DOI: 10.13530/j.cnki.jlis.170010

中国历史地理数据在图书馆数字人文项目中的开放应用研究

夏翠娟

摘　要　利用文献调研、数据建模、比较研究、实验研究等方法，在调研和比较 CBDB 项目、复旦大学史地所、台湾地区“中研院”GIS 研究中心等中国历史地理数据库的建设及利用情况的基础上，分析图书馆数字人文项目建设中涉及历史地理数据时遇到的困境、目标和需求，探索在图书馆的数字人文项目建设中引入现代历史地理学的研究成果，利用知识组织和规范控制方法重组历史地理数据，实现历史地理数据在图书馆领域开放应用的目标，拉近历史地理学与人文社会科学研究者等图书馆用户之间的距离。本文重点研究历史地理数据在图书馆领域内的开放共享以及提供数据开放应用服务的方法，探索更为通用和大众化的时空数据模型及基于知识本体和关联数据技术的实现方案。图 4。表 3。参考文献 24。

关键词　历史地理学　数字人文　地理信息系统　关联数据

分类号　G254

The Opening and Application of Chinese Historical Geography Data in Digital Humanities Projects of Libraries

XIA Cuijuan

ABSTRACT

The modern Historical Geography, which integrates History, Geography, and GIS technology, forms a new research method based on “space,” and develops professional and strong historical-geographical data, technologies and platforms. It provides a new perspective for Digital Humanities and cannot be ignored. The purpose of this paper is to improve the application of modern History Geography in the libraries´ digital humanities projects with the reorganization of Chinese historical and geographical data by the knowledge organization and authority control methods which the library field is good at, and then to narrow the gap between specialized historical geography data and humanistic research in practical applications.

This paper investigates the application of modern historical geography in humanities research through the methods of literature research and comparative research. Spatially Integrated Humanities and Social Science which is driven by spatial analysis and GIS technology have become the hotspot and frontier in the field of Humanities and Social Science research. This paper investigates the similarities and differences of data modeling, data utilization, data open application in Internet environment of CBDB, CHGIS, CTSS, Sino-family-tree GIS, GeoNames, Getty Thesaurus of Geographic Names and other Chinese historical geo-

本刊“青年学术论坛”特约稿(Special contribution for the Youth Academic Forum sponsored by this Journal)

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)

databases (sets).

The problems, goals, and needs of the application of historical and geographical data in libraries' digital humanities projects are also analyzed based on literature research and experimental research. Although the libraries have laid a certain foundation for the digital humanities with the full text and metadata of digital resources, the lack of historical and geographical data related to the literature has made it difficult to integrate and associate resources in the multidimensional space-time framework. However, because of the professionalism and complexity, the existing GIS platforms are difficult to be directly applied to the digital humanities projects in the library. Therefore, Ontology and Linked Data technology are recommended to support historical geography data modeling, publication and open application in the construction of "Chinese historical geography knowledge base".

The innovation of this paper is as follows: Firstly, on the basis of researching the data model of CHGIS, CTSS, CBDB and Sino-family-tree GIS, aiming at ease of use and extensibility, the spatial-temporal model of historical and geographical data is designed by using ontology method, which clearly distinguishes between the place and place name, clarifies the change of place names in different temporal, and the relationship between these changes and locations. The concept of "event" is introduced to model the changes of place or place names, and the time data associated with places and place names was encapsulated in "events". Secondly, the paper puts forward the construction plan of "Chinese historical geography knowledge base" based on Linked Data, and provides the technical realization guide on the data publication and data open application, and puts forward three kinds of Linked Data consumption technology: "content negotiation, Restful API and SPARQL Endpoint".

The spatial-temporal data model of Chinese historical geographic data proposed in this paper mainly adopts the knowledge organization and the authority control method in the field of the library. It is different from the data modeling method in the field of historical geography. It needs more practice to be further improved during the process of large historical geography data mapping and conversion. 4 figs. 3 tabs. 24 refs.

KEY WORDS

Historical Geography. Digital humanities. GIS. Linked Data.

0 引言

随着大数据、文本挖掘、地理信息系统(GIS)、信息可视化、虚拟现实(VR)、增强现实(AR)等现代信息技术的飞速发展及其在人文研究领域的深入应用，"数字人文(Digital Humanities)"以其跨学科、跨领域的特点，在研究方法和研究手段上成为传统人文研究的有力补充和强劲推动力。其中，以地理信息系统(GIS)技术为依托的"空间分析法"在经济学、历史学、语言学、人类学等人文社会科学领域得到了深入应用，逐渐形成了一套以"空间"为切入点的新型研究方法[1]，吸引了众多领域的研究专家和学者，为传统的人文社会科学研究拓展了研究视野，提供了新的视角。

在人文社会科学研究领域，"空间"被视为一种社会建构，是被赋予了人类社会及文化意义的自然—人文综合景观空间，而不仅仅是传统地理学意义上的自然地域空间[2]。在此视域下，空间的各种要素会随着时间的改变而改变。现代历史地理学将"空间"与"时间"结合起来，引入现代信息技术和定量分析、比较研究、科学统计等手段和方法，利用GIS技术的空间数据采集、时空数据建模、多层地图叠加功能，试图重现不同时间切面中的地理景观，研究其与历史、

社会、自然之间的关系，探索发展演变的规律。现代历史地理学为人文研究贡献了多维时空框架下的数据、技术、平台与方法。历史地理数据及其相关技术、平台和方法的利用，已成为数字人文中不容忽视的要素和不可或缺的一部分。

近年来，图书馆尤其是大中型研究型图书馆纷纷试水数字人文，哥伦比亚大学、斯坦福大学、加州大学洛杉矶分校等高校图书馆和相关院系成立了数字人文研究中心，国内的武汉大学、北京大学图书馆、上海图书馆也在积极研究和探索图书馆提供数字人文研究的支撑方法和路径，并积极开展数字人文项目建设，以为人文研究提供更好的服务。图书馆从事数字人文项目建设有着天然的优势：图书馆作为人文研究所需原始资料的保存和服务机构，经历了近 20 年的数字化建设后，积累了大量数字资源和高度结构化、规范化的元数据记录，奠定了数字人文项目建设的数据基础。然而，在图书馆元数据方案设计和元数据著录规范中，忽视了历史地理数据的系统化加工和整理，也缺少引入现代历史地理学研究成果的意识，对多维时空架构下的数据处理和数据展示造成了障碍，成为图书馆开展数字人文项目建设的瓶颈之一。

