

东南海疆研究数据库的构建与发展策略*

郑咏青 王 爽 (厦门大学图书馆)

摘 要 东南海疆区域具有典型的海洋文化特征,厦门大学图书馆着力收集以族谱、碑刻、田野调查、地方史料为主的民间资料和中外文地图资料,构建了独具特色的“东南海疆研究数据库”。构建过程包括主题大纲的确定、数据库文献类型及收录范围的确定、元数据标引中主题词表与地区名称表的定制、用户需求的分析以及系统架构等。该数据库内容丰富、持续更新、技术完善,但也存在数据收集不均衡、使用率低的问题。针对不足,未来该数据库将分级定位地理区块,全面收藏各类资源;并尝试以核心人物和热点问题为线索构建数据库专题;探寻灰色文献收集渠道;同时依托厦大研究学者众多的优势,借助校友力量采集数据;以专题服务为重点,参与院系教学科研;更须着力提升用户体验,宣传推介数据库的使用。

关键词 东南海疆 区域研究 特色数据库 发展策略

DOI: 10.13663/j.cnki.lj.2015.03.007

The Construction and Development Strategies of "Database on Southeast Coastal Area Studies"

Zheng Yongqing, Wang Shuang (Xiamen University Libraries)

Abstract Southeast coastal area has typical characteristics of marine culture. Xiamen University Libraries focus on collecting folk data and map data in Chinese and foreign languages mainly in the formats of genealogy, inscription, field survey and local historical data, and has thus constructed a unique "Database on Southeast Coastal Area Studies". The construction process includes determination of the main subjects, genres and collecting scope of documents, customization of thesaurus and list of area names in metadata indexing, as well as analysis of user demands and system architecture. The database is abundant in content, updated continuously with improved technology, but also has shortages of imbalanced collection and low usage rate. In order to overcome the shortages, the database will be classified by geographical areas in future and will collect various resources comprehensively. It will try lining up database projects by core figures and hot issues. Collection channels of gray literature are explored. Meanwhile, data is collected by virtue of numerous outstanding researchers of Xiamen University, whereas scientific research in colleges and schools also contribute to the information services. User experience should be greatly enhanced, and the usage database should be publicized and promoted with vigorous efforts.

Key words Southeast coastal, Area studies, Characteristic database, Development strategies

1 东南海疆研究数据库的构建背景

东南海疆系指我国东南沿海苏浙闽粤台港澳琼等省区以及东海、南海等海域,与日本及东南亚多国有着漫长的海洋国界线。该区域具有典型的海洋文化特征。历史上,百越民族曾经在此区域活动,秦汉以来,中原文化交融于此,宋元时期的泉州是世界贸易大

港,明末的漳州月港被定为中国帆船的发船地,清初的厦门则被定为往南洋贸易的发船地,近代的五口通商亦都在东南海疆^[1]。当代的东南沿海,则是中国经济最活跃、对外交往

* 本文系 CALIS 三期专题特色库项目阶段成果(项目编号:4401-FJ-201)

最频繁的地区,近年来,为加快地方文化建设,不少沿海城市都提出了建设“海洋文化名城”的发展战略。同时,在地缘政治经济背景下,我国与日韩及东南亚五国有海洋争端。有关东南亚的区域研究在国际上已经成为研究热点。

区域研究是人文社会科学的一个重要研究领域。厦门大学在台湾与东南亚研究、区域社会经济史、海洋社会经济史、移民与海外华人、海疆史地及现代国际关系等方面研究成果显著,相关院系在国际经济法学、区域经济、海洋与环境科学、海洋与海岸带发展等领域研究也取得了丰硕成果。为了更好地开展区域研究的文献保障工作,厦门大学图书馆于2010年设立了区域研究资料中心,着力收集以族谱、碑刻、田野调查、地方史料为主的民间资料和中外文地图资料,构建独具特色的“东南海疆研究数据库”(以下简称“该数据库”)。

2 东南海疆研究数据库的构建策略

2.1 构建内容

该数据库的建设主要包括以下内容:

