

# “DPLA（地普乐）”：美国数字公共图书馆

□ 宋玉武 / 美国国会图书馆亚洲部 华盛顿 20540

摘要：美国数字公共图书馆是美国图书馆、档案馆、博物馆及相关机构联合发起的一个开放性、分布式数字公共图书馆建设项目。文章从它的建立缘起入手，概述了该项目。

关键词：美国数字公共图书馆，数字仓储，数字图书馆

DOI: 10.3772/j.issn.1673—2286.2012.10.003

“如果我们可以把人类送上月球，为什么不能把国会图书馆送上光纤网路？”——引自DPLA项目成员

## 1 缘起

自20世纪80年代末90年代初以来，美国基本实现了图书馆的自动化、网络化。1993年，美国开始启动数字图书馆的建设。以后的互联网技术、信息科学、信息存储与检索、超媒体、人机互动、人工智能以及在计算机科学其他领域的发展使数字图书馆成为可能。经过近二十年的努力，成百上千个数字图书馆已在美国建立。美国数家大型机构曾主持了几个大规模的数字图书馆项目：

### (1) American Memory (美国记忆)

American Memory (美国记忆)是由美国国会图书馆于1994年10月13日与多家单位合作建立的网络资源库。它是一个美国历史资源

数字图书馆。资源库收集的数百万件藏品基本都是历史类材料，如地图和图片。数据主要来自国会图书馆下属的两个部门，地理/地图部，另一个是图片部。尽管美国记忆有多家单位参与，但其他单位所占的比例很小。美国记忆网络资源库收集的内容，与国会图书馆整个馆藏相比，还只是很小的一部分。

### (2) HathiTrust数字资源库 (<http://www.hathitrust.org>)

HathiTrust数字资源库是一个大规模的数字内容资源库(digital repository)。HathiTrust于2008年10月由50多个美国研究图书馆(哈佛、耶鲁、普林斯顿、斯坦福、哥伦比亚、印第安纳、密歇根等大学)联合建立。资源库由印第安纳大学和密歇根大学管理。HathiTrust截至2012年1月，总共拥有超过1000万种图书，其中公共领域图书超过270万种。

### (3) Internet Archives (互联网档案馆)

Internet Archives (互联网档案馆)是一个实施书籍数字化的非营利性机构。每天扫描1000多册书，并链接谷歌图书和其他来源的数字版书籍。截至2011年5月，其拥有的公共领域的图书约280万，超过了谷歌图书的公共领域的书籍。

### (4) Google Books (Google数字图书馆)

Google公司于2004年12月14日宣布，Google将与五所图书馆——美国纽约公共图书馆、哈佛大学图书馆、斯坦福大学图书馆、密歇根大学图书馆以及牛津大学图书馆合作，将这些图书馆的图书扫描制成电子版，供读者免费检索、阅读。Google将根据图书的版权情况提供不同层次的服务。用户可以阅读公共领域的图书的全文。对于受版权保护的图书，Google将与出版商或作者建立合作关系，根据合作协商的程度，在网上提供图书的一部分或全文。

尽管上述数字图书馆极大地

改进了人们对信息的获取,然而它们基本上都各自为政,缺少联结,好像导弹发射井一样,此外还有重复。如何解决“导弹发射井”和重复问题成了人们的新的挑战。2010年底,美国图书馆、信息界的人士提出了新一代数字公共图书馆(DPLA)的构想。今天DPLA数字公共图书馆尚未诞生,不过基本理念已经开始成型,基本技术模型也在建立中。

DPLA与其他大型数字图书馆不同在于:其他大型数字图书馆主要侧重图书,而DPLA除了书籍,将囊括其他类型的信息资源——报纸、期刊、手稿、图片、音像、视频、档案文献、博物馆的藏品,以及众多格式的其他资源。在DPLA框架下,美国国会图书馆、美国国家档案馆、史密森博物馆学会三家最近联手启动了Modeling a Digital Collaboration for America's National Collections(美国国家馆藏数字化协作)项目。这个项目充分反映了DPLA力图整合资源,使百川归海,达成一站式信息服务(One-Stop Shopping)的理念。