本文试图探索在图书馆的数字人文项目建设中引入现代历史地理学的研究成果，利用知识组织和规范控制方法重组历史地理数据，实现历史地理学的资料、数据、工具、平台在图书馆领域深入应用的目标，拉近历史地理学与人文学社会科学研究者之间的距离；重点研究历史地理数据在图书馆领域的协同建设和开放共享的目标和需求，探索更为通用的大众化的时空数据模型和基于知识本体和关联数据技术的实现方案。

1 现状调研

1.1 现代历史地理学在人文社会学中的研究和应用现状

现代历史地理学是融合了历史学、地理学和现代信息技术的交叉学科，并随着 GIS 技术的快速发展催生了新的研究前沿。王晓光认为：目前国际数字人文研究的前沿和典型应用包括“历史学方面的基于 GIS 的历史地理可视化”[3]。实际上，基于 GIS 的历史地理可视化在数字人文中的应用远不止历史学。自 2009 年至 2016 年，汇聚了海内外地理学、历史学、语言学、人类学、社会学、信息技术等领域专家学者的“空间综合人文学与社会科学论坛”连续举办，反映了“空间综合人文学与社会科学（Spatially Integrated Humanities and Social Science）”逐步成为人文学、社会科学研究的热点和前沿[4]。

GIS 技术为现代历史地理学注入了新的活力，表现为一系列大型历史地理信息系统的建设[5]。2001 年，复旦大学历史地理研究中心启动了中国历史地理信息系统项目，建立了中国历史地理信息系统（CHGIS）[6]，台湾地区“中研院”发布了“中华文明之时空基础架构（CCTS）”和“台湾历史文化地图（THCTS）”两个平台[7]，成为备受瞩目的中国现代历史地理学与人文学结合的成果。它们的共同特点是在整合数字化地图影像资料的基础上，同时进行历史地理数据的标注、提取和结构化组织，形成集原始地图影像资料和历史地理数据于一体的信息化支撑平台，为特定主题的人文研究项目和学者提供数字化地图影像资料、结构化的历史地理数据、GIS 工具软件支持。这两个系统的建设代表了中国历史地理信息系统从地图影像资料库到数据支撑平台的发展趋势。

上述支撑平台除了具有浏览、查询和下载功能外，主要是面向历史地理学的专业人员、有一定计算机技术素养的人文社会科学研究人员和为应用系统的开发人员提供专业的历史地理数据下载和应用程序接口。如学者徐永明利用 CHGIS 和“中国历代人物传记资料库”（CBDB）的历史地理数据、人物社会关系数据和 GIS 工具软件，可视化地呈现汤显祖的社会关系在地理空间的分布情况，使大量分布在文献资料中的数据得到了直观、有序的呈现[8]，这是有技术素

养的人文研究人员利用 CHGIS 等历史地理信息化支撑平台的一个典型案例。又如“中国历代人物传记资料库(CBDB)”利用 CHGIS 的历史地理数据，在人物传记资料的组织和展示上融入“空间分析法”，将历史上人物的出生、生活、任职、游历、田产、死亡、安葬等活动置于多维时空架构之中，为研究者发现问题、解决问题提供了全新的方法和手段，这是一个为其他专题资料库提供历史地理数据支撑的成功案例。同样，台湾地区“中研院”在 CCTS 的支撑下，开发了大量专门领域的应用系统，为特定专题研究提供数据和平台支撑，如黄河泛滥分析、明清江南市镇研究、傅斯年图书馆人名权威资料库(人名规范档)、苏轼文学地图等，是现代历史地理学应用于人文学、社会科学的典型范例。

1.2 中国历史地理信息系统建设及应用现状

历史地理信息系统的建设无论在国际还是国内，已上升到了国家的高度。如国际上美国地质调查局下属地名委员会的地名数据库、加拿大地名数据库、澳大利亚国家地名委员会的澳洲地名库，新西兰国家地理理事会的地名数据库等。国内如我国民政部敦促各省市建立了本地地名数据库；国家测绘局也相继建成了各种比例尺的全国地名数据库[9]；中国社会科学院于 2009 年 9 月发布了“中国社会科学综合地理信息服务平台系统”，借助 GIS 技术展示河流的变迁、历次人口的迁徙、经济地带的发达及没落、各民族融合和分离的过程[10]；一些高校的专业院系和研究机构也在历史地理数据库的建设上取得了成就，上文提到的 CHGIS 和 CCTS、TH-CTS 是其中的主要代表，南京师范大学的“华夏民族家谱地理信息系统”——简称“华夏家谱 GIS”，则是为家谱文献在统一时空模型下的组织和呈现而开发，与图书馆利用历史地理信息系统的考虑有着相似之处。

随着互联网时代的到来，历史地理信息系统也遇到了瓶颈与困难，学者陈刚对此做了总结，其中重要的一条是历史地理学界对新技术的理解、掌握和运用缺乏足够的认识[3]。在国内，历史地理信息系统建设起步于 2000 年前后，在不到二十年的时间内，信息技术的发展日新月异，数据成为第一生产力和第四研究范式，互联网思维席卷全球，系统中的数据能否在互联网上方便获取和易于利用，变得尤为重要。互联网意味着协同、共享和开放，而越是专业的就越有开放共享的价值。现有的历史地理信息系统能否满足万维网环境下的开放应用需求？针对这一问题，笔者对基于关系数据库构建的 CHGIS 和 CCTS、整合了 CHGIS 历史地理数据的 CBDB、用于特定文献资源组织和呈现的“华夏家谱 GIS”，以及基于关联数据技术发布的 Getty 地理名词表、GeoNames 数据集进行调研，主要从数据建模、数据库的利用方式、历史地理数据在互联网环境下的开放应用等几个方面来考察。

