



# 日本藏书印数据库的开发利用及其启示\*

□万亚萍

**摘要** 开发藏书印数据库,是我国图书馆藏书印资源数字化整合的未来方向。日本业已建成的藏书印主题数据库中,以国文学研究资料馆的“藏书印数据库”最具代表性。文章在调查分析该数据库的数据采集方式、著录体例、检索功能、用户情况的基础上,思考藏书印数据库的利用价值,并结合我国实际,提出我国图书馆进行藏书印数据库建设应遵循“三步走”的路径,即开发馆藏印记数据库、整合域外汉籍藏书印资源、打造全球汉籍藏书印知识库。

**关键词** 藏书印 数据库 日本

**分类号** G256

**DOI** 10.16603/j.issn1002-1027.2020.05.009

藏书印,又写作藏印、藏章、收藏印、鉴藏印、钤印、印记等,广义的藏书印包括印章本身以及钤盖在书籍字画等上面的印记图像。本文主要从狭义上来讲,专指钤盖在书籍上的印记图像。

藏书印是推断一部典籍流转经过的重要线索和直接依据。在数字人文视域下,通过网络信息技术等现代科技手段整合海量藏书印信息,构建递藏关联图,显得十分必要。藏书印数据库,即运用现代技术对藏书印记图像进行采集加工,并以数字格式进行存储,是纸质藏书印资源的集成化、有序化和利用便利化,也可称作钤印主题知识库、印记资料库等。目前,国内外已开发了多个规模不一、功能各异的印记主题数据库,如浙江图书馆开发的“中国历代人物印鉴数据库”<sup>[1]</sup>、中国台湾地区“中研院史语所”傅斯年图书馆的“数位典藏资料库整合系统·印记库”<sup>[2]</sup>、日本国文学研究资料馆的“藏书印数据库”<sup>[3]</sup>等。但目前国内学者针对国内外藏书印数据库的专门研究并不多见,仅有少数学者在研究中涉及印记著录规范<sup>[4]</sup>、印记图像提取技术<sup>[5]</sup>等。拙文“经验及启示:日本藏书印资源整理开发现状考察”<sup>[6]</sup>中曾提及日本的藏书印数据库建设,但也未能展开。他山之石,可以攻玉。中日藏书印资源开发有着明显的共性与差异,日本的藏书印数据库建设经过多年发展,业已趋于完善。研究日本藏书印数据库的开发

利用经验,对我国今后藏书印资源数字化工作的开展具有重要的借鉴意义。

## 1 日本的藏书印主题数据库建设现状

在日本的藏书机构,藏书印数据库的开发已成为其藏书印资源数字化整合的主要方式,目前已开发有多个印记、实体印章或藏书印谱相关的主题数据库。经笔者调查整理,将诸数据库的情况简述如表1所示。

通过对上述数据库的对比分析可以看出,印谱数据库注重文献价值,以提供研究资料为目的;印章展示数据库是以鉴赏功能为主;印记数据库则是兼具鉴赏价值、史料价值、工具书等的研究知识平台。日本已公开运行的几个印记数据库中,以日本国文学研究资料馆“藏书印数据库”的资源最丰富,且该数据库整合国内外日本古代典籍的藏书印资源,试图打造日本藏书印知识集成数据库,实现书籍流转路径可视化功能<sup>[13]</sup>,其开发利用经验对我国藏书印数据库建设具有一定的借鉴意义。

## 2 日本国文学研究资料馆“藏书印数据库”

日本国文学研究资料馆“藏书印数据库”于2012年3月公开运行。截至2019年3月,记录数据总量已达到39686条,印影数目35148个<sup>[14]</sup>。下文

\* 天津市艺术科学规划项目“日本藏书印资源数字化开发及应用研究”(编号:DJ16003)的研究成果之一。  
万亚萍,ORCID:0000-0002-0093-2399,邮箱:wanyaping@163.com。



