



中华人民共和国档案行业标准

DA/T 80—2019

政府网站网页归档指南

Guidelines for archiving webpages from government websites

2019-12-16发布

2020-05-01实施

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本指南由国家档案局提出并归口。

本指南起草单位：国家档案局技术部、上海中信信息发展股份有限公司。

本指南主要起草人：黄丽华、陶水龙、聂曼影、王大众、王熹、冯剑波、袁嘉新、孙源。

引 言

网站作为机构和组织信息发布和对外服务的重要窗口,形成了大量具有凭证和查考价值的网页信息内容,需要及时归档和安全保存。

为指导国家机关及其档案部门规范开展网页归档工作,促进实现网页信息的有序归档和长期保存,制定本指南。

政府网站网页归档指南

1 范围

本标准规定了政府网站网页归档的总则,网页的归档范围和网页档案的保管期限,网页归档的收集、整理、移交接收和网页归档功能模块建设的一般方法。

本标准适用于各级人民政府及其部门、派出机构和承担公共服务职能的事业单位在互联网上开办的政府网站的网页归档,其他单位的网页归档工作可参照执行。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 18894—2016 电子文件归档与电子档案管理规范

GB/T 33190—2016 电子文件存储与交换格式 版式文件

DA/T 47—2009 版式电子文件长期保存格式需求

3 术语和定义

下列术语和定义适用于本文件。

3.1

网页归档 **webpage archiving**

将网站中对国家和社会具有保存价值的网页信息进行收集、整理、移交接收的过程。

3.2

网页档案 **webpage record**

对国家和社会具有保存价值并归档保存的网页信息。

3.3

网页归档功能模块 **webpage archiving functional module**

对网页信息的收集、整理、移交接收等归档过程进行管理的功能模块。

4 总则

4.1 网页归档原则

网页归档应遵循及时性、完整性、可用性和安全性的原则。

4.2 网页归档责任主体

网站主办单位应对网页信息的生成、发布和归档实行全过程管理,确保归档网页信息的真实性和完整性。

4.3 网页归档责任划分

网站主办单位负责本单位网站的网页归档,明确责任部门具体落实网页归档的相关工作,包括:

- a) 档案部门负责制定网页归档的规章制度,指导监督网页归档工作,接收和保管网页档案,提出网页归档功能模块需求;
- b) 网站管理部门负责网页信息的收集、整理和移交等工作;
- c) 信息化部门负责建设网页归档功能模块,为网页归档工作提供信息化支持。

5 归档范围和保管期限

5.1 归档范围

网站主办单位应综合分析网页信息的重要程度和保存价值,制定本单位网页信息的归档范围。网页归档范围应包括反映网站整体面貌的网站首页及栏目首页,反映本单位职能和网站功能的信息发布类页面、解读回应类页面、办事服务类页面、互动交流类页面,以及其他具有保存价值的页面。网站主办单位制定网页信息的归档范围时可参见附录 A。

5.2 保管期限

网页档案的保管期限划分为永久、定期 30 年和定期 10 年。网站主办单位制定网页档案的保管期限表时可参见附录 A。

6 收集

6.1 收集的时间

应合理设置网页信息的收集时间,根据页面更新频率确定合理周期定期收集,当页面发生较大变化时应及时收集,包括:

- a) 网站首页或栏目首页可在展现形式发生较大变化时进行;
- b) 信息发布类页面可在网页信息发布时或更新时;
- c) 解读回应类页面可在解读回应完成后;
- d) 办事服务类页面可在事项办理完毕或服务结束后;
- e) 互动交流类页面可定期收集并在交流结束或交流主题关闭后。

6.2 收集的内容

应将网页信息收集完整,一般只收集本网站内的网页信息,包括:

- a) 网页信息收集时应完整收集页面中的可视静态内容,包括文本、静态图片等,页面中的动画、音视频、脚本、URL 等内容在保证真实性、可用性和安全性的情况下也可进行收集;
- b) 对于有附件的页面,收集时应将附件与页面作为一件同时收集;
- c) 当一项内容因篇幅较长而分为多个页面显示时,应将该项内容所对应的多个页面作为一件同时收集;
- d) 网页信息收集时应将页面的元数据同时收集,包括标题、发布时间、来源、关键词、作者、摘要、网址等,具体参见附录 B。