在数据建模方面，南京师范大学在构建“华夏家谱 GIS”的过程中，陈旻研究了家谱 GIS 的时空支撑架构，将地名作为实体对象来研究，深入细致地分析了古今地名实体的空间特征、空间归属关系、行政隶属关系等[11]；胡颖在其硕士毕业论文中将具有空间特征的地理实体从地名中剥离开来，详细深入地分析了地理实体和地名的时间特征和空间特征，以及各自在生命周期中的变化情况[12]；温永宁等提出家谱 GIS 需要构建一个为所有家谱中记录的时间和空间信息提供支持的时空描述框架，使得家谱中的所有事件和人的活动能够在统一的空间和时间模型下展开，该模型强调的是家谱文献中时间与空间的统一表达[13]。CHGIS 数据模型的基本功能是描述各政区之间的隶属关系和政区界线，并表达出名称的变化以及行政区域合并、分置、新建、撤销等形成的界线变化，允许用户按他们需要的时间和地区重新组合数据库中的数据。此外，数据也需要反映一个行政单位地理形态的前后变化过程和其本身变化对其他部分的影响，能按时间检索行政区域的变化是 CHGIS 空间—时间模型设计的基本概念[14]。CBDB 沿用了 CHGIS 的做法，依靠两类空间实体“地址”和

"地点"(经纬度)。"地址"这一实体作为有地名的历史"场合"——即空间中有特定名称的行政辖区,可作为其他行政辖区的一部分,其位置由 x、y 坐标(即经纬度)的交汇点来确定,如果边界或者名称中任何一个改变,就要建立一个新的地址[15]。但由于历史地理数据的时空关系错综复杂,CHGIS 和 CCTS 虽然对地名的各种时空属性和历史变迁进行了丰富的描述,但对同一地理空间的、在不同历史时期的、各种不同地名之间的关系没有建立清晰的数据模型,缺乏地名的规范控制。Getty 地理名词表和 GeoNames 注重的是地名的行政归属关系和地理空间属性的描述,但缺少与地名相对应的时间序列数据,因而缺乏不同时代各种地名间相关关系的建模和形式化表达[16]。

在历史地理数据的利用方式方面,华夏家谱 GIS 在网站上提供古今地名和历史纪年数据查询,在功能上主要是为用户提供参考,历史地理数据建设的目的是为本系统中家谱文献的查询和展示服务。CHGIS 的数据在网站上可下载,但其数据格式为专业的.map 格式,或关系数据库.mdb 格式,结构复杂,需掌握一定的专业知识和计算机技术才可对数据进行挖掘和操控,而这点给大部分人文学者造成了一定的障碍。CHGIS、CCTS 作为专业的历史地理信息化支撑平台,与人文学者和普通大众之间还存在着一定的距离。首都师范大学的周丙锋、周文业基于 CHGIS 开发的中国历史地理数字化应用平台,就试图拉近这一距离,该应用平台提供易于操作的用户界面、自动生成专题历史地图等功能,可根据用户需求建立专题历史地理信息系统,将分析软件提供的功能嵌入平台之中,一定程度上降低了人文学者利用 GIS 技术的门槛[17]。

在互联网环境下历史地理数据的开放应用方面,CCTS 和 CBDB 的做法值得推荐,虽然在数据建模方面,CHGIS、CCTS 等现有历史地理信息系统采用的是以关系数据库为主的技术,这意味着其底层数据存储在封闭的关系数据库中,但提供大量应用程序接口(API)供开发人员调用。这些应用程序接口是建立在互联网的 HTTP 协议上的 RestfulAPI,在数据调用层面符合互联网环境下数据开放应用的需要。而 Getty 研究中心的地理名词表(Getty Thesaurus of Geographic Names)和 GeoNames 则是遵循关联数据技术标准来进行数据编码和在 Web 上发布的,每一个地理名词都被赋予一个 HTTP URI,实现了地名在全网域范围内的唯一标识和定位,关于一个地名的更多信息如所属行政区域、经纬度等以 W3C 的推荐标准——RDF(资源描述框架)格式编码,具有跨平台跨系统的特性,便于机器读取和处理,这种基于关联数据的技术框架为地名在互联网环境下的规范控制奠定了基础。HTTP URI 使得地名数据在标识和访问时即与互联网紧密融合,而 RDF 在数据格式上具备了通用性和开放性,能很好地满足互联网环境下的数据开放共享需求。由于关联数据技术的采用,GeoNames 加入了"关联开放数据项目(Linked Open Data Project)",已成为在 Web 上应用最为广泛的数据集之一[18]。

2 历史地理数据在图书馆数字人文项目中开放应用的问题、目标和需求

2.1 问题

对于承担着社会教育和知识传播功能的图书馆来说,利用数字人文方法更好地组织和呈现馆藏的大规模数字化文献资源,为人文学者提供更精准的服务,是图书馆进行数字人文建设项目的特点和基本出发点。现代历史地理学是数字人文中最重要的方法和手段之一,然而,现有历史地理信息系统中的数据、技术、方法,对图书馆领域的用户而言存在着较高的技术门槛。原因在于:一方面,高校图书馆和研究型图书馆尤其是公共图书馆面对的用户群体不仅是研究人员,还有本科生及社会大众,需要将专业的数据转换成为大众的、通用的知识,并降低 GIS 工具应用的门槛;另一方面,大部分图书馆的人力资源有限,缺少专业的历史地理人才和

信息技术人才，这就使得共享和开放变得尤为重要，尤其是在互联网环境中的开放应用。

图书馆在长期的资源编目实践中，积累了大量有着规范结构的元数据记录。在图书馆领域应用最为悠久和广泛的 MARC 和 DC 元数据规范中，也有简单的对于地理空间的著录，如 MARC 的 650、651、043 字段，DC 核心元数据元素集中的 dc: spatial 属性。但这些字段和属性的取值是字符串组成的文本，缺乏经纬度等地理空间属性，也缺少地名在不同时间中变化和关联的信息。近年来图书馆界推出了“资源描述框架(RDA)”和旨在取代 MARC 的新的书目框架格式 BIBFRAME，对地名的处理更进了一步，引入计算机科学中面向对象的思想，将特定的地理空间看作现实中真实存在的实体对象，并从实体对象中抽象出概念。例如 rda: Place 和 bf: Place，目的是与书目数据中的各种概念(如作品、载体表现、单件)建立关联，对于历史地理数据的容纳仍然没有足够的重视，但在内容框架上具备了引入历史地理数据的条件。在具体应用中，可以为这些概念扩展历史地理相关的属性，如空间存续时间范围、经纬度等。可惜的是，RDA 在中文编目领域还没有进入实施阶段，BIBFRAME 也尚处于实验性探索和研究阶段，虽然上海图书馆采用 BIBFRAME2.0 作为核心数据模型，扩展了 bf: Place，增加了经纬度、行政区域归属等属性，基于此模型对家谱文献中抽取的谱籍地名进行建模，并以关联数据技术在 Web 上公开发布了地名词表，也提供开放应用程序接口供开发人员调用，但该词表缺少地名的时间序列数据，只有今地名，没有古地名。

