

政府网站信息资源归档保存的概念与模式研究

赵玉

(西北工业大学档案馆 陕西 西安 710072)

[摘要]笔者首先从政府网站信息资源的内涵、种类和特点界定了“政府网站信息资源”的概念,其次说明了“政府网站信息资源归档保存”的含义及目前的保存模式和典型项目,并分析了各个模式的优缺点,以便更好地了解世界各国政府网站归档保存项目的理论和实践进展,为下一步的研究打好基础。

[关键词]政府信息资源;网站信息资源;资源归档保存

[DOI] 10.12252/j.issn.2096-6288.2021.03.2161

一、政府网站信息资源的概念

1. 政府网站信息资源的内涵

政府信息资源是社会民主进程与公众精神生活的基石,通常被大多数学者认可和引用的“政府信息资源”定义为:一切产生于政府内部或虽产生于政府外部但对政府活动有影响的信息资源的统称。^[1]学者何欢欢^[2]据此认为“政府网站信息资源”是指以政府网站为传播媒介,由政府、其他机构或者个人产生的与政府各种活动有关的信息资源的总称。总之,学界普遍认同的政府网站信息资源就是指在政府网站上发布或传播的信息资源。

笔者认为,政府网站信息资源是指在政府网站上发布或传播的信息资源及其相关元数据。首先,政府网站信息资源与其他载体的政府信息资源相比,它有易逝性、不稳定性、不可控性等特点;其次,大部分政府网站信息资源都是原生信息,没有固定的实体,编号、来源等难以记录;最后,从各国的网站信息资源归档保存的实践来看,都尽可能保存网站的原始背景和内容信息。

2. 政府网站信息资源的种类

政府网站信息资源根据分类标准的不同会划分为很多种类。根据笔者对中央人民政府网站的统计,按照信息内容的不同,政府网站信息资源主要可以分为以下几类:

第一,国家政策类信息。如政府发布的文件、法律法规、政策解读、国务院公报、白皮书、地方督察等内容。

第二,新闻类信息。主要是国务院领导人发表的重要讲话、重要活动、时政热点、部门消息、地方政府新闻等内容。

第三,领导人信息。主要是国务院领导集体参加的会议、活动、出访和发表的讲话等内容。

第四,民生服务类信息。根据对象的不同,分为公民、企业、外国人和社会组织四类,针对每类服务对象又有不同的服务项目。

第五,数据类信息。主要是由国家统计局统计的经济类数据,如国内生产总值GDP、居民消费者价格指数CPI、生产者价格指数PPI、工业生产增长速度等内容。

第六,问政类信息。主要是政府与公民的互动情况,如回应公众关心的问题、征集意见、回复网友的留言等内容。

第七,站点地图类信息。主要是连接到相关网站的信息,如国务院部门网站、地方政府网站、驻港澳机构网站等内容。

总之,无论哪种类型的政府网站信息资源,都有各自的作用,都可能会有特定的用户需要访问与利用,因此都需要归档保存。

3. 政府网站信息资源的特点

笔者认为,政府网站信息资源具有以下特点:

第一,权威性。首先,政府网站上面发布的信息,都是经过政府相关部门确认的信息;其次,政府网站是发布大政方针的平台,其掌握的信息资源质量和数量,与其他网站信息资源相比,具有更高的可靠性。

第二,时效性。信息时代,最讲求效率。政府网站一般都会在第一时间发布新闻与热点事件,并且就公众最关心的问题给予解答与回应,为公众提供更好的便民服务和用户体验。

第三,综合性。与各种专题网站不同,政府网站是大型的门户网站,信息资源丰富多样,政治、经济、文化、军事以及社会民生等问题,无所不包,尽力为公众提供“一站式服务”。

第四,不稳定性。传统的政府信息资源一般依托于纸张、光盘或缩微胶卷等载体保存,电子文件一般也有对应的实体档案,有固定的编码和放置位置,被妥善保管。而政府网站信息资源有相当一部分是原生信息(如网友互动等),也就是只存在与网站上,没有对应的实体或者电子文件,而网站的安全风险大,数据复杂,链接随着时间也会出现无效的情况,再加上技术的不断发展,导致政府网站信息资源的归档保存难度较大。

二、政府网站信息资源归档保存的概念

笔者在查找文献资料的过程中发现,由于信息和网络技术的发展,国外的网站信息资源归档保存的实践早于我国,国外学者研究也较早,对应的英文为Web Archive(简称WA)。而国内学者在学习和研究国外主要项目和理论的基础上,对Web Archive有不同的理解,对此有多种翻译,除了“网站信息资源归档保存”这个说法,还有学者用“网络档案馆”“网站保存”“网站信息资源长期保存”“网站存档”等词来表达。

首先,“归档保存”强调对政府网站信息资源的档案化管理,是一种行为和过程;而“长期保存”强调通过各种手段确保政府网站信息资源长久的可用性和真实性,是一种状态和目的,归档是其手段之一。其次,图书馆的对象是各种电子出版物,如报刊、图书等,而档案馆则是各国政府文件的保存者,因此政府网站信息资源也应该由档案馆保存。

因此在本文的研究环境下,笔者倾向于“归档保存”的说法,但不排斥其他说法。综上,在本文的研究环境下,“政府网站信息资源归档保存”是指对政府网站信息资源进行采集、整理、存储和提供利用的过程,其目的是使用户能够在足够长的一段时间里访问归档网站的内容,即使原网页已不存在。