表1 日本免费公开的藏书印主题数据库

类别	数据库名称	数据来源	数据著录	功能价值
印记数据库	国立国会图书馆·藏书印的世界 <sup>[7]</sup>	遵从著名人物、藏书家、特色印章的原则,从日本国立国会图书馆藏书中,选出30位藏书家的印记进行展示	人物(人名、职业、履历、藏书);印影(印文、大小);载体书籍	鉴赏;印文解读工具;通过藏书印探寻藏书家的踪迹
	日本国文学研究资料馆·藏书印数据库 <sup>[8]</sup>	藏书印谱;日本国文学研究资料馆馆藏日文古籍文献;日本高校、公共图书馆藏古籍文献、电子资源;国外日文古籍网络资源	印主(ID、印主名及读音、履历);印影(ID、形状、印文、字数、释义);载体书籍(ID、书名、著者、出版信息、分类号、典藏地)	印文解读工具;书籍的流转变迁史迹可视化呈现;考证散逸藏书群;数字人文研究平台
	九州大学·藏书印图像 <sup>[9]</sup>	九州大学附属图书馆所藏日文古籍文献,以狩野、近藤、西田、樋口、逍遥、萩野等文库为主	印主(印主名、职种、年代、人物考证);印影(形状、印文、印文字数);书籍关联信息	鉴赏;馆藏溯源;印主考证
印谱数据库	东京国立博物馆·馆藏印谱数据库 <sup>[10]</sup>	东京国立博物馆所藏约2300种7500册藏书印谱(包括中国各朝代古铜印谱、中国近代名人印谱、日本近代名人印谱等)	陈列品编号、图像编号、印谱类别、印谱题名、编著者、时代、尺寸	开放存取的文献资料
印章展示数据库	怀德堂的小宇宙·怀德堂印章展示 <sup>[11]</sup>	从大阪大学怀德堂文库所收存的200余颗江户时代的印章中选出15颗进行展示	解说项(印主名、印文释义)、印面图像、印组属性、重量	3D技术应用;鉴赏
	泊园文库·泊园印章集 <sup>[12]</sup>	泊园文库现存幕末、明治时期的170余颗印章,以书法家藤泽南岳的藏品为主	序号、印章图像、印面图像、侧款、印影、印文、备考	鉴赏;印主考证

将从数据采集、著录体例、检索功能、用户情况几个方面对该数据库作扼要介绍。

### 2.1 数据采集

现行日本国文学研究资料馆“藏书印数据库”是在2006年5月进行公开试验运行的“藏书印数据库”的基础上制作而成的。原库以三村清三郎、横尾勇之助所编《藏书印谱》《续藏书印谱》和丸山季夫、安藤菊二所编《藏书名印谱(第1—4辑)》为数据源,扫描著录了419名印主的713个印影,可通过印文、印主名等检索项查看印影及印主信息。该数据库主要实现了印记的鉴赏功能,针对学术研究的价值不大。在此基础上,2010年提出了创建“新·藏书印数据库”构想,希望通过藏书印数据库考证版本价值、揭示书籍的流转变迁史迹,也可以通过同一印影(或同一印主)的藏书复原散逸藏书群等。正因为如此,新数据库中采录的藏书印,大大超出藏书印谱的收存范围,除了一些墨迹不清、印文残缺的藏书印外,采购印、图书租赁印、书店印等此类不在印谱的选录范围内的印记,也被收录其中。虽然这类商业化的印记普遍缺少鉴赏性、趣味性,但它对于考究书籍的流传史十分有价值。此外,该数据库中采录了一些藏书票、书店票以及写本制作相关人物的印

记和书画落款等,这极大丰富了其资源外延。

另一方面,日本国文学研究资料馆“藏书印数据库”的载体文献来源也十分广泛。除了本馆馆藏古籍资源和业已出版的藏书印谱外,也采集国内外各高校、公共图书馆的日本古籍文献纸本或电子资源。针对馆内外不同存储形式的藏书,其采集方式也有所不同。直接采集自馆藏纸本原书的,如国文学研究资料馆和大阪大学附属图书馆的藏书,参照《馆藏日文古籍目录》上的数据,对原书封面、扉页、正文、序跋、封底等全面审查,逐一扫描著录藏书印。针对网络公开资源,如早稻田大学图书馆和京都大学附属图书馆等的数字化古籍资源,采集时则是通过浏览镜像画面查找并著录藏书印信息。而对于外馆藏小册子、图录等上面的藏书印,仅采集有网络全文资源的,如东京大学附属图书馆霞亭文库、立教大学图书馆江户川乱步旧藏资料等。

### 2.2 著录体例

日本国文学研究资料馆“藏书印数据库”的主要存取数据包括人物信息、印影信息和书籍信息三类,各类分别建立子数据库。各子数据库下又有诸多细目,如图1所示。其中,印影信息的著录最为关键,为此也制定了诸多著录体例。例如,印影采集时,原



则上是一个印影对应一条数据,对于连珠印(联珠印)等,则把两个独立的印记作为一个整体来采集,并生成相应数据;对于只有图像没有文字的印影,在著录印文项时用“[ ]”来标注相应文字;对于图文组合的印影,一般在印文项只著录文字信息,图像信息则著录在“形状”项中;对于印影的形状、大小的著录,以工具测量辅以目测、方便检索为基本原则,如工具测量边长为 $2.2 \times 2.3$ 的印影,被视为正方形;对于不能辨识的印文,用“□”表示,而对于文字数目不确定的,则用“■”表示。