综上所述，现有的历史地理信息系统难以直接应用于图书馆的数字人文项目建设，图书馆虽然在数字化资源全文和元数据上为数字人文奠定了一定的基础，但在历史地理数据的储备上基本是缺失的。因而难以实现基于空间尤其是多维时空架构的资源整合和关联，特别是对大规模文献资源进行内容分析和数据结构化时，需要提取、匹配其中的地名，而在各种古籍、档案资源中，地名是以资源所在时代的实际情况出现的。例如，同一地理空间可能以不同的地名出现在不同时代著述的古籍中，如何对这些地名进行合并和消歧，实现互联网环境下地名的规范控制，是亟待解决的问题。

2.2 目标和需求

历史地理学虽然有着鲜明的跨学科特性，但也有着强烈的专业性，存在着较高的应用门槛，需要专门的知识背景。如果采用“拿来主义”直接用于图书馆的数字人文项目建设，在缺乏足够的历史地理专业人才和 GIS 技术尖端人才的情况下，存在极大的困难，应分步骤、有选择地实现有限目标，以解决图书馆数字人文项目建设中最迫切的问题。因而需要呼吁大中型图书馆利用图书馆领域擅长的规范控制和知识组织方法和 Web 技术，在现有专业性历史地理信息系统的基础上，建设适用于图书馆领域的历史地理知识库(以下简称“知识库”)，在具备通用性、易用性、便捷性的同时，满足互联网环境下历史地理数据开放应用的需求，以为更多的中小型图书馆提供历史地理数据服务。需要特别指出的是，本文提到的“中国历史地理知识库”，只包括历史地理数据，不包括历史地图影像资料，但需考虑与历史地图影像资料库的接口。

笔者在近年来上海图书馆数字人文平台的设计和开发中总结出以下两种历史地理知识库的应用场景。

(1)在数据加工清洗阶段实现大规模半自动化的地名提取和校准。例如：家谱中的迁徙数据大多以古地名出现，无法与该地名对应的空间建立关联，进而与空间在不同时代所对应的不同地名尤其是今地名建立关联；因此需要知识库提供方便快捷的古今地名对照服务和地理空间数据提供服务。

(2)在数据的可视化呈现时实现不同时代历史地图的多图层叠加展示。例如：盛宣怀的 9 万多封书信，其中发信地和收信地是重要的信息，但这些地名是晚清和民国早期所用，与今地

名多有区别，需要从知识库中获取与地名相对应的空间的经纬度信息，实现更精确的查询，并在不同时代的历史地图上可视化地展示。

基于上述应用场景，知识库应满足以下需求。

(1) 功能需求

①地名规范控制。知识库首先应提供互联网环境下的地名规范控制服务，这要求每一个历史上曾经出现过的地名，都应该在互联网上被标识、被定位、被访问，也就是说，每一个地名都应有 URI(统一资源标识符)。

②历史地理数据提供服务。当访问地名的 URI 时，可获取关于该地名存续的时间范围、空间范围、治所名称及其经纬度数据，以及与其他地名的行政归属和空间归属关系。

③古今地名对照服务。可根据地名存续的时间范围、空间范围、治所名称等关键信息提供与该地名在时间序列上有同一关系的其他地名及其相关时空数据。

(2) 技术需求

①采用开放的数据模型。数据模型的设计，需要考虑到与现有的历史地理数据模型兼容，以支持多源数据的融合、混搭；同时要求有良好的可扩展性，便于数据的修改和增补。

②采用标准化的、通用性强的数据编码格式。数据的编码格式，与数据共享的便利性密切相关。标准化的、通用的数据编码格式有助于数据在异构系统间的传输和互操作，也有利于数据在不同应用开发环境中读取和处理。

③基于 Web 提供开放数据服务。Web 是互联网的主要载体，提供了随时随地的数据访问环境。以 Web 的基础架构 HTTP 协议为依托提供数据应用程序接口(API)，是互联网环境下数据开放应用的常规选择。

3 面向数字人文的“中国历史地理知识库”建设方案研究

3.1 中国历史地理数据建模——基于本体的时空模型设计

笔者在对 CHGIS、CTSS、CBDB、华夏家谱 GIS 的数据模型进行调研的基础上，以通用性、易用性、便捷性为目标，以开放性和可扩展性为原则，采用知识本体方法设计历史地理数据的时空模型，以满足图书馆数字人文建设项目中历史地理数据开放应用的需求。

本体(Ontology)是特定领域内概念及概念间关系的形式化定义。概念是同一类实体对象特性的抽象，一般用“类(Class)”来表示，现实世界中同一实体对象在不同的领域可能会被定义成不同的类，有不同的内涵和外延，但在同一领域内应该得到共识。概念的各种特性被称为“数据属性(Data Property)”，概念间关系则被称为“对象属性(Object Property)”，属性一般以域(Domain)和范围(Range)来界定，一个属性的“域”所指向的概念，决定了该属性可用来描述何种概念，“范围”则定义了属性的取值约束，数据属性的范围一般是文本或数值，对象属性的范围是某个概念对应的实体对象。而“形式化”则强调了机器可读，本体是现实世界的模拟，其目的是把现实世界中的对象和其属性特征数据化，使机器可读，是现实世界与机器世界沟通的桥梁。由于本体采用了面向对象的思想，具备了良好的可扩展性，又有 W3C 等机构推动相关标准规范和应用指南的制订，并能很好地结合语义网和关联数据技术，具备了良好的开放性，近年来成为一种有效的知识组织方法，广泛地应用于数据建模。

本文所提出的历史地理数据时空模型，将空间分为“地点(Place)”和“地名(PlaceName)”两种概念，借鉴华夏家谱 GIS、GeoNames、Getty 等机构的做法，将地点和地名都作为对象，地点对应着现实世界中存在着或存在过的空间实体，而地名则是人类赋予空间实体的名称(代号)。地点具有空间特征、属性特征、时态特征等信息，在空间上，地点可以是点、线、面的二维空间，也可以是有高度的三维空间。地名则有对应的行政区域范围、聚落、治所、行政区域归属关系、存续时间范围等特征信息，以及新建、更名、治所迁移、撤销等生命周期信息。