三、网站信息资源归档保存的不同模式

根据保存对象范围的不同，网站信息资源归档保存的模式可以分为以下三种情况。

1. 全面保存模式

全面保存是指没有特定的范围，网站上的所有内容全部都需要保存。这种保存模式工作量大，不可能靠人工完成，一般是通过“爬虫”和“快照”等方式定期自动抓取。典型项目包括加拿大国家图书和档案馆（Library and Archives Canada, 简称LAC）的Government of Canada Web Archive（GCWA）和瑞典皇家图书馆（Kungliga biblioteket, 简称KB）的Kulturarw³。

2. 选择保存模式

选择保存是根据一定的标准对网站进行筛选、鉴定，选择有历史价值、文化价值、经济价值和研究意义的内容进行收集和保存。这种保存模式针对性强，工作量相对较小，可采用人工与自动抓取结合的方式进行。典型项目包括澳大利亚PANDORA和美国的政府网站信息计划。

3. 专题保存模式

专题保存模式是基于某一特定的主题或者事件，将网站上所有与该主题或者事件相关的信息资源集中保存。严格来说，它是选择保存模式的一种方式。它比选择保存模式的针对性更强，可以直观地反映事件的全貌。典型项目如美国国会图书馆的密涅瓦项目（MINERVA, 全称为Mapping the Internet Electronic Resources Virtual Archive）。

目前，很多政府网站归档保存项目都包括专题保存模式。因为这种方式把关于某一事件或者主题的信息资源汇总到一起，用户不再需要在归档保存的所有网站里查找与此事件相关的信息，而是通过对这一个专题里信息的浏览与查找，即可了解事件的脉络。

四、对我国政府网站信息资源归档保存项目的启示

国家图书馆的中国政府公开信息整合服务平台其实已经是我国第一个专门的政府网站归档保存项目，其表现可圈可点，但由于其存在的问题以及与国外类似项目的差距，还有进一步完善的空间，因此笔者在本文提出一些建议。

（一）关于保存对象

1. 中央政府网站信息资源

根据分层模式，国家图书馆应收集中央人民政府及国务院部门、地方政府、驻港澳机构、驻外机构、媒体、中央企业等网站上的公开信息，并且对于重大事件应进行专题保存。所有采集到的信息经过整理、编目，为用户提供分类浏览和检索的功能，便于用户利用。

2. 社交媒体文件

目前我国运用最广泛的社交媒体有微信、微博、抖音等平台，政务微博早在2009年就已出现，近年来更是成为政府发布消息、与公众互动沟通的重要平台，许多信息都会第一时间发布到微博平台上。微信、抖音等短视频媒体虽是后起之秀，发展势头却异常迅猛，大有赶超微博之势。笔者查证，国务院、发改委、财政部等等中央部委也纷纷开通账号，推送政务

信息。笔者认为中央政府部门在社交媒体平台上的信息也应属于归档范围之内，但不是所有信息都要保存，比如：推送信息的正文及转向政府网站的链接、对用户的回复应被保存，而转发消息、指向外部的链接等则不需保存。

（二）关于责任主体

《中华人民共和国政府信息公开条例》（国务院令第711号）第25条明确规定：“各级人民政府应当在国家档案馆、公共图书馆设置政府信息查阅场所，并配备相应的设施、设备……行政机关应当及时向国家档案馆、公共图书馆提供主动公开的政府信息。”笔者认为，该条例赋予了国家档案馆和公共图书馆采集政府公开信息的权利，虽然没有明确提出政府网站信息资源，但是政府网站信息资源也应在政府公开信息范围内。网站信息资源归档保存工作如果想要顺利开展，就必须协调好政府网站信息资源的提供方、保存方和使用方等责任主体之间的关系。

1. 保存方：图书馆和档案馆等保存机构

图书馆和档案馆在保存信息资源方面有丰富的经验和优势，应当作为政府网站信息资源的长期保存责任者，其主要职责是确保被保存信息的真实性、完整性、可信性和安全性，对采集到的资源进行整理和分类，并提供检索工具，为用户提供利用。

2. 提供方：政府机构

作为政府网站信息资源的提供方，政府有责任确保其信息资源的短期保存，为保存方的采集工作和长期保存工作打好基础。此外，政府机构还应该尽可能采取标准化的格式和技术环境，向保存方提供长期保存所需的元数据，为保存方在项目启动初期提供相关数据以供测试，许可保存方的爬虫软件进入自己的网站等。

3. 使用方

政府网站的使用方一般包括公众、研究人员、政府机构自身和其他组织等。用户在利用被保存的政府网站信息资源时，会有不同的信息需求，但是都应当注意信息资源的使用权限，确保合理合法的利用。

如前文所述，不管是国外的网站归档保存项目如UKGWA项目、PANDORA项目、密涅瓦项目等，还是中国目前的中国政府公开信息整合服务平台，都不是孤军奋战，而是展开了广泛的合作，已达成资源共享、技术共享和控制成本的目的。

政府网站信息资源归档保存的重要性已经被越来越多国家的保存机构认识到。回顾我国政府网站信息资源归档保存的相关理论和实践，可以发现目前仍处于向国外学习的起步阶段，未来还需在经济、技术和法律方面做更多努力。政府网站信息资源应该被作为数字记忆的一部分，为整个社会而留存。

参考文献

- [1] 查先进. 论政府信息资源管理及其发展动向[J]. 中国图书馆学报, 2002(04): 37-39.
- [2] 何欢欢. 政府网站信息资源保存体系研究[D]. 武汉大学, 2010.