

# **Preservation Action Plan: Web Records**

## **National Archives and Records Administration (NARA)**

Plan Date: 20200629

Template: 201907

### **Electronic Record or Digital Surrogate Types and Associated Formats**

A collection of information, documents, or database that is sent from a server to a browser via Hypertext Transfer Protocol (HTTP) when a URL has been activated and meets the definition of Federal record and is provided via an agency's web site.

Component parts of web content may include other record types where it may be necessary to address the essential characteristics of those record and data types within a separate series of records. However, similar to the record type "email," when encountering another record type within web records, they are not treated as a separate record type.

### **Essential Characteristics of Web Records**

The rapid evolution of the web, the complexity of the web, and the variety of ways agencies make use of it, limits defining characteristics only in the manner in which NARA can currently accept the transfer of web content.

The presentation of web content, although it may be intended to be fixed, is most often changed or mutable, as is the content over time. The website's appearance is variable, as it is altered for the user depending on user preferences, browser characteristics, network limitations, frequency, and topic of use.

The Structure and Appearance/Layout are the most important characteristics. The manner in which elements are organized, interrelated, and displayed and can be found in one or more of the following: source code, record layout, table and frame structure, and linkage, site map, and hypertext.

Web content is either static or dynamic. The static portion of web content does not maintain additional behavior characteristics beyond hypertext/internal links. Dynamic "deep" web content is usually managed through the use of databases and style sheets, which are component parts of the web. If static, capturing the source code of web content generally will encompass content and aspects of the appearance, and structure characteristics. If dynamic, you may encounter

another record type, such as a database. Defining characteristics for these component parts should be addressed within a separate series for that component and its record type.

## Appearance

| Name            | Definition  | Function Description   |
|-----------------|---|--|
| Layout/Inline   | Inline or embedded layout and look and feel of page content. Includes but is not limited to: <ul style="list-style-type: none"><li>• Style</li><li>• Format Elements</li><li>• Class</li><li>• Heading</li><li>• List</li><li>• Table</li><li>• Form</li><li>• Canvas</li><li>• SVG (Scalable Vector Graphics)</li><li>• Client-side Script instructions</li></ul>                          | Part of the page/source code   |
| Layout/External | External file(s) that identifies layout and design elements, or are required for layout. Includes but is not limited to: <ul style="list-style-type: none"><li>• Linked objects/files to be embedded in the layout, such as images, video, audio, dynamic or static output from external applications/APIs/web services</li><li>• Server Scripts</li><li>• Cascading Style Sheets</li></ul> | External to the page, must be captured to accurately capture full content and layout |

## Structure

| Name                       | Definition  | Function Description  |
|----------------------------|---|---|
| Links to External Sites    | Significant links should be described and documented in external documentation. Linked content in a different domain should be redirected and considered an essential characteristic only if associated with the agency or externally hosted for the agency through a formal agreement. | All links are part of the page/source code.<br>Externally referenced content (e.g., accessed via hyperlink) that resides in a different domain and is not managed for an agency under a formal agreement will not be accepted for transfer. |
| Site Organization          | Logical organization of web content, including navigation, embedded objects or actions.   | Part of the page/source code and documented through an external Site Map file   |
| Schema for Dynamic Content | If Database driven, the schema for the database is required. See the essential characteristics for Database Records.  | The schema should be present because if there are serious shortcomings in descriptive and technical documentation then according to NARA guidance the web records should not be considered for transfer.                                    |

## Behavior

| Name              | Definition  | Function Description   |
|-------------------|---|--|
| Intended Function | The purpose, audience, and functionality of the site. | Accompanying documentation that describes the intended audience and functionality. |

## Context

| Name                 | Definition   | Function Description         |
|----------------------|--|------------------------------|
| Descriptive Metadata | Information contained within the record (intrinsic) that refers to the | Part of the page/source code |

|                        |  |  |
|------------------------|--|--|
|                        | intellectual content of material and aids discovery of such materials. Includes but is not limited to:<br>Title<br>Meta/Author<br>Meta/Description<br>Meta/Keywords<br>Caption/Subject/Date/Event/Transaction can all add value to the record. |  |
| Crawl/Capture Metadata | Metadata about a web capture (date, mechanism, etc.) needs to be preserved along with the website.   |  |

## Current NARA Transfer Guidance for Web Records

### [Bulletin 2014-04](#)

- Preferred:
  - None specified
- Acceptable:
  - Web ARChive Format (WARC)
  - Archive File Format (ARC)

## Current NARA Format(s) for Public Access and Reference for Web Records

Formats for Public Access are those made available online through the National Archives Catalog. Formats for Reference are defined as those made available to researchers upon direct requests for digital copies.

Format(s) Available for Public Access: Content created or delivered for public access in the Catalog is delivered primarily in the following file formats: PDF (Textual and Image), JPEG (Textual and Image), MP3 (Audio), and MP4 (Audio/Video) and ASCII (Datasets). Other file formats may be present depending on when they were added to the Catalog. For Legislative web content made available through the Center for Legislative Archives, public access is provided through <https://www.webharvest.gov/>

Format(s) Available for Reference: When available, records may be delivered to researchers in the formats in which they are preserved.

### **Comments and Notes**

Where possible, access should be provided in the native formats for the site as captured, including HTML and all associated media sites and files required for look and feel. This can be provided in a container format, e.g., WARC or ARC.