

亚历山大数字图书馆 ——互联网上的地理信息资源中心

□ 刘燕权 / 美国南康涅狄格州立大学 美国康州纽黑文 06515
孙秀良 / 北京化工大学图书馆 北京 100029
池煜 / 中建材集团进出口公司 北京 100048

摘要: 亚历山大数字图书馆项目 (Alexandria Digital Library, ADL) 是美国最大的、持续时间最长的数字图书馆研究项目之一, 自1994年起连续得到了美国国家科学基金会 (NSF) 的美国数字图书馆倡导计划 (Digital Library Initiative, DLI) 长达十年的资助。该项目的重点在于设计并运行其颇有特色的地理参考数字图书馆。ADL是一个分布式参考资料数字图书馆, 以各种数据来源的地图、图像和地理定位文献为主要收集对象, 实现公众对地理空间信息对象的访问和搜索, 是互联网上的地理信息资源中心。文章对该数字图书馆的项目背景、信息资源组织、服务和技术特点等进行了介绍和评价。

关键词: 数字图书馆, 亚历山大项目 (ADL), 美国国家科学基金会 (NSF), 美国数字图书馆倡导计划 (DLI), 地理参考资源, 地理空间信息, 分布式数字图书馆

DOI: 10.3772/j.issn.1673—2286.2013.11.012

1 项目概述

1.1 背景介绍

亚历山大数字图书馆 (Alexandria Digital Library, 简称ADL, 网址<http://www.alexandria.ucsb.edu/>) 是美国国家科学基金会 (NSF) 资助的美国数字图书馆倡导计划 (Digital Library Initiative, DLI) 项目之一, 它由美国加州大学圣巴巴拉分校 (University of California at Santa Barbara, 简称UCSB) 戴维森图书馆的地图和图像实验室 (Map and Imagery Laboratory, 简称MIL) 于1994年创建。该项目的重点在于设计并运行其颇有特色的地理参考数字图书馆。ADL在接受NSF的美国数字图书馆一期工程 (DLI-1, 1994-1998)

资助的基础上, 继续接受其二期工程 (DLI-2, 1998-2001) 的资助, NSF的资助一直延续到2005年。在二期工程中, 该馆启动了亚历山大数字化地球图书馆原型 (ADEPT) 项目。目前, ADL的运营由戴维森图书馆 (Davidson Library) 资助并支持, ADL的技术研究方面分别接受USCB和斯坦福大学的美国国会图书馆的数字保存计划资助。ADL作为研究机构、开发机构以及教育机构, 收集来自图书馆、学术机构、政府机构、私营企业等各种数据来源的地图、图像和地理定位文献, 汇聚了各种机构的地理参考数据, 这一方式类似于古埃及的亚历山大数字图书馆 (它是当时知识和学习中心), 因此, 该项目取名为“亚历山大数字图书馆”, 意为构建互联网上的地理参考资源中心。

1.2 项目使命

ADL的目标是构建一个汇集地理学相关信息资源的分布式数字图书馆, 并实现公众对地理空间信息对象的访问与搜索; ADEPT的目标是将ADL构建的地理参考资源成功引入到本科和研究生的教育中。UCSB戴维森图书馆的MIL团队和开发团队肩负着完成ADL重要使命的任务, 它们建立了一个分布式理参考资料数字图书馆, 开发相应的技术支持该数字图书馆的发展, 设计、建立并评价该项目的测试系统, 最终实现数字图书馆的正常运营。截止1999年, ADL已经成为正式运营的数字图书馆, 能够为广大的用户群体提供明确的访问和搜索服务。目前, ADL仍致力于收集世界各地的地理参考信息资源, 并为

公众提供该资源的访问和搜索。

1.3 项目意义

地图是人类保存下来的最早文献之一,一直以来都以物理馆藏进行存储。但地理相关资源的存储有其自身的特点。其资源类型十分独特,收集价值高,存储占用空间大,地图的编目比较复杂。因此,物理馆藏不易保存地理相关文献,而且不方便用户的检索和利用。ADL为用户提供了一条查找地理参考资源的捷径。用户不用到特定的物理馆藏去实地查找资料,只需通过远程访问,就可以方便地查找自己感兴趣的内容,扩充了其资源的可用性;数字存储可以避免物理馆藏带来的弊端,提高图书馆和文献的管理效率;分布式系统可以加强与用户的交互能力,便于不断完善用户界面的设计;还能够通过地理坐标实现信息检索,可以将具体的地理位置表达成地理坐标(即经度和纬度),对于图书馆界来说,这是一个全新的概念。ADL极大地扩展了用户搜寻地理学相关资料的能力,降低了用户筛选地理参考资源的成本。