明确区分地点和地名，有助于厘清地名本身因不同因素导致的在不同时间范围中的变化情况，以及这些变化与地点之间的关系。地点是随着地名的存在而存在的，地名一定有对应的地点，地点也必有至少一个地名来指代。一般来说，地点和地名是一对一的关系，地名的任何变化都会导致新的地点产生。但为了模型的简化，在本模型中，只有当地点的空间特征发生变化时，才会产生新的地点。而地名不再只是地点的一个文本型的属性，它本身也是一个概念，是人类在人文历史活动的影响下赋予的名称。依照陈旻的分析：地名一般由“专名”和“通名”组成，如“上海市”是由专名“上海”和通名“市”组成，专名是人赋予的名称，有特别的含义，通名指代了地名的类别，如“省”“府”“州”“道”“县”“镇”“村”等。在本模型中，将地名看作从“符号”中抽象出来的概念，地名如在字面上保持一致，则被认为是同一地名实例，如“上海市”和“上海县”作为两个不同的地名实例，而历史上出现过多个“安昌县”，在地点上相距甚远，但由于在文字表征上完全一致，则被视为同一地名实例，而其不同的政区归属、治所及相对应的时间范围则由地名实例的各种属性来区分，其地理空间特征则由不同的地名实体来区分，因而同一地名实例在不同的时间中可能对应不同的地点实体。

图 1 表达了本模型的基本思路。其中，圆角矩形代表本体中“类”的概念，带箭头的虚线代表本体中用来表达概念特征和概念间关系的“属性”，箭头的出发点即为该属性的“域”，箭头的指向点即为该属性的“范围”。若从地点出发，可找到该地点在不同时间的不同地名，若从地名出发，则可找到该地名在不同时间中对应的不同地点。

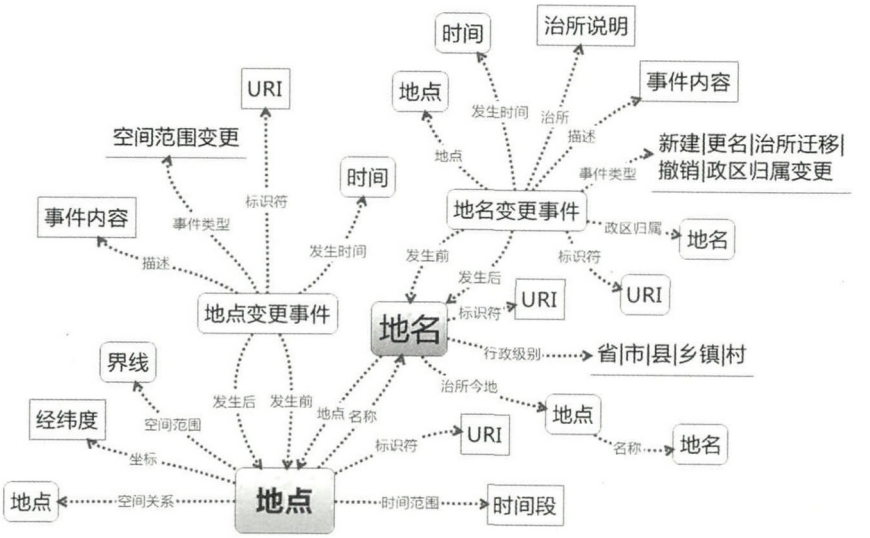

图 1　基于本体的历史地理数据之时空模型

在本模型中，还引入了“事件”的概念，用于对地点或地名的变化情况建模。事件有时间、事件内容、事件类型等要素，而与事件相关的治所和地点则是指事件发生时相关地名所对应的治所和地点，这样就通过事件这一概念为地名所对应的地点赋予了时间属性。事件有“地名变更事件”和“地点变更事件”两种。地点变更事件一般指空间范围的变更，其结果导致新的地点产生，比如某一地名所对应的地点因某种原因导致空间范围的扩大或缩小，此时就会产

生一个新的地点实体，而这个新的地点实体通过原地点实体与地名建立关联关系；地名变更事件包括“新建、更名、治所迁移、撤销、政区归属变更”等情况，只有更名和新建可能导致新的地名产生，其他事件只会导致地名特征的变更。通过事件的“发生前”或“发生后”这两个属性指向的地点或地名为线索，找到与之相关的变更事件。用此模型对与“安昌县”这一地名实例相关的地名变更事件进行梳理，结果如表 1 所示。

表 1 与“安昌县”相关的地名变更事件举例[19]

内部 ID	事件类型	发生时间	相关地名政区归属	相关地名所在地点	发生前	发生后	治所说明	事件内容
1	新建	公元 526 年	南新蔡郡	地点 1#		“安昌县”	今湖北省阳新西北	南朝梁普通七年(526 年)置，治今湖北省阳新西北。属南新蔡郡。
2	撤销	公元 589 年	南新蔡郡	地点 1#	“安昌县”		今湖北省阳新西北	隋开皇九年(589 年)废
3	更名	公元 911 年	福州	地点 2#	“长乐县”	“安昌县”	今福建省长乐市	五代梁乾化元年(911 年)改长乐县置，治今福建省长乐市。属福州。
4	更名	公元 948 年	福州 URI	地点 2#	“安昌县”	“长乐县”	今福建省长乐市	汉乾祐元年(948 年)属吴越福州，复改名长乐。

由于中国历史纪年的复杂性，时间也作为从时间实例中抽象出来的概念来处理。图 2 是时间模型，将“年号纪年”作为时间实例的基本单位，包括起止公元年、国号、朝代、年号、帝王姓名、帝王谥号等属性，由“朝代+年号”或“朝代+国号+年号”来唯一确定一个时间实例，如“明洪武”为一个时间实例，其起止公元年为“1368”—“1398”，帝王谥号和姓名为“太祖”“朱元璋”；又如三国时期，有魏蜀吴三国并存，就需要加上国号才能唯一确定一个时间实例，“三国魏太和”为一个时间实例，其起止公元年为“227”—“233”，帝王谥号和姓名为“明帝”“曹叡”。通过年号起止年属性可实现中国历史纪年与公元纪年之间的对照与转换，若要定位到具体某一年或获取这一年的干支纪年，则由算法来实现，比如要想知道明洪武二年对应的公元年，只需将其起始时间 1368 加 1，即可得到明洪武二年为 1369 年。对于地点、地名或相关事件的时间属性取值，可不直接引用此时间模型的数据，既可用历史纪年表达，也可用公元纪年表达，利用此时间模型提供的历史纪年与公元纪年之间的相互转换服务，来获取公元纪年

