

North Carolina Department of Cultural Resources

Standard for Automated Web Site Capture

Government agencies and entities increasingly utilize Web pages to document their actions as well as to publicize their policies and activities. General Statute 132.1 identifies public records as "...all documents, papers, letters, maps, books, photographs, films, sound recordings, magnetic or other tapes, electronic data-processing records, artifacts, or other documentary material, regardless of physical form or characteristics, made or received pursuant to law or ordinance in connection with the transaction of public business by any agency of North Carolina government or its subdivisions." (Emphasis added.) As indicated by this statute Web pages constitute public records. Additionally, Chapter 125.11 of the *General Statutes of North Carolina* states that the State Library "shall be the official, complete, and permanent depository for all State publications, regardless of format." (Emphasis added). As the statute indicates, the authority for collecting and making these publications accessible rests with the State Library of North Carolina. Additionally, GS 121-4 (3), give the North Carolina State Archives authority to "preserve and administer, in the North Carolina State Archives, such public records as may be accepted into its custody." Many government authorities utilize the Web to publish and disseminate information that they once printed on paper. This legal background forms the basis for the Department of Cultural Resources' (DCR) capture of government Web sites.

Because of its stated commitment to capture, preserve and provide access to state government Web sites, the staff of the DCR appraises Web sites to determine if they qualify as a digital asset appropriate for capture. First, potential Web sites are appraised via a set of yes/no questions to determine if DCR should collect the content [see **Determination of Which Web sites Should be Captured** below]. This first step is based on practical limitations and the stipulations of G.S. 132.1 and 125.11. Once a site is determined to be in scope, the staff utilizes the **Web Site Macro-appraisal Score Table** to assess the frequency of capture.

Frequency of Capture

DCR has endorsed automated capture of Web sites versus manual capture as the best method of collecting Web sites, because of its reliability, ease of use, more effective capture, and better retrieval mechanisms. The staff utilizes the Web site Macro-appraisal Score Table to rank Web sites for capture. (See Step II below). Depending on the outcome of these rankings, capture takes place at various intervals. DCR captures Web sites with higher appraisal scores more frequently than those with lower scores. Each site's score is re-evaluated at regular intervals. Outside influences, such as increased media interest, may cause the scoring for a particular site to be re-evaluated on an as-needed basis.

The frequency of Web site changes and revisions may vary from agency to agency. Some agencies that publish fairly static versions of policies, publications, or images may not change their Web site design for months at a time. However, agencies with more advanced services (interactive forms, streaming video, etc.) might undergo major changes numerous times each year. Recognizing these differences, DCR will capture a Web site at least once a year, and more frequently for those sites that change often. For those agencies that have a high litigation risk, DCR staff recommends that the agency audit every change to the site, cite the date that change occurred and whether or not the agency posted that change as part of the official Web site. This audit document (or log) clarifies, for legal, historical, and informational purposes, the agency's position at a particular point in time.

Determination of Which Web sites Should be Captured

STEP I: In Scope or Out of Scope

1. All Web sites selected for inclusion are sites that contain official state government information, either hosted on government Web servers or hosted by private companies working for the agency.

Example:

http://www.ncbar.com	North Carolina State Bar	State Government Web site, hosted on a state server
http://www.ncartmuseum.org/	North Carolina Museum of Art	State Government Web site, hosted on a private server

2. Private non-profit organizations and for-profit organizations are NOT collected, even if they assist some government agency.

Example:

http://www.nczoo.com/	North Carolina Zoological Society	Nonprofit organization supports zoo activities, but is separate from the NC Zoo
http://www.mulkeyinc.com/	Mulkey Engineers and Consultants	For-profit firm that contracts with the Dept. of Transportation

3. In general, captured Web sites will contain only state government information and do not contain a combination of public and private information. However, some private Web sites may be considered for capture if they contain significant state government information and assist in the formation of government policy.

Example:

http://erc.northcarolina.edu/	North Carolina Education Research Council	Provides research support to North Carolina policy and law makers; members appointed by high-level elected officials
---	---	--

4. Web sites that are aliases of the main domain address are not collected because all information on aliases would duplicate information already captured.

Example:

http://www.dor.state.nc.us	North Carolina Department of Revenue	Main domain address for state agency
http://www.dornc.com	Alias of North Carolina Department of Revenue	Mirror site or alias of www.dor.state.nc.us

Step II: Web site Macro-Appraisal Score Table

Once the staff determines that a Web site is in scope, they utilize the Web site Macro-Appraisal Score Table to assign points based on value (see table below). These points will help staff to determine the level of legal and/or managerial value and the potential historical value associated with a Web site. The risk level under which the agency currently operates determined future recordkeeping actions. Each Web site is assigned a different number of points, depending on whether it falls in the low, medium, or high value category for each of seven characteristics. Scores range from 7 to 21, with the higher score indicating more evidential, legal or historical value. However, scoring will not be a substitute for sound archival appraisal and good judgment.

All agency Web sites are included in the automated capture program unless they contact DCR to opt out of the program, or technical challenges preclude capture using automated methods. The [General Schedule for State Agency Records](#) governs records retention and disposition for record types that are common in state agencies. According to this schedule, web sites have a permanent retention with ultimate transfer to the custody of North Carolina State Archives. Agencies unable (for technical reasons) to participate in the automatic capture of Web sites should evaluate their own Web site using the [Procedures for Manual Collection of Web-Based Activities](#) and the [Web Site Content Assessment Table](#) (http://www.ah.dcr.state.nc.us/records/e_records/default.htm#web) developed by the Office of Archives and History. This method requires agencies to evaluate their site based on a legal and managerial risk assessment, and to send “snap shot” copies of their Web site to DCR for appraisal and possible archival storage. For all agencies participating in automated capture, DCR will determine whether the Web site is in scope and assign a score using the **Macro-Appraisal Score Table**.

Macro-Appraisal Score Table

	1 point – Low value	2 points – Medium value	3 points – High value
1: Size	Small Web site with few directories	Large Web site with multiple directories	Large Web site with multiple directories from multiple offices
2: Originality	Mainly copies of paper publications	Combination of original material and copies of paper material	Substantially original material not available in paper form
3: Frequency of update	Static Web site with time-insensitive material; information rarely changes	Web site updated on a regular basis; information current and fresh	Web site information changes frequently; timeliness and currency of information extremely important
4: Historical value	Information is ephemeral; little value beyond the present day	Some original information with historical value; partially documents decisions/trends of government and programs	Rich original information; statistics/facts document government decisions/trends and programs
5: Evidential value	Information never litigated and little or no legal risk	Potential for litigation	Substantial liability
6: Public interest	Low public interest; little press coverage	Medium public interest; some press coverage; potentially controversial information	High public interest; frequent press coverage; public interest groups monitor sites
7: Governmental interest	Little legislative, executive, or judicial interest	Some legislative, executive or judicial interest	Active and vigilant legislative, executive, or judicial interest