

资源采集 管理政策——虚拟馆藏选择评价刍议之二

[作者] 吴云标

[单位] 浙江大学

[摘要] 在《影响虚拟馆藏利用的因素分析——虚拟馆藏选择评价刍议之一》(载本刊 2001 年第 5 期)一文的基础上,提出建设虚拟馆藏的资源采集和馆藏管理政策。

[关键词] 馆藏建设, 虚拟馆藏, 信息资源采集, 网络信息资源

1 虚拟馆藏资源的采集

1.1 选择准则

从《影响虚拟馆藏利用的因素分析——虚拟馆藏选择评价刍议之一》(载本刊 2001 年第 5 期)对影响虚拟馆藏利用的主要因素的分析可以看出,确立完善且可行的虚拟馆藏资源的选择准则并非易事。这里所说的准则,是指能使评价过程形式化,具有较强的可操作性。从选择准则的评价层次来看,选择准则可分为两部分:一部分是说明虚拟馆藏中的所有资源都必须具备的品质,也就是 Internet 资源入选虚拟馆藏的基本条件;另一部分是说明在有不止一种资源满足入选虚拟馆藏基本条件的情况下,考虑其它因素进行优选。不过,哪些因素决定了入藏的基本条件,哪些被考虑为优选因素,不存在统一划界,因为对不同的虚拟馆藏和不同的用户群体,其内容是不相同的。另外,从选择准则的评价可操作性来看,显然有难易之分。譬如,要评价手上的资源与虚拟馆藏的用户社会相关状况,有时可能很容易,但许多情况下并不容易,而对于虚拟馆藏的用户,社会是否具备条件供他们使用这一资源,却很容易定夺,只要知道他们所具有的软硬件设备情况就可以了。

1.2 评价步骤

不同的虚拟馆藏需要不同的选择准则,对期望从建立虚拟馆藏获得较多利益的中小型图书情报信息服务机构而言,由于缺乏所需的相关人才,在评价 Internet 信息资源时会有较多困难,建立评价准则更加不易。因此,我们的目的是建立一个一般框架,作为各个虚拟馆藏选择评价准则的基础。虚拟馆藏资源的选择评价可以分两步:初选和细评。这样,那些技术力量不足、难以投入较多精力进行资源评价的中小型图书情报信息服务机构,就能把主要力量放在初选上,建立起一个虽然粗糙,但会对本馆用户有用的虚拟馆藏。分两步进行评价,不仅体现了逐步完善的思想,而且也便于各类型图书情报信息服务机构能根据自己的情况,作出适当的调整修改,各取所需。

所谓“初选”,就是图书情报信息服务机构在建设虚拟馆藏时,根据本单位用户的信息需求,为自己确定一个虚拟馆藏的范围准则,规定网络信息资源的采集范围和标准。确立范围准则的一个重要思想是:范围准则是基本准则与简单准则的交集,是入藏的基本条件、必

要条件和共有条件；同时，它们又必须简单易行，不需要较多的专门知识和技术，容易确定。笔者将就确立范围准则所要考虑的问题，提出一个框架，作为“初选”所需的范围政策，供读者修改参考。

任何图书情报信息服务机构，在确立范围准则（或基本准则）时，首先要考虑的是自身的目标和用户对象，这同处理实体馆藏一样。对确立范围准则而言，网络信息资源与传统信息资源相比，有两个问题尤其值得注意：虚拟馆藏的内容特色和虚拟馆藏的可用性。无论一个机构有多大，实力有多强，也无法建立一个包罗整个 Internet 信息资源的虚拟馆藏。有特色就是别人有的当然可以有，别人没有的我也有，但不要去重复劳动，这是由 Internet 的特点所决定的。其次，Internet 作为一种媒介，其信息资源的可进入性和可用性对虚拟馆藏建设也具有重要意义，因为它影响或决定了虚拟馆藏的可进入性和总的可用性。由于影响因素众多，确定一种网络信息资源是随时可得，还是由于服务器的不可靠或要求进入服务器的数量过大或其它原因而只能间歇可得，需要随着时间的推移来评价。

1.3 Internet 信息资源的格式

在给出范围政策以前，还需对 Internet 上信息资源的类型与格式作一简单介绍。国际互联网信息资源的标准格式是：

PDF (Adobe 的可移植文档格式 Portable Document Format)：用电子手段呈现印刷型文献的一种格式，以使它们能如印刷物一样地精确显现——由 Adobe Acrobat 软件所使用。