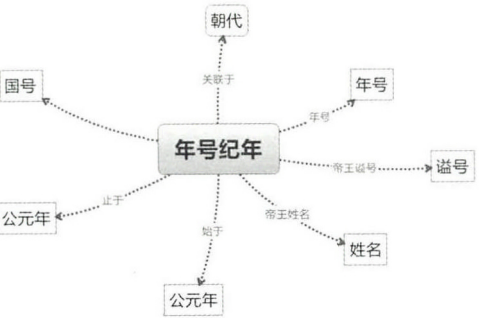

图 2 中国历史纪年的时间模型

对应的历史纪年或历史纪年所对应的公元纪年。

地点、地名、事件、时间等对象均以统一标识符(URI)来唯一标识,以实现对象在互联网环境下的访问、定位和关联。

3.2 中国历史地理数据开放应用的技术实现方案

上一节中提出了基于本体的时空数据模型,其功能实现需要相关技术的支持。在W3C提出的语义网框架中,本体处于数据模型的层面,其目的是将现实世界中的实体对象数据化,使得机器能够对实体对象的特征和实体对象间的关系进行查询、计算甚至简单的推理。要达到这一目的,需要相应的底层实现技术来支持,本节从数据的编码、数据的发布和开放应用两个方面来说明中国历史地理数据开放应用的技术实现方案。

3.2.1 数据的编码

为了历史地理数据在更广泛的范围内得到利用,建议历史地理数据的编码借鉴GeoNames的做法,采用W3C的推荐标准——RDF作为数据抽象模型和数据编码格式。RDF标准规范体系包括以"主体—谓词—客体"组成的三元组为最小单位的RDF抽象数据模型,和RDF/XML、Turtle、N3、JSON等一系列满足不同数据传输或处理目的、适应不同应用开发环境的数据编码格式,也叫序列化(Serialization)格式。"主—谓—宾"是知识组织与描述的通用模型,与本体的"对象—属性—属性值"一脉相承,有着很好的通用性和兼容性。RDF的各种编码格式是W3C的推荐标准,可很好地支持异构系统间的数据交换和传输,也可方便地被各种流行的编程语言处理。更重要的是,编码后的RDF数据可以存储在专用的RDF存储库而非关系数据库中,这样的RDF存储库也被称为"图数据库(Graph Database)"。

与关系数据库相比,图数据库是以三元组而非记录为数据的最小单位,以主体作为节点,以可重复、不限量的属性作为节点的分支,如果属性所指向的客体是另一个实体对象,则该客体又可作为另一组属性的主体,如此循环往复,成为相互关联的网状图形,如图3所示。这样的结构决定了数据的开放性和可扩展性。一方面,往某一个节点上增加属性和属性值时不会影响节点本身和整个数据库;另一方面,RDF存储库所用的RDF数据查询语言SPARQL,具有跨网域查询的功能,可对互联网上位于不同网域的数据源进行联邦查询,超越了关系数据库只能在局域网内对同一数据库进行查询的限制[20]。

3.2.2 数据的发布和开放应用。

Berners Lee提出开放数据的五星标准:将数据发布到Web上为一星,以机器可读的格式(如EXCEL)提供数据为二星,数据格式为非专业的机读格式(如CSV)为三星,采用开放的数据标准(如RDF、SPARQL)为四星,为数据建立更多的外部关联为五星[21]。为了实现中国历史地理数据在更广泛的范围内开放应用,笔者建议参照该标准进行数据发布。

以本体、RDF和关联数据为技术框架,可以很好地实现开放数据的五星标准,也有利于互联网环境下的规范控制[22]。关联数据是一种在Web上发布数据的方法,以HTTP URI(可理解为遵循Cool URI[23]稳定性、永久性原则的URL)作为各种对象的统一标识符(URI),例如"安昌县"以<http://data.library.sh.cn/placename/anchangxian/>作为URI,即可实现全网域范围内的唯一标识和定位(访问)。地名一旦被赋予HTTP URI,就具备了发布到Web的条件,并可方便地与Web上的其他数据集如GeoNames中的地名建立关联。关联数据要求数据以RDF序列化格式编码,可存储于本地RDF存储库中,以SPARQL进行数据查询,可联合查询本地RDF存储库和已发布在Web上的地名数据集如GeoNames,与本体结合,可以在不同的对象之间建立可被机器理解的关联关系。整体技术框架如图4所示。

