

面向人文研究的“数据基础设施”建设 ——试论图书馆学对数字人文的方法论贡献*

夏翠娟

摘 要 数字人文不仅革新了人文研究的范式,还意味着全新的知识生产方式和知识交流模式。联合各领域的研究者和研究机构,包括各类美术馆、图书馆、档案馆、博物馆(GLAM)等文化记忆机构建设数字人文基础设施,以支撑新的研究范式、知识生产方式和知识交流模式,已经成为共同的呼声。探讨如何建设数字人文基础设施,厘清各种责任主体尤其是 GLAM 机构在数字人文基础设施建设中扮演何种角色、承担何种任务、做出何种贡献,有利于帮助图书馆找准角色定位,发挥自身优势,成为整个数字人文基础设施建设中不可或缺的有机构成单位。利用文献调研、需求分析、案例调研、实践验证等方法,将“数据基础设施”从“研究基础设施”中剥离出来并作出定义;分析面向人文研究的“数据基础设施”的一般性功能需求和技术规范,包括跨机构的共建共享、跨网域的开放获取、跨领域的知识融通、跨时空的版本迭代四个方面。以上海图书馆的数字人文实践为例,从原理和实践两个层面论证图书馆的书目控制、规范控制、知识组织、文献循证方法和关联数据、知识图谱、大数据等技术结合起来,有望为“数据基础设施”的建设提供一定的方法论贡献。联合各类机构加强新技术的应用,推进数据开放互联和知识融通,建设公共性、开放性、持续性的“数据基础设施”,是 GLAM 机构在数字人文大潮中应尽而可尽的责任。图 2。参考文献 28。

关键词 数字人文 数据基础设施 书目控制 规范控制 知识组织 文献循证

分类号 G254

The Construction of “Data Infrastructure” for Humanities Research: The Methodological Contribution of Library Science to Digital Humanities

XIA Cuijuan

ABSTRACT

Digital Humanities is not only the innovation of the research paradigm of humanities research but also means a brand-new mode of knowledge production and knowledge communication. Therefore, there comes a trend that it needs to combine researchers and research institutions in various fields, as well as the cultural memory institutions such as Galleries, Libraries, Archives, Museums (GLAM) to build Digital Humanities infrastructure to support new research paradigm, knowledge production and knowledge communication. This

* 本文系国家哲学社会科学基金一般项目“支撑城市记忆项目的‘数据基础设施’理论建构与实践探索”(编号:19BTQ007)的研究成果之一。(This article is an outcome of the project “The Concept and Application Exploring about the Construction of Data Infrastructure to Support Urban Memory”(No. 19BTQ007) supported by National Social Science Foundation of China.)

通信作者:夏翠娟,Email:cjxia@libnet.sh.cn,ORCID:0000-0002-1859-6979 (Correspondence should be addressed to XIA Cuijuan,Email:cjxia@libnet.sh.cn,ORCID:0000-0002-1859-6979)

paper aims to explore how to build Digital Humanities infrastructure, clarify the roles, tasks and contributions of various responsible subjects, especially GLAM institutions, so as to help libraries to identify the role, orientation, advantages and become an indispensable organic unit in the whole Digital Humanities infrastructure. By using the methods of literature research, requirements analysis, case study and practical verification, the concept “Data Infrastructure” is separated from “Research Infrastructure” and defined. It is considered that the “Data Infrastructure” of Digital Humanities is based on the “data”, which fully embodies the characteristics of large scale, long time coverage, wide geographical scope, small description granularity and multiple dimensions of “data”. It is an open, public and sustainable “middle-end” data platform which is independent of specific application development and research in specific fields and between the “background” of information infrastructure and the “front-end” of research in specific fields. Based on the investigation of the current situation of theoretical research and practical application of infrastructure in the fields of humanities and GLAM, this paper analyzes the general functional requirements and technical specifications of “Data Infrastructure” for Digital Humanities research, including four aspects: cross institutional co-construction and sharing, cross network open access, cross domain knowledge consilience, and cross space-time version iteration. Taking the Digital Humanities practice of Shanghai Library as an example, this paper demonstrates that the combination of Bibliographic Control, Authority Control, Knowledge Organization, Evidence-based practise methods and Linked Data, Knowledge Graph, and Big Data technology is expected to provide some methodological contribution to the construction of “Data Infrastructure”. At the same time, that the GLAM institutions have accumulated massive, structured and standardized open data and resource integration experience lays the foundation for the construction of “Data Infrastructure”. Finally, the following conclusions are drawn: the specific research work should be handed over to the humanities researchers. GLAM institutions are committed to strengthening the application of new technologies, promoting data openness, interconnection and knowledge consilience, building a public, open and sustainable “Data Infrastructure”, and providing new research paradigm, knowledge production and knowledge communication to support humanities research. Those are what GLAM institutions should and can do in the Digital Humanities tide. However, this paper puts forward the concept, definition, general functional requirements and technical specifications of “Data Infrastructure”, the demonstrated contribution of library methodology, and the role of institutions in the construction of Digital Humanities Data Infrastructure, which need to be discussed in depth with more GLAM practitioners and researchers in specific fields and further tested during the practice in the future. 2 figs. 28 refs.

KEY WORDS

Digital humanities. Data infrastructure. Bibliographic control. Authority control. Knowledge organization. Evidence-based practice on literature.

1 引言

1.1 背景

随着数字人文的兴起,大数据、语义网、机

器学习、GIS、数据可视化、VR/AR/MR 等技术被广泛地应用到人文社科领域的研究过程中,更好地支持基于大规模数据、事实、知识的量化研究和循证实践,催生了文本分析法、社会网络关系分析法、空间分析法等新型研究方法,为人文

社科领域同时提供了用于宏观研究的“望远镜”和个案研究的“显微镜”,被认为革新了人文社科领域的研究范式。但数字人文带来的影响远不止如此,由于开源、众包等理念的引入以及数字出版和开放获取的渗透,数字人文已经从根本上触及人文社科领域的知识生产方式和学术交流模式的革新^[1]。对于如历史学、文学、语言学、艺术、人类学等人文社科领域的单个学科来说,数字人文提供的方法共同体(Methodology Commons),具备了将不同学科的学者汇聚到一起的条件,尽管各自研究的问题大相径庭,但仍可在方法论和研究范式的层面互相参考、借鉴、启发,对促进跨领域的对话、弥合学科之间的界限起到了积极的作用,更进一步地推进了技术、资源、数据和知识的分享和重用。

