

# 美国珀尔修斯数字图书馆 ——探索古文明的窗口

□ 刘燕权 / 美国南康涅狄格州立大学 美国康州纽黑文市 06515

高颖 尹涛 / 北京师范大学管理学院 北京 100875

李丽凤 / 中央财经大学信息学院 北京 100081

**摘要:** 美国珀尔修斯数字图书馆 (Perseus Digital Library, PDL) 是一个以收藏古希腊罗马时期古典文献资料为主的人文科学研究型数字图书馆, 旨在完善人类文明记录, 包括语言资源、物质工艺品、历史空间等, 让人文科学尽可能智能地被不同语言和文化环境的人们所接受。文章重点从资源组织、技术特征、界面设计、服务特点、评价和建议等方面对珀尔修斯数字图书馆做了评述。

**关键词:** 希腊-罗马时期, 珀尔修斯, 数字图书馆, 数字图书馆先导计划项目II, 美国国家科学基金会 (NSF), 美国国家科学数字图书馆 (NSDL)

DOI: 10.3772/j.issn.1673-2286.2010.11.009

## 1 概述

美国珀尔修斯数字图书馆 (Perseus Digital Library, 以下简称PDL) 隶属于美国塔夫茨大学 (Tufts University), 是一个以收藏古希腊罗马时期古典文献资料为主的人文科学研究型数字图书馆。网址是<http://www.perseus.tufts.edu/hopper/> (图1所示为PDL主页)。PDL是一个非盈利性的机构, 建馆的最高理想是收录各类人文思想资料及文献, 包括文字文献、艺术品和反映历史时空的各种文献, 并将所有资源开放, 无论是任何语言或文化背景的人, 都能直接在线免费获取其所有文献知识<sup>[1]</sup>。1987年建馆伊始, PDL将资料收集的重点放在古希腊文明古典文献资料上, 随着古希腊馆藏日渐丰富, PDL也不断地收集其他相关的人文科学资料, 如古罗马文明、文艺复兴时期、近代早期英格兰文学、美国内战时期、伦敦的历史和地理等。经过二十多年的不断努力, 现在PDL已成为一个以古希腊文明资料为主的服务型人文科学数字图书馆, 并朝着它的最高理想迈进。

Perseus项目始建于上世纪八十年代的哈佛大学, 是由塔夫茨大学古典文学系的Gregory Crane教授 (至今仍是PDL的主要负责人) 发起并计划实行的。Crane教

授当时还是哈佛大学古典文学系研究生, 在学习和研究过程中, 古典文献资料的查找、梳理和联系占据了大量的精力, 因此, Crane教授决定创建一个方便古希腊文献研究学习的数字图书馆。



图1 PDL主页

在得到了一些资助的情况下, 1985年至1988年, Perseus进入了长达三年的筹备期, 于1987年正式开始了古希腊馆藏的建设, 并得到了Annenberg/CPB基金会的资金支持。Annenberg基金是私人家族基金, 由Walter Annenberg建立, 主要资助对象是非盈利性机构, 基金致力于从通信媒介方面提升公众福祉。

之后, 1992年, 耶鲁大学发行了CD-ROM光盘格

式的Perseus 1.0版本，它采用了苹果公司的超媒体应用软件Hypercard，这时的Perseus严格来说只能算作一个数字图书馆的雏形。1993年Perseus移至塔夫茨大学并于1995年在网上发布运行；1996年，再次发行CD-ROM光盘格式的Perseus 2.0版本；自此，PDL逐渐稳定地发展起来。

在PDL的发展初期，Perseus的目的是尽可能多地数字化古希腊语文献及其英译文本，并呈献给大众。在资料的收集过程中，PDL的研究人员发现，仅仅收录古希腊语文献及其英译文本并不能够满足研究需要，因此1997年，Perseus逐步地增加了拉丁语文本和文艺复兴时期的资料，扩充了其馆藏。