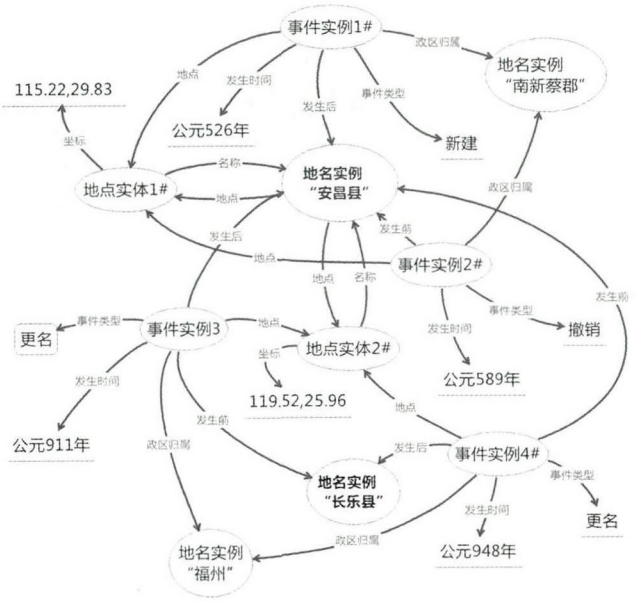

图 3 基于 RDF 图数据模型的地名编码示例

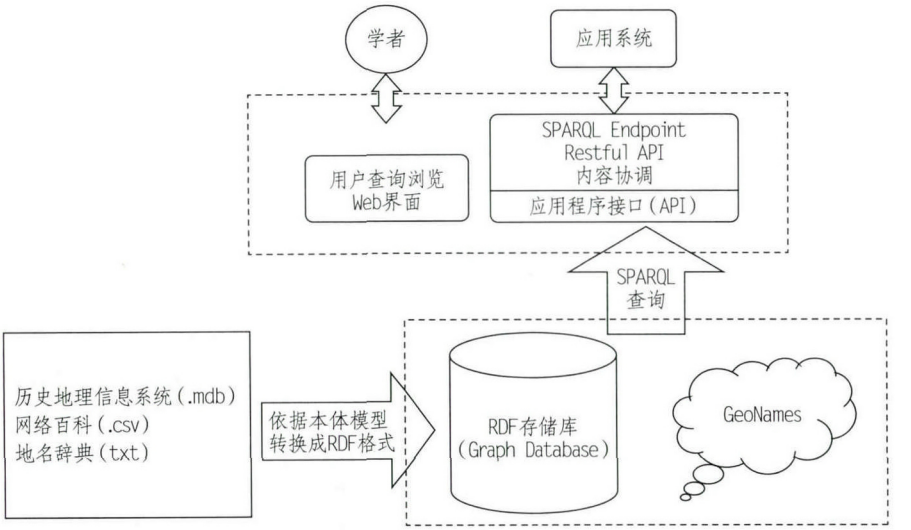

图 4 中国历史地理数据开放应用的技术框架

在 Web 上除了为人文学者和图书馆的读者提供查询和浏览界面之外，还要为其他应用系统提供历史地理数据的开放应用服务，这是本知识库建设的重要目标和需求之一，因而设计可被计算机应用程序调用的数据应用程序接口(API)就显得尤为重要，以关联数据为基础的数据服务接口技术也被称为关联数据消费技术。关联数据的消费接口有多种方式，如 DBPedia、FreeBase、GeoNames 等大型数据集均提供 SPARQL Endpoint、Restful API、定制开发工具包等多种多样的数据消费接口，以下三种方式基本可以满足程序员不同技术层次的需求。

(1) 内容协商。访问特定地点、地名的 HTTP URI 时，可获得其详细的 RDF 数据。当用普通的

浏览器访问时，系统返回供人阅读的 HTML 页面，当用语义浏览器或程序访问 HTTP URI 时，系统按照请求方通过 HTTP Header 传送的关于内容格式的请求返回相应序列化格式的 RDF 数据，如 RDF/XML、RDF/Turtle、JSON-LD 等。

(2) Restful API。是一种轻量级的 Web Service 技术框架，基于 HTTP 协议提供应用程序接口供程序调用，一般表现为包含各种输入参数的 URL。参数的数量、调用方法和返回数据的结构和格式由开发人员事先定义。这种方式因其简便性和跨平台性，已逐步成为数据应用程序接口的主流方式，可被多种程序语言如 C、Java、PHP、Python 等调用。

(3) SPARQL Endpoint。为熟悉 RDF 专用查询语言 SPARQL 的开发人员调用，与 Restful API 相比，可为开发人员提供更多的灵活性。要求开发人员对本体有着详细的了解。

表 2 以中国历史地理数据的 Restful API 为例，说明如何找到已知地名在不同时间范围内的不同地点及其他特征信息。表 3 以中国历史纪年的 Restful API 为例，其主要目的是实现中国历史纪年和公元纪年之间的相互转换。

表 2　中国历史地理数据应用程序接口实例

API 实例	功能说明
http://localhost:8080/webapi/data/安昌县?key=YourAPIKey	已知地名，找到地名在不同时间范围内的不同地点及其他相关地名，包括政区归属地地名和在某个历史时间的曾用名。
返回结果	
地名	“安昌县”
地点 1#	时间：526—589 坐标：115.22，29.83
政区归属	时间：526—589 归属地名称 “南新蔡郡”
地点 2#	时间：911—948 坐标：119.52，25.96
政区归属	时间：911—948 归属地名称 “福州”
其他名称	时间：911—948 其他名称 “长乐县”

表 3　中国历史纪年数据应用程序接口示例

API 实例	功能说明
http://localhost:8080/webapi/data/明?key=YourAPIKey	返回明朝的公元起止年
http://localhost:8080/webapi/data/明洪武?key=YourAPIKey	返回年号明洪武的公元起止年
http://localhost:8080/webapi/data/明洪武2年?key=YourAPIKey	返回明洪武二年的公元年
http://localhost:8080/webapi/data/1369?key=YourAPIKey	返回 1369 年的年号纪年

4　总结与展望

现代历史地理学、GIS 技术、数字人文、图书馆的数字图书馆建设发展到今天，已具备了相互渗透、彼此融合、互相提供方法与工具的条件。本文在调研现代历史地理学与人文社会科学各自的进展及二者之间关系的基础上，发现

现代历史地理学与数字人文的结合已成为数字人文研究与应用的热点和前沿，针对图书馆数字人文项目建设中存在的难点及问题，建议在图书馆的数字人文项目建设项目中引入现代历史地理学及 GIS 技术提供的数据、方法和工具，并利用图书馆领域擅长的知识组织和规范控制方法重组中国历史地理数据，以拉近专业性较强的历史地理信息系统与人文社会科学研究者和普通大众的距离，同时促进历史地理数据的开放利用。以此为目标和需求，站在现代历史地理学和数字人文研究者的肩膀上，借鉴其研究成果，设计了基于本体的历史地理数据时空模型，并提出利用关联数据技术在 Web 上发布历史地理数据和提供应用程序接口的技术实现方案，供同行批评指正。

作为国内最早关注并实施数字人文项目建设的图书馆之一，上海图书馆近年来正在建设旨在成为人文研究数据基础设施的一部分的数字人文服务平台[24]，以“人物、地点、时间、事件”为维度，重组上海图书馆丰富多样的馆藏历史文献资源，为读者和研究人员提供更精准的服务，为其他中小图书馆提供开放数据服务，“中国历史地理知识库”是其中极其重要的一部分。本文提出的时空数据模型和技术实现方案，将会在“中国历史地理知识库”的建设中进行试验验证和不断调整，并在此时空模型的基础上设计本体词表，进行形式化编码后发布到 Web 上，供同行及相关学者参考。

参考文献

[1] Nash P R，Asencio K. GIS and spatial analysis for the Social Sciences: coding，mapping，and modeling [M]. London: Routledge，2008: xiii-xvi.

[2] 陈刚. “数字人文”与历史地理信息化研究 [J]. 南京社会科学，2014(3)：136-142.(Chen Gang. Digital Humanities and informationization studies for historical geography [J]. Social Sciences in Nanjing，2014(3)：136-142.)