新技术带来的数据存储规模的扩大、分布式计算和自主学习能力的增强,使得数据成为一种支持研究、创新和服务大众的新型能源^[2]。互联网上越来越多的开放数据集可被按需使用,数据将与水、电、气等一样可以开放获取,作为即插即用的公共资源,联合各领域的研究者和研究机构,包括各类美术馆、图书馆、档案馆、博物馆(GLAM)等文化记忆机构建设数字人文基础设施,已经成为共同的呼声^[3]。然而,到底什么是数字人文基础设施,如何建设?各种责任主体尤其是GLAM机构在其中扮演何种角色、承担何种任务、做出何种贡献?厘清这些问题,有助于图书馆找准角色定位,发挥自身优势,成为整个数字人文基础设施建设中不可或缺的有机组成单位。

1.2 “数据基础设施”与“研究基础设施”

“基础设施(Infrastructure)”一词源于水、电、煤、交通等生活公共设施,随着信息时代的到来,衍生出满足人们信息交流、网络接入等基本需求的“信息基础设施”(Information Infra-

structure)。而“Cyberinfrastructure”一词则被美国国家科学基金会(NSF)用作互联网时代支撑研究活动的数字环境的总称,包括数据采集、数据存储、数据管理、数据集成、数据挖掘、数据可视化等分布在互联网上、超出单一机构范围的计算和信息处理服务^①,在英国被称为“e-Science”或“e-Research”,在欧盟则被称为“Research Infrastructures”。数字人文基础设施被归为一种支持数字人文研究活动的“研究基础设施(Research Infrastructure)”,包括数字化的文献资源、数据库、工具平台、支持知识生产和信息交流的网络空间等,需由数字出版商、文化记忆机构、研究机构、研究者、IT支持人员共同参与构建。本文提出的数字人文“数据基础设施”(Data Infrastructure),作为数字人文研究基础设施的一部分(见图1),试图聚焦于数字人文研究基础设施中的内容、数据和知识部分的生产和组织,探讨如何构建独立于具体应用开发和领域研究问题之外的数据层,及其长期保存、共建共享和开放利用的技术规范。

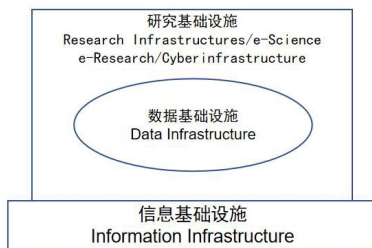


图1 “数据基础设施”与“研究基础设施”

“数据基础设施”以“数据”为基本的组成单位。“数据”最初的定义是“可被计算机传输和存储的信息”^②,对于人文研究来说,“数据”可理解为可被机器处理的信息单元,如文献或实物资源对象、概念、人物、机构、团体或其结构化的描述信息(包括变量、数值、文字符号或事实等)。数字人文“数据基础设施”,除了具备“基

① <https://en.wikipedia.org/wiki/Cyberinfrastructure>

② <https://en.wikipedia.org/wiki/Data>

基础设施”的开放性、公共性和可持续性,还应充分体现数据规模大、覆盖时间长、地域范围广、描述粒度小、维度多等特点,以支持全网域(Web-scale)的数据调度、融合和自动化分析、统计,以及数据可视化。同时还应独立于具体应用开发和特定领域研究之外,遵循通用的数据组织规范和开放共享规范,成为介于信息基础设施“后台”和特定领域研究“前台”之间的“数据中台”^[4]。

本文试图通过分析“数据基础设施”建设的一般性功能需求和技术规范,结合上海图书馆的实践,来探讨 GLAM 等文化记忆机构在建设数字人文研究基础设施中的角色、侧重点和着力点,以及互联网时代的书目控制、规范控制、知识组织、文献循证等图书馆学方法在数字基础设施建设过程中对数字人文的方法论贡献。

2 文献调研

在 GLAM 领域,关于“基础设施”的提法由来已久。肖珑从国家战略的角度出发,认为人文社会科学是“文献依赖型学科”,而文献资源保障体系建设是我国人文社会科学学科发展需要的软性基础设施,并对其需求、现状及其宏观战略规划进行论述^[5]。赵生辉在总结了国外档案领域数字人文项目具备“多重功能复合”“多元学科属性”“面向知识整合”等特点的基础上,提出我国档案界应参与“国家数字人文基础设施建设工程”前期规划,建设“若干体现我国特色文化资源的数字人文仓储”,形成各类资源“相互补充、互相联系、交相辉映”的格局^[6]。这些研究体现了图书馆界和档案界在人文研究基础设施建设中对文献资源的重视。

近年来,数字人文的外延不断扩大,成为涵盖数字化资源、数据技术、平台和工具、研究环境的“大帐篷”,对“基础设施”的认识也随之变化,开始关注除了数字化资源建设之外的工具、平台、环境、制度等重要因素。刘炜等深入地探讨了数字人文基础设施的定义、构成,认为数字

人文基础设施“是指在数字环境下为开展人文研究而必须具备的基本条件,包括全球范围内与研究主题相关的所有文献、数据、相关软件工具、学术交流和出版的公用设施及相关服务等”,并对图书馆如何建设数字人文基础设施,和建设国家层面的数字人文基础设施提出了建议和意见^[7]。

除了 GLAM 机构,人文社科研究者们也认识到建设基础设施的必要性和重要性。2018 年 3 月,来自汉学领域的 100 余名世界各地的数字人文学者参加了上海哈佛中心举办的“中国历史研究的网络基础设施国际研讨会”(International Conference on Cyberinfrastructure for Historical China Studies),共同探讨如何连接各种资源库、专题数据库、工具平台,建设面向中国历史研究的网络基础设施。在 2018 年上海图书馆国际论坛的数字人文分会场,哈佛大学包弼德教授以哈佛燕京历代人物资料库(CBDB)为例,发表了题为“数字人文与中国研究的网络基础设施建设”的报告,分析了数字人文的方法和工具为人文研究带来的智能和理论层面的飞跃,为了避免重复建设,连接信息孤岛,提出了联合建立“全球智慧数据平台”的愿景^[8]。

对数字人文基础设施建设中数据层面的整合和互操作的关注也开始显现。2018 年 12 月在台湾地区的数字人文大会上,德国马普研究所介绍了“RISE”(Research Infrastructure for the study of Eurasia)项目。该项目旨在通过定义一套标准化的 API 和元数据描述框架,帮助数据库提供者更好地描述内容,提供标准化的数据访问接口,以构建整合汉学领域各种数据库和研究工具的统一平台,为研究者回答有哪些资源库、资源库中有什么文献、文献如何获取等问题。其主要目的是在现有知识产权框架下,在元数据规范和技术应用的层面实现多源异构资源的一站式获取。