(1) 构建数据库软硬件平台,以实现动态管理和发布本领域的相关数据,支持文本、图像、音频、视频等多种媒体信息;

(2) 制订文献资料描述元数据规范,以使各类数据能被正确标引,规范表达;

(3) 整理收集有关我国东南地区及东南亚各国的政治、经济、军事、法律、外交、社会、移民、历史、地理、文化、宗教、民俗、民族、考古、语言、文学、艺术、海洋环境与生态等方面的相关资料,包括书目、题录、文摘等二次文献,期刊论文、会议论文、专著及手稿等的全文,历史图片及实物照片等图片资料,音频视频及网络原生数字资源等多媒体资料。其中,与台海及东海、南海相关的族谱、碑刻、田野调查、地方史料等民间资料和中外文地图资料是建设期间的收集重点;

(4) 标引加工所收集的各类资料;

(5) 探索资料收集与共享模式。

2.2 构建步骤

(1) 制订主题大纲

为便于数据的加工标引,笔者在数据库建设之初制订了主题大纲,该大纲在数据库软件设计时以表单形式体现,在标引界面通过“显示主题”和“显示地名”的下拉菜单显示。数据标引人员可根据文献内容勾选相应的主题与地名,还可根据所标引的文献数量多寡在“词表管理”菜单中对其进行增删管理。

(2) 确定数据库文献类型及收集范围

本数据库选取的文献类型主要包括:国内外期刊论文、会议论文、专著图书及专著图书析出文献的书目信息。书目信息是学术研究的基础。故应以收集齐备为原则。除报道正式出版物之外,还应揭示非正式出版物如内部刊物、手稿、舆图、档案等。期刊、专著图书、会议论文、学位论文的全文。此类属精选文献,应努力反映既往的最高依据水平或最新成果,选取范围主要为高频被引文献、著名专家学者的论著或热点研究的相关论著。图片资料。本领域内有关人物、习俗、曲艺、工艺及碑刻、实物等方面的图片分藏于各单位及个人,其中不乏珍贵文献,宜着力有选择地收集。音频视频资料。除收集概况资料之外,将着力收集具有强烈地方文化和海洋文化气息的戏剧音乐、民俗方言、民间文学、民间技艺等方面的资料。包括电子剪报和站点导航在内的网络原生数字资源。

舆图资料,收集国内外各类单张地图、地图集。

(3) 数据加工

在加工格式方面,文本文件采用PDF格式,图像文件为JPG格式,音、视频文件为MP3、RM格式。在数据加工的质量控制方面,须做到对象数据链接准确,文本信息尽量通过应用现有软件或编程实现自动抓取与转换,扫描图片尽量以较高的分辨率存档。

(4) 元数据标引

在规范应用方面,笔者采用DC核心元素集,统一制订了元数据输入工作表单,便利

数据的加工标引。该表单包括题名、创作者、其他责任者、出版者、出版时间、描述、主题、地名、学科代码、关键词、分类号、类型、语种、来源、关联、权限及全文链接等字段。数据格式采用《CALIS 专题特色数据库信息描述元数据规范》及《CALIS 特色库管理型元数据规范》，以使各类数据能被正确标引、规范表达。同时，笔者根据东南海疆区域研究相关主题制订了专门的主题词表（见表 1），对于地区（Coverage）字段也拟定了针

对性的详细地区表（见表 2）。此外还设置了独有的专题（Topic）字段，将部分数据按专题组织，现有专题四个：陈笃生家族、东南亚华文文学、钓鱼岛问题、South China Sea。在编码体系方面，学科代码采用“授予博士、硕士学位和培养研究生的学科、专业目录”中的学科及代码、分类号采用“中国图书馆分类表”、类型采用“专题特色数据库信息资源名称规范列表”等，未对元数据集进行扩展。

