

●[美]John Wilkin 撰 何欢欢译

展望未来:数字图书馆技术的挑战和机遇*

摘要 我们长期将数字图书馆作为一个独立的世界来构建,这个独立的世界中包括数字对象的创建、管理、发现、传递以及利用等。这种相对封闭的模式与高度发达的开放网络世界背道而驰。在网络世界中,用户可以通过大型的检索系统,如 Google 和 Amazon 来发现信息,还可以在管理这些信息资源的系统之外对其进行利用。当用户都涌向网络世界时,数字图书馆就面临着打破其固有封闭状态的挑战,以便适应开放的网络世界,否则将失去其实用性。当前,许多条件都给数字图书馆提供了改变现状、提高实用性的机遇。本文将讨论数字图书馆现存的典型模式以及该模式带来的后果,讨论数字图书馆工作改革的方式,并简要探讨数字图书馆建设中的成功技术实例。

关键词 数字图书馆 挑战 机遇

分类号 G250.76

ABSTRACT For too long, we have conceived and built digital libraries as self-contained worlds consisting of digital object creation, management, discovery, delivery and use — walled gardens in the midst of thriving communities. This model runs counter to an intensively networked world where users encounter information through large discovery systems like Google and Amazon, and where “use” takes place in a variety of environments outside of the systems that manage those resources. As users shift to a networked world, digital libraries are challenged to break with their history of stubborn insularity, to adapt to that world or lose relevance. There is much that is on the horizon that holds promise for helping digital libraries change and increase their relevance. I will discuss the current paradigm of digital libraries and the consequences of that paradigm, will discuss the way that we can adapt our work in digital libraries, and will briefly examine specific examples of technologies that can contribute to more successful digital library efforts.

KEY WORDS Digital library. Challenge. Promise.

CLASS NUMBER G250.76

1 问题——数字图书馆现状

长久以来,数字图书馆都作为独立的世界存在,不仅数字图书馆之间相互隔绝,它们与外部的网络世界也相互隔绝。数字图书馆的建设包括数字对象生命周期的一系列过程:创建、管理、发现、传递以及利用等。其中,存在以下问题:

1.1 各自为政

这个问题由来已久。即使在美国数字图书馆联盟中,也存在这一问题。美国数字图书馆联盟是一个由 30 个机构组成的联盟,但是,这些机构馆藏的登记方式却各不相同,约有 500 种截然不同的馆藏登记方式。登记方式的差异导致这些馆藏信息无法作为一个整体进行检索,甚至在许多情况下,一个机构内部的馆藏信息也无法作为整体实现检索。

1.2 结构体系混乱

几乎每个机构都有自身独特的体系结构,这些体

系结构都是依据不同的指导思想来构建的。

1.3 软件差异

软件差异主要体现在数字图书馆自产软件的使用上。各机构在努力维护各自体系结构的同时,都大力开发自产软件,因为在图书馆界流行这样的说法:“如果不能自行开发软件,你就无法在数字图书馆界立足”。尽管有些软件在数字图书馆界也能够成为主流软件,例如 Greenstone、Fedora、DSpace 以及密歇根自行开发的 DLXS 等,但是仍然有许多自产非主流软件的出现和使用,这些软件的开发基础薄弱并且应用的范围较小,导致数字图书馆建设所使用软件的巨大差异。

1.4 服务共享性差

除了信息检索和显示这两项最主要的服务之外,其他服务几乎没有任何的一致性。唯一例外的就是“开放文档计划”的元数据发现(OAI metadata expo-

* 本文为中国国家自然科学基金/美国国家科学基金委联合资助;中美政府信息资源管理研讨会(NSFC 70373048)成果。作者授权翻译并在本刊发表。英文稿见本刊网站(www.ztxb.net.cn)。

sure)。即便如此,在众多实践中,仍然没有通过“开放文档计划”实现元数据的共享。

许多证据表明,上述问题导致数字图书馆的孤立存在:

①OAIster*中大约70%的记录内容无法被Google检索到。

②最近,我们从一个在线列表中随机挑选了23个数字图书馆,并在Google中检索它们的内容。有17个数字图书馆的部分或者全部内容无法在Google中检索到,至少有2个数字图书馆的内容在Google上完全无法找到,即74%的数字图书馆都存在着重要内容无法被Google检索的问题。

