“崛起美国”数字图书馆
——记录美国成长的足迹

刘燕权[1]，田硕[2]
（1. 美国南康涅狄格州立大学，纽黑文市 06515，美国；2. 铁道党校，北京 100088）

摘要：“崛起美国”数字图书馆（Making of America Digital Library, MOA）是一个收集美国内战前到重建时期（1850-1877）社会历史一手资料的数字图书馆项目。馆藏资源涵盖教育、心理、社会、宗教、科学技术和美国历史等多个学科领域，是通过数字技术保存的有关美国基础设施发展方面重要原始资料的代表性项目。本文对该图书馆的建设及现状进行了综合性评析，包括项目概述、资源组织、技术特征、服务特点等，并给出评价与建议。

关键词：MOA；数字图书馆；美国社会历史；光学字符识别技术
中文分类号：G250.7
DOI：10.3772/j.issn.1673-2286.2015.06.011

1 概述

“崛起美国”数字图书馆(（Making of America Digital Library，以下简称MOA）是一个收集美国内战前到重建时期（1850-1877）社会历史一手资料的数字图书馆项目，网址是：http://quod.lib.umich.edu/m/moagrp/[1]。MOA由密歇根大学和康奈尔大学共同研究开发，并得到了安得鲁•梅隆基金会（Andrew W. Mellon Foundation）的资助，是通过数字技术保存以便于人们使用有关美国基础设施发展方面重要原始资料的代表性项目。其馆藏资源主要涵盖教育、心理、社会、宗教、科学技术和美国历史等学科领域，主要任务是集合科研机构和国家企业等力量开发研究在更大范围内推广数字资源的筛选、转换、存储、检索和使用的统一标准和通用协议[2]。

2 数字资源及其组织

2.1 资源范围及种类

MOA数据库资源主要包括图书和期刊两种文献类型，并以图片和文本为主要存在形式。自1995年项目起步至今，MOA馆藏中十九世纪版本的图书约10,000本、期刊文章约5,000篇，资源总量约3,818,000 个页面、12,000卷图书。

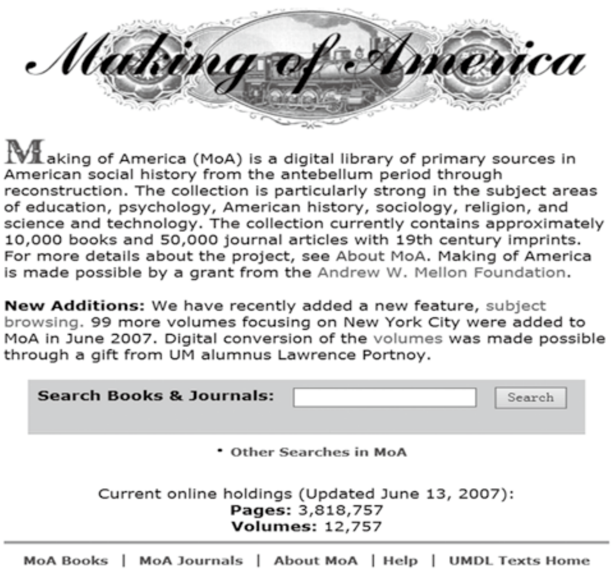

图1 MOA首页

由MOA的主界面（见图1）可知MOA的资源主要分

布在期刊和图书两个数据库中。用户既可以在主界面通过综合检索获取资源，也可以直接进入相应的数据库检索。同时MOA还包含一个资源接入点——密歇根大学数字图书馆文本资源主页（UMDL Texts Home），供用户检索使用[3]。

2.2 UMDL Texts

UMDL Texts（见图2）是密歇根大学可供MOA使用的数字期刊和电子图书资源服务产品，目前包括118个资源集合，共202,490个文本。UMDL Texts 中的数字资源是以资源集合形式组织和排序的，用户可以通过浏览获取资源，或是进入资源集合的二级目录，点击资源条目链接到达详细资源首页进行检索。此外，UMDL Texts还提供有密歇根大学数字图书馆产品服务全部资源的链接（DLPS List of All Collections）。但需要注意的是，UMDLTexts对资料集合的描述与其中包含的资源条目并非完全吻合，并且同一资源条目也有可能出现在多个资料集合中。此外，UMDLTexts的数字资源有些是需授权使用的，如只限于密歇根大学的师生和工作人员登录使用。

