

《专题：数字资源长期保存》

日本国会图书馆因特网资源搜集、保存和提供计划

日本国会图书馆 著 余敏 编译

日本国会图书馆（The National Diet Library, NDL）早在 1998 年 4 月就制定了“日本国会图书馆电子图书馆构想”（National Diet Library Electronic Library Concept）。该构想指出，因特网上可获取的信息资源都应该被搜集、存档并由图书馆加以提供。2000 年 NDL 制定了“电子图书馆服务基本实施计划”（Basic Implementation Plan for Electronic Library Services），并于 2002 年起进行项目试点。2004 年 2 月，NDL 制定出“日本国会图书馆 2004 年中期规划”（National Diet Library Digital Library Medium Term Plan for 2004），从此步入数字档案建设时期。

基于以上提到的各项计划，我们已经持续的建立了对因特网信息资源搜集、存档计划。以下是我们目前正在运作的主要项目：

1. 因特网信息资源搜集、存档以及提供的试验性项目

伴随着 2002 财政年关西馆的启用，NDL 就开始了因特网信息资源的搜集、存档、提供的相应项目的试点。以下三个项目的实行正是为了发现并设计合理的运营模式。

● 网页存档计划 WARP（Web Archiving Project）(<http://warp.ndl.go.jp>)

WARP，作为 NDL 的子项目，是一个在版权持有者允许的情况下，对因特网资源进行有选择性的搜集以存档并保证公众获取的计划。该项目始于 2002 财政年的试点，并于 2006 财政年步入实际运营阶段。

在 WARP 项目中，NDL 将搜集以下站点资源：国家行政机构；地方政府；市，镇，村，以及合并前和合并后的议会；公共利益公司、组织,以及转变成独立机构前的国立大学；以及诸如 2002 年世界杯足球赛的各类活动。NDL 同时也搜集那些由以上提到的各个组织在网络上出版的各类期刊。

以上提供的这些信息资源是由一个经过版权人授权的网页爬行机器人进行搜集的。对于每个站点和网络期刊，我们都赋予其相应的元数据作为一个单元对象（标题）以便和相应的不同日期搜集的不同版本（实体）联系起来。截至 2007 年 1 月份，我们已经存档了以下的因特网资源：

类型	标题数量	版本数量	文件数量 (百万)	数据量 (GB)
在线期刊 (共)	1,499	6,837	4.57	560
网络站点 (共)	1,922	7,521	54.37	3,403
国家机关	38	298	8.85	683
地方政府	8	98	8.31	495
合并的城镇和农村	1,687	6,095	21.33	1,436
公共利于公司、组织	85	780	12.73	615
大学	71	106	2.61	155

活动	26	102	0.47	18
其他	7	42	0.07	1
总计	3,421	14,358	58.94	3,963

表 1 WARP 内容 (截至 2007 年 1 月)

● **对日本网站进行全面搜集，保存以及存档的调查**

该项调查是从 2004 年的 10 月到 2005 年的 3 月，调查目的在于研究对日本网站进行搜集、保存以及存档的可行性和相应方法论。调查的对象主要是那些包含日本域名的国内网站。根据此项调查的结果，我们计算出日本网站的总体数据量接近 18.4TB,总文件数达到 4.5 亿。更多的有关该项调查的信息，请参见 NDL 网站：

http://www.ndl.go.jp/en/aboutus/bulkresearch2005summary_e.html

● **数据库导航服务：Dnavi (Database Navigation Service) (<http://dnavi.ndl.go.jp>)**

大量的互联网络资源，事实上是以网页机器人无法抓取的数据库的形式存在的。这导致大量的资源无法通过搜索站点而被搜索到。Dnavi 能够为用户浏览日本的数据库提供门户导航。该项目始于 2002 财政年的试验计划，并于 2006 财政年与 WAPP 项目共同进入实际运营阶段。截至 2006 年 1 月份，Dnavi 拥有的数据库数量已达 9,600 个。

2. 法定呈缴委员会的报告

2002 年 3 月，NDL 馆长曾向法定呈缴委员会咨询其对于如下问题的看法：在日本境内发行的网络电子出版物是否应该纳入呈缴制度的范围内？如果否，应该以何种标准，何种方法对其进行搜集？

2004 年 12 月，委员会向馆长提交了标题为“网络电子出版物采集系统构想”（Concept of the Acquisition System for the Networked Electronic Publications）的报告。该报告指出将网络电子出版物纳入呈缴制度是不符合呈缴制度的基本原则的，但同时该报告也给出了在呈缴制度之外进行网络电子出版物搜集的方法框架。该框架的纪要和全文可参见：
http://www.ndl.go.jp/en/aboutus/deposit_council_book.html

3. 创建数字档案

NDL 于 2004 年 1 月制定了“日本国会图书馆 2004 年中期规划”（National Diet Library Digital Library Medium Term Plan for 2004），该规划明确了 NDL 数字图书馆服务的建设目标，其中之一就是要把 NDL 建设成为日本数字档案的主要基地。

NDL 一直致力于开发作为数字档案基础设施系统的 NDL 数字档案系统（NDL Digital Archive System，NDL DA System）。该系统的主要作用在于确保从搜集、组织、通过保存数字信息加以提供，到确保数字资源的长期保存和可获得性的整个流程。这里所说的数字信息不仅包括因特网信息资源，还包括那些通过光盘出版的电子出版物以及那些传统的报业出版物的数字化形式。NDL 数字档案系统建立在基于开放档案系统（the Open Archival Information System，OAIS）参考模型（ISO14721:2003）的基础上，其中包含三个层，分别是应用层、保存层以及海量存储层。数字信息通过元数据编码及转换标准 METS（Metadata Encoding and Transmission Standard）被封装成可以长期保存的格式。我们计划采用元数据对象描述框架（Metadata Object Description Schema，MODS）作为创建描述性元数据的标准。我们正在努力促进系统的开发以实现在 2009 财政年实现 NDL 数字档案保存系统的正式运行，目前的 WAPP 系统将会被转换成 NDL 数字档案保存系统。

NDL 始终在尝试创建日本数字档案门户。在 2004 年 NDL 数字档案门户的系统原型开始试运行访问的基础上，2007 年我们将展开全面的服务。

NDL 的目标在于在未来法律允许的范围之内搜集使用日本域名的因特网信息。为了将

这个愿景变为现实，目前的立法需要更加强调解决诸如知识产权在内的现实问题。尽管在短期之内该问题无法得到解决，但是我们将继续努力以实现相关的立法。

为了实现这些数字信息资源的搜集及长期保存，目前还需解决不少技术问题，如重放系统的过时（the obsolescence of playback systems）等。NDL 一直在进行不同的保存技术的可行性研究，如电离和仿真等。NDL 希望能够与其他的数字档案馆包括日本图书馆、档案馆以及其他的私营机构和研究者共同努力推进标准化的形成。

我们将继续努力进行因特网资源的搜集和保存，推动标准的形成和广泛合作，从而对日本和国际社会有所贡献。

编译自：National Diet Library's Projects of Collecting, Preserving and Providing Internet Resources. <http://www.ndl.go.jp/en/publication/cdnlao/058/582.html>. [2007-4-5]