1998年，PDL获得美国数字图书馆先导计划项目（二期）的2,800,000美元的资助，于1999年正式扩展馆藏资源收录范围，从希腊语古典文献及其英译文本外延至其他相关人文领域的文献资源。美国数字图书馆先导计划项目由美国国家科学基金会（NSF）、美国国防部高等研究计划署（DARPA）及美国国家航空与宇航局（NASA）等共同启动，1998年进入二期工程，主要侧重于以教育教学为中心、重在数字化图书馆的内容和对收藏实践的研究<sup>[2]</sup>。至此，作为美国国家科学数字图书馆（NSDL，#9817484）项目之一，Perseus的总体目标转变为：创建一个以古希腊文明资料为主的服务型人文科学数字图书馆。

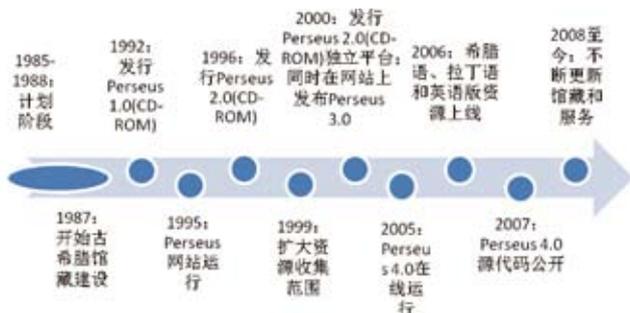


图2 PDL发展时间线<sup>[3]</sup>

从此PDL进入一个快速发展、不断完善馆藏资源及其利用手段的阶段：

2000年发布Perseus 2.0（CD-ROM光盘版）独立平台版本；同期将网站更新为Perseus 3.0版本；

2005年将网站更新为Perseus 4.0版本，期间经历了1年的过渡期，直到旧版本退出舞台；Perseus 4.0也称Perseus Hopper，整个网站基于Java语言架构，较之前基于Perl语言的版本，Hopper的交互性更强，其源

代码也在2007年发布，也更符合其资源开放与公众分享的目标；

2006年，发布希腊语、拉丁语和英文文献资源。所有资源均基于文本编码倡议（TEI，Text Encoding Initiative）标准的XML格式，采用知识共享协议（Creative Commons）的方式共享并署名。

PDL从1985年提交项目申请至今，已经走过20多年的历程，在许多个人和机构通过各种方式的不断支持下，它灵活适应新技术发展，持续更新和扩展馆藏资源，其自适应和发展能力，已使其成为人文研究中，特别是古希腊文明的研究中不可或缺的数字图书馆之一。

## 2 数字资源及其组织

### 2.1 资源范围及种类

经过二十多年的建设，PDL的数字资源非常丰富，除古希腊、古罗马、美国19世纪时期和文艺复兴时期等不同时期、不同语种的古典研究文本文献资料外，还收录有艺术品、考古文献资料、词典等多种类型资料。截至2010年4月，PDL共收录约有1,517万字的文本文献<sup>[4]</sup>，其中大部分是公元前四至五世纪的古典文献原文及译本，这些文本所使用的语言包括英语、希腊语、拉丁语、阿拉伯语、德语、意大利语及古英语等；馆藏还收录有30000多张图像、1000多张地图和一份数量众多的雕刻品目录<sup>[5]</sup>；此外，PDL还提供学科相关博物馆和考古学站点收藏的图像和文选的连接服务。

PDL网站的数字资源分为七个资料库：

（1）古希腊—罗马资料：主要是研究古希腊和罗马时期的一次文献和二次文献，这个资料库是珀尔修斯的重点馆藏，也是PDL从建馆之初就开始建设的数字馆藏，共包括159种关于古希腊及罗马时期的古典文献及相关研究资料，几乎涵盖所有目前已知并遗存的古希腊和古罗马经典，其中很多书都有希腊语、拉丁语原本以及英语译本，如荷马史诗、柏拉图对话以及几大剧作家作品的古代注疏集等。在本库中，还包括一个子库——艺术品和考古文献资料，里面共收录1305种古硬币，1909个花瓶、2003个雕塑、179处遗址、140块宝石和424个建筑物等的详细描述及图片等相关信息，如希腊神殿、希腊古迹等。其中每个目录