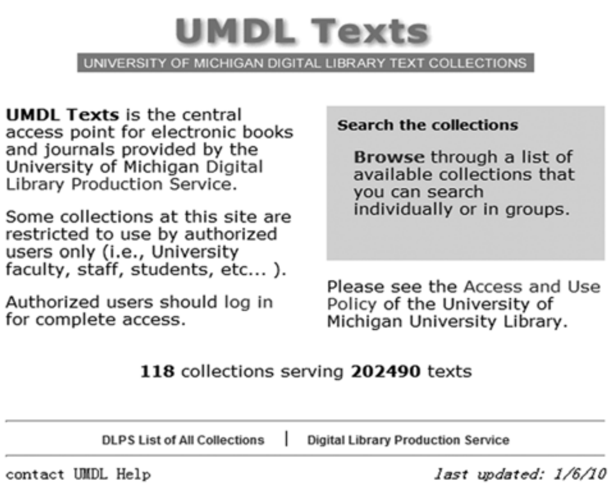

图2 UMDL Texts Home界面

2.3 资源的搜集

（1）第一阶段

1995年，密歇根大学和康奈尔大学合作开发了一个主题相关的数字图书馆，即MOA创建的初始阶段。两所大学合作对美国战前时期到重建时期的社会历史资料进行筛选和扫描等深度加工。其中密歇根大学筛选重点在教育、心理、美国历史、社会、科学技术和宗教等领域的专著，康奈尔大学筛选重点在战前到重建时期的定期刊物，包括有广泛受众的期刊和特定目标受众的期刊（如农业）。MOA的初始阶段，馆藏资源有1,500,000张图片，相当于5,000册原始图书。

（2）第二阶段

1999-2000年间，密歇根大学图书馆历时8个月制定基准和方针，向用户提供可在线访问的高质量电子书。这些策略和方法也成为MOA起步阶段的基础，包括数字资源的验证、描述和记录方法等。约有2,347,000页内容又加入到MOA中，主要来源于密歇根大学图书馆布尔储存设备中的一组丰富的候选资料。此外，他们还转换了1850-1876年间7576卷美国出版公司出版的英文资料，这些资料大部分是人文和社会科学方面的。

（3）第三阶段

在MOA起步阶段后期，又添加了一些 新功能，如“主题浏览”。2001年，7,500卷图书加入到MOA中；2005年11月，MOA进行了一次资源更新，并计划后期要进一步加入有关美国内战和纽约的资源；2007年6月，99卷有关纽约的文献资料被加入到MOA中。这些书卷的电子化转换还得到了密歇根大学校友劳伦斯•波特诺伊的大力支持。

2.4 资源的组织

MOA主要采用元数据技术组织其数量庞大、种类繁多的馆藏资源，即对数字化对象进行描述，使数据的基本属性能被揭示；并选用SGML这一标准化编码语言来描述这些元数据。在图片数据描述上，SGML不仅清晰地描述了资源属性，还可以实现向前翻页、向后翻页、分页符和指向包含原始资源全部信息的图片功能。

此外，为了便于资源检索，MOA采用的本地联机编目系统是NOTIS，即MOA的MARC记录是基于USMARC通讯格式的NOTIS。这些记录有着多样化的来源，包括供应商和OCLC，所以这些标准的使用反映了它们的出处：pre-AACR1，AACR1和AACR2（AACR：英美编目条例）。在MOA的编目过程中，也努力使专著编目条例记录达到现行国家标准，虽然这其中也不可避免有一些缺失的目录标识[4]。

3 技术特征

3.1 界面设计

MOA检索主页以白色为基调，简洁清新。所有的功能布局直观便捷，一目了然。主要功能模块包括：资源库主要内容简介，资源库更新动态，图书和期刊综合检索框和“其它检索”链接，资源库最近的更新时间和总量介绍，相关功能链接（包括“MOA图书”、“MOA期刊”、“关于MOA”、“帮助”和UMDL Texts Home）。

