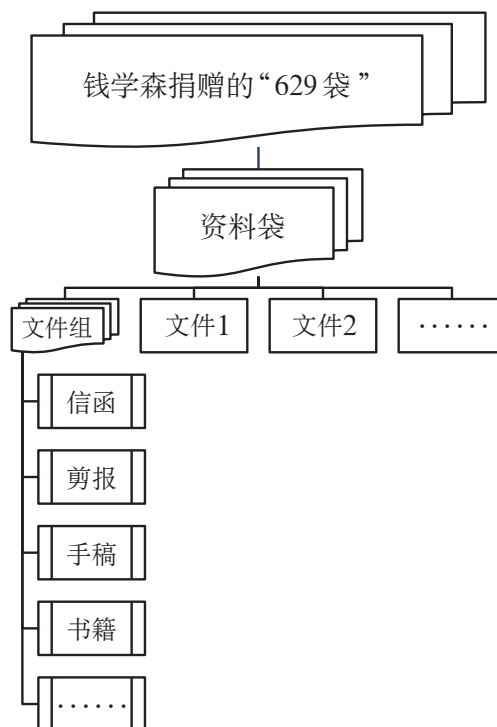


图1 “629袋”内部文件层次及关系(从形式上)

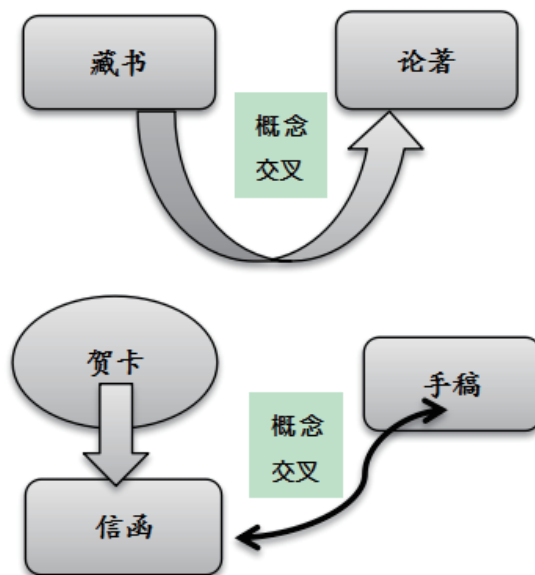


图2 按照档案实体性分类存在概念交叉示意图

界定。为此,我们对资源本身进行重新分析,并严格区分资源类型,详见表1。

分类是藏品整理与科学管理的基础工作。尤其在计算机辅助管理的情况下,更加显得尤为重要。以上分类区别于档案的一般分类,也不同于人物档案通常采用的“大块分类”,使得资源本身的交叉尽量消失,从而有助于元数据方案的制定与深化。

3.2 “629袋”的多级整理层级

“629袋”时间跨度为20世纪50年代至2004年前夕,是钱学森自己看书读报、与人学术交往过程中形

表1 钱学森图书馆馆藏资源类型定义表(部分)

资源类型	定义
书籍	钱学森本人购买、他人赠送钱学森并由钱学森收藏,存放在办公室、家中的图书、期刊。语种有中、英、德、俄等。此部分来源通常为钱学森家属捐赠。馆方自行购买的除外。
信函	钱学森与他人的各类通信,包括信件、便笺、贺卡、明信片、请帖等。
手稿	钱学森在研究、创作、讲座等活动中形成的书写文字记录,可以是草稿、未定稿、校改稿,也可以是论文。手稿是未出版的公开发表物,其每一次修订都作为一个版本,因此须对其著录到每一种的修订稿。
剪报	钱学森亲自剪贴或专门保留的报、刊。以单独一份、一页的剪报居多,也有若干份相关剪报钉在一起的情况,我们以套计数,目录里面详细罗列各剪报情况。
笔记	钱学森晚年读书看报时所作的摘抄、随想等文字记录,没有确切时间或者不是以日期作为记录顺序的。
打印资料	以打印形式呈现的材料。
证件	钱学森出席政协会议、人大会议等场合,代表其身份的文件,如出入证、通行证等。
证书	由机关、学校、团体颁发给钱学森以表彰其成就或授予其荣誉的文件。
实物	钱学森创建或使用过的,证章除外。
音像资料	以磁光介质记录的有关钱学森的视频、音频材料。