## 2 简介

2010年12月份,美国哈佛大学博克曼互联网和社会中心(Berkman Center for Internet & Society)首先提出“DPLA(地普乐)”项目计划,即美国数字公共图书馆(Digital Public Library of America, DPLA)<sup>[1]</sup>,该项目在美国Alfred P. Sloan基金会的资助下成立。2011年,英国所属的Arcadia基金会加盟DPLA。这两个基金会认捐500万美元,计划在2013年4月以前将资金筹措完毕,使数字公共图

书馆开始启动。

美国数字公共图书馆项目计划基于美国总统汤姆斯·杰佛逊的理念:知识是公众的财产。该项目旨在探索如何建设一个开放的、分布式的在线资源网络,囊括美国的图书馆、大学、档案馆以及博物馆的文化遗产,并让世界上的每个人免费地获取这些丰富馆藏。

2011年春,DPLA正式成立并设立了工作指导委员会。成员包括美国国会图书馆的副馆长Deanna Marcum、哈佛大学图书馆馆长Robert Darnton、Internet Archives(互联网档案馆)的创始人Brewster Kahle、旧金山图书馆馆长Luis Herrera、Alfred P. Sloan基金会副总裁Doron Weber、Public.Resource.Org主席Carl Malamud等人。

DPLA召集了图书馆、大学、博物馆、档案馆、文化机构、州和地方政府、出版商、作者和私营行业以及基金会等相关机构,讨论建设美国数字公共图书馆的可能性,及改善公众获得网上资源的战略。同时DPLA讨论了数字公共图书馆的模式、内容和范围、财政/商业模型、管理、法律、技术、时间表等各方面的问題。

经各方通力合作,DPLA的概念、特点基本成形<sup>[2]</sup>:

(1) DPLA有如一个资讯社区中心,让使用者可以上网取得各地、多种格式的馆藏资料。

(2) DPLA将是一个分布、开放式的系统,并在一个全球性的数据环境下运行。DPLA允许无条件自由访问或复制。

(3) DPLA将汇总美国图书馆现有的数据。DPLA不是一个单一的数据库,它将包括大型图书

馆已数字化的公共领域的书籍,比如HathiTrust(美国高校图书馆建立的一个共享式数字仓储项目)、Internet Archives(互联网档案馆)等数字馆藏。这些公共领域的数字馆藏独立于Google等商业公司。

(4) DPLA含有丰富的政府资源。除美国国会图书馆的American Memory和Chronicling America(50个州已经数字化的主要报纸)等外,亦包括各级政府的公开信息。

(5) DPLA与普通公共图书馆的服务有本质的不同。由于著作权法,DPLA将不包括近期出版物。首先在DPLA上线的是公共领域已经数字化的作品。陆续成为无版权的作品将循序数字化并链接至DPLA。

(6) DPLA将服务于一个多元社会。它将保持相对的独立自主性。为了免于政治压力,DPLA将不会设在精英学府或政府机构内。

(7) DPLA在可能的情况下将全部使用现有的免费或开放的元数据。

(8) DPLA将提供相关工具 and 应用程序接口API(Application Programming Interface)。

(9) DPLA将提供一个可自由参与的平台,鼓励各界人士积极参与与创新,开发新工具与新服务,为全人类作贡献。

DPLA自成立以来举行了一系列重要活动:

2011年3月,DPLA设立了六个委员会(Workstreams)专门讨论相关问题并提出解决方案。六个委员会包括:

- 用户与参与(Audience & Participation)
- 内容和范围(Content & Scope)

● 财政/商业模型 (Financial/Business Models)

- 管理 (Governance)
- 法律 (Legal Issues)
- 技术 (Technical Aspects)

2011年5月, DPLA宣布参考Internet Archives (互联网档案馆)、public.resource.org、HathiTrust、美国国会图书馆的American Memory以及Europeana和另外一些国家的国家数字图书馆项目。

2011年5月, DPLA宣布进行“测试冲刺赛 (Beta Sprint)”的活动——即一项可供网路展演的软件设计比赛活动, 向社会各界征求对DPLA的理念、技术平台、技术工具、用户界面等方面的意见和方案。参赛机构可提交原始技术模型, 接受专家评估。