3.2 扫描技术

MOA进行资源数字化加工的首要技术工作就是纸质文献资源的扫描。MOA的扫描技术流程包括将需要扫描的易损图书按照其老化程度、装订损坏程度进行划分，然后按照600dpi的分辨率，采用CCITT组4压缩成TIFF格式的图片。考虑到需要采用多样化格式保存这些资料，MOA还将这些TIFF图片用无酸纸打印出来装订存档。另外，由于诸多单个用户不支持TIFF格式图片，MOA还将扫描的图像转换成GIF格式提供给用户使用，并将TIFF格式图片存储和制成硬拷盘，以备在线版本出现问题时的不时之需。

3.3 转化技术

MOA项目最主要的技术工作是页面图像向文字文本的转化，这个过程保证了文献资源全文检索的实现。MOA使用施乐公司的ScanWorx进行资源转换，采用光学字符识别技术（OCR）以实现页面图像和文字文本的一对一转换，并且这个过程是全自动的，几乎不需要人工干预。同时，由于ScanWorx存在着不能提供最佳水平的自动化和不具备保留单个页面必要信息的功能，MOA又引入Perl5来开发一系列的脚本：①创建基于目录结构和命名规则的脚本来保留单个页面的必要信息；②管理ScanWorx处理这些脚本；③提供出错信息，以便工作人员可以识别需要重新扫描的问题文本，或是进行人工干预。这两项技术的联合使用，保障了MOA资源转换的准确度。

3.4 资源的保存和检索技术

MOA使用“即时”保存图像格式，即MOATIFF格式的图片在转换过程中就直接复制到磁盘保存，并且使用CCITT组4压缩，不会损坏图片的分辨率和图像质量。

MOA的检索技术主要是CGI脚本，它拓展了来源于HTI的模板，管理来自表格的信息，并解决了使用开放文本搜索引擎面临的检索语言的问题。同时，CGI还进行数据分类和显示屏幕上用户观看的检索结果。

4 服务特点

4.1 用户类型及权限

从MOA的建立和资源库内容的介绍中可以看出，MOA的目标用户大体可以划分为三类：

①专业用户。主要包括从事美国历史研究，特别是从事美国南北战争到重建时期基础设施发展方面研究的相关专业人员。这些用户可以充分利用MOA的各种数据资源及其多种版本进行研究和工作。

②教育用户。主要包括学校教师及学生，特别是密歇根大学和康奈尔大学的师生。这些用户可以利用MOA的资料辅助教学和拓展学习。

③普通用户，即大众。MOA面向所有人开放其馆藏资源，任何对美国1850-1877年历史感兴趣的用户均可登录网站浏览、检索和下载感兴趣的相关资料和数据。MOA的资源涵盖教育、心理、科学技术等多个领域，侧重于社会科学，易于理解和掌握，适合普通用户兴趣阅读和参考阅读。

MOA数据库大部分资源的使用没有任何权限限制，用户无需登陆注册就可以直接浏览、检索和全文下载数据库中的资源[5]。

4.2 服务方式

（1）浏览服务

浏览服务，即用户通过简单的浏览页面获取资源的服务，这是MOA后期新加入的服务功能。MOA的图书数据库向用户提供了“标题”、“作者”、“主题”三种途径的浏览服务，用户根据需要直接浏览按照字母顺序排列的资源“标题”、“作者”或“主题”，即可找到所需资源。期刊数据库除了提供上述三种途径的浏览外，还可以按“卷数”和“发行”浏览。用户既可以在图书和期刊联合浏览页面浏览资源，也可以在图书和期

刊单独浏览页面浏览资源。此外，MOA的UMDL Texts Home还向用户提供了“资源集合”浏览，即按照首字母排序的资源集合目录进行资源浏览。

（2）检索服务

MOA向用户提供了多种检索方式，可以简单划分为综合检索和多选择检索。MOA在首页为用户提供了一个十分醒目的检索框，在此可进行初步的简单检索；点击这个检索框下方的“其它检索”就进入到MOA的多选择检索页面（如图3），包括基本项检索、布尔检索、相似检索、书目检索和历史检索。不同类型的用户可以根据需要选择合适的检索方式。例如，初级用户可以使用“基本检索”进行简单的限定词检索，而较为高级的用户可以使用相似性检索、书目检索和布尔检索进行较为专业和复杂的检索。