一个成功的数字图书馆,必须确保用户能够在Google或者其他网络搜索引擎中检索到其内容信息。但是现实却与此相反:数字图书馆的信息孤立而且自成一体,即使是用各个系统中最高级的检索功能或者工具都无法将其检索,这就导致大部分用户逐渐与数字图书馆脱离。

2 数字图书馆的外部世界

2.1 高度网络化

Google、Amazon、Flickr、iTunes以及其他搜索引擎已经在用户的检索领域中占据了主导地位。正如Lorcan Dempsey所说,“Google、Amazon以及EBay的大量计算和数据平台已经成为了用户上网的主要工具”,它们已经成为了用户上网所使用的一整套工具体系中无可否认的中心力量。OCLC环境扫描(OCLC environmental scans)指出,用户首先会考虑使用Google和Amazon所提供的服务,其次才会考虑寻求数字图书馆所购买或创建的服务。

2.2 日渐学术化

三大最受关注的新兴工具为:OCLC的WorldCat Local, Google Book Search以及Google Scholar。它们的出现扩展了学术信息资源发现的范围。但是,当前类似于Amazon和Google的主流网络搜索服务无法完全满足学术信息资源的发现需求,因此,需要使用专门用于学术信息资源检索的计算和数据平台对其进行补充。

2.3 日益远离图书馆世界

数字图书馆建设壁垒增多,但是缺乏相互联接的

枢纽。而外部的网络世界却在逐步诞生诸如Amazoogle的协作开发工具,这些已经出现以及将要出现的工具都将影响图书馆世界设计和构建系统的方式。如果图书馆界忽视这些最新的进展成果,继续选择使用陈旧的工具来构建不实用的系统,那么构建出来的系统就会失去效用,仅仅能够为现实中不存在的“理想用户”服务。

新的资源、技术以及其他条件为新一代数字图书馆系统的构建提供了机遇。图书馆界必须认识到强大的网络层以及作为优秀研究型图书馆所必须承担的职责和需求,应当根据环境的变化,重新对图书馆进行构思和定位。在重新构思和定位的过程中,不仅应该意识到用户的新目标和需求,还应该认识到数字图书馆对于图书馆资产长期管理必须承担的特殊职责。

外部网络世界也有其缺陷和不足,即使Google Scholar发展到最高级阶段,它也不可能包含我们花费800万美元为大学教育所购买的所有电子资源。同样地,Google Book Search也无法检索到我们为了学术交流而进行的活动部署安排。

3 重建数字图书馆世界的原则

数字图书馆的重建必须改变投资方向。本文没有为投资方向的转移描绘一个确切的蓝图,仅对数字图书馆世界重建的主要原则进行了探讨:

3.1 与网络服务达到平衡

这是构建数字图书馆的最重要原则之一。数字图书馆必须在充分利用网络提供资源和服务的基础之上,寻求各种方式实现数字图书馆与网络服务的一体化,确保在合适的领域进行投资,所谓“合适的领域”,即在这个领域中,数字图书馆能够完全或部分取代网络服务,发挥主要作用。

例如,Google作为信息检索的主要工具,能够满足大部分用户检索信息的需求,数字图书馆不必模仿其功能,而应当:

(1)将数字图书馆的信息资源放到网上,使其能够被Google检索到;

(2)弥补Google在满足用户信息需求方面的缺陷。首先,对于受法律保护或者涉及隐私的信息,有必要采取一定措施,阻止其被Google检索。其次,在

* OAIster是密歇根大学图书馆开发维护的一个优秀开放存取搜索引擎,提供了各种学术数字资源的一站式检索。目前,收集了来自935家学术机构的15,475,946条数字资源,包括图书、期刊、音频、图像、电影、数据集等资源。这些资源通常是其他搜索引擎无法找到的隐性资源。

构建弥补 Google 功能的服务时,必须在保持灵活性的同时具有战略上的稳定性。

3.2 开放

数字图书馆开发的工具必须具有开放性,能够在不同数字图书馆的开发和构建中发挥作用。对于现有的方案、工具以及服务,要善加利用。此外,所开放的数据必须是他人开发工具和服务所需要的数据,因为开放几乎无人问津的资源是没有价值的。