2011年10月, DPLA在华盛顿美国国家档案馆召开会议。

2011年10月, DPLA宣布与Europeana (欧洲数字图书馆) 进行合作<sup>[3]</sup>, 双方决定建立联合数据库。美国与欧洲两地用户可随时随地获得两个系统的丰富馆藏。联合数据库将包括数以百万计的书籍、报纸、期刊、手稿、图片、音像以及众多格式的其他资源。双方拟长期合作, 建立可互操作的数字模型、共享的源代码、资源规范及可开放获取的馆藏资源。DPLA和欧洲数字图书馆希望能促进同全球伙伴一起创建国际网络。

根据计划, DPLA将于2013年4月正式启动。届时DPLA将拥有一个相当规模的数字仓储基础设施。随着更多资金的注入, DPLA将实现与其他国家的大型数字图书馆的可互操作联网。

### 3 技术模型实例

2011年9月, DPLA评选了六个“测试冲刺赛” (Beta Sprint) 技术模型以供研发人员进一步参考。

#### 3.1 Modeling a Digital Collaboration for America's National Collections (美国国家馆藏数字化协作项目)<sup>[4]</sup>

在DPLA框架下, 美国国会图书馆、美国国家档案馆、史密森博物馆学会三家联手启动Modeling a Digital Collaboration for America's National Collections (美国国家馆藏数字化协作) 项目。

作为三足鼎立的国家级大馆, 美国国会图书馆拥有1.44亿件馆藏, 包括超过460种语言的3,300万种编目图书和其他印刷材料, 及超过6,300万件手稿。美国国家档案馆拥有100亿件藏品。史密森博物馆学会馆藏文物、艺术品和标本的总数超过1.37亿件。

美国国家馆藏数字化协作项目旨在建立一个用于示范的搜寻检索

系统, 以检索美国国会图书馆、美国国家档案馆、史密森博物馆学会三处的馆藏。作为项目起点, 三馆利用史密森博物馆学会现有的网络信息系统: Enterprise Digital Asset Network (EDAN) 建立了“多面检索汇总系统 (a faceted search aggregator)”。该汇总系统汇集了各种在线内容, 并在单一界面上显示检索结果。工作人员使用了Enterprise Digital Asset Network (EDAN) 元数据架构 (metadata schema) 以测试数据映射 (data mapping) ——即在两个以上的数据库间建立数据元素的对应关系, 最终达成数据集成汇总。测试内容包括数据迁移 (data migration)、数据抽取 (extract)、数据转换 (transform)、数据加载 (load)、数据清洗 (data cleaning)、数据集成 (data integration) 等。

在项目进行过程中, 美国国会图书馆、美国国家档案馆的工作人员选取了各自馆中十件左右的藏品电子版及记录 (国会图书馆的10条记录来自表演艺术百科全书, 包括音乐作曲家勃拉姆斯和几件19世纪