ASCII 美国信息交换标准码(美国信息互换标准代码 American Standard Code for Information Interchange)：一种文本文件格式。

GIF (可交换的图像文件格式 Graphics Interchange Format)：用于互联网(WWW)上压缩图形的一种标准。

JPEG (ISO/IEC 联合摄影专家组 Joint Photographic Experts Group)：用于互联网(WWW)上压缩图形的一种标准。

PNG (可移植的网络图像文件格式 Portable Network Graphics)：压缩图形的一种文件格式，被期望取代 GIF 格式。

MPEG (运动图像专家组 Moving Picture Experts Group)：压缩数字视频和音频的一种标准。

REALAUDIO (实时广播)：播放声音文件的一种手段，当文件下载时同时播放，而不必先下载并储存以后才播放。

1.4 资源类型的划分

1.4.1 根据资源格式划分

文本(text)：文摘、文章、电子邮件、词典、主页、索引、手册、单行本、小册子、诗、会议文献、丛书。

图像(image)：动画、电影、照片、图表。

声音(sound)：环绕声、音效、音乐、道白、演讲。

软件(software)：程序(即软件)。

数据(data)：数字的、空间的、光谱的、统计的、结构化文本(比如书目数据和数据库报告)。

交互性的(interactive)：聊天、游戏、多媒体、虚拟现实。

1.4.2 根据 Internet 协议类型划分

区分资源类型的另一种方法是通过辨别它所使用的 Internet 协议。这里是一些不同的协议的例子：

HTTP (超文本传送协议 Hypertext Transfer Protocol)：互联网上交换文件的标准和规则。

FTP (文件传送协议 File Transfer Protocol)：允许人们访问和下载在一个远程的 FTP 服务器上拥有的文件，并且将文件从一个计算机传送到另一个计算机。

Mail to：能通过互联网交换储存于计算机的电子邮件信息。

Telnet 用于远程联接服务的标准协议：能使我们访问在另一台计算机上拥有的文件(当已经得到允许时)。

1.4.3 根据资源的来源

原始信息。若网站的服务器上确实拥有信息，并这些信息是由网站的所有者创造和生产的，那么这种网站就包含原始信息。大部分链接可能是内部的，即指向同一服务器上的信息。

二手信息。如果网站与其它地方的信息链接，并且这些信息是在其它地方创造的“第三方”信息，就像虚拟馆藏做的那样，那么这种网站就包含二手信息。大部分的链接将是外部的——将带读者浏览位于其它服务器上的信息。

2 范围政策

2.1 主题内容

对目标用户，什么主题内容是适当的？主题内容的界定应有所侧重，因服务对象是综合性的，故应根据重点对象、形势需要等考虑设立特色栏目。

是否有需要审批的主题？如有这样的主题，应根据主题内容对其进行审批。

2.2 可接受的资源类型

对目标用户，哪种类型的资源是适当的？

对目标用户，信息应是学术性的还是通俗性的？

对目标用户，哪种格式的资源是适当的？

是否可选择只包含一个链接列表的资源（二手资源）？应考虑选择有专人维护或自动更新的资源。

2.3 可以接受的来源

在学术的、政府的、商业的、工业的、非营利性私人资源中哪些资源是可以接受的？

由个别网迷（譬如学生）制作和维护的网页可以接受吗？

2.4 可接受的难易程度

哪种水平的资源对目标用户是适当的，材料的宽度、深度应定位在什么水平上？划分参考：研究或实践用信息资源：国际先进水平，国内先进水平，一般水平；学习教育用信息资源：大专及以上水平，中学水平，小学及以下水平。

2.5 时间要求

是否需要证明被选择的网站是持久，或被期望是持久的？一般而言，用户要求时间性强、新颖的网站。

过时的信息材料可以选择吗？

临时性的信息资料可以选择吗？如果需要临时性材料，最好单立栏目，像电子邮件信息或新闻组布告。

2.6 广告

包含广告的资源可以接受吗？注意是否存在任何形式的广告审查，比如是否需要检查广告的合法性。

2.7 费用

是否只收集免费访问的信息资源？

访问费用负担的限制如何？有许多中外文专业期刊在网上能浏览到每期的目录，甚至每篇文献的摘要，但获得全文，需要注册并付费。如果在一定时期内，用户群体注册获取全文的费用低于在网下用传统方式购买的费用，则集体注册显然是个省钱的办法，更何况其具有优越的时间性。