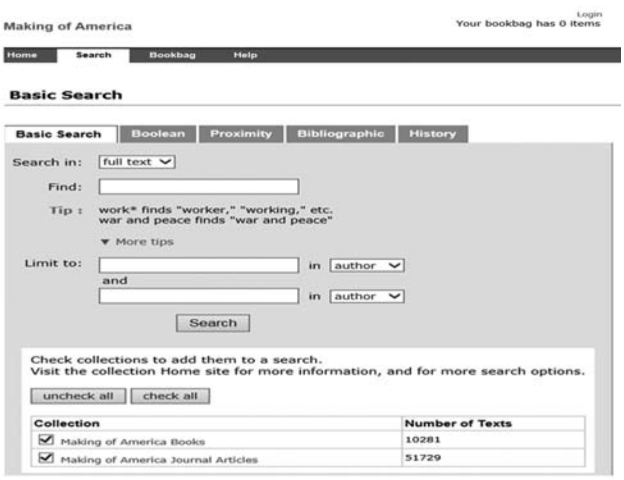

图3 MOA的多选择检索页面

（3）收藏服务

MOA向用户提供“书包（bookbag）”收藏功能，即用户如需收藏、保存所需资料的检索结果页面或目录以便下次使用，则可点击检索结果详细列表中“add to bookbag”选项，将这些记录保存到“bookbag”中，下次使用时只需要点开“bookbag”按钮查看即可，当然这需要用户先在MOA进行账户注册。

5 项目发展与维护

一直以来，MOA致力于提高OCR的准确率，即使用“高级OCR”（高级识别技术的一个软件包，使用5台OCR引擎提高图文识别准确率）对部分或全部文本进行再加工。基于对分页和文本结构的关注，MOA会在文本转化的完成阶段，选择性的插入“高级OCR”技术。随着更加精确OCR技术的植入，也使得MOA的检索工具更加有效。

进一步拓展MOA的规模以及使之融入其它主要研究型图书馆等工作和计划都在进行中。数字图书馆联盟的代表机构也已经准许将开发MOA作为多机构合作建立分布式数字图书馆的一个范例。目前，MOA资源还有以下几个方面的工作要做：

- 密歇根大学图书馆得到安德鲁•梅隆基金会的资助，将会继续增加MOA的资源量，并建立在线文件系统。
- 密歇根大学与康奈尔大学的MOA资源将进一步整合。
- 人文科学文本行动（The Humanities Text Initiative）作为密歇根大学数字图书馆产品服务的一部分，将会承接OCR的校对工作和基于用户需求的精确标注工作。同时，人文科学文本行动将会陆续编码诗歌资源到MOA中，作为数据库中美国诗歌项目的补充。
- 密歇根大学图书馆将会把数字转化技术引入到其易损书目保护工程中。随着这些资料的转化，更多新的内容将会加入到MOA中。

未来，MOA希望能够得到更多机构的资助和与更多机构合作，以便可以为MOA加入更多重要内容[6]。

6 评价与建议

MOA的优势体现在以下几个方面：

（1）馆藏资源丰富、质量较高。MOA是研究美国从南北战争到重建时期社会、历史、文化的极佳数字资源，并且其包含的欧洲各国历史以及著名作家的经典著作也非常之多，如英国湖畔派诗人的全集、休谟的英国史（6卷本）、基佐的法国文明史（8卷本）等等。同时，MOA遵循严格的工作流程和较高的技术标准，这保障了其资源库的高质量。1999年，美国国会图书馆保存委员会就曾参照MOA电子化了10卷《园与森林》杂志。

（2）检索效率高。首先，MOA的检索简便易用。例如，用户键入“Kickapoos”一词进行检索，结果列表就会显示在页面上，也就是说用户通常只需要点击两次就可以检索到包含检索词的扫描页。其次，MOA检索方法多样，根据不同水平用户需求提供有基本项检索、布尔检索、相似检索、书目检索和历史检索。并且，MOA