成的原始记录。它不仅反映了钱学森科学思想的形成轨迹,而且是钱学森留给世人的一笔文化与智力财富。我们对其整理遵循多级整理原则。

首先,“子全宗”的确立。将钱学森捐赠的“629袋”作为“钱学森全宗”的“子全宗”看待。因为,这批材料是钱学森晚年读书看报的真实写照,呈现的是钱学森与各界学人学术交流的记录。最重要的是,钱学森对其进行系统的收集、整理并分类。仔细审视其中的材料,我们发现它存在着复杂的文件之间的关系,如图3所示。

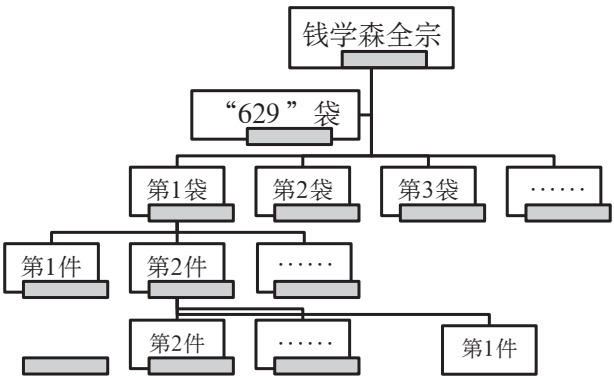


图3 钱学森特色馆藏“629袋”组织结构示意图

其次,各级的确立,包括“物”层级。简单地说,钱学森捐赠的“629袋”存在以下几层结构:子全宗级、案卷级、文件级、子文件级。鉴于博物馆的缘故,我们整理时将每袋中的文件视为“物”来揭示。例如,钱学森用于放置相关主题文件的资料袋,在档案著录中,通常将其视为案卷级信息,实际操作中,我们将其也视为“物”,与其中的文件同级,是

collection-level层级的著录。

3.3 “629袋”元数据描述著录项的确定

著录工作作为档案现代化管理的重要环节,其内容包括著录项目、著录格式、著录标识、文字、著录级别、著录来源等。本文将着重阐释著录项目,分为全宗级、子全宗级、文件级、子文件级,以示多级著录在我国人物档案著录中的应用,如图4所示。

(1) 全宗级著录项目的确定

包括:名称、数量、背后故事等。

以钱学森全宗为例,单位可为卷,也可为长度,还可作为件/套。背后故事与整个馆藏,甚至馆的总体定位、宗旨等密不可分。该层较为宏观。

(2) 子全宗级著录项目的确定

包括:名称、数量、时间、征集经过、流传经过、背后故事等。

以钱学森“629袋”为例,时间跨度为“629袋”中最早文献的时间至最晚文献的时间。对其征集、流传经过的描述是对整个子全宗级的描述,避免对每件文献重复的描述。

(3) 案卷级著录项目的确定

包括:名称、主题、数量、时间、关联等。

以钱学森“629袋”中每袋为例,我们需要描述其主题、数量、时间跨度及与其他相关案卷的关联。在“629袋”中,存在着围绕同一主题的若干案卷组合的情况。

(4) 文件级著录项目的确定

包括:定名、数量、尺寸、完残状况、关联信息等。

以钱学森“629袋”每袋中的每件文献作为描述对象。需要注意的是,这里的“定名”要参照博物馆文物藏品的定名规范来进行,为后期的文物定级工作打下基础。对尺寸、数量、完残状况进行一一描述。同样地,文献间存在着大量而复杂的关联关系,需要用关联信息加以描述。照此层级,也能较好地处理档案与文物在整理、管理、描述、利用等方面的矛盾。这里需要指出的是,关于此层的描述,应当将其视为“物”层面加以理解,这与国际著录标准与实践中所强调的collection-level一致,也与博物馆一般围绕“物”本身管理及研究的做法不谋而合。(见图4)