2 资源组织

2.1 资源范围与类型

ADL除了拥有地图、图像等普通的地理参考资源,还收集航拍照片、栅格地图、数字高程模型、遥感图像和卫星照片等稀缺资源。整个图书馆拥有36个数据库、1000件音像资料和缩微资料、15000多幅地图。收藏大约700万份文献目录记录,包括地理空间对象存储库

以及地名索引,提供的数据量达10万亿字节,且以每年1.2—2万亿字节的速度增加。ADL地名表收录了4,437,405个覆盖全球的地理实体和所涉及的5,947,611条地名、别名或历史地名以及与之相关的地理空间信息,形成了一个庞大的地理信息系统。

2.2 数据来源

ADL的数据来源十分广泛,汇聚了各图书馆、学术机构、政府机构、私营企业等各种机构的地理参考数据来源。其中,重要的数据来源机构有戴维森图书馆的地图和图像实验室(MIL)、美国图像地图机构(National Imagery and Mapping Agency, NIMA)、美国地质调查部门(U. S. Geological Survey, USGS)和美国中央情报局(U. S. Central Intelligence Agency, USCIA)等。目前,ADL继续加大对全世界地理信息资源的收集,新的数据来源机构有国家地理空间档案馆(The National Geospatial Digital Archive, NGDA)等,收集的新资源包括新西兰和南加州的航拍地图,希尔维斯特市地质幻灯片和墨西哥、危地马拉和伯利兹城的玛雅森林地区的卫星图像等。

数据来源的评价方面,ADL有着严格的把关和收录标准。首先,内容尽量涉及具有代表性的地理空间信息资源;其次,内容应该具有明确的价值,而且便于用户访问;第三,材料是独特的,注重“局部”区域;第四,必须遵循ADL的研究和发展使命,实现分布式的信息收集方式;最后,还要尽可能降低ADL的运营负担。ADL团队较高

标准的审核机制,保证了资源的可利用性。

2.3 元数据方案

ADL采用以文件集为单位(相对于以单个文件为单位)著录的元数据结构。ADL所接受的查询是利用按照层次关系限定的词汇、语句、数值、符号、日期等标准的图书馆学语义以及地理空间界定词等扩展的语义对象来定义的。ADL编制了地名索引,地名辞典是其重要部分。ADL专门设计了一个描述地名类型的主题词表——《地理特征主题词表》(Feature Type Thesaurus, 简称FTT)。FTT有助于对地名进行归类和检索、区分同名的地理实体。FTT是严格按照树形等级结构组织排列的,最深处有5层。它收录了210个正式词和1046个非正式词,分属行政区划(administrative areas)、水文特征(hydrographic features)、陆地(land parcels)、地文学特征(physiographic features)和区域(regions)6个词族^[9]。词族是指一组具有等级关系的主题词集合,6个族首词分别作为树形结构的根节点,其他所有主题词都依据属分的上下位等级关系,组织到父节点(上位词)和子节点(下位词)中。ADL的桶式(bucket)框架,为跨越不相似馆藏检索而设计,特别要求在馆藏的物件类型和文件格式描述成分中使用带有等级结构的词集,目的在于将这些术语用于物件元数据和地理词典条目的描述内容。在自然科学教学方面,ADL设计和建制了概念空间,概念空间模型实际上扩展了叙词表模式,对知识领域中概念集的表达更加全面和完善。

3 信息服务

3.1 目标用户

ADL 不仅为一般用户提供信息服务, 而且为本科生和研究生的教学提供服务, 面向的服务对象比较广泛: 有研究人员、教育者、学生以及所有对地理学相关领域感兴趣的用。它不仅为大众提供适合自己的地理参考资料, 而且为教学人员提供很多方便易用的教学参考资料。另外, 一些更稀缺、有特点的航拍图片、栅格地图和卫星图片等资源也适合各类人群的认识了解。