入口都有文字描述此目录中物品的概要信息，目录信息和关键词从标准资源中抽取，用于检索；

(2) 阿拉伯语文献资料：包括英文版阿拉伯语字典、古兰经相关的四本著作和阿-英词典，总共约有29万字的阿拉伯语文本和86万字的英语文本；

(3) 日耳曼语文献资料：主要包括研究日耳曼民族的相关资料。古诺尔斯语和古英语均属日耳曼语种，古诺尔斯语大部分尚存文学作品写于12世纪中叶的冰岛，PDL中收录的古诺尔斯语文本主要是萨迦(saga)，萨迦是一种史诗、传奇，有冰岛人的传奇故事和神话文学，也有与宗教相关的典籍等。这类文献总计约英语文本45万字、古诺耳斯语48万字和古英语22万字；

(4) 美国19世纪时期的历史文献资料：包括美国内战等19世纪时期的相关文献，共计近6000万字；

(5) 文艺复兴时期资料：包括早期现代英国文学的一次文献和二次文献，有莎士比亚全集及研究文献、英国文艺复兴时期历史等，总计近700万字英语文本；

(6) 理查蒙德时报资料：在这类中收录19世纪中叶<sup>①</sup>(1860.11.1-1865.12.30)理查蒙德时报的文本资料，总计近2000万字；

(7) 杜克纸莎草本文献资料库(the Duke Data Bank of Documentary Papyri, 简称DDBDP)：包括从分散在纸莎草、陶片和木片等上发现的所有希腊和拉丁文本的数字化资源，主要是有关埃及托勒密王朝和罗马时期的资源。

此外，PDL还以CD-ROM的方式出版发行其资料库，但都有期限限制，内容不能保持与网站同步。

## 2.2 资源组织

### 2.2.1 一般组织形式

PDL馆藏资源根据资源主题分成的七个资料库，分列成行，简单浏览。资源页面分三级：馆藏资源主页、资源库页面和资源内容页面。馆藏资源主页列出七个文献资料库及其简要描述；资源库页面中由上至下依次列出库中收录的资源条目，条目用著者项和题名项标示；资源内容页面即资源项目的主要内容。

PDL不仅数字资源丰富，而且向用户提供从多种角度揭示文献的方法。如上文中所介绍的途径查找或浏览文献、二次文献、图片等时，PDL同时提供了这些文献集的时间分布、地区分布、语言分布等。这是对文献元数据的一种深度挖掘，实现容易，可对研究者查找相关资料提供便利条件。

### 2.2.2 特殊组织形式

如果说在馆藏资源主页和资源库页面这两级页面中并不能看出PDL的馆藏有什么特色，用户查找资料只是简单地逐级进入，那么在第三级资源内容页面中PDL的馆藏特色则突出展现。用户可通过点开资源项目进入资源内容页面，页面上方有书本和章节的条形位置示意图，用蓝色标示出当前页面所属项目以及章节；同时在章节内容的左边栏以树状列表形式显示资源所属资料库级、资源大纲，用户可根据需要点击相应章节进行浏览。

除此之外，PDL还采用知识管理的方式来组织馆藏文本资源，如在希腊语著作中，通过页面右方的工具链接，方便地对对应英译本中的章节段落；还有，在文中选中任意希腊语或拉丁语单词可以进行检索，查看相关词典、注释、参照条目、单词分析以及统计所选单词在著作中出现的次数和位置等。这些知识管理方式的加入让PDL的馆藏资源更好地被揭示，为研究工作提供更多便利，也成为PDL区别于其他数字图书馆的特点之一。

## 2.3 资源免费获取

PDL支持开源内容，是“开放内容联盟”的贡献者之一，其馆藏数字资源对公众开放，所有数据均可在线访问获取，所有文本资料和数据资源都可从Perseus 4.0以XML形式免费下载（但一些内容受到知识产权许可协议的限制）。同时，PDL一直与耶鲁大学出版社合作发行珀尔修斯希腊文献资源（光盘版），用户也可购买光盘使用资源。除了馆藏资源，PDL网站还将其源代码发布至SourceForge对所有人开放，并提供免费下载。

<sup>①</sup> 19世纪时，由于工业革命促使欧美社会在技术上与经济上产生巨大进步与变革，自然科学学科逐渐成形，同时，社会科学也受到深远影响，各种学科、思潮的产生与碰撞，是古典人文逐步转向现代人文的转折时期。因此，19世纪是人文研究中关注的重要时期，这也是PDL收录美国19世纪和理查蒙德时报19世纪中叶资料的原因。