表 1 “东南海疆数据库”主题词表

主 题	Subject	备 注
华侨华人		
福建人、闽南人	Fujianese	
文化、教育、宗教与社会		
城市化	Urbanism	
社会问题	Social Issues	毒品问题等
社会服务	Social Services	
文学、艺术	Art & Literature	
佛 教	Buddhism	
教 育	Education	
伊斯兰教	Islam	
劳工问题	Labor	
语言文字	Language	
移 民	Migration	难民入此
人 口	Population	
民族志	Race	民族问题入此
新闻传媒	Press & media	
民 俗	Folklore	
哲 学	Philosophy	
五缘文化	Five-relationship Culture	五缘指亲缘、地缘、神缘、业缘和物缘
体 育	Sports	
基督教	Christianism	
道 教	Taoism	
经 济		
经济史	History of Economy	

(续表)

主 题	Subject	备 注
工业经济	Industry	
投 资	Investment	含 FDI , 外商直接投资
跨国公司经营活动	Multinational	
贸 易	Trade	出口加工业入此
贸易史	Trade history	
交通运输业	Transportation	
旅游业	Tourism	
农业和渔业经济	Agriculture & Fisheries	RuralEconomy 农业和渔业经济
银行、资本市场与金融	Bank, Capital Market & Finance	Debt 债务入此
经济合作	Economic Co-operation	含南北经济合作、南南经济合作等
东南亚经济危机	Economic Crisis	
财政、税收	Finance & Taxation	
经济政策与管理	Economy Policy & Management	
企业经济	Enterpreneur	
科技发展与实用	Science Development & Utility	
地球、海洋、大气	Earth Science, Marine Science, Atmospheric Science	
数学、物理、化学	Mathematics, Physics, Chemistry	
环境与地理		
环境保护	Environmental Protection	
名胜古迹	Scenic & Historic Site	
自然资源	Natural Resouces	
材料科技	Materials Engineering and Science	
机电工程	Electrical and Mechanical Engineering	
电信、电子工程	Telecommunications, Electronic Engineering	
其他工程技术	Other Engineerings	
航空航天	Aerospace Engineering	
化工环保	Chemical Engineering, Environmental Engineering	
能源工程	Energy Engineering	
建筑、土木、水利	Construction, Civil Engineering	
医学、医学工程	Medicine, Medical Engineering	
信息系统与自动化	Information Systems and Automation Engineering	
食品科技	Food Science and Technology	
生物科学与技术	Biological Science and Technology	
历 史		
公元 1951 年—2000 年	From 1951 to 2000	
公元 2001 年以后	After 2001	
人物资料	Biography	
古代史	Ancient History	

(续表)

主 题	Subject	备 注
公元 1950 年以前	Before 1950	殖民地史入此
国际关系		
地区安全	Regional Security	
领土争端	Territorial Dispute	
国际条约	Treaty	
区域合作	Regional Co-operation	Multilateralism 多边主义入此
对外关系史	History of IR	
政治、军事		
政治问题	Political Issues	Decolonization, Nationalism, Separatism 去殖民化、民族主义、分离主义入此
政治改革	Political Reform	
共产主义运动	Communism	
法 律	Law	
军 事	Military	西方国家军事基地入此

(续表)

表 2 “东南海疆数据库”地区 (Coverage) 名称表

地 区	Coverage	备 注
东南亚		
东盟 (含 AEC)	ASEAN	全称“东南亚国家联盟”，又译“亚细安”
亚太经合组织	APEC	
文 莱	Brunei Darussalam	
柬埔寨	Cambodia	
印度尼西亚	Indonesia	历史名词：新几内亚，东印度
老 挝	Laos	
马来西亚	Malaysia	
緬 甸	Myanmar	
菲律宾	Philippines	
新加坡	Singapore	
泰 国	Thailand	
东帝汶	Timor Leste	
越 南	Vietnam	
中国南方		
云 南	Yunnan	
浙 江	Zhejiang	
广 西	Guangxi	