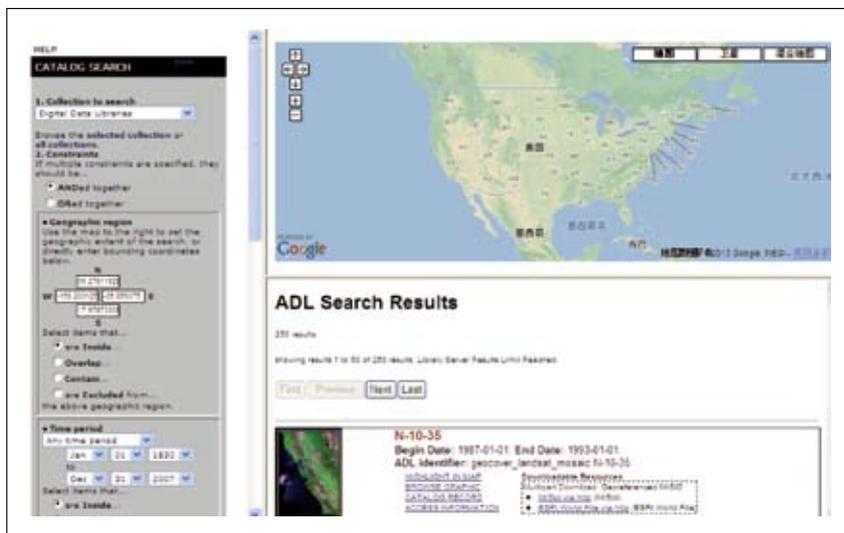


图1 数字图书馆互联网客户端 (Web Client)

3.2 服务方式

ADL 的服务方式比较多样, 基本功能齐全, 包括检索、浏览, 同时还能为教学提供服务, 能满足用户的不同需求。

3.2.1 检索功能

ADL 可以通过两种不同的方式来检索——数字图书馆互联网客户端 (Web Client) 和数字图书馆地名辞典客户端 (Gazetteer Client), 两个客户端的功能都比较齐全。

互联网客户端的检索主页如图1所示, 分为三个主页区域, 检索选项位于主页的左侧, 地图浏览器位于右侧上方, 右下方是结果显示界面。地图的浏览器可以在世界地图上选择其中任何一个正方形的网格, 进行搜索特定的位置。用户也可以输入经纬度坐标找到相应要搜索的地域。除了这些选项, 用户还可以通过选择时间、对象类型、对象格式、关键字、出版商、排序方



图2 数字图书馆地名辞典客户端 (Gazetteer Client)

法等等不同的选项来筛选得到的结果。这个搜索界面还包括一个“快速地名搜索 (Quick Placename Search)”, 例如, 用户输入城市名称, 浏览器的相应网格内会出现所查询的地图。

数字图书馆地名辞典客户端 (Gazetteer Client) 的检索界面分为三个主要区域, 检索框位于主页面左侧最明显的位置, 结果列表垂直位于检索框的下方, 主页的右侧

是地图浏览器, 如图2所示。采用地名辞典搜索时, 用户可以在搜索浏览器上使用放大/缩小功能来操作地图, 这非常类似网页客户端的功能。通过搜索名录只能进行一次有效的会话, 这个会话会持续30分钟。其他的搜索选项包括选择位置、地名 (单词、词组)、类型特征、地点的状态 (以前、现在) 和识别码。当执行完了一个搜索后, 系统会给出一个识别码。如果用户返回

地名辞典系统,他可以输入这个识别码,这样用户就可以回到以前所搜索的结果。

3.2.2 浏览功能

服务模式中还有一种地理特征词表 (Feature Type Thesaurus) 的工具,该工具和地理辞典客户端没有整合在一起。这个工使用户能够找到正确的带有特定地理特征的词,比如“农业功能”替代“农场”。

ADL的浏览功能主要体现在描述地名类型的主题词表FTT(如图3)上。浏览功能可通过主题词首字母顺序或树形等级结构实现。在FTT的主页中点击按照主题词首字母从A到Z的顺序排列的超链接,读者可以按需或顺序依次浏览。点击FTT主页上的族词索引,可进入树形等级结构浏览,下设6组具有等级关系的主题词集合,最深有5层,用来提供各种主题资源的选择浏览,用户可以按照从大类到小类逐层深入地查找信息。

3.3 在线教学资源服务

ADL将数字图书馆资源整合到课堂及课程教学中,支撑的新教学模式。ADL的教学资源管理提供概念术语集检索和全记录。虚拟学习环境是ADEPT的组成部分,提供了一个“基于数字图书馆技术的被资源整合学习环境”。在这里,学习资料被组织、存取、表现,在某种程度显示为科学概念的相互关系。虚拟学习环境让教师在不同的学习环境来生产、使用和重用学习材料。图4为加州大学圣巴巴拉分校的本科生学习自然地理。