6.3 归档文件格式

应以符合 GB/T 33190—2016 或 DA/T 47—2009 要求的版式文件格式保存页面。页面的附件如果不能保存为版式文件,应保存为符合 GB/T 18894—2016 要求的归档文件格式。

7 整理

7.1 分类

7.1.1 网站主办单位应根据本单位的实际情况制定网页档案分类方案。

7.1.2 同一个单位应保持分类方案的一致性和稳定性。

7.1.3 同一单位不同网站的网页档案应分别分类。

7.1.4 同一网站内的网页档案可按照“年度——保管期限——类别”或“年度——类别——保管期限”进行分类。

7.1.5 同一类目下的网页档案一般按形成时间顺序排列。

7.2 编号

7.2.1 应根据分类排列方案编制网页档案的档号。

7.2.2 档号的结构宜为：“全宗号-档案门类代码-网站代码-年度-保管期限代码-类别代码-顺序号”或“全宗号-档案门类代码-网站代码-年度-类别代码-保管期限代码-顺序号”，上、下位代码之间用“-”连接。

7.2.3 档号各组成部分编制要求如下：

- a) 档案门类代码：网页档案的档案门类代码为“WY”；
- b) 网站代码：为2位阿拉伯数字，门户网站为01，其他网站顺序编号。单位只有一个门户网站时，可不编制网站代码；
- c) 年度：网页信息采集归档的年度，以4位阿拉伯数字标注公元纪年，如“2017”；
- d) 保管期限代码：永久为“Y”，定期30年为“D30”，定期10年为“D10”；
- e) 类别代码：以网页档案类别名称的首字母作为类别代码，如网站及栏目首页类为“LMSY”，信息发布类为“XXFB”，解读回应类为“JDHY”，办事服务类为“BSFW”，其他类为“QT”等，网站主办单位可采用网站栏目或根据本单位网页类别的实际情况设置网页档案的类别，设置相应的类别代码并以汉字标注类别名称；
- f) 顺序号：为4位阿拉伯数字，同一类目下从0001开始顺序编号。

示例：2017年采集归档的某单位门户网站首页，保管期限为“永久”，档号为“XXXX(全宗号)-WY-01-2017-Y-LM-SY-0001”。

7.3 数据组织

7.3.1 网页档案可采用层级文件夹的形式进行存储，在“网页档案”文件夹下，依次按不同的网站、“年度-保管期限-类别”或者“年度-类别-保管期限”建立层级文件夹。

示例：2017年采集归档的某单位门户网站下永久保存的信息发布类网页的存储文件夹为“\网页档案\门户网站\2017\Y\XXFB\”。

7.3.2 应采用档号或其他唯一标识符作为组成要素为网页档案的计算机文件命名，并保持网页档案各组成部分之间的联系。

8 移交接收

8.1 移交时间与方式

整理完毕的归档网页信息，应实时或定期以电子形式向档案部门移交，定期移交时间一般不晚于次年6月底。

8.2 移交要求

网站管理部门应清点、鉴定归档网页信息的内容及元数据的真实性和完整性,然后向档案部门移交。

8.3 接收要求

档案部门对网站管理部门移交的归档网页信息的内容及元数据进行清点和检测,检测合格后登记接收,并记录移交接收行为,完成网页归档过程。

9 网页归档功能模块要求

9.1 通则

应建设网页归档功能模块,实现网页信息的收集、整理、移交接收等功能。

9.2 归档模块功能设置

网页归档功能模块基本功能参考《电子档案管理系统基本功能规定》中的相关要求设置,并符合如下要求:

- a) 网页归档功能模块应能收集网页信息并将其转换成符合国家有关标准的文件格式进行归档和保存;
- b) 网页归档功能模块应能捕获和记录网页信息及网页信息归档过程的元数据,包括网站名称、网站域名、标题、栏目、发布时间、来源、关键词、作者、摘要、网址、采集时间、采集人等,具体参见附录 B;
- c) 网页归档功能模块应支持采用相关技术手段保障归档网页信息的安全。