### 3 技术特征

#### 3.1 数据采集、存储和输出

PDL最初发布的CD-ROM光盘版只允许在Mac机上使用,之后,随着CD-ROM光盘版独立平台的发布及互联网的应用,用户服务范围大大扩大。只要拥有一台联网的计算机,即可方便地访问PDL,不仅可以免费获取其资源,还能进行互操作。

早期的PDL在数字化时使用SGML(Standard Generalized Markup Language,标准通用标记语言)来标识文献,而在当时,苹果公司的HyperCard系统不能处理SGML标识,直到几年后由于机读系统发展,PDL的努力开始得到丰厚的回报。随后,在元数据出现之前,PDL项目组就已开始尝试使用专门的元数据架构来标引一些资源项目,最初他们使用面向对象的手段来管理多媒体数据,以一个花瓶为例,尽管它有多个缩略图,但是主要款目信息只有一个,其中包含以下信息:收藏、概述、器皿、形状、区域、日期、年代、描述、关键词等。

现在,PDL遵从OAI协议标准(Open Archives Initiative Protocol for Metadata Harvesting),采用统一的都柏林核心元数据架构。元数据结构为:内容(Context),种类(Type),概述(Summary),日期(Date),尺寸(Dimensions),地区(Region),时期(Period)。Anne Mahoney指出,数字图书馆大多使用SGML或XML标识,符合文本编码倡议(TEI)的文档定义类型(DTD)<sup>[5]</sup>。使用这些标识的数字图书馆都遵照Stoa公约,标识需要标识的特征,使用相应的属性值同时在文档头文件中写入相关元数据。

2006年起,PDL进入“第四代”文本资源的整合,这些文本资源不仅整合了转录文本和原始页面的图像,还包括其他一些工具接口(如语态和语法分析、人物和地点的索引、特定词语的特定意义查询等)<sup>[6]</sup>。以下是PDL第四代资源在2008-2009年间的一些整合:

(1) 将希腊语和拉丁文原始资源转成可扩展的基于TEI标准的XML格式记录。

(2) 根据FRBR(Functional Requirements for Bibliographic Records)模型规范,采用XML MODS和MADS格式著录数据,来描述不同版本、译文、注释、索引和其他实体数据等。

(3) 在规范文本服务协议(Canonical Text Services,CTS)下,PDL的元数据支持通过书、章节、诗节或其他惯用学术引用为入口。这种元数据方便用户从原文引文处直接生成动态链接,跳转到相应的原文数字图像页面或XML格式记录。

#### 3.2 技术架构

PDL网站自1995年正式在线发布后,经历了两个阶段的发展。网站早期版本用Perl语言编写创建,但随着时间的推移,图书馆馆藏越来越丰富,Perl编码逐渐显露出它的弊端,这使得PDL越来越臃肿,对源码的修改变得不太容易,也不利于源码的开放和共享。在2000年,Perseus团队决定新建一个数字图书馆管理系统,希望能够实现互操作性、结构化和开源的功能。同时在2002年,Perseus将馆藏转移至塔夫茨的FEDORA存储库中,而FEDORA对其中的数据系统也有一定的要求,因此,Perseus开始了新系统的开发与设计。现在PDL的数字图书馆管理系统,即Perseus 4.0,也称Perseus Hopper,是基于Java技术、使用开源软件J2EE编写的系统,网站更精简、运行更顺畅,同时也更具有互操作性,与时代的发展要求更切合。另外,PDL是按Firefox的页面格式开发的,支持各种操作系统和浏览器平台。

#### 3.3 界面设计

网站的系统界面设计对于用户的学习起到重要作用。它是与用户进行交互的平台,直接影响用户利用图书馆资源的主动性、有效性和用户对图书馆的评价。

PDL网站界面简洁大方,黑白灰的主色调切合了古老而悠久的馆藏资源,功能划分十分明显。主页上包括了主要的网站功能,如全文检索(位于页面右上角)、清晰的导航栏、热门资源、特色资源等。部分导航栏下还增设二级导航栏,进行服务的二级细分。其中,选择“Collections/Texts”中的七种资源的某一类后,右上角快速检索工具的检索范围自动缩减到该类文献集中,便于人们准确获取所需文献。在页面的右边也有关于该类文献的各种语言的单词统计量,以便于了解该类资源的收藏情况。主页左半面上方显示更新以及招聘的信息,帮助用户了解图书馆的工程建设