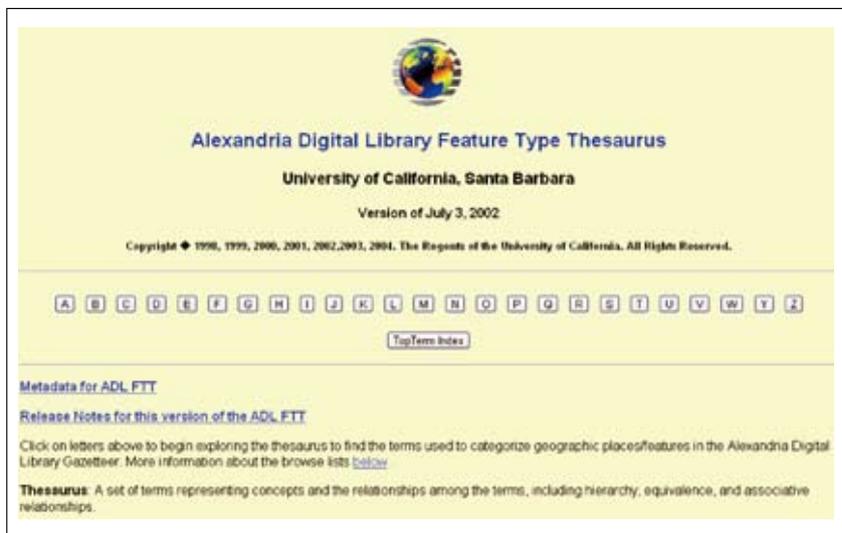


图3 地理特征主题词表界面



图4 加州大学使用“虚拟学习环境”

4 技术特征

4.1 界面设计

亚历山大数字图书馆的主页设计十分简洁、清晰大方(图5)。主页上方是标题“Alexandria Digital Library Project”,标题文字配以世界地图的图片作为背景。右上方是ADL的联系方式、反馈选项、站内搜索入口和主办方UCSB的网址链接。下方是网站的主题界面,基本占用了整个篇幅的版面,右侧是ADL的简介,左侧是ADL的四个重要标题入口,分别包括ADL运营界面、ADL研究界面、ADL检索界面和

ADL地理特征主题词表。主页上未设计搜索框,也未涉及项目各个部分相关介绍的菜单。这种简洁的设计风格给用户一目了然的感觉,令用户迫不及待想进入各个主题入口去了解ADL的详细内容。

4.2 ADL的知识组织系统

4.2.1 地名辞典 (gazetteers) 的设计和运行

地名辞典一直是ADL中的重要组成部分。地名辞典是含有名称和类型划分的有关地点的地理参考字典,辞典中将地理位置词间的相互



图5 亚历山大数字图书馆主页 (<http://www.alexandria.ucsb.edu/>)

关系通过地理表达方式以及明确的关系类型来表示,地名辞典可以扩展到包括事件(如飓风)和有命名的时间阶段,在第二种情况下地理参考数据变成时间跨度。地名辞典还支持对文本文献的地理切分,在这种情况下,一篇文献中关于某地区的内容被表达成经纬度,以此将这些文本文件转变成适合于数字图书馆的物件,从这个意义上来说,地名辞典已成为ADL结构中的一个组成部分,如图6。

4.2.2 叙词表 (thesauri) 的编制

叙词表是表达概念及其等同、等级、相关关系的词汇集。地理辞典的条目就是根据《ADL地貌特征叙词表》来组织的。概念空间模型扩展了叙词表模式,对知识领域中概念集的表达更加全面和完善。地名辞典、叙词表以及概念空间都是用某种标识 (labels) 表达概念的,表达概念的词语都通过有关特征以及与系统中的其他词语的关系来定义,或者通过指向外部资源来定义(图7)。