开放固然很重要,但是并非所有工具和服务都必须开放。正如 O'Reilly 在他对 Web 2.0 出现的分析报告中所写到的,“我们在 Amazon 和 Google 的体系结构中能看到它们构建各种服务的机制都是相互联系的,但是,没有人能看到他们基础服务的代码。”

3.3 开源

数字图书馆应当采取开源许可模式来运作。由于许多开源软件都是封闭孤立的,没有太大的作用。当选择使用开源软件时,必须挑选良好的开源软件成果——即事实上已经开放并且已经被使用的成果。数字图书馆开源的目标应当是开放服务,而不在于开源行为本身。

3.4 整合

整合是构建数字图书馆系统的最重要特征之一,在所有或者大部分系统中都必须实现最优的整合。根据整合程度的不同,可以将整合区分为紧密的整合和松散的整合。这两种方式各有优劣,在对资源做进一步整合之前,必须经过仔细慎重的考虑。

整合能够为用户带来更高的效率以及更可靠的结果。例如,通过整合远程和本地的信息资源,用户就能够使用元搜索来发现远程和本地的资源,他们只需访问一个单独的物理或者虚拟数据库就能获取服务器上的数据。

密歇根大学的 MBooks* 数据库已经开始显示出整合的价值,用户依靠 Aleph X-Server 就能够访问所有书目信息。

3.5 迅速及时开发并提供可靠的服务

在经过详细部署之前迅速开发新一代图书馆体系结构是不切实际的。管理者和开发者之间的矛盾

在于:一方认为图书馆浪费大量时间去开发徒有其表的编码,却没有为用户提供服务和产品;另一方则认为急于求成提供的服务和产品质量必然下降。这两种极端的观点都是错误的。

Google 实施“永久测试版”(perpetual beta)计划的最新实践以及日新月异的发展前景都强调迅速及时提供服务和产品的必要性,但是服务质量的可靠性也很重要。二者相辅相成,缺一不可。这是衡量利用现有资源所取得成果的重要指标,能够确保将资源用于合理的用途,同时避免急于求成的行为。

4 方向与前景

以上述原则为基础,数字图书馆应当及时利用技术变革带来的机遇,充分利用现有的资源,从以下几个方向努力:

4.1 形成统筹性的整合环境

只有在良好的大环境之下,才能实现信息资源的最佳整合。

4.2 使用最佳工具

在整合过程中,应当善于利用良好的工具,Aleph X-Servers 以及 MetaLib 就是这类工具的典型代表,使用这些工具能够实现我们所设想的整合。

4.3 共享成果

如果具有提供共享资源的团队,就可以利用这些团队所开发的共享成果。

5 进展实践

大量实践证明,重建图书馆世界的原则已经开始产生效用:

5.1 开源行动

Evergreen 和 LibraryFind** 仅仅是现有图书馆技术模型的形式之一,但是它们源自于图书馆业界及其开源的事实预示了其良好的发展前景。

5.2 网络层次的服务

WorldCat Local 的发展尽管有明显的缺陷,但是极具发展前途。由于各种原因,它无法与 Endeca、Primo 以及其他新一代的检索工具竞争,但是它努力使自己在这类检索工具中脱颖而出。(下转第 13 页)

* MBooks 系统能够提供访问 Google 扫描资料。MBooks 使用户可以通过 OPAC (Mirlyn) 中的全文检索,找到图书并链接到图书页面 (page turner)。在那里可以浏览、打印、放大、旋转页面,也可以做书内检索,并提供链接到 Google 的图书搜索 (那里有可能提供整书的 PDF 文件下载)。

** LibraryFind 是美国俄勒冈州立大学图书馆研制的开放源代码的元搜索软件 (<http://libraryfind.org>),使用类似于 Google 搜索引擎的简洁应用界面,提供通用检索、图片和书籍等检索,支持用户界面的个性化,集成了 OpenURL 解析器,允许全文资源链接,采用两层缓存系统提高检索反应速度。