由欧洲众多高校和研究机构的历史学、考古学、社会学等多个人文研究领域协作构建的“赛莎特全球历史数据银行”(Seshat: Global His-

tory Databank)^①是一个跨领域的大型历史人文数据基础设施,目前集成了 211 611 个数据点(Data Points),采用关联数据技术,将不同来源和领域的的数据经过清洗转换后以 RDF 格式存储在三元组存储库(Triple store)中,数据集之间可以在不同的粒度(资源集合、元数据、资源对象、内容中的实体)上建立关联,同时也可以方便地与外部数据集关联^[9]。与 RISE 项目相比,SESHAT 项目不仅解决了多个数据集的一站式获取问题,还利用关联数据技术解决了跨领域的数据融合问题。

3 “数据基础设施”建设的需求和技术规范

对于各类 GLAM 机构来说,“数据基础设施”建设注重的不仅是传统馆藏资源的数字化、服务的有序化,而更强调资源描述信息和内容的数据化以及数据间逻辑关系的形式化(机器可处理)。同时要求在各机构之间实现跨机构的共建共享,在资源对象和数据单元层面实现跨网域的开放互联、跨领域的知识融通和跨时空的版本迭代。

3.1 跨机构的共建共享

由原文化部、财政部共同发起并组织实施的三项重大公共数字文化工程(简称“三大工程”),即“全国文化信息资源共享工程”“数字图书馆推广工程”和“公共电子阅览室计划”,使图书馆尤其是公共图书馆在数字资源建设和服务的跨机构共建共享方面取得了一定的成就:在光纤网络、系统软件、资源建设、服务模式、运行机制上形成了一定的标准,初步完成了信息基础设施建设;在资源建设方面也制定了一致性的加工标准,如文本、图像、音频资源的加工标准和各类元数据规范等,建设了数千 TB 的数字资源和分布式的资源库群^[10],为“数据基础设施”的建设奠定了一定的基础。高校图书馆

经过中国高等教育文献保障系统(CALIS)、大学数字图书馆国际合作计划(CADAL)等项目在跨机构的共建共享方面也取得了类似的成就。然而,“数据基础设施”建设对跨机构的共建共享提出了更高的要求。

(1)覆盖各类 GLAM 机构。国外 GLAM 机构呈现出鲜明的一体化趋势,如由美国国会图书馆牵头的“美国记忆”项目、“欧洲数字图书馆(Europeana)”项目、“威尼斯时间机器”项目就是由各类 GLAM 机构共同参与的。国内的“三大工程”、CALIS、CADAL 主要涵盖的还是图书馆,尚未涉及档案馆、博物馆等其他文化记忆机构。同一主题的文獻资源由于载体类型、管理机制的不同或流藏际遇的差别而分散在不同种类的机构中,而数字人文“数据基础设施”面向的是各类特定主题的研究问题,需要超越资源的物理位置和载体的限制。例如由国家古籍保护中心发起的古籍普查工作,已经超出了图书馆的范围,将博物馆的古籍也纳入其中,构建了“全国古籍普查登记基本数据库”。一些特色文献资源的总目编纂工作如《中国家谱总目》除了图书馆、博物馆、档案馆等机构,还囊括了社会或民间团体如宗亲会 and 个人的收藏,基于此上海图书馆构建了“中国家谱知识服务平台”。这样的跨机构数据库有必要成为“数据基础设施”的有机组成部分。

(2)在资源对象和数据单元层面实现跨机构的共建共享。面向人文研究的“数据基础设施”需要支持基于大规模、细粒度、长时间、全量级、多维度数据的调度、融合、分析、统计、可视化。目前国内 GLAM 领域的共建共享停留在资源库层面,初步实现了资源库建设的整体规划和跨机构访问,但尚未深入到单个文献资源对象的层面,例如 FRBR 中的第一组实体对象:作品、内容表达、载体表现、单件,更毋庸说深入到文献中的数据单元层面,例如 FRBR 中的第二组实体对象:概念、主题、实体及其结构化描述数

① <http://seshatdatabank.info/seshat-about-us/seshat-who-we-are>

据。仅仅资源库层面的共建共享无法满足资源对象层面的共享和数据单元层面的计算。要在资源对象和数据单元的层面实现跨机构的共建共享,需要细化描述对象的粒度,不仅描述资源对象,还应描述数据单元,并采用唯一的、全局的、永久的标识符对资源对象和数据单元进行标识和定位,支持跨机构的访问和调度。与 ISBN、DOI 相比,HTTP URI 是一种可在互联网上唯一标识和定位的统一资源标识符。

3.2 跨网域的开放互联

哈佛燕京历代人物资料库(CBDB)一直被作为数字人文研究基础设施的典范,为大量的研究者提供了数据支撑。CBDB 在向广大研究者开放之初,采用的是整库下载的方式,后来开始提供基于互联网 HTTP 协议的应用程序开发接口(Restful API),辅以翔实的数据结构说明书,在互联网上供开发人员编程调用。这种方式也被 CText 项目、上海图书馆数字人文项目等采用,已逐渐成为数字人文领域数据库或平台建设的默认需求。然而,“数据基础设施”跨网域的开放互联不限于此,可以总结为以下三个方面。

(1) 支持跨网域的开放访问。在互联网时代,跨网域的开放访问是“数据基础设施”的基本要求。“数据基础设施”中的数据有着作为公共资源的公共性,需要超越时间和空间的限制,超越系统(如图书馆自动化系统)、网域(如机构局域网)、领域(如图书馆领域)的限制,实现在全网域(Web-scale)范围内的开放访问。这意味着数据可以在任何时间、任何地点被访问,支持即插即用。

(2) 支持机器(应用程序)调用的开放数据服务。“数据基础设施”提供的服务不仅仅是支持人通过浏览器的查阅,更重要的是为机器提供开放数据服务,支持开发人员编写应用程序调用数据,根据客户端请求返回所需的数据,支持按需利用,满足多层次、个性化的数据消费需求。Restful API 是目前数字人文项目常用的开

放数据服务方式,能在一定程度上满足机器调用数据的需求。面向机器的开放数据服务可分为两个层面:一是数据实例的层面,二是数据结构定义(Schema)的层面。目前类似 CBDB 这样的数字人文平台提供的 API,数据可被机器处理,但数据结构的定义和说明则主要供人读。