具有检索结果排序功能，用户可以按照标题、作者、日期以及使用频率对检索结果进行排序。

（3）资源易获取和使用。首先，MOA没有过多的权限门槛，普通用户不需要注册和登陆就可以获取MOA的大部分资源；其次，MOA中所收图书的作者都是至今已故70年以上，不涉及版权保护问题；再有，MOA拥有密歇根大学数字图书馆延伸服务（DLXS）这个强大的后台支撑系统，MOA中的资源获取和阅读不需要任何特殊的浏览器或阅读器，可提供TIFF图片、PDT和TXT文本格式供用户使用。

同时，MOA有以下几个方面亟待改进和提高：

（1）资源更新不及时。通过MOA首页介绍可以看到，MOA在2007年更新以后，至今没有更新过，资源的更新频率较低。虽然MOA中的数据并不是时效性很强的资源，但是资源的定期更新和扩充完善还是十分必要的。可考虑开放存取，实现公众开放上传资源，这样能获取更多宝贵资源充实其馆藏。

（2）资源下载速度较慢。尽管MOA数据库可以全文下载图书或期刊，并可以将图片转换成文本或PDF格式，但是这些过程往往需要很长时间。若有用户不愿意付出这个时间成本，那MOA资源的使用率会大打折扣。

（3）用户服务略显单薄。除了浏览、检索和简单的收藏这些基础功能外，没有参考咨询、互动服务等其它基础服务，更没有自己的特色服务。

（4）MOA数字资源在转化过程中存在一定的差错率。使用OCR技术进行资源转化不可避免地存在一定的差错率，特别是对一些破损页、褪色页、艺术字以及字体歪斜的页面进行转换差错率更高[7]。

参考文献

[1] Making of America [EB/OL]. [2014-10-11]. http://quod.lib.umich.edu/m/moagrp/.

[2] About MOA [EB/OL].[2014-10-11]. http://quod.lib.umich.edu/m/moagrp/about.html.

[3] Shaw, E.,Blumson, S.Making of America: Online Searching and Page Presentation at the University of Michigan.[J]. D-Lib Magazine, 1997.

[4] Sam Stavis. Online Resources Chronicles 19th Century America[N].. The Michigan Daily,1998-03-12.

[5] Bonn, Maria S. Building a Digital Library: The Stories of the Making of America. [EB/OL]. [2015-05-25]. http://www.lib.umich.edu/digital-library-production-service-dlps/building-digital-library-stories-making-america.

[6] Nineteenth Century in Print: a Distributed Digital Library Collaboration [EB/OL]. [2015-03-01]. http://memory.loc.gov/ammem/ndlpcoop/moahtml/ncpcollab.html.

[7] 吴美美, 林珊如, 黄慕萱, 叶乃静. 数字图书馆/博物馆评鉴指针建构探讨 [J].图书信息学刊, 1999.

作者简介

刘燕权，男，博士，美国南康涅克州立大学教授，研究方向：数据挖掘、数字图书馆等，E-mail：liuscsu@gmail.com。
田硕，女，硕士，铁道党校助理工程师，E-mail：tianshuo2288@163.com。

Making of America Digital Library——Documenting American Social History

Yan Quan LIU[1], TIAN Shuo[2]
(1. Southern Connecticut State University, New Haven, CT 06515, USA;
2. Railway Party School, Beijing 100088, China)

Abstract: Making of America (MOA) is a digital library of primary sources in American social history from the antebellum period through reconstruction. The collection is particularly strong in the subject areas of education, psychology, American history, sociology, religion, and science and technology. It represents “a major collaborative endeavor to preserve and make accessible through digital technology a significant body of primary sources related to development of the U.S. infrastructure.” The paper made an extended review on the digital library's construction and current development, including project background, resources organization, technological structures and service features. Comments and suggestion were also given.

Keywords: MOA; Digital Library; the Social History of the U.S; OCR

（收稿日期：2015-06-01；编辑：雷雪）