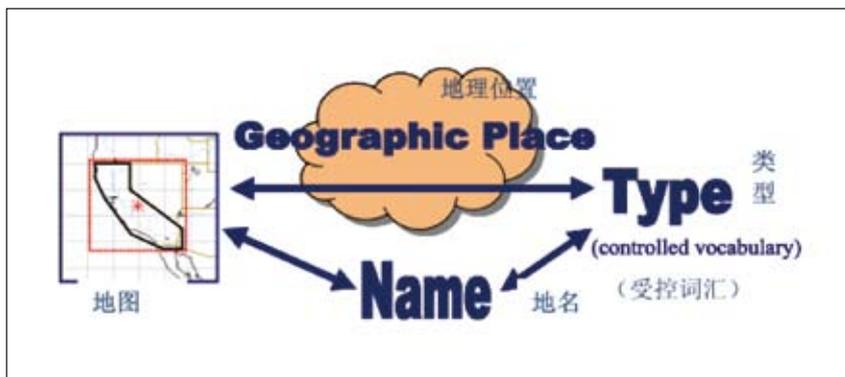


图6 地名辞典的组成部分和相互关系

4.3 ADL的服务协议

为了提供方便查找各种不同结构的地名辞典的通用程序,ADL制定了地理辞典服务协议(ADL Gazetteer Service Protocol),以此支持按照地理名称的基本原则性款项(名称、足迹、类型、关系)进行查询,并且将查到的资料按照标准格式送到客户端,所有这些所要求的就是要安装一个地理辞典服务器,该服务器要能接受特定的XML提问式并能发送按特定标准结构做的报告。

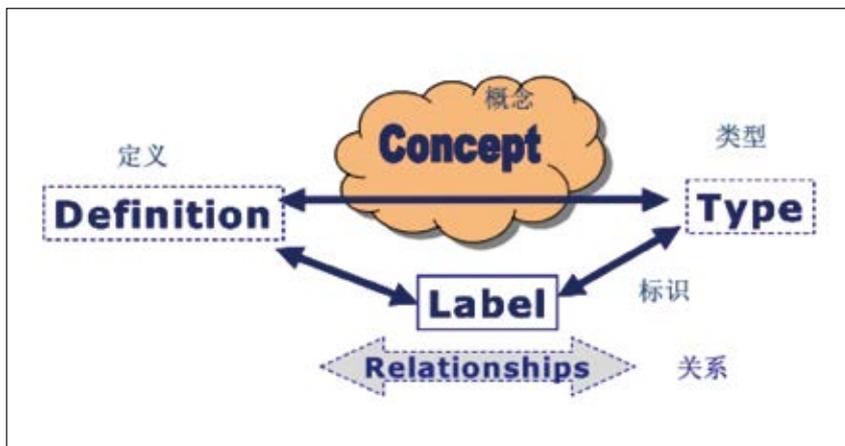


图7 知识组织系统的组成部分以及相互关系

为了支持叙词表和规范文档在ADL中的应用,本项目又制定了ADL叙词表服务协议(ADL Thesaurus Service Protocol),作为自动联网查询各种不同词表的通用协议。就像地理辞典服务协议一样,所有这些所要求的就是要建一个叙词表服务器,该服务器能识别XML提问式并向用户端发出按特定标准结构制作的报告。

4.4 技术支持

ADL使用XML语言编写,同时采用超文本链接语言HTML,以便于容易兼容ADL软件。图书馆通过HTML客户端访问方式提供资源和地名辞典、特殊的信息管理工具,例如专门对应地理学特征语义网之类的(for classing type of geographic features)地质地貌特征辞典,以及可下载的软件代码。地球模型基于语义网结构和本体,后者可以进行属性约束,让资源处于高度结构化。符合逻辑推理的知识库中,本体结构支持分布式典藏。

基于支撑ADEPT的虚拟学习环境,开发了一个“亚历山大数字图书馆的中间件”,中间件的开发由Java和Python语言来完成。建立虚

拟学习环境有三个重要的工具,分别是编目工具、概念系统和讲义编制,其目的是要便捷地为数字地理资源编目、创建概念模型、构造结构化的教学讲义。

5 评价与建议

ADL是第一个收集地理学相关地图、图片等资源的数字化图书馆项目,也是美国数字图书馆工程(DLI)第一期的6个项目之一。ADL凭借其独特的资源优势受到地理学研究者的推崇与青睐。其提供高质量的地理参考资源、促进地理参考资料的数字化、保存和共享,对研究地理学的研究者和学习者都起到了很大的帮助,已成为专业类数字图书馆研究、教学和学习的典范。

ADL的优势显而易见:收藏价值高,资源覆盖面广,分布式,权威性高,技术支持及时,拥有专业特色的地名辞典和地理特征主题词表,能进行地理坐标定位表达,而且能够实现数字图书馆的在线教学资源组织服务。另外,ADL与传统图书馆服务紧密结合在一起,满足图书馆基础技术设施为支持信息资源而不断发展的需要。