(3) 支持资源对象和数据单元层面的开放互联。整库下载、Restful API 的方式都存在一定的不足之处。首先,整库下载的方式需要用户细致深入地了解数据结构,且用户将数据获取到本地后不能实时更新;而 Restful API 则往往会受到服务器端性能、返回数据量的制约。其次,整库下载、Restful API 主要面向的用户群体是具备一定开发能力的研究团体或个人,“数据基础设施”的用户还包括各种文化记忆机构,其需求主要是调取本机构缺乏的数据与已有的数据融合,为本机构的用户提供更丰富的查阅、展示服务,对数据的实时性有较高的要求。因而支持资源对象和数据单元层面的开放互联可作为前两种方式的补充,即其他机构通过 API 查询所需的资源对象或数据单元,获取其 HTTP URI、元数据或对象数据(扫描影像、全文文本、音视频),在本机构的应用中提供其他机构资源对象或数据单元的链接或详情。这样各机构可以在保有本机构的数据资产(数据存储在本机构的服务器上)的情况下,通过互联网为其他机构提供实时的数据服务,同时增加本机构资源的可见度,互相导流。

3.3 跨领域的知识融通

数字人文“数据基础设施”需要支持面向特定研究主题的多源异构数据的融合,在数据融合中产生新的知识。在知识管理领域的 DIKW 模型中,知识(Knowledge)位于数据(Data)和信息(Information)之上、智慧(Wisdom)之下。知识是从数据的定量到定性的分析,通过归纳、演绎、比较等手段进行挖掘,并融入已有的知识体系而得到,面向决策和行动。而特定研究主题的数据往往来源于不同的机构,采用不同的领

域描述规范和多样性的数据编码格式存储于不同的数据管理系统中。一方面,这种融合往往发生在机器与机器之间,这就要求采用一致的知识表示方法,以支持来自不同机构的数据单元跨越各自的系统、平台、格式的限制,基于互联网进行数据的调度、传输和交换。另一方面,需要有高层的、抽象的知识共享模型,超越领域元数据规范,在特定应用场景中进行知识的分享、重组和再生。

(1)采用一致的知识表示方法以支持跨领域的数据传输和交换。不同机构的资源对象描述元数据元素和著录规则往往存在着差异,例如图书馆系统的大量数据是基于 MARC 标准的,而档案馆系统的数据则是基于 EAD 标准的,同样是 MARC,还有 MARC21、UNIMARC、CNMARC 之分。这为系统与系统之间跨领域传输和交换数据带来了障碍。图书馆领域以往通过制定行业标准在系统和数据编码格式层面来解决这一问题,但涉及跨领域的数据传输和交换,就不仅仅是数据格式层面的互操作,还需要借助更高层面的“知识表示”方法,达到领域知识的共享和融通。“知识表示”是把人类知识表示成机器能处理的数据结构和系统控制结构的策略,是人工智能和认识科学中最热门的研究课题之一^[11]。目前应用较为广泛的知识表示方法有谓词逻辑表示法、语义网络表示法、面向对象表示法、框架结构表示法、基于本体的知识表示法等,其中基于本体的知识表示法被认为有助于实现领域知识的共享和重用,促进跨领域的知识交流^[12]。基于 W3C 的推荐标准 RDF 的知识表示方式,随着关联数据、知识图谱技术的成熟和普及,在 GLAM 中得到了广泛的应用。

(2)采用统一的知识建模方法以支持跨领域的知识分享和重组。“知识提取和建模”(KCM,简称知识建模)是一种跨学科的方法,用于在数据中抽象出可共享的知识模型、提取知识并将其转换为可重用的格式,以便保存、改进、共享、替换、聚合和重新应用它,提供了一种从本地专有解决方案到将知识模型嵌入到跨领

域的大型计算解决方案转变的方法^[13]。知识建模方法在知识管理、语义网、人工智能等领域应用广泛,相关技术有决策树、贝叶斯网络、人工神经网络、基因进化模型、统计模型、知识图谱等。其中语义网领域的知识图谱技术采用本体工程的方法,为领域概念、术语给予明确的可被机器理解的定义,在不同领域的概念和术语之间建立映射,实现跨领域的知识共享和融合。GLAM 领域因应互联网时代的趋势所制定的各种元数据标准规范(主要是 DC 和基于 DCAP 的各种扩展)和数据模型(BIBFRAME, EDM, CIDOC-CRM)是一种基于本体的知识建模方法^[14]。

3.4 跨时空的版本迭代

数字人文不仅革新了人文研究的范式,也改变了人文领域的知识生产方式,这是互联网和大数据技术提供的红利。在分布式网络环境下,包括领域专家、图书馆员、程序员、界面设计人员在内的虚拟松散跨学科团队一起合作和互动,随时随地进行知识分享和再生产。数字出版降低了数据库商更新数据的成本,数据库更新的频率极大提高。GLAM 机构的工作人员日常知识生产和整序工作的成果也可以频繁导入数据库,除此之外,领域专家和研究者、终端用户也成为出版商和 GLAM 机构知识生产流程的一部分。数字人文“数据基础设施”的开放性,不仅体现在数据的开放访问,也体现在数据的开放更新上,版本迭代超越时间、空间的限制,知识的生产不再以原始生产者为中心。

(1)支持实时版本更新和历史版本追溯。新的数字环境使得数据的更新和发布变得空前便捷。“数据基础设施”以数据为基本单位,其中的每一个数据单元、资源对象的描述元数据元素和值,以及彼此之间的关联关系都可以独立地完成增、删、改的整个生命周期,每一次更新的记录均可保存。一个资源对象或数据单元、多个资源对象组成的资源集合和多个数据单元组成的数据集,不同的版本可以方便地汇

集在一起,追溯其每一次更新的前后变化,由谁在何时何地进行的更新都历历在案,展现不同版本比照的证据,这些证据将供后来者进行数字世界的循证研究。

(2) 支持用户贡献内容。在传统的知识生产环境下,用户往往作为知识的被动接收者,而新的数字环境正在重建 GLAM 机构的知识生产者与用户之间的关系。数字人文“数据基础设施”的用户包括文化记忆机构、研究机构、研究团体和个人、程序员、数据艺术家、第三方公司等,在与数据的互动过程中,用户可能会发现已有数据的错误、不足,或收集新的数据,或产生新的想法,这些都可以作为新的知识更新到“数据基础设施”之中,甚至用户的访问记录、使用偏好都可以利用大数据技术纳入到“数据基础设施”中,从中挖掘出有价值的知识。数字人文“数据基础设施”需要开放相应的数据更新接口,支持用户贡献内容。例如 CBDB 除了提供数据查询的 API 外,还提供数据更新的 API。上海图书馆建设的“历史文献众包平台”,支持用户对各类手稿、抄本进行元数据标引和全文抄录。