纸,并为古籍量身定制修复用纸,在纸面光滑度和薄度方面均有重大突破。用这类超薄修复纸进行修裱工作,不仅透析度高可保持文字原貌,还可减少修裱后书页厚度,因此很受博物馆、图书馆与档案馆的欢迎,在台湾“适应市场能力强,经济效益很好”^[6]。

创建中国文献用纸产研基地,不仅可生产与研究文献修复用纸,其传统造纸的手工艺还可成为宣传中华文化的观光之地。国外许多纸张生产商也看准了这一商机,发起了中性纸张的生产运动,例如加拿大许多纸张生产商就曾与加拿大保护研究所等文化机构合作,共同研制适用于保存重要文献的纸张。此外,澳大利亚、德国、日本、荷兰、瑞士等都有这类实例^[7-8]。因此,创建我国文献用纸产研基地是必要和可行的,也是保护更多重要文献和今后可能成为古籍的文献的重要战略。

参考文献:

[1] 古籍修复难题亟待突破[N].南方日报,2007-10-13.
 [2] 大学生,莫患学习“短视症”[N].新闻观察,2005-01-4(科教卫版).

[3] 谈中等职业教育改革与发展[OL]. [2008-01-08]. http://www.moe.gov.cn/edoas/website18/fangtan_050926_liuzs.jsp
 [4] 徐娜. 图书馆开拓经费来源之我见[J]. 图书馆建设, 1999(3).
 [5] 刘家真. 文献保护学[M]. 武汉:武汉大学出版社, 339-340.
 [6] 李忠正. 台湾造纸工业访问见闻[J]. 纸和造纸,1998(2).
 [7] 树火全球资讯[OL]. [2008-01-08]. <http://www.su-hopaper.org.tw/chinese/otherlinks/otherlinks.htm>.
 [8] 刘家真. 文献遗产保护[M]. 北京:高等教育出版社,2005.

刘家真 武汉大学信息管理学院教授,博士生导师。通讯地址:武汉大学信息管理学院。邮编 430072。

程万高 武汉大学信息管理学院博士生。通讯地址同上。

(收稿日期:2008-01-10)

(上接第7页)

5.3 灵活的体系结构

Fedora 存储系统已经被设计成为适用于整合的系统。围绕该系统成长起来的团队就是它成功的证据之一。VITAL 系统以及 Fedora 与 VTLS 图书馆管理系统整合的实例都是我们所希望达到的目标之一。

5.4 模块化

模块化的成功实践是 OAI 扩展,它能够支持对象的再利用和交换(OAI-ORE)。正如 OAI 网站所提出的那样,它们致力于开发一系列规范,实现分布式数据库之间数字对象相关信息的交流。这些规范包括数字对象标识方法以及便于存取和吸收数字对象的数据库服务方法。这些规范将促使新一代的交互式数据库服务产生,这种服务能够发挥分布于不同地点的数据库中数字对象的内在价值。我们对此抱有很大期望,期待变革能够产生重要和深远的影响。

6 结论

当前,图书馆是一个承担了一系列资源管理(既包括纸质资料管理,也包括数字信息管理;既包括本地信息管理,也包括远程信息管理)职责的组织,这

些管理行为能够支撑日渐成熟和发展的网络服务,从而使用户能够有效地发现和使用他们有权使用的信息资源。图书馆必须改变对于技术的思考方式,对技术进行战略上的合理部署,把握相关技术之间的有机联系。目前,由于纸质资料的固有属性以及图书馆发展的历史原因,图书馆可能会被大量的本地纸质资料管理所牵绊。但是,与此同时,图书馆必须熟练掌握数字信息管理方法,并将数字信息管理的各个流程(例如接收、公开、传递、利用等)以及能够实现这些流程的最佳工具进行概念化和标准化,最终为构建数字图书馆服务。

数字图书馆和数字档案馆分别面临着许多不同的基础性问题,但是他们所面临的挑战却有相似之处。我希望,我所提出的问题以及解决办法能够对档案领域也有所裨益。

John Wilkin 美国密歇根大学图书馆副馆长。

何欢欢 武汉大学信息管理学院博士研究生。通讯地址:武汉大学信息管理学院。邮编 430072。

(收稿日期:2008-05-22)