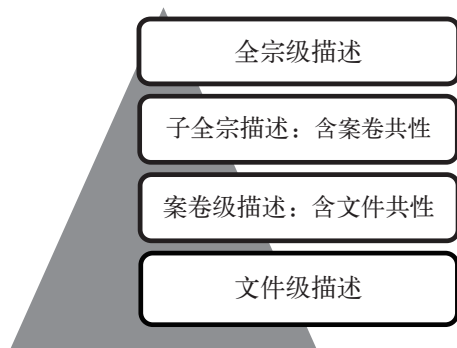


图4 钱学森特色馆藏“629袋”元数据描述层级示意图

档案多级著录,是提升档案著录质量、建设档案数据库、提高档案检索效率的关键要素之一,在当前乃至今后依旧是我国档案工作亟待解决的问题之一^[18]。档案“多级著录”这一新思想在我国档案著录实践中的最终实现将对我国档案著录体系的发展与完善起到极大的推动作用^[19]。长期以来,我国学者一直探索档案多级著录规则的制定与实施。人物档案作为一类专门档案,尤其像钱学森“629袋”这类价值特殊的私人档案,具有明显的多级结构特征,本文是对该理论与思想的应用尝试,希望能为人物档案的著录研究提供更多的参考。

本文系国家社科基金项目“大数据背景下档案数据管理理论重构、技术选优与实践创新研究”(编号:18BTQ092)、上海市社科规划青年课题“跨LAM名人数字资源整合模式研究”(编号:2018ETQ004)研究成果。

参 考 文 献

[1] 赵红艳,康鑫.近30年来我国档案著录研究综述[J].

档案,2013(1):13-16.

- [2] 尚珊,高文静.国内外档案著录准则比较研究[J].档案学通讯,2014(2):100-104.
- [3] 王大青.中国档案著录标准与国际档案著录标准比较研究[J].档案管理,2012(4):42-44.
- [4] 应海燕,乐淑芳.《档案著录规则》适用性初探[J].兰台世界,2006(9S):20-21.
- [5] 赵芳.对我国《档案著录规则》再修订的思考[J].山西档案,2009(5):34-36.
- [6] 澳大利亚国家档案馆网站档案著录与检索初探[J].北京档案,2015(3):39-41.
- [7] 徐锦.“名人档案”研究综述[J].档案,1993(1):31-33.
- [8] 王利伟.关于人物档案基本问题的思考[J].山西档案,2012(3):65-68.
- [9] 蒋友钧.人物全宗档案的管理[J].浙江档案,1993(11):13-15.
- [10] 刘伟.基于本体的数字图书馆语义互操作[D].复旦大学,2006.
- [11] 张洁,彭佳,郑巧英.图书馆学与档案学双重视角下的科学手稿组织方法研究[J].图书馆杂志,2016,35(3):74-79.
- [12] Holmes O W. Archival Arrangement—Five Different Operations at Five Different Levels [J]. American Archivist, 1964, 27(1): 21-42.
- [13] 段荣婷.论数字时代的档案多级著录[C]//中国档案学会第六次全国档案学术讨论会论文集.2002:159.
- [14] 李和平,李超,刘波.档案检索系统中的多级著录研究[J].兰台世界,2009(5):11-12.
- [15] 杨惦南.钱学森先生关于档案的两次论述及推论——兼论档案是最典型的记忆工具[J].档案学研究,1990(3):56-60.
- [16] 摘自国家社科基金项目《钱学森手稿整理与研究》结项报告之序言部分。
- [17] 冯惠玲,张辑哲.档案学概论[M].中国人民大学出版社,2006:15.
- [18] 马寅源.国内外档案多级著录的比较研究[J].档案学研究,2017(2):51-56.
- [19] 段荣婷.论数字时代的档案多级著录[J].档案学研究,2002(s1):159.