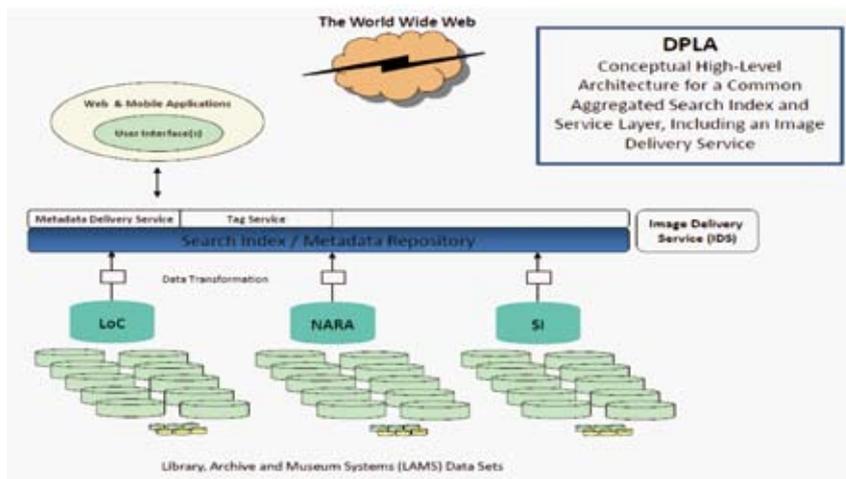


图1 技术高端概念图 (检索索引、元数据库、服务层、图像传递服务) 及图书馆、档案馆、博物馆馆藏数据集

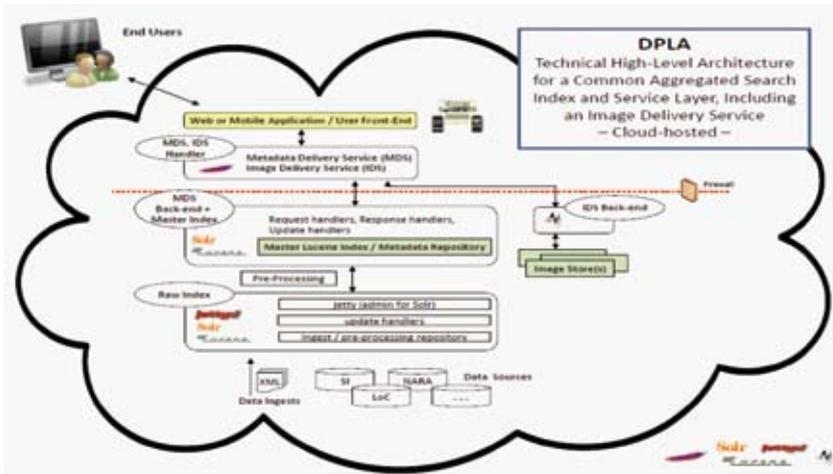


图2 技术高端建构图 (检索索引、服务层、图像传递服务——云储存)

的乐谱手稿。国家档案馆提供了19世纪的照片、专利、图纸和通信的记录), 然后转交给史密森博物馆学会工作人员, 并由他们手工输入史密森博物馆学会拥有700万藏品的数据库Enterprise Digital Asset Network (EDAN)系统中。该系统采用了云技术。它包括综合搜索索引、服务层、图像传输服务等。

通过史密森博物馆学会馆藏检索中心界面 (Smithsonian Collections Search Center) 即可看到三馆如何互动操作, 在同一平台上显示检索结果。

```

Guidelines for the geoLocation tag

<geoLocation>
  <L1 type=[Continent | Ocean]>
  <L2 type=[Country | Nation | Sea | Gulf | Bay | Sound]>
  <L3 type=[State | Province | Department | Country | District | Republic | Sea | Gulf | Bay]>
  <L4 type=[County | Island]>
  <L5 type=[City | Town]>
  <Other type = [anything: examples = Neighborhood, Street, Desert, Park, etc.]>
  <points label=[text] dates="yyyy-yyyy">
  <point>
  <latitude type=[decimal | degrees]>
  <longitude type=[decimal | degrees]>
  </point>
  <point>
  <latitude type=[decimal | degrees]>
  <longitude type=[decimal | degrees]>
  </point>
  <point>
  <latitude type=[decimal | degrees]>
  <longitude type=[decimal | degrees]>
  </point>
  </points>
</geoLocation >
    
```