4 图书馆学的方法论贡献

“数据基础设施”所要求的数据规模大、覆盖面广、时间跨度长,及其开放性、公共性、可持

续性等特点,决定了由社会公共制度支持、长期致力于资源保存、文化遗产和知识整序的 GLAM 机构必将成为其主要责任主体,而以盈利为目的的出版者成为其上游数据提供方,专门领域的研究人员则在面向特定问题的研究过程中发现、创造新的知识,为其提供高质量的、专业的、特定领域的数据补充。

图书馆的书目控制和规范控制、知识组织、文献循证等方法,与语义网、大数据、关联数据、知识图谱等新技术结合后,焕发出新的生命力,催生出大量的智慧数据^[15],有望在数据层面将不同文化记忆机构、社会团体、个人的文化记忆资源,以及研究机构和领域专家的研究成果和知识,与有形的物质文化遗产有机地整合、有序地组织起来,构建互联互通的“数据基础设施”,形成独立于具体应用开发和特定领域研究的“数据中台”。上海图书馆在多年的数字人文项目建设中初步形成了一套将图书馆学方法和新技术结合起来建设“数据基础设施”的路径和技术框架(见图2)。该框架用基于本体的知识建模和知识表示方法(知识组织),将各种不同的文献知识库(书目控制)和基础知识库(规范控制)的数据融合起来。其中基础知识库通过 Restful API 的方式为文献知识库提供服务,以 HTTP URI Link 的方式提供基于资源对象和数据单元层面的数据链接;将文献服务(查阅服务)和

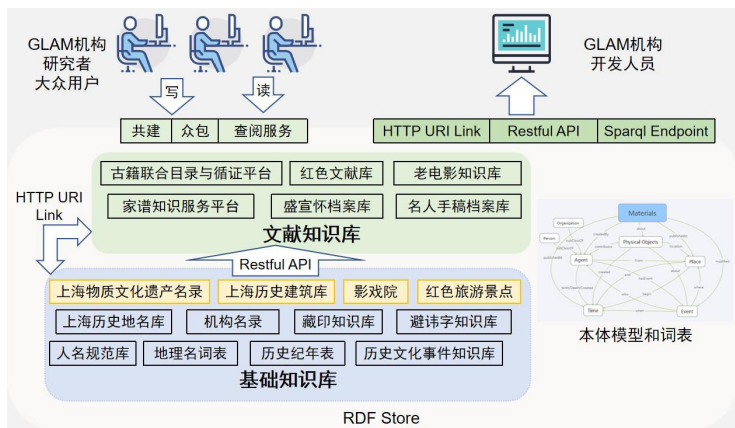


图2 上海图书馆历史人文“数据基础设施”

开放数据服务 (HTTP URI Link, Restful API, Sparql Endpoint) 与“数据基础设施”分离开来,使得不同的服务(具体应用开发和特定领域研究)可以共享“数据基础设施”提供的数据,同时新的服务应用开发过程中对已有的数据进行丰富和补充;在提供查阅服务的过程中以包包的方式支持 GLAM 机构共建,支持研究者、大众、第三方公司贡献内容,通过提供开放数据服务支持 GLAM 机构和开发人员链接、调用、丰富数据。“数据基础设施”的作用在上海图书馆为用户提供家谱、古籍、手稿、档案、电影等各种新型文献服务和支撑四届“开放数据应用开发竞赛”中得到了一定的印证。

4.1 书目控制

书目控制是现代图书馆的核心职能之一,20 世纪 50 年代,“控制人类已出版的全部文献”成为图书馆的使命和职能。“全面记录并保存人类所有的文献”和“通过记录达到对文献的有效检索和广泛利用”被认为是书目控制的两个基本任务,制订国际化的编目条例和实施全球范围内的联合编目成为书目控制的方法和手段。到了世纪之交,图书馆的大量文献被数字化,互联网上的原生数字化资源与日俱增,“文献”的种类不再限于纸质资料而扩展到多媒体资源和原生数字化资源,研究数字文献书目控制的“数字目录学”等议题得到高度重视^[16]。2008 年美国国会图书馆发布的《书目控制未来报告》草案提出“书目控制未来将是合作性的、分散的、世界性的和基于万维网的”^[17],试图将图书馆的书目控制方法延伸到互联网。在编目规则层面, AACR2 升级为 RDA,充分借鉴 FRBR,基于概念来区分书目记录和实体对象,定义概念、属性及其关联关系,并对各类规范取值词表概念化、规范化。在知识建模和知识表示方面,充分利用语义网的本体、RDF 及关联数据技术。美国国会图书馆推出书目数据的本体模型及词表 BIBFRAME 以替代 MARC,全球最大的书目控制机构 OCLC 基于互联网上应用最

为广泛的 Schema.org 本体词表开发了书目扩展 SchemaBibExt。与此同时,美国、英国、瑞典、芬兰、韩国、日本的国家图书馆以及 OCLC 将自己的书目数据以关联开放数据 (Linked Open Data) 技术发布为可在互联网上开放获取的“关联书目数据” (Linked Bibliography Data)。

图书馆界的这些努力在数据、方法和技术实施路线上为建设跨网域开放互联的“数据基础设施”奠定了基础。上海图书馆基于关联数据技术构建的“中国家谱知识服务平台”,将全球 597 家家谱收藏机构、社会团体和个人的家谱收藏目录整合到一个平台之中;“中文古籍联合目录及循证平台”汇集了全球 1 400 余家机构的馆藏古籍目录;“盛宣怀档案知识库”包含了 15 万余种档案目录,涉及近 3 万名晚清民国时期名流政要的信函、电报、合同、章程、日记、实物等;此外还有上海市文献联合编目中心的 150 余万种书目数据。上海图书馆将这些不同格式的书目数据转换为 RDF 格式,产生了近 3 亿个 RDF 三元组,基于此构建了同时服务于普通大众和人文研究者的文献知识库,并发布为关联开放数据。所有的资源对象、规范实体、术语词表被赋予可在互联网上唯一标识和定位的 HTTP URI,支持跨网域的开放数据互联,试图实现跨机构的书目控制,同时在互联网上提供开放数据服务。将其作为上海图书馆历史人文“数据基础设施”的有机组成部分(见图 2 中“文献知识库”)。

4.2 规范控制

规范控制原是图书馆编目业务工作流程中重要的一环,目的是“集中同一责任者的不同著作,汇集同一著作的不同版本或译本,提高文献检索率”^[18],主要包括名称规范控制(人、机构、地名)和文献题名的规范控制,通过在文献编目的过程中建立并使用著者规范档和统一题名来实现。合并异名、区分同名,是图书馆“名称规范控制”工作的主要任务。在互联网环境下,图书馆的规范控制在理念、方法和技术上经历了