受版权保护的资源怎么处理？

2.8 技术

对目标用户，哪些界面技术是适当的？用户界面技术主要有表格、图像、框架、数据库展现、cgi 脚本、Java 应用程序等等。当用户利用计算机来使用虚拟馆藏时，其设备的技术水平可能不一致，这就需要根据大多数情况来确定应采用的技术。

目标用户使用哪种连接方式，并且它会如何影响选择？当用户通过自己的计算机来利用虚拟馆藏时，电缆连接速度较快较稳定，而拨号连接较差，可能影响较大文件或界面技术复杂的信息资源的获取。

2.9 注册

是否选择被访问之前必须先进行用户注册的资源？

是否选择可联机注册的资源？

是否选择被访问之前用户必须签订书面合同的资源？

2.10 安全

当用户有必要在 Internet 上传送机密信息时，一种安全码系统的提供或者密码技术是否会影响选择？

2.11 专门需求

用户有任何会影响信息资源选择的特别要求吗？比如对伤残用户，是否有大号印刷或声音交替的选择。

2.12 间隔水平

资源在什么间隔水平会被选择及编目？是在整个网站水平，还是在某个或几个专门栏目水平，抑或在一个个网页水平；是在新闻讨论组水平，还是在新闻讨论组文章的水平等等。

2.13 资源说明

选择一种资源应该包含哪些基本信息？在虚拟馆藏目录中，当需要对被选择的资源进行描述时，所需要的最少信息量是什么？一般而言，如选择整个网站，应对这个网站作基本介绍，如该网站的主办者或拥有者是谁，在行业中的地位如何，有什么特色等等；如选择一个栏目，应有这个栏目的基本介绍，如栏目主持人、栏目特色等等；如选择某篇国际互联网文献，应知道著者、获取方式、最近更新情况等等。

是否只选择带有专门的元数据的信息资源？除非你的虚拟馆藏非常专门，或者能自动分类，使用搜索引擎等，否则大多数网页不提供主题词或关键词之类的元数据。

2.14 地理限制

虚拟馆藏是否选择本单位、本地区产生的信息？

是否需要挑选或排除来自一些特别的国家、地区的信息，或者世界各地的信息？

2.15 语言

对目标用户，哪种语言的资源是可以接受的、适当的？

3 管理政策

明确了虚拟馆藏的范围选择政策，就容易建立起相应的范围选择准则，从而建设一个虚拟馆藏。但这还不够，建设虚拟馆藏，还需要在使用过程中进行维护管理，制定虚拟馆藏的管理政策。虚拟馆藏的管理，重要的一条，就是追踪用户使用虚拟馆藏的情况，根据用户的意见和要求，利用各种网上信息资源，对虚拟馆藏进行调整、增加和剔除。下面所列的问题和范围选择政策中所列的问题一样，可以被用来确定相关准则，甚至有些可以直接为准则所用。这些问题考虑了这样的事实，一种资源在某时某点可以被选择，但在另外的某时某点或许就不应被选择。随着虚拟馆藏的增加，用户的需求可能会变化，虚拟馆藏的特色、覆盖范围和平衡也可能会改变，这就会影响选择过程。

3.1 虚拟馆藏范围和平衡

虚拟馆藏中还缺少什么？

与已经在虚拟馆藏中的其它资源相比，欲增加或剔除的信息资源是否具有相对使用价值？

在整个虚拟馆藏的内容范围内，欲增加或剔除的信息是独一无二的还是重复的？

3.2 Internet 资源的可得性

欲增加或剔除的信息在虚拟馆藏以外可以得到吗？

与有关这一主题的其它可得信息资源相比，该信息资源是否具有相对价值？能寻找比该信息资源具有更多价值的其它资源吗？

如果要在虚拟馆藏中增加或剔除一个网站，这个网站包含了不容易得到的信息吗？

3.3 图书馆资源的可得性

这些资源是否提供其它媒体（报纸、电台、广播、出版等等）或其它信息源不容易得到的信息？

这些资源是否提供其它载体类型（图书、期刊、胶片、光盘等等）不容易得到的信息？或者在许多图书馆所不拥有的信息？

这些资源是否提供用户自身上网访问所难以找到的信息，或者只能在图书馆中进行访问的信息？