图3 XML位置标记



图4 美国数字公共图书馆检索平台界面屏幕截图

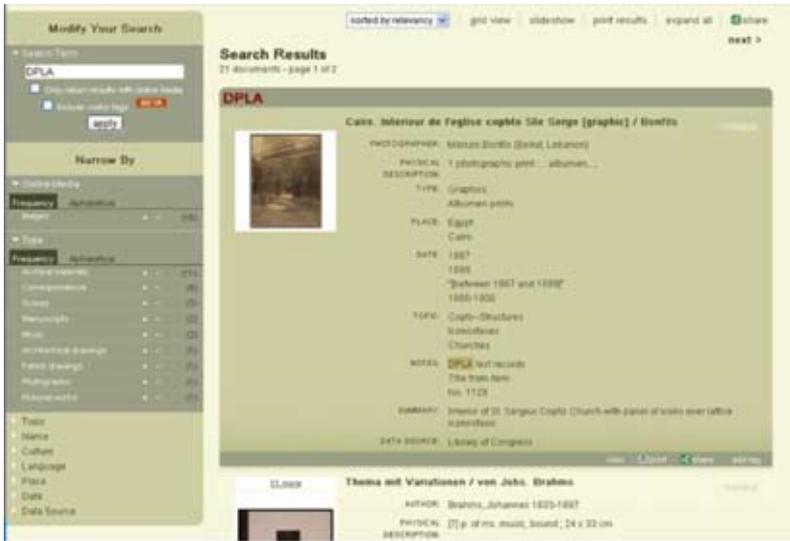


图5 检索结果列表视图

图6 检索结果网格视图

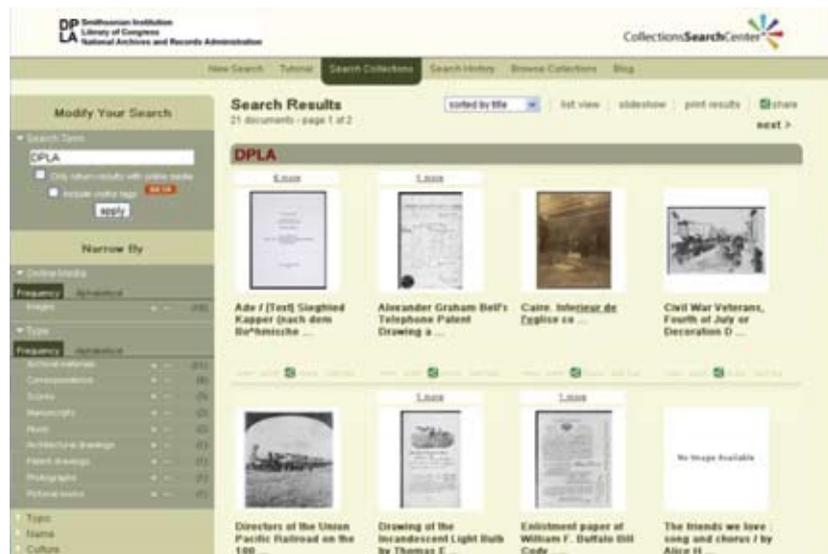


图7 美国国会图书馆藏品检索结果概述





图8 美国国家档案馆藏品检索结果概述



图9 史密森博物馆学会藏品检索结果概述

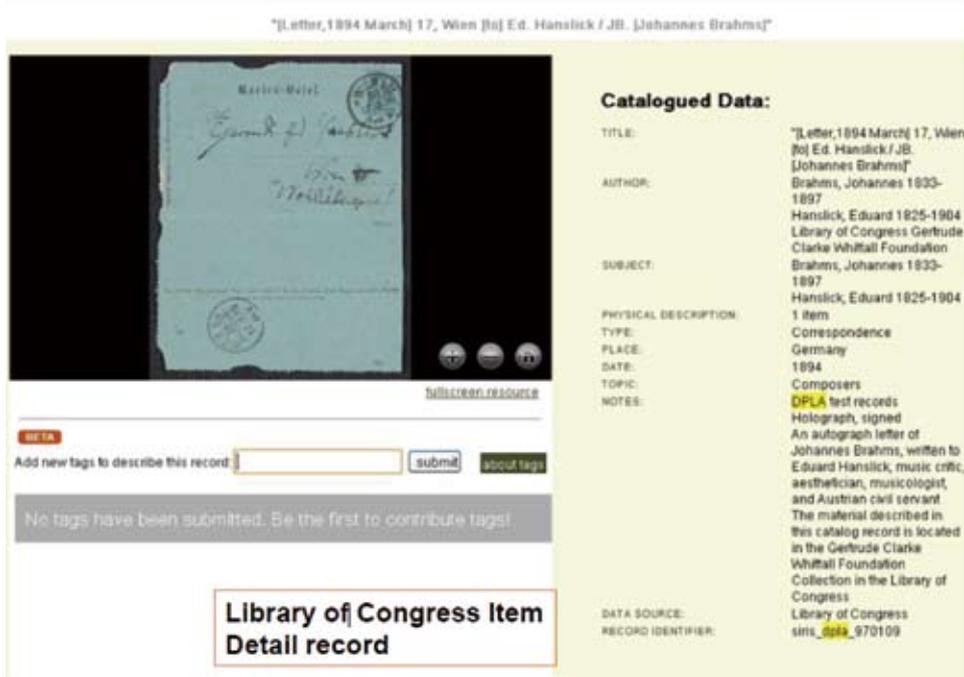


图10 美国国会图书馆藏品检索结果详述