一系列的变革。以人名为例,传统的人名规范控制,以人名为基本单元,在计算机系统中作为字符串编码,以关键词匹配的方式提供索引和检索服务。随着语义网、关联数据和知识图谱技术的发展,一个具体的人,被作为“人”这个抽象概念在现实世界中的实例对象(实体, Entity),人名和人的生卒年、生平大事,以及人之间和各种关系是“人”这个概念的属性。这一思想也体现在“IFLA 图书馆参考模型”(IFLA-LRM)和美国国会图书馆的关联书目数据模型——BIBFRAME 2.0 的设计中^[19],并逐步成为互联网环境下名称规范控制的发展趋势,多国的国家图书馆联合构建了国际虚拟规范档(VIAF)等基于实体管理的大型关联开放数据集^[20]。这样的方式正好契合了数字人文的典型研究范式。

在数字人文研究中,“人”是历史、人文、社会研究中的重要因素,是除了时间和空间之外的另一个重要维度,个人的生卒时间、地点及在生命旅程当中的重要事件、工作经历、社交活动、作品著述等,是构成整个人类社会人文和历史的基本单位。以人的个人信息为研究资料,以个人或群体为研究对象,是人文研究的常规范式,以群体的社会关系为基础的“社会网络分析法”,是数字人文研究的基本方法之一。

图书馆长期的名称规范控制工作形成的名称规范档,结合互联网环境下基于实体管理的名称规范控制方法,按照数字人文研究的要求进行重组和丰富后,可成为“数据基础设施”的有机组成部分。同样的方法还可以推广到机构、地点、时间、事件、实物等,建立各种实体规范库,以构建开放的、以知识服务为目的的人文数据基础设施,保证人文研究的持续性、一致性和高效性,各种以实体为基本单位的规范库正是数据基础设施建设中重要的一环^[21]。上海图书馆的“人名规范库”就是在图书馆的人名规范档的基础上,补充了大量的人物资料,转换为结构化的语义数据,并与“地理名词表”“历史纪年表”“事件库”“建筑库”等一起作为历史人文研

究的“数据基础设施”的有机组成部分。这些基础知识库在资源对象和数据单元的层面,基于本体定义的关系,通过各自 HTTP URI Link 相互关联,作为各类 GLAM 资源和数据的链接中心,既起到规范控制的作用,同时还以 Restful API 和 Sparql Endpoint 的方式,为其他机构和开发人员提供开放数据服务(见图 2“基础知识库”)。

4.3 知识组织

“知识组织”源于图书情报学的分类法和主题法,作为图书馆学情报学理论和实践研究的热点和前沿,随着信息技术的发展历久弥新,在理论、方法、技术上都有了长足的进步。简单知识组织系统(SKOS)、主题图、大众分类法、自动标引技术等 in 图情领域得到了卓有成效的应用,并引起了知识管理、人工智能、语言学等领域的关注^[22]。在最初的分类法和主题法的基础上,图情领域的知识组织经历了两次大的变革和升级:一是以 DC 为代表的元数据方案设计方法,二是基于语义网本体的知识建模和知识表示方法。

元数据方法使得图情领域的知识组织不再仅仅是给文献分类和主题标目,而形成了更为规范、科学和灵活可扩展的多维度的描述和揭示方法,并可适用于各种各样的特色文献资源、多媒体资源和网络资源。在过去的 20 余年中, GLAM 以 DC 元数据方案设计方法,制定了覆盖纸本、多媒体资源、实物、网络资源等各种资源类型的元数据描述标准、扩展规范和编码规范,使元数据元素具有明确的、一致性的定义,积累了大量标准规范的元数据记录。

引入语义网技术和本体方法后,图书馆知识组织方法的适用范围进一步扩大。“语义”(Semantic)一词在语义网的语境下意为“机器与机器之间可相互理解的语义”。为数据赋予机器可理解的语义后,将主观的知识客观化、将隐性的知识显性化,使知识成为可计算、可分析、可支持推理和决策的语义数据。语义网提供了一套完整的知识建模和知识表示规范,超越了

元数据以资源对象为中心的知识组织方式,不仅描述资源对象,还描述资源对象中隐含的各种知识节点(人/机构、地、时、事、物),尤其擅长描述其中错综复杂的关联关系。这些知识节点超于资源对象而又为其所共有,因而可以作为建立在资源对象之上的知识层,承担着链接不同领域的资源对象、实现跨领域知识共享的任务。本体方法将原有的分类主题词表、不同资源的元数据标准,建模成统一的、抽象的概念体系,形成可共享的术语词表,以标准规范的形式化语言(RDFs, OWL, SKOS)和序列化格式(RDF/XML, Turtle)编码,成为机器可理解的客观知识^[23]。欧洲数字图书馆(Europeana)的EDM(Europeana Data Model)即是一种基于本体的跨GLAM领域的知识模型。上海图书馆的数字人文平台采用基于本体的知识建模和知识表示方法后,将RDFs形式化编码的本体发布在互联网上,在以Restful API方式提供开放数据服务的同时,也开放机器可读的数据结构定义。

关联数据作为语义网的轻量级实现方式,在各国以国家图书馆为首的“关联书目数据”(Linked Bibliographic Data)项目中得到了深入的应用,使图书馆的书目数据和规范数据成为跨网域开放互联的开放数据集。知识图谱作为语义网在产业界的成功应用模式,吸收了W3C关联开放数据(Linked Open Data)运动的成果。2012年,最大的关联开放数据集Freebase被Google收购,成为Google知识图谱的第一桶金。“知识图谱”(Knowledge Graph)在图情领域往往被误解,与一种也被翻译为“知识图谱”(Knowledge Map)的可视化技术混淆。“知识图谱”(Knowledge Graph)在本质上是一种知识组织方法,为搜索引擎提供大规模、高质量、有着丰富关联关系的实体库,在索引之上建立知识层,在关键词匹配上辅以概念匹配以极大地提高搜索引擎的查全率和查准率^[24],基于“知识图谱”(Knowledge Graph)进行知识组织的数据也可以生成可视化图表,但只是这种知识组织方法的外部表现形式。知识图谱(Knowledge Graph)最

初是语义网技术和搜索引擎技术的结合,现已在人工智能领域得到关注应用^[25],成为符号主义的一种方法论基础,为自动化、大规模、多源异构数据的知识提取和知识融通提供了方法和技术支撑^[26]。

4.4 文献循证

循证实践(Evidence-Based Practice),也叫“循证学”,本意是“基于证据的实践”。“循证实践”脱胎于医学领域的“循证医学”,并成为一种研究方法向人文社科领域延伸,逐渐形成了循证教育学、循证管理学、循证经济学、循证犯罪学、循证软件工程、循证图书馆学等,其核心思想是强调研究和决策应建立在客观证据的基础上,而非仅仅是个人的经验。