具有争议性，用户能通过投票选出他们所认为正确的格式项。

(2) 用户个性化设置。PDL提供用户偏好设置工具，可根据自身的喜好，设置检索结果显示方式、默认显示语言、定制文本的链接格式等。这些个性化设置以Cookie的形式保存在电脑中。

(3) 交互式文本。PDL中的数字资源文本都用工具自动标引，与馆藏中所有引用文本双向参照，用超文本链接方式提供交互式访问，方便读者联系研究相关参照引文。

### 4.2.3 帮助与说明服务

PDL网站通过“帮助信息中心”页面给用户提供帮助与说明服务，包括用户帮助、网站信息与联系方式、馆藏信息介绍三个部分。用户帮助主要是提供给用户利用Perseus的基本方法和途径，帮助用户快速熟悉PDL的馆藏资源及其工具，包括用户快速导航、版权申明、常见问题（FAQ）、文本使用帮助、词汇工具使用帮助、Perseus历史档案等项目；另外，用户可在网站信息与联系方式查找网站相关及联系信息，亦可通过馆藏信息介绍得到对PDL主要馆藏的概览，这些帮助对于新用户无疑是一座与PDL建立良好联系的桥梁。

## 5 评价和建议

PDL是数字图书馆事业发展的先驱，算得上是老牌的数字图书馆之一。它在发展过程中不断创新、努力应用新技术、把握时机、紧跟时代的步伐，一直将数字资源的建设和揭示作为它的工作重点，这些都是我们应该学习和借鉴的地方：

(1) 维护的及时性。在PDL的网站公告中，我们可以看到PDL团队更新和维护频率非常快，仅2010年2月就更新了3次信息，从图2中的大事记也可看出PDL一直不断地紧跟社会文化的发展。

(2) 丰富的古希腊—罗马资源文献。作为PDL的特色馆藏，为研究古希腊、古罗马历史创造了良好的条件。现在，PDL的文献资源仍不断增长，并致力于收录所有古典人文学科资源，为读者开放更多更好的广泛资料。

(3) 新技术的应用。PDL是超文本技术发展的先

驱，随着时间的推移和新技术的出现，PDL果断地将网站替换为基于Java的操作平台，采用了主流字段标准，如XML和TEI，同时在资源揭示中也应用了GIS和数据挖掘，这些技术上的更新表明PDL为了能够向读者提供更广泛资料的目标不断挑战自我、努力进取。

(4) 资源免费共享。从1995年网站在互联网上发布，PDL所有资源都免费开放，任何人都可访问其网站无偿获取，很好地体现了PDL的共享性原则。

(5) 数字资源的揭示。PDL在数字资源的揭示上下了很大功夫，包括各种分布、GIS的应用、对著作中单词的分析等，表现出PDL对文献元数据内容的挖掘要比我国普遍深入得多，其中有些揭示方法技术上实现并不困难，也不需要增加许多工作量，但取得的效果却是有目共睹，值得我们学习借鉴。

尽管PDL有很多优势，但仍有一些不足之处值得我们思考：

(1) 界面和导航。PDL的界面，严格来说，不能让我们很快地把其资源、功能弄清。其中PDL主页上大部分是声明相关内容，没有一个清晰的资源和工具集合的划分，未能很好地实现主页作为门户给用户进行导航的功能，显得似乎PDL创建者更加倾向于把PDL作为一个项目来展示。

(2) 查询工具。字典查询工具是最常用的工具，应该放在更显著的位置，而不是必须通过点击页面右上角检索框下的“All Search Options”才能进入。在实际使用中，用户可能更希望能够便捷地访问所有检索工具，因此作者认为，PDL可以考虑在导航栏上添加“Tools”一项，提供更明显的入口。