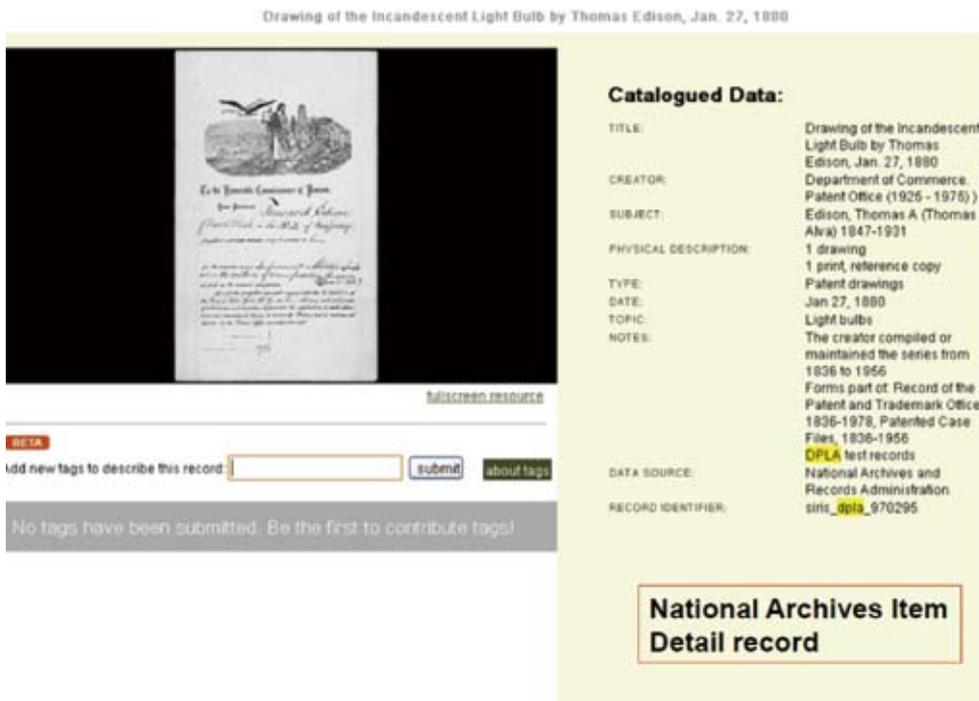


图11 美国国家档案馆藏品检索结果详述

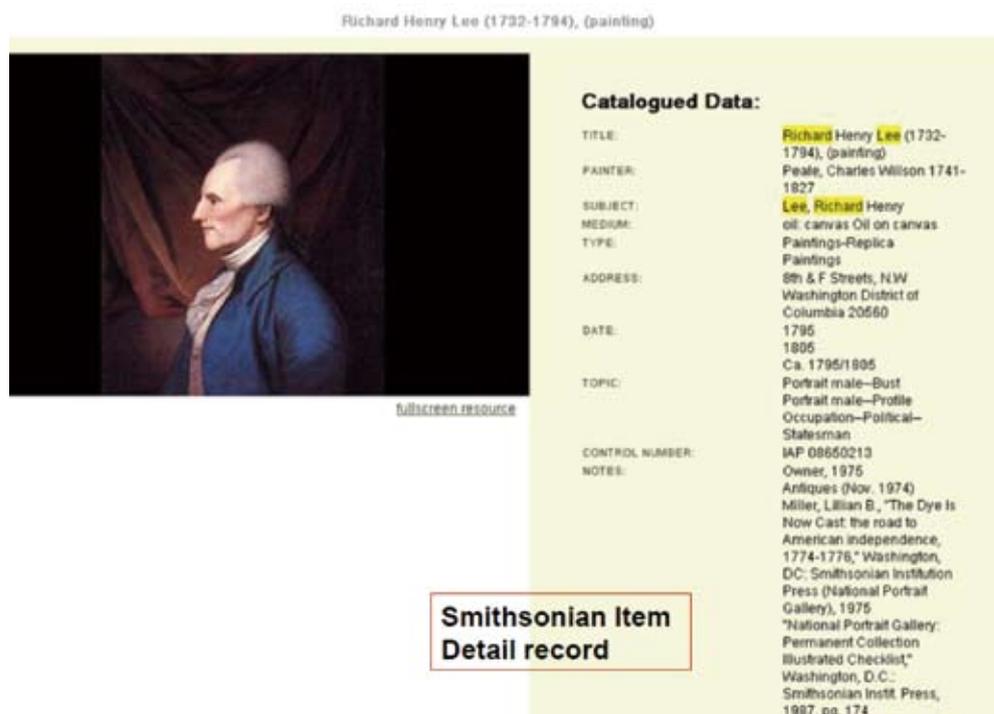


图12 史密森博物馆学会藏品检索结果详述

### 3.2 DLF/DCC: DPLA Beta Sprint<sup>[5]</sup>

DLF/DCC: DPLA Beta Sprint

系统由Digital Library Federation (数字图书馆联盟)与University of Illinois – Urbana Champaign