图书馆领域的循证实践,是以历代文献中的内容作为依据,将科学的研究方法与研究人员的工作经验结合起来,解决特定研究问题的过程,是基于文献的循证,也叫“文献循证”^[27]。传统的目录学、版本学、校勘学研究就是一种典型的基于文献的循证实践,但这种传统的文献循证,其过程包括搜集资料、阅读资料、比较遴选、分门别类、深度阅读、提炼观点、发现未知、查找更多资料,对于研究者的个人经验和能力还有较强的依赖性。在数据技术的加持下,图书馆学的文献循证方法可演变为数据驱动的循证实践,其证据来源不再仅仅是需要依靠人工翻阅的文献,还包括在文献中提取的可被机器处理的各种知识。

上海图书馆的“中文古籍联合目录及循证平台”除了包括5家合作机构的古籍文献的书目数据和全文影像、1400余家机构馆藏的现代古籍联合目录数据外,还融合了历代的古籍目录数据(官修目录、史志目录、藏书楼目录、私家目录)、版本数据,以及各种在内容上独立、在数据底层相互关联的著者和批校题跋者知识库、藏印知识库、避讳字知识库、刻工知识库(见图2中“基础知识库”),其中著者和批校题跋者在“人名规范库”中统一管理和维护^[28]。这样的

知识库基于本体进行知识建模和知识表示,基于关联数据技术以 Restful API 方式提供开放数据服务,在资源对象和数据单元层面实现基于 HTTP URI 的开放互联,天然就成为“数据基础设施”的一部分。

数字人文中的文献循证不仅表现在对过去的各种文献证据的循证,还表现在对当前和未来在“数据基础设施”不断的跨时空版本迭代中发生的各种变化进行循证研究。上海图书馆的家谱、古籍、人名规范库等向登录用户开放了留言功能,向专家用户开放了数据更新功能,向其他机构、团体个人和第三方公司开放了共建共享通道(见图2),鼓励用户对每一个资源对象的描述元数据和在内容中提取的数据单元进行评论和修改,并记录每一次修改的详情,包括谁在何时何地、将什么数据修改为什么数据,用于支持数据维护追溯和研究者的循证研究。

5 结语

数字人文研究是数据驱动的研究,需要在一个开放共享的环境下,获得大规模、跨网域、跨领域、覆盖地域广、时间跨度长、多维度、小粒度、自解释、富语义的数据,来支持定量统计分析、灵活多维聚类、多重证据对照、溯源循证、全景式的数据可视化。在人文研究领域中,各研究机构、研究团体和个人为支持特定的研究主题收集整理加工了大量的数据,构建了众多的专门领域数据库。但这些数据库之间缺乏相互的共建共享机制和数据互操作功能,缺乏跨领域的顶层知识建模设计和底层知识表示规范,

同时由于研究项目的结项或研究者兴趣的转移而缺乏长期保存和可持续发展的保障。而开放性、公共性、持续性正是 GLAM 机构的特点和优势。各类 GLAM 机构在千百年来已经积累了海量的、结构化、规范化的开放数据,虽然仍然存在着信息孤岛与重复建设并存的问题,但已经在跨机构和跨领域的数据整合方面做出了长期的努力,积累了宝贵的经验。

本文提出“数据基础设施”的概念,将其从“研究基础设施”中剥离开来,试图分析面向人文研究的“数据基础设施”的一般性功能需求和技术规范,论证图书馆的书目控制、规范控制、知识组织、文献循证方法和关联数据、知识图谱、大数据技术结合起来,有望为“数据基础设施”的建设提供一定的方法论贡献。GLAM 机构应该在数字人文热浪中认清认准自身的定位,找准侧重点和着力点,将具体的研究工作交给领域研究者,以为人文研究提供开放的、可长期保存和持续更新的数据支撑和数据服务为核心,致力于数据采集加工、知识整序和跨机构、跨网域、跨领域的数据融合和知识融通,抓住机遇,迎接挑战,在互联网和大数据时代发挥更大的作用。而建设公共性、开放性、持续性的“数据基础设施”,正是 GLAM 机构在数字人文大潮中应尽而可尽的责任。然而,本文关于“数据基础设施”概念、定义、一般性功能需求和技术规范,所论证的图书馆学方法论贡献,以及 GLAM 机构在数字人文基础设施建设中的侧重点和着力点等结论,需要与更多的 GLAM 机构从业人员和具体领域人文研究者进行深入讨论,并在将来的实践中进一步验证。

参考文献

- [1] 安妮·伯迪克,约翰娜·德鲁克,彼得·伦恩费尔德,等. 数字人文:改变知识创新与分享的游戏规则[M]. 马林青,韩若画,译. 北京:中国人民大学出版社,2018:52-55. (Burdick A, Drucker J, Lunenfeld P, et al. Digital humanities[M]. The MIT Press, 2012.)
- [2] 刘伟,叶鹰. 数字人文的技术体系与理论结构探讨[J]. 中国图书馆学报, 2017, 43(5): 32-41. (Liu K W, Ye F Y. Exploring technical system and theoretical structure of digital humanities[J]. Journal of Library Science