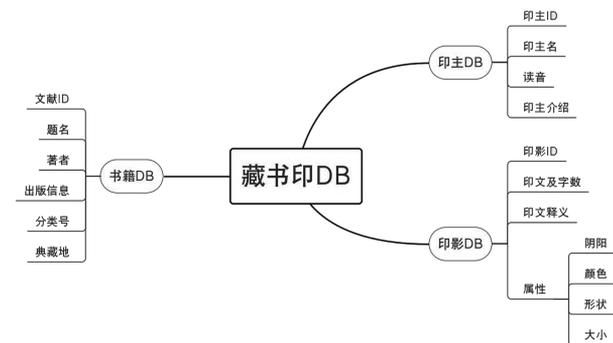


图1 日本国文学研究资料馆“藏书印数据库”存取资源类目

### 2.3 检索功能

目前,日本国文学研究资料馆“藏书印数据库”具备浏览检索、简单检索及高级检索三种检索功能,检索项主要包括典籍ID、书名、藏书印ID、印文字数及轮廓、人物ID、藏书印主、印文、印文出现位置、形状、书法字体、篆字部首等。而且,检索结果有多种排列顺序可供选择。

数据库的检索界面附设了分类检索书签目录,分为职业、朝代、藏书印ID、藏书印文、形状、人物ID、藏书印主、典籍ID等诸多类别。各类目下又有细目,并精确标明藏书印记录条目。例如,职业大类目下又细分为9类,共计55种职别;朝代大类目下又细分为古代、中世、近世、幕末明治、近代、现代6个时期;形状类目下又细分为有框、无框、圆形、井桁形、羽毛形、云朵形、不规则形等368种。此外,该数据库的简单检索功能在支持关键词检索的同时,为保证检准率,还可以附加藏书印文或印主名称。而使用其高级检索时,还可限定藏书印的尺寸、印文颜色、形状、印影轮廓、印文的行数或字数、印主职业以及文献载体的典藏库等,完全实现了多样化检索。

### 2.4 用户情况

用户的认可度也是判断一个数据库成功与否的标准之一。那么,从用户量来看,日本国文学研究资

料馆“藏书印数据库”无疑是一个成功案例。青田寿美等人曾对该数据库的用户访问量进行过统计分析<sup>[15]</sup>。2012年的总页面浏览量为2883次,检索总条目为31043条。而到了2015年,总页面浏览量达到14509次,检索总条目为50180条,不论是页面浏览量还是检索条目数量均呈现大幅度上升趋势。在2016年,总页面浏览量有所回落,为10037次,但检索总条目则达到68091条。可以说,自2012年公开运行后,其年度检索总条目一直保持上升趋势。

另一方面,基于该数据库的文史学术研究逐渐增多。早在2011年,大贯俊彦就曾发表了利用藏书印数据库进行的文化、文学相关研究设想<sup>[16]</sup>。当时,数据库还处于馆内试运行阶段。大贯俊彦提议利用该数据库中东京帝国大学图书馆的旧馆藏图书的藏书印资源,研究关东大地震后东京帝国大学图书馆的复兴过程,以及支援其馆藏再建的藏书家、寄赠者们的信息。2012年,该藏书印数据库开放运行以来,青田寿美等人发表了一系列研究成果,例如“以日本国文学研究资料馆藏藏书印数据库为起点、走向书籍·人物·知识的互联网+”<sup>[17]</sup>、“基于藏书印数据库的数字人文研究展望”<sup>[18]</sup>、“大阪府立大学图书馆史溯源——基于藏书印的考察”<sup>[19]</sup>等。

## 3 思考:藏书印数据库的利用价值

详细著录载体书籍信息,是日本国文学研究资料馆“藏书印数据库”等印记数据库的重要特征。通过印文、印主与文献载体信息的相互关联,揭露典籍递藏经过、印主梗概等关联信息,生成的海量藏书印知识主题库,将是搭建藏书史研究知识平台的重要基础。正因为如此,藏书印数据库在印文解读、揭示递藏关系、实现跨学科研究知识平台等方面具有重要价值。

### 3.1 印文解读“工具书”

藏书印对于文学历史等人文社会科学研究的价值不言而喻,但长期以来,这一价值始终未被充分挖掘。其原因无怪乎藏书印文、藏书印主的辨识难度太大。尤其是非文字学专业背景、不懂篆刻的研究人员,在辨识藏书印文这一环节时就感到十分棘手,更别提基于印文信息对藏书印主进行考证。即便是求助于篆刻字典等,也未必能够准确解读印文。试想,若整合全国藏书印资源建成一个汉籍藏书印数据库,由专家对采集到的海量印记信息源进行解读并规范著录。那么,无形之中,印记数据库便可以成



为功能庞大的藏书印事典。它有可能取代传统的书法字典,成为解读印文的重要网络工具书。而且,依托印记数据库强大的检索功能,有效利用印文、印文字数、印影形状、书法字体、篆字偏旁等各检索项,将能够大大缩短研究者查找、辨识印文的时间,提升研究效率,对于从事文学历史研究的普通科研人员而言,大有裨益。