9.3 归档功能模块维护

信息化部门应定期维护网页功能模块,确保功能模块的有效运行。

附 录 A

(资料性附录)

归档范围与保管期限参考表

表 A.1 给出了归档范围与保管期限参考表。

表 A.1 归档范围与保管期限参考表

序号	归 档 范 围	保管期限
1	网站及栏目首页页面	
1.1	网站首页	永久
1.2	一级栏目首页	30 年
1.3	二级栏目首页	10 年
2	“信息发布类”页面	
2.1	经济、社会、历史、地理、人文、行政区划等介绍性信息	永久
2.2	机构设置、主要职责和联系方式等机构职能信息	永久
2.3	本地区、本部门、本机构的负责人信息以及重要讲话文稿	永久
2.4	本地区、本部门出台的法规、规章、应主动公开的政府文件以及相关法律法规等	永久
2.5	本地区、本部门政务要闻、通知公告、工作动态等需要社会公众广泛知晓的信息	30 年
2.6	转载上级政府网站、本级政府门户网站发布的重要信息	10 年
2.7	政府信息公开指南、政府信息公开目录和政府信息公开工作年度报告	永久
2.8	人口、自然资源、经济、农业、工业、服务业、财政金融、民生保障等社会关注度高的本地区本行业统计数据	永久
2.9	行政权力清单、财政资金、公共资源配置、重大建设项目、公共服务、国有企业、环境保护、食品药品安全、推进社会组织、中介机构等信息	永久
3	“解读回应”类页面	
3.1	由文件制发部门、牵头或起草部门提供的重要政策文件的解读材料,包括各种形式的解读、评论、专访等	永久
3.2	公布涉及本地区、本部门的重大突发事件的客观事实,并根据事件发展和工作进展发布的动态信息	永久
3.3	相关业务部门对社会公众关注的热点问题作出权威、正面的回应	30 年
3.4	相关部门针对涉及本地区、本部门的网络谣言发布的辟谣信息	30 年
4	“办事服务”类页面	
4.1	发布本地区、本部门政务服务事项目录的网站在线服务资源清单	永久
4.2	细化的规范办事指南,包括依据条件、流程时限、收费标准、注意事项、办理机构、联系方式等信息	30 年
5	“互动交流”类页面	
5.1	网民意见建议和问题咨询的审看、处理和反馈信息	10 年
5.2	开展专项意见建议征集活动并在网站上公布采用情况的信息	30 年
5.3	针对近期社会热点和本部门重点工作开展访谈活动的音视频、图片、文字记录等信息	10 年

附录 B

(资料性附录)

网页归档基本元数据参考表

表 B.1 给出了网页归档基本元数据参考表。

表 B.1 网页归档基本元数据参考表

编号	元数据名称	英文名称	约束性	可重复性	元素类型	数据类型
M1	网站名称	site name	必选	不可重复	简单型	字符型
M2	网站域名	site domain	必选	不可重复	简单型	字符型
M3	标题	article title	必选	不可重复	简单型	字符型
M4	副标题	article subtitle	可选	不可重复	简单型	字符型
M5	栏目	column	可选	不可重复	简单型	字符型
M6	发布时间	publication date	必选	不可重复	简单型	日期型
M7	来源	content source	必选	不可重复	简单型	字符型
M8	关键词	key words	可选	不可重复	简单型	字符型
M9	作者	author	可选	不可重复	简单型	字符型
M10	摘要	description	可选	不可重复	简单型	字符型
M11	网址	url	可选	不可重复	简单型	字符型
M12	采集时间	collection date	必选	不可重复	简单型	日期型
M13	采集人	collector	必选	不可重复	简单型	字符型

参 考 文 献

- [1] GB/T 29194—2012 电子文件管理系统通用功能要求
 - [2] 国务院办公厅. 国务院办公厅关于印发政府网站发展指引的通知. 2017 年 5 月 15 日
 - [3] 国家档案局. 电子档案管理系统基本功能规定. 2017 年 12 月 15 日
-