虽然亚历山大数字图书馆有很多优势,但在发展建设中也有些应注意的问题:一、用户查询时,为使用正确的主题词来检索,需要预先熟悉词表内容,给用户造成了额外的负担;二、标引员在对地名进行标引和归类时,从FTT中选择恰当的主题词是一项十分繁重的任务;三、FTT词表本身也需要不断维护和更新,在应用中随时更新和增补新的主题词;四、高级检索界面比较复杂,给新用户制定检索策略增加了难度;五、ADEPT的教学服务部分还未对公众开放;六、页面上有一些链接打不开,比如站内搜索,而且不能访问ADL研究部分的ADL地名辞典(ADL Gazetteer)网页。需要ADL进一步完善,否则必将影响其数字图书馆的信息更新和优化服务。

总而言之,ADL是世界上第一个地理参考资源的分布式数字图书馆。它致力于成为世界地理参考资源的中心,推动地理学的专业发展及学生的教育进步。作为一个专业数字图书馆,ADL为公众带来了丰富、便利、有用的资源,成为地理学专业用户必不可少的一个信息和服务中心。

参考文献

- [1] 乔欢. 美国数字图书馆项目评价及评价项目研究[J]. 图书馆杂志, 2005, 24(2): 59-63.
- [2] GOODCHILD M F. The Alexandria Digital Library Project. D-Lib Magazine [J]. 2004, 10(5).
- [3] The Alexandria Digital Library [EB/OL]. [2013-08-10]. <http://www.alexandria.ucsb.edu/>.
- [4] BURYCHKA H. Digital Library Review: Alexandria Digital Library [J/OL]. Digital Library Review Paper, 2006-07-31 [2013-08-10]. <http://www.alexandria.ucsb.edu/>.
- [5] Operational ADL [EB/OL]. [2013-08-10]. <http://www.alexandria.ucsb.edu/adl/>.
- [6] Research ADL [EB/OL]. [2013-08-10]. <http://www.alexandria.ucsb.edu/research/index.htm>.
- [7] ADL webclient [EB/OL]. [2013-08-10]. <http://webclient.alexandria.ucsb.edu/>.
- [8] Globetrotter: ADL Library Webclient [EB/OL]. [2013-08-10]. <http://clients.alexandria.ucsb.edu/globetrotter/>.
- [9] Classic: ADL Library Webclient [EB/OL]. [2013-08-10]. <http://clients.alexandria.ucsb.edu/webclient/index.jsp>.
- [10] Feature Type Thesaurus [EB/OL]. [2013-08-10]. <http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/index.htm>.
- [11] ANCONA D, SMITH T R. Visual explorations for the Alexandria Digital Earth Prototype [C]// The Second international workshop on visual interfaces to digital libraries, ACM+IEEE Joint Conference on Digital Libraries, 2002-07-18, Portland, Oregon, USA.
- [12] HILL L L, DOLIN R, FREW J, et al. Collection metadata solutions for digital library applications [J]. Journal of the American Society for Information Science, 1999, 50(13): 1169-1181.

作者简介

刘燕权, 男, 博士, 毕业于美国麦迪逊大学信息图书馆学院, 现任南康涅狄格州立大学教授。他近年的出版物主要集中在数字图书馆、数据挖掘、多媒体技术、软件管理、国家信息基础结构、信息存储、图书馆统计及管理。E-mail: liuscslu@gmail.com

孙秀良, 男, 博士, 北京化工大学图书馆副研究员。E-mail: sunxl@buct.edu.cn

The Alexandria Digital Library (ADL) – The Center of Geographically Referenced Materials on the Internet

Yan Quan Liu / Southern Connecticut State University, New Haven, CT, USA, 06515
Sun Xiuliang / Beijing University of Chemical Technology Library, Beijing, 100029
Chi Yu / China National Building Materials & Equipment Import & Export Corporation, Beijing, 100048

Abstract: The Alexandria Digital Library (ADL) is one of the biggest and the longest digital library research projects in USA, headquartered at the University of California at Santa Barbara with funding through NSF DLI for ten years. The ADL is a distributed digital library with collections of geo-referenced materials, providing not only services to its collections and gazetteer but also their information management tools, such as the Feature Type Thesaurus for classing types of geographic features. This review examines various aspects of the library, including the project's background, organization of resources, service features, and technologies. Authors' comments are also given.

Keywords: Digital library, Alexandria Digital Library (ADL), Digital Library Initiative (DLI), Geographically referenced materials, Distributed library, NSF

(收稿日期: 2013-10-21)