[3] 王晓光.“数字人文”的产生、发展与前沿 [C]//全国高校社会科学科研管理研究会组.方法创新与哲学社会科学发展.武汉：武汉大学出版社，2010：207-221.(Wang Xiaoguang. The origin，development and frontier of Digital Humanities [C]//National Institute of Scientific Research Management for Social Science of Universities. Methodological innovation and development of Philosophy and Social Sciences. Wu Han: Wuhan University Press，2010: 207-221.)

[4] 林珲，张捷，杨萍，等. 空间综合人文学与社会科学研究进展 [J]. 地球信息科学学报，2010，8(2)：30-37.(Lin Hui，Zhang Jie，Yang Ping，et al. Development on spatially integrated Humanities and Social Science [J]. GEO-Information Science，2010，8(2)：30-37.)

[5] 潘威，孙涛，满志敏. GIS 进入历史地理学研究10年回顾 [J]. 中国历史地理论丛，2012，27(1)：11-17.(Pan Wei，Sun Tao，Man Zhimin. The review of GIS entered into Chinese Historical Geography since 2000 and outlook [J]. Journal of Chinese Historical Geography，2012，27(1)：11-17.)

[6] 葛剑雄. 中国历史地理学的发展基础和前景 [J]. 东南学术，2002(4)：31-39.(Ge Jianxiong.The development basement and prospect of Chinese Historical Geography [J]. Southeast Scholarship，2002(4)：31-39.)

[7] 廖泫铭，范毅军. 中华文明时空基础架构：历史学与信息化结合的设计理念及技术应用 [J]. 科研信息化技术与应用，2012，3(4)：17-27.(Liao Hsiung-Ming，Fan I-Chun. Chinese civilization in time and space: the design and application of China Historical Geographic Information System [J]. E-science Technology & Application，2012，3(4)：17-27.)

[8] 徐永明. 中国古典文学研究的几种可视化途径——以汤显祖研究为例 [J]. 浙江大学学报：人文社会科学版，2016.(4)：1-21.(Xu Yongming.Some visualization approaches to the study of Classical Chinese literature: a case study on Tang Xianzu [J]. Journal of Zhejiang University(Humanities and Social Sciences Online Edition)，

2016.(4):1-21.)

[9] 曹睿. 面向家谱GIS的古今地名时空数据模型研究[D]. 南京:南京师范大学,2009:3-9.(Cao Rui. The research on spatio-temporal model for Genealogy GIS[D]. Nanjing:Nanjing Normal University,2009:3-9.)

[10] 吴陆. 社科院携手SuperMap首开社科领域GIS应用先河[J]. 中国测绘,2009(5):86-86.(Wu Lu.GIS application in Social Science as precedentby academy of Social Sciences and SuperMap[J]. China Surveying and Mapping,2009(5):86-86.)

[11] 陈旻. 华夏家谱GIS建设关键技术研究[D]. 南京:南京师范大学,2009:35-54.(Chen Min. The research on key technologies of Sino-family-tree GIS[D]. Nanjing:Nanjing Normal University,2009:35-54.)

[12] 胡颖. 家谱GIS中古今地名的时空关系研究[D]. 南京:南京师范大学,2008:7-17.(Hu Ying. The research on spatio-temporal modeling for Genealogy GIS[D]. Nanjing:Nanjing Normal University,2008:7-17.)

[13] 温永宁,闾国年,陈旻,等. 华夏家谱GIS的数据组织与系统架构[J]. 地球信息科学学报,2010,12(2):235-241.(WenYongning,Lü Guonian,Chen Min,et al. Data organization and system architecture of Sino-family-tree GIS[J]. Journal of Geo-Information Science,2010,12(2):235-241.)

[14] 复旦大学历史地理研究中心. CHGIS数据说明[EB/OL]. [2016-12-10]. http://yugong.fudan.edu.cn/views/chgis_data.php.(Center for Historical Geographical Studies of Fudan University. Instruction of data[EB/OL]. [2016-12-10]. http://yugong.fudan.edu.cn/views/chgis_data.php.)

[15] 傅君劢. 中国历代人物传记资料库用户指南[EB/OL]. [2016-12-06]. http://www.zggds.pku.edu.cn/006/cbdb/usersguide.pdf.(Fuller M A. The China biographical database user's guide[EB/OL]. [2016-12-06]. http://www.zggds.pku.edu.cn/006/cbdb/usersguide.pdf.)

[16] Berman M L. Historical gazetteer system integration: CHGIS, Regnum Francorum, and GeoNames[EB/OL]. [2016-12-10]. http://www.people.fas.harvard.edu/chgis/gazetteer/AAG_GazIntegration_Revised_2014.pdf.

[17] 周丙锋,周文业. 基于中国历史地理信息系统CHGIS的中国历史地理数字化应用平台[C]//中国地理信息系统协会年会,2007:556-580.(Zhou Bingfeng,ZhouWenye. The Chinese historical geography digital application platform based on CHGIS[C]//Chinese Historical Geographical Information System Annual Conference,2007:556-580.)

[18] Schmachtenberg M,Bizer C,Paulherm H. State of the LOD Cloud[EB/OL]. [2016-11-26]. http://lod-cloud.net/state/.

[19] 戴均良. 中国古今地名大词典[M]. 上海:上海辞书出版社,2005:22-50.(Dai Junliang. China ancient names dictionary[M]. Shanghai:Shanghai Dictionary Press,2005:22-50.)

[20] Harris S,Seaborne A,Prud'hommeaux E. SPARQL 1.1 query language[EB/OL]. [2016-12-08]. https://www.w3.org/TR/2012/PR-sparql11-query-20121108/.

[21] Berners-Lee T. Linked Data[EB/OL]. [2016-12-12]. https://www.w3.org/DesignIssues/LinkedData.html.

[22] 刘炜,张春景,夏翠娟. 万维网时代的规范控制[J]. 中国图书馆学报,2015,41(3):22-33.(Liu Wei,Zhang Chunjing,Xia Cuijuan. Authority control for the Web[J]. Journal of Library Science in China,2015,41(3):22-33.)

[23] Sauermann L,Cyganiak R,Ayers D,et al. Cool URIs for the semantic Web[EB/OL]. [2016-12-18]. https://www.w3.org/TR/cooluris.

[24] 刘炜,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.)

夏翠娟 上海图书馆高级工程师。上海 200031。

(收稿日期:2017-01-26)