(伊利诺伊大学香槟分校)合作研发,旨在演示文献资料及元数据的组织与展示的特殊途径。

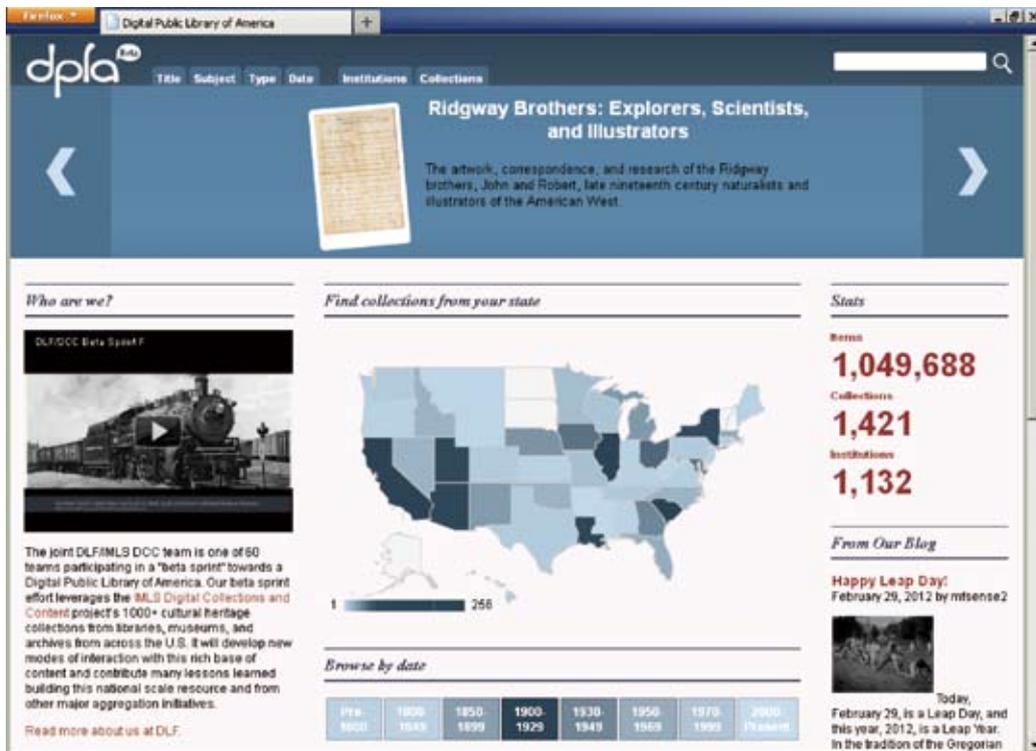


图13 DLF/  
DCC: DPLA  
Beta Sprint

### 3.3 extraMUROS<sup>[6]</sup>

extraMUROS由哈佛大学metaLAB、Harvard Library Lab与Media And Place (MAP) Productions合作研发。extraMUROS使用开放性HTML5平台,力图建立一个多媒体虚拟图书馆。

### 3.4 Government Publications (政府出版物): Enhanced Access and Discovery through Open Linked Data and Crowdsourcing<sup>[7]</sup>

Government Publications (政府出版物)项目由The University of Minnesota (明尼苏达大学)、The Committee on Institutional Cooperation与HathiTrust合作研发,旨在建立一个数字化网络系统,将100万件政府文献数字化。

### 3.5 Metadata Interoperability Services (元数据互操作服务系统, MINT)<sup>[8]</sup>

Metadata Interoperability Services (元数据互操作服务系统, MINT)是希腊雅典国家技术大学(MINT at the National Technical University of Athens)的一个项目。他们利用网络平台以集结各个不同的线上资料及元数据。