地 区	Coverage	备 注
海 南	Hainan	
南中国海	South China Sea	
台 湾	Taiwan	
福 建	Fujian	
广 东	Guangdong	
贵 州	Guizhou	
上 海	Shanghai	
江 西	Jiangxi	
江 苏	Jiangsu	
东 海	East China Sea	
美 洲		
拉丁美洲	Latin America	
美 国	United States	
加拿大	Canada	
亚 洲		
香 港	Hong Kong	
日 本	Japan	
中国大陆	China (Mainland)	中国与东盟关系入此
中国台湾	Chinese Taiwan	
东 亚	East Asia	

(续表)

地 区	Coverage	备 注
韩 国		Korea
印 度		India
朝 鲜		Korea(DPRK)
澳 门		Macao
欧 洲		
葡萄牙		Portugal
俄罗斯	Russia	含前苏联
西班牙		Spain
英 国		United Kingdom
法 国		French
荷 兰		The Netherlands
欧 盟	EU	全称“欧洲联盟”(European Union), “欧共体”(European Communities)入此
大洋洲		
新西兰		New Zealand
巴布亚新几内亚		Papua New Guinea
澳大利亚		Australia
国际组织		
世界银行		World Bank
非政府组织		NGOs
OECD		OECD
亚洲自由贸易协定		AFTA
亚洲开发银行		ADB
世界贸易组织		WTO
太平洋经济合作理事会	PECC	前身为“太平洋经济合作会议”
国际货币基金组织		IMF

(5) 需求分析

本数据库的目标用户群为首先为厦门大学的相关院系的师生,其次为国内外本领域研究人员、党政部门决策人员、文化艺术工作者及关注本领域研究的一般用户。笔者针对该数据库的潜在用户进行了需求调查。针对该调查的分析表明,在服务功能方面,数据库应具备较

强的检索功能: 能够提供题名、作者、主题、关键词、摘要、地区及分类号的基本检索及高级组合检索; 同时提供一、二级主题的浏览。检索命中或浏览的结果在同一窗口显示简要信息,详细信息通过点击题名的超链接在新窗口显示; 用户能够多种途径对检索结果根据各自所需进行排序; 采用 Web 方式进行发布。用户只需登录网站即可访问。用户的反馈意见通过电子信箱与管理人员联络。设置留言讨论版、动态信息发布栏以及文献传递链接等模块,以扩展数据库的服务功能; 提供访问统计和工作量统计。

系统架构

最终的系统架构方案搭建在 Windows 平台上,数据存储 in SQL Server 2008 中,程序使用 .net 编写,检索采用 Lucene 搜索引擎,并提供了数据导入、数据编辑、管理及发布、OAI 接口等完整的服务功能。数据结构详见图 1。

3 数据库的发展优势与不足

3.1 内容丰富

该数据库现有数据 33 万余条,全文数据库近 29 万条,全文数据库占数据总量的 87%。其中包含厦门大学图书馆此前建成的“东南亚及闽台研究数据库”所有内容,厦门大学东南亚研究中心的“东南亚数据库”部分内容,以及厦门大学台湾研究中心的大量台湾研究资料。1946 年 5 月成立的“海疆学术资料馆”所收集的研究南洋、台湾及闽南地方史与海外交通史问题的资料在 1950 年 9 月并入厦门大学“南洋研究馆”,这些资料目前仍大部分藏于厦门大学图书馆和厦门大学南洋研究院,已按计划逐步进行数字化,整合到数据库中。

3.2 持续更新

厦门大学图书馆配备了专门的标引人员和技术人员负责该数据库的建设与更新。为充实数据库,近年来相关人员多方搜集相关资源,网罗各类资料,重点收集第一手资料: 请有摄影专长的馆员实地拍摄厦门大学校园实景照片,记录校园发展变化,历时一年,拍摄原片 1000 多张,精选入库; 将厦门大学台湾研究院院长陈在正教授漳州地区田野调查族谱