- in China, 2017, 43(5): 32-41.)
- [3] Benardou A, Champion E, Dallas C, et al. Cultural heritage infrastructures in digital humanities[M]. Abingdon, Oxon: Routledge, 2018: 1-12.
- [4] 高洪福. 又到了打造“铁腰”的时候!——“IT 生存法则”之数据中台初探[J]. 网络安全和信息化, 2019(9): 24-25. (Gao Hongfu. A preliminary study of “IT Survival Rules” in data center[J]. Security & Informatization, 2019(9): 24-25.)
- [5] 肖珑. 人文社会科学繁荣发展的软性基础设施建设[J]. 图书情报工作, 2011, 55(1): 5-9. (Xiao Long. Building soft infrastructures for the prosperous development of social sciences and humanities[J]. Library and Information Service, 2011, 55(1): 5-9.)
- [6] 赵生辉. 国外档案领域数字人文项目的实践与启示[J]. 浙江档案, 2015(9): 14-17. (Zhao Shenghui. The practice and enlightenment of foreign archives digital humanities project[J]. Zhe Jiang Archives, 2015(9): 14-17.)
- [7] 刘伟, 谢蓉, 张磊, 等. 面向人文研究的国家数据基础设施建设[J]. 中国图书馆学报, 2016, 42(5): 29-39. (Liu Wei, Xie Rong, Zhang Lei, et al. Towards a national data infrastructure for digital humanities[J]. Journal of Library Science in China, 2016, 42(5): 29-39.)
- [8] 包弼德, 夏翠娟, 王宏甦. 数字人文与中国研究的网络基础设施建设[J]. 图书馆杂志, 2018(11): 169-176. (Bol P K, Xia Cuijuan, Wang Hongsu. The digital humanities and a cyberinfrastructure for China studies[J]. Library Journal, 2018(11): 169-176.)
- [9] Francois P, Manning J G, Whitehouse H, et al. A macroscope for global history seshat global history databank: a methodological overview[J/OL]. Digital Humanities Quarterly, 2016, 10(4) [2020-01-10]. <http://www.digitalhumanities.org/dhq/vol/10/4/000272/000272.html>.
- [10] 丁宁宁. 从“三大公共数字文化工程”的实践看中国数字图书馆发展[J]. 图书馆理论与实践, 2017(3): 80-83. (Ding Ningning. The development of Chinese digital library from the practice of the three major public digital cultural projects[J]. Library Theory and Practice, 2017(3): 80-83.)
- [11] 张钦. 人工智能中知识表示方法之比较[J]. 科教文汇, 2008(27): 279-280. (Zhang Qin. Comparison of knowledge representation methods in artificial intelligence[J]. The Science Education Article Collects, 2008(27): 279-280.)
- [12] 徐宝祥, 叶培华. 知识表示的方法研究[J]. 情报科学, 2007(5): 690-694. (Xu Baoxiang, Ye Peihua. Research on the method of knowledge representation[J]. Information Science, 2007(5): 690-694.)
- [13] Makhfi P. Introduction to knowledge modeling[EB/OL]. [2020-02-02]. http://www.makhfi.com/KCM_intro.htm#What.
- [14] 张梅, 郝佳, 阎艳, 等. 基于本体的知识建模技术[J]. 北京理工大学学报, 2010(12): 1405-1408, 1431. (Zhang Mei, Hao Jia, Yan Yan, et al. Ontology-based knowledge modeling[J]. Transactions of Beijing Institute of Technology, 2010(12): 1405-1408, 1431.)
- [15] 曾蕾, 王晓光, 范伟. 图档博领域的智慧数据及其在数字人文研究中的角色[J]. 中国图书馆学报, 2018(1): 17-34. (Zeng Marcia Lei, Wang Xiaoguang, Fan Wei. Smart data from libraries, archives and museums and its role in the digital humanity researches[J]. Journal of Library Science in China, 2018(1): 17-34.)
- [16] 柯平. 数字目录学——当代目录学的发展方向[J]. 图书情报知识, 2005(3): 18-22. (Ke Ping. Digital bib-

- liographic science;trends of current bibliography science[J]. Document, Information & Knowledge, 2005(3): 18-22.)
- [17] 顾森. 关于《书目控制未来报告》草案[J]. 国家图书馆学报, 2008, 17(1): 76-78. (Gu Ben. About the draft of *Report on the Future of Bibliographic Control* [J]. Journal of the National Library of China, 2008, 17(1): 76-78.)
- [18] 丘东江. 图书馆学情报学大辞典[M]. 北京: 海洋出版社, 2013. (Qiu Dongjiang. Great dictionary of library and information science[M]. Beijing: China Ocean Press, 2013.)
- [19] 胡小菁. BIBFRAME 核心类演变分析[J]. 中国图书馆学报, 2016(5): 20-26. (Hu Xiaojing. Evolution of BIBFRAME core classes[J]. Journal of Library Science in China, 2016(5): 20-26.)
- [20] 胡小菁. 规范控制: 从名称选择到实体管理[J]. 数字图书馆论坛, 2018(1): 2-7. (Hu Xiaojing. Authority control: from selection of a name to entity management[J]. Digital Library Forum, 2018(1): 2-7.)
- [21] 夏翠娟. 以连接开放资料服务为基础的数字人文平台建设方案研究[J]. 图书馆学与资讯科学, 2017(4), 43(1): 47-70. (Xia Cuijuan. Building a digital humanities platform by using linked open data services[J]. Journal of Library and Information Science, 2017(4), 43(1): 47-70.)
- [22] 卜书庆, 刘华梅, 王广平. 近年来国内知识组织研究热点综述[J]. 中国索引, 2010, 8(1): 2-12. (Pu Shuqing, Liu Huamei, Wang Guangping. A review of the research hotspots of knowledge organization in current China [J]. Journal of the China Society of Indexers, 2010, 8(1): 2-12.)
- [23] 韩燕, 何琳, 彭爱东. 近10年我国传统知识组织系统关联化研究述评[J]. 图书情报工作, 2017, 61(24): 135-143. (Han Yan, He Lin, Peng Aidong. Review on the research of open link of knowledge organization system in China in recent ten years[J]. Library and Information Service, 2017, 61(24): 135-143.)
- [24] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016(4): 589-606. (Xu Zenglin, Sheng Yongpan, He Lirong, et al. Review on knowledge graph techniques[J]. Journal of University of Electronic Science and Technology of China, 2016(4): 589-606.)
- [25] 符山, 吕艾临, 闫树. 知识图谱的概念与应用[J]. 信息通信技术与政策, 2019, 299(5): 17-20. (Fu Shan, Lv Ailin, Yan Shu. The concept and application of knowledge graph[J]. Information and Communications Technology and Policy, 2019, 299(5): 17-20.)
- [26] 漆桂林, 高桓, 吴天星. 知识图谱研究进展[J]. 情报工程, 2017, 3(1): 4-25. (Qi Guilin, Gao Huan, Wu Tianxing. The research advances of knowledge graph[J]. Technology Intelligence Engineering, 2017, 3(1): 4-25.)
- [27] 刘伟, 林海青, 夏翠娟. 数字人文研究的图书馆学方法: 书目控制与文献循证[J]. 大学图书馆学报, 2018, 36(5): 116-123. (Liu Wei, Lin Haiqing, Xia Cuijuan. Bibliographic approach to digital humanities: authority control and evidence-based practices[J]. Journal of Academic Libraries, 2018, 36(5): 116-123.)
- [28] 夏翠娟, 林海青, 刘伟. 面向循证实践的中文古籍数据模型研究与设计[J]. 中国图书馆学报, 2017, 43(6): 16-34. (Xia Cuijuan, Lin Haiqing, Liu Wei. Designing a data model of Chinese ancient books for evidence-based practise[J]. Journal of Library Science in China, 2017, 43(6): 16-34.)

夏翠娟 上海图书馆研究员。上海 200031。

(收稿日期: 2020-02-07)