### 3.2 藏书印主题知识关联网络

在藏书史上,“聚与散是一对辩证存在的孪生物”<sup>[20]</sup>。然而,学者们往往对于“聚”研究较为深入,从大量史料中旁征博引,研究涉及藏书家生平、收藏兴趣、聚书途径、藏书保护措施等等。而对藏书的“散逸”则往往点到为止,对散逸的原因分析较多,而对散逸图书的去向则语焉不详,存在不全和不实的缺憾。究其原因,在于茫茫书海中无从查找散逸藏书的蛛丝马迹。这一研究瓶颈正是需要通过开发印记数据库来打破,将相对独立的藏书印图像和印主信息、书籍信息等通过技术手段关联呈现出来,为获得十分宝贵的隐性知识——典籍流转历程创造了条件。通过对数据库的深度挖掘,构建藏书印主题知识关联网络,从而揭示典籍流转路径、探求藏书集来源、复原散逸的藏书集合等。如前所述,日本学者大贯俊彦、青田寿美等人已进行过诸多实践,不再赘言。理论上讲,历史上曾经存在的、已经散逸的藏书集合,只要书籍未被毁损,数据得到采集,都有可能通过藏书印数据库集成平台进行复原。

### 3.3 跨学科数字人文研究平台

近年来,数字化技术在史学研究中普遍应用。在数字人文的浪潮下,“数字史学”“e-考据”等新兴词汇不绝于耳。各类历史资料数据库被不断地开发建立,投入使用。如何在符合学科特点的基础之上,更有效地利用信息技术,以推动史学相关研究的深入发展乃是新的挑战。不仅是历史学,诸多人文社会学科都受到了数字人文研究的冲击,研究学者需要搭建各种跨学科研究信息平台。而以古籍为研究基础的学者们,也需要一个共同的平台。另一方面,随着数字人文研究领域及理念的不断深入,近年来,众多图书馆均开展了数字人文研究的重要实践。未来的图书馆“不再只是研究的旁观者,而是研究的参与者与合作者”<sup>[21]</sup>。以古典籍为核心的数字人文研究平台,必须由资源所有者——图书馆来构建。而藏书印数据库作为古籍资源库的重要组成部分,可以将艺术学、历史学、文学、社会科学、图书馆学、计

算机科学等众多领域的研究者聚集到一起,打通数字人文研究领域的各个环节,促进跨界融合和对话,降低数字人文的研究成本,有效促进学术资源的共享共建。因此具有重要的学术价值。

## 4 启示:我国图书馆藏书印资源数字化开发的路径

如上所述,藏书印数据库在印文解读、揭示递藏关系、实现跨学科研究知识平台等方面具有重要的利用价值。因此,开发藏书印数据库,也是我国图书馆藏书印资源数字化整合的未来方向。我国图书馆收存的古籍资源极其丰富,建设特色馆藏数据库的技术业已相当成熟,研究学者群体也大量存在,藏书印数据库开发的理想条件似乎已然具备。然而同时,我们也不得不面对现实,藏书印数据库的建设是需要资源方、技术方和研究者多方协作才能完成。目前我国大多数图书馆缺少藏书印数据库建设所需的人力、财力和物力,仅有少数有实力的图书馆能实施数据库的开发,但也缺乏从史料中挖掘深层次细节信息的数字技术。由此来看,我国藏书印资源数字化整合开发尚任重道远,需循序渐进、逐步深入。笔者以为,我国图书馆藏书印资源数字化开发可以遵循三步走的原则,即首先从馆藏古籍印记数据库的建设着手,分散采集,统筹建库,然后整合国外汉籍藏书印资源,最后是打造全球汉籍印记知识库。

### 4.1 分散采集,统筹建库

加大馆藏印记资源整理力度,开发馆藏印记数据库,是目前我国古籍藏书印资源数字化开发的重点。单个馆藏藏书印资源的开发,对资源和技术都有较高的要求,需要多领域专业人士的协同工作。首先要进行印记数据的采集、分析、管理与数据挖掘等一系列复杂数据的处理;然后设计一种立体的、动态的、实时交互的表现形式;最终创建对应的可视化算法及技术实现手段。这对于人力、财力欠缺的部分图书馆来讲,无疑是难于实现的。笔者以为,我国印记数据采集和数据库建设需统筹管理,即以省市为单位分别采集数据后,由主导部门进行数据整合,建成全国性的藏书印数据库资源平台。

我国于2014年10月10日正式开通的“全国古籍普查登记基本数据库”,随着古籍普查登记工作的推进,陆续发布更新普查登记数据。截至2018年9月,已累计发布169家单位所藏672467部6541261册古籍的普查数据<sup>[22]</sup>。该数据库的元数据涵盖普查编号、索书号、题名、著者、版本、册数、存缺卷、收