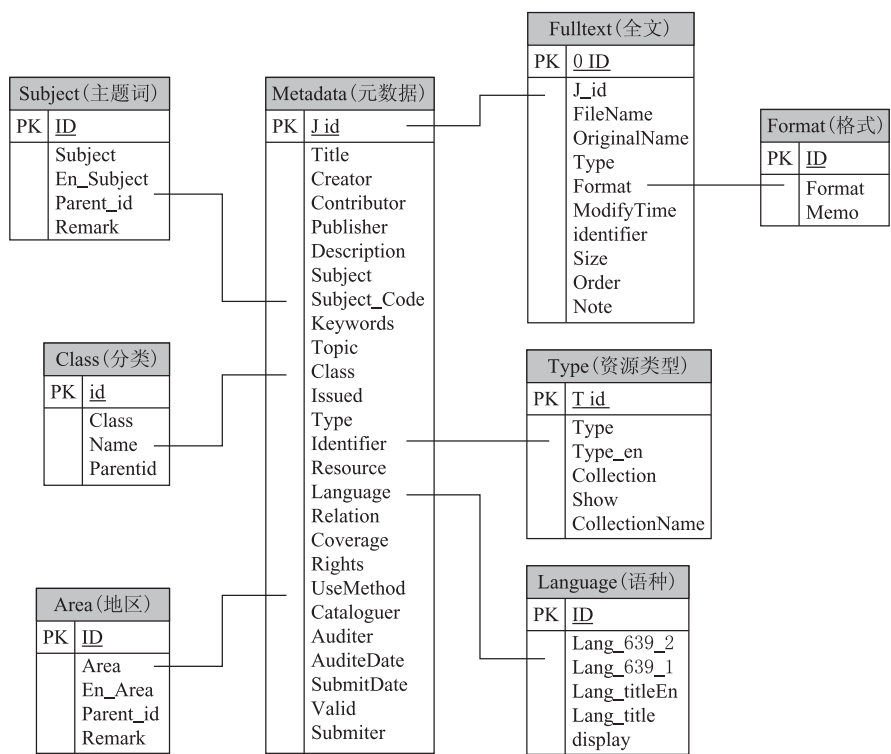


图 1 “东南海疆研究数据库”的数据结构

进行复印资料整理，数字化族谱文档 28.8GB，合为 129 册族谱入库；配合厦门华侨博物院“笃行一生继往开来——一个华人家族与新马社会”展览^[2]，搜集新加坡侨领陈笃生家族专题资料 184 篇，包括新闻、照片、视频、族谱等多种类型，均为第一手资料。扫描古地图 100 多幅入库，地图来源为本馆藏书：《中华古地图珍品选集》、《中国古代地图集（清代）》；编辑刊印内部资料《钓鱼岛问题资料汇编》，后续形成“钓鱼岛专题”数据 700 多篇；与厦门大学南海研究院合编电子期刊《南海导报》，依托图书馆资源，搜集有关南海问题的中英新闻、会议预报和书目信息，相关资料悉数收入数据库中。不断更新的各类资源拓展了数据库的内容，提升了数据库的学术内涵。

3.3 技术完善

该数据库采用基于通用数据库平台的自建数据库系统，实现动态管理和发布本领域的相关数据，支持文本、图像、音频、视频等多种媒体信息，同时采用《CALIS 专题特色数据库

信息描述元数据规范》，在数据加工、元数据标引等方面符合规范。

除了以上优势外，该数据库建设也存在一些问题。主要有三点：未能均衡收集特色资源。目前的资源收集策略偏重于网络资源，主题相对集中于台湾政情与科技报告方面。该数据库所服务的研究领域对非正式出版的文献如谱牒、手稿、票据、碑拓、图片、口述资料等的需求较高，而这些资料在该数据库中尚未达到一定规模。数据加工标引与资源收集速度未能同步。由于精力所限，相关人员未能对该数据库进行必要的宣传和推介，使用率不高。

4 数据库发展策略

经过数年建设，该数据库已有 33 万条数据，涵盖东南海疆人文、社科及海洋环境与生态等多学科的相关文献。作为一个区域研究综合数据库，在数据量达到一定规模之后，今后的发展须注重以下几个方面：