### 3.6 ShelfLife/LibraryCloud<sup>[9]</sup>

ShelfLife/LibraryCloud是哈佛大学图书馆创新实验室(Harvard Library Innovation Lab)制作的一个技术模型。它能够根据图书的元数据显现出虚拟的书架。ShelfLife告诉读者如何使用图像书架,搜寻信息。LibraryCloud是支持ShelfLife的后台元数据服务器。

## 4 结语

系统论有一个著名的定理,叫“整体大于部分之和”。其涵义有如古代寓言:一根筷子一折就断,十根筷子捆在一起,就无法折断。这就是说整体功能大于部分功能相加总和,合作效率大于分散效率相加总和。

美国数字公共图书馆的实现将会使异地多种数字馆藏整合到一起,它发挥了“整体大于部分之和”的作用。这样可以使信息资源的开放性利用在时间和空间上得到极大的延伸,为整个人类的科学、文化活动的的发展作出突破性的贡献。

作为新一代数字图书馆, DPLA对中国数字图书馆建设将会提供一些启示。另外在实践方面,特别是在技术平台、模型、技术工具、用户界面领域也可提供可参照的依据。也许有一天,中国



图14 ShelfLife

图书馆、信息界人士能与有关方面合作，在中国国家数字图书馆工程（National Digital Library Project of China）和中国大学数字图书馆国际合作计划（China Academic Digital Associative Library, CADAL）基础上研发出：

- 中国数字公共图书馆 Digital Public Library of China (DPLC)
- 亚洲数字公共图书馆 Digital Public Library of Asia (DPLAsia)
- 世界数字公共图书馆 Digital Public Library of the World (DPLW)

我们期待着“地普乐（DPLA）”变成“普天同乐”。

#### 参考文献

- [1] Digital Public Library of America, DPLA [EB/OL]. [2012-06-10]. <http://dp.la>.
- [2] HAWKINS D. The Digital Public Library of America (DPLA) [EB/OL]. [2012-06-10]. <http://www.theconferencecircuit.com/2011/11/04/the-digital-public-library-of-america-dlpa>.
- [3] EUSCREEN. European and DPLA Announce Collaboration [EB/OL]. [2012-06-10]. <http://blog.euscreen.eu/?p=2181>.
- [4] MARCUM D, FERRIERO D, GWINN N, et al. Digital Collaboration for America's National Collections [EB/OL]. [2012-06-10]. <http://www.sil.si.edu/dpla.cfm>.
- [5] Digital Library Federation, University of Illinois - Urbana Champaign. DLF/DCC: DPLA Beta Sprint [EB/OL]. [2012-06-10]. <http://dp.la/entries/dlfdcc-dpla-beta-sprint/>.
- [6] metaLAB, Harvard Library Lab, Media and Place (MAP) Productions. extraMUIROS [EB/OL]. [2012-06-10]. <http://dp.la/entries/extramuiros/>.
- [7] The University of Minnesota, the Committee on Institutional Cooperation, HathiTrust. Government Publications: Enhanced Access and Discovery through Open Linked Data and Crowdsourcing [EB/OL]. [2012-06-10]. <http://dp.la/entries/government-publications-enhanced-access-and-discovery-through-open-linked-data-and-crowdsourcing>.
- [8] MINT at the National Technical University of Athens. Metadata Interoperability Services [EB/OL]. [2012-06-10]. <http://dp.la/entries/metadata-interoperability-services/>.
- [9] Harvard Library Innovation Lab. ShelfLife/LibraryCloud [EB/OL]. [2012-06-10]. <http://dp.la/entries/shelflife-and-librarycloud/>.

#### 作者简介

宋玉武（1958-），美国国会图书馆亚洲部图书馆员。研究方向：中国研究、数字图书馆。E-mail: [yson@loc.gov](mailto:yson@loc.gov)

#### DPLA: Digital Public Library of America

Song Yuwu / Asian Division, Library of Congress, Washington, DC 20540, U.S.A.

Abstract: Digital Public Library of America (DPLA) is a joint project by American libraries, archives, museums, and other stakeholders to develop a digital public library providing free access to information for all. This article starts with the rationale for the project and discusses its concepts and prototypes. It is hoped that the introduction of DPLA to China would help inspire the Chinese counterparts to develop sophisticated digital libraries in China.

Keywords: Digital Public Library of America, DPLA, Digital repository, Digital library

(收稿日期: 2012-07-11)