(3) 页面的构建。由于PDL采用最简单的列举式显示资源，这种方式在文献资源数量较多时容易导致页面的载入速度过慢，显示不全甚至打不开。在浏览Art & Archeology artifact时，浏览器也会出现缩略图无法显示的问题。作者认为，在资源的显示上，PDL可以以一种更好的方式来呈现。

(4) 特殊用户群的资源获取服务。PDL的优势在于它的易获取性，任何人可从互联网的终端方便地获取资源，这也是PDL的目标之一。然而，由于PDL的多语言性和工具的复杂性，对于某些特殊用户群体的需求可能会不能很好地满足，如某些有障碍的群体<sup>[6]</sup>，这一点也是比较难以解决的。

(5) 光盘（CD-ROM）的滞后性。与另一个希腊文献数字图书馆——Thesaurus Linguae Graecae（以下简

称TLG)相比, TLG由于专注于希腊语相关文献, 在希腊语方面的研究资料相对丰富一些; 并且由于PDL为非盈利性机构、所有资源均可在线免费获取, 因此光盘版的Perseus相对TLG和PDL网站版资源来说并不好

用, 光盘内容不能与最新资源保持同步。

相信珀尔修斯数字读书馆在不断扩大视野, 完善自我中, 会成为一个非常有价值的研究人文科学的窗口。

#### 参考文献

- [1] Perseus Digital Library - Research[EB/OL]. [2010-04-15]. <http://www.perseus.tufts.edu/hopper/research>.
- [2] DIGITAL LIBRARIES INITIATIVE - PHASE II[EB/OL]. [2010-04-17]. <http://www.nsf.gov/pubs/1998/nsf9863/nsf9863.htm>.
- [3] Perseus version history[EB/OL]. [2010-04-17]. <http://www.perseus.tufts.edu/hopper/help/versions.jsp>.
- [4] Perseus Digital Library - Collections&text[EB/OL]. [2010-04-17]. <http://www.perseus.tufts.edu/hopper/collections>.
- [5] RYDBERG-COX J A, CHAVEZ R F, SMITH D A, et al. Knowledge Management in the Perseus Digital Library[EB/OL]. [2010-04-17]. <http://www.ariadne.ac.uk/issue25/rydberg-cox/>.
- [6] MARCHIONINI G. Evaluating Digital Libraries: A Longitudinal and Multifaceted View[J]. *Library Trends*, 2000, 49(2):307.
- [7] PREECE E, ZEPEDA C. The Perseus Digital Library: A Case Study[EB/OL]. [2010-04-17]. [http://repositories.lib.utexas.edu/bitstream/handle/2152/6836/perseus\\_case\\_study.pdf?sequence=5](http://repositories.lib.utexas.edu/bitstream/handle/2152/6836/perseus_case_study.pdf?sequence=5).

#### 作者简介

刘燕权, 毕业于美国麦迪逊大学信息学院, 现在南康涅狄格州立大学任教。曾在各类书刊杂志上发表80余篇专业论文。他近年的出版物主要集中在数字挖掘、数字图书馆、信息存储、数字化及多媒体技术、国家信息基础结构、图书馆统计及管理。通讯地址: Southern Connection State University, USA 06515。E-mail: liuscsu@gmail.com

#### Perseus Digital Library: A Path to Explore the Classics of Greco-Roman & Western Cultures

Yan Quan Liu / Southern Connecticut State University, New Haven, CT, USA, 06515

Gao Ying, Yin Tao / Department of Management, Beijing Normal University, Beijing, 100875

Li Lifeng / School of Information, Center University of Finance and Economics, Beijing, 100081

Abstract: As a digital library for the humanities in the US DLI Phase 2, the Perseus Digital Library (PDL) hosted in Tufts University is a scalable, broad-based, interdisciplinary digital library for the humanities. PDL offers classical texts, images, maps, video objects, as well as tools for searching and manipulating the resources, aiming at engineering interactions through time, space and language. This paper provides an overview of PDL, including its organization of resources, technologies employed, and services provided. Comments and suggestions for the future improvements are also given.

Keywords: Digital Library, NSF, Humanity, Ancient Greek and Roman materials, Digital Library Initiative II, National Science Digital Library (NSDL)

(收稿日期: 2010-09-24)