- (1) 以专题服务为重点，参与院系教学科

研。馆员应深入与院系师生交流,寻找合作项目,协助院系进行项目资料的查找与汇聚,同时可将专题服务资料收入特色库,如前述与厦门大学南海研究院联合编辑电子期刊《南海导报》的合作模式;另可借鉴新加坡国立大学图书馆做法,馆员参与人文学院师生每年开展的教学实践和田野调查活动,协助收集整理第一手资料,助力师生完成科研课题,也将原始材料及研究成果入库存档和传播。

(2) 分级定位地理区块,全面收藏各类资源。针对东南海疆地域范围较广,数据类型丰富,来源途径众多,资源宽泛的特点,界定收藏范围较为关键,未来须将该区域分级定位,将大区域拆分成一个个小范围,如福建区域,可依行政区域划分,如厦门、福州、龙岩等,也可依地域风土区分,如莆仙地区、闽南地区、闽北山区、福州五区八邑等,立足当地风土人文,全方位收集当地文献。长期的计划是尽量收集齐全东南海疆区域正式出版的研究成果,以及经由网络途径可获取的信息资源,关注即将出版的相关研究成果,形成可持续发展的区域研究资源特色。

(3) 以核心人物和热点问题为线索构建数据库专题。着手挖掘整理东南海疆区域核心人物事迹,拟编制核心人物名册,逐一追溯人物发展轨迹,全面梳理人物生平、事件、贡献、时代背景、家庭关系、历史评价、中外影响等,收藏人物日记、手稿、著作、后人研究成果等资源,后续可建成区域核心人物库。该区域热点问题频出,如南海问题、钓鱼岛问题、两岸关系问题等,以问题为主导收集文献资料较贴近时代需求,数据的时效性强,受关注程度也将随之提高。

(4) 探寻灰色文献收集渠道。馆员可针对性地联络以区域人文建设为主要职能的政府机

关和社团组织,如本地文化局、建设局、方志办、侨联、民间社团等,了解这些部门组织是否有灰色文献,其出版规律和内容品质如何,有选择地收集文献质量高、对区域研究有帮助的资料入库。可采用征集、交换、捐赠和适当购买等方式获取,也可借阅资料,由图书馆协助进行数字化收藏传播,而后原件归还。

(5) 依托厦大研究学者众多的优势,借助校友力量采集数据。退休学者部分已淡出研究领域,早期研究资料留之无用,弃之可惜,而对图书馆而言这些资料仍弥足珍贵,馆员可广泛宣传特色库的收集宗旨和范围,与学者深入沟通过了解其研究资料的内容,评判其回收利用价值,决定是否收集入藏。馆员对本校学者的研究领域较为了解,沟通交流相对顺畅,而这些学者大多居住本地,资料搬运相对便捷,可操作性强,但难点在于资料评价与后期的整理、数字化加工以及标引。

(6) 提升用户体验,大力宣传推介数据库的使用。馆员可针对研究型人群进行宣传推广,保持馆员与数据库用户的互动沟通,开展定期不定期的数据库使用调研,及时关注用户在数据收集或数据库使用方面遇到的问题,并针对性地予以解决,使该数据库成为用户科研的必备平台。

5 结语

构建“东南海疆研究数据库”,依托厦门大学“侨、台、特、海”的区位与学科优势,以台海工作及维护东海、南海主权服务为重心开展各类文献资源的收集加工,对促进祖国统一、维护国家主权和领土完整,具有重要的现实意义。笔者将不断思考项目发展策略,完善数据库内容,提升数据库的学术价值,助力相关领域的学术科研。

参考文献

- [1] 庄国土. 论17-19世纪闽南海商主导海外华商网络的原因[J]. 东南学术, 2001(3):64-73.
- [2] 陈笃生家族专题展在厦们展出 展示新马侨胞奉献[EB/OL]. [2014-03-24]. <http://world.chinanews.com/zgqj/2011/08-03/3230923.shtml>.

郑咏青 厦门大学图书馆区域研究资料中心, 馆员。
E-mail: cutedang@gmail.com 福建厦门 361005
王爽 厦门大学图书馆信息技术部, 馆员。 福建厦门 361005

(收稿日期: 2014-04-03)