Key considerations before initiating a web archiving program

Policy considerations:

- Why do we need this program/project? What is the goal for initiating web archiving at my institution?
- Rights issues are rarely clear-cut and risk assessment is often necessary. Will we seek permission from site owners to archive their site?
- Will we provide open/public access to our web archive collections?
- What is the timeframe for initiating a web archiving program?

Scope considerations:

- What will this collection(s) consist of? Review the mission of the collecting organization and consider how a proposed collection further supports that mission.
- Identify your target audience for use of the collection and consider potential secondary uses/users.
- If you are also collecting print and non-digital materials, consider how this born-digital collection might complement those resources.
- What does *not* fit within this collection? Cite specific exclusions.

Resources/Stakeholder considerations:

- Which staff will be responsible for this work and what is their level of expertise?
- What is the budget for the implementation of this collection?
- What frequency of capture will be sustainable for these collections (or will the objective be solely one-time capture)?
- Does the organization presently use a web harvesting technology in some capacity?
- What is our capacity to curate and manage the collection during the entire lifecycle?

Sample collecting statements

To begin defining your own web archiving program, it is most helpful to start by revisiting your organization/institution's overarching mission and consider how it may encompass materials born to the web:

> The Library of Virginia's mission is to preserves the legacy of Virginia's culture and history and provides access to the most comprehensive information resources for and about Virginia. Specifically, the Library collects the archival records of the executive, legislative, and judicial branches of Virginia state government and Virginia-related materials that are of a private nature and that may assist researchers in discovering more about the history and lives of Virginia citizens. The Library's Web collection is designed to capture Web sites that mirror records that are already represented in Library's State and Private Papers collections. For more information, please click on the policies listed below... [See more from the Library of Virginia]

Specifically, consider the direct and secondary stakeholders of your collections to articulate what use or usefulness your web archives will serve and decide what belongs in your collecting "scope":

> The Archive of the Washington and Lee University School of Law Website is a project of the Lewis F. Powell, Jr. Archives. It was established in 2011 to collect and preserve the contents of the ever changing law school web presence...Researchers and law school administrators can use this collection to trace changes in publicly accessible information and to retrieve information that is no longer currently available on the web." [See more from Washington & Lee University School of Law]

These needs and uses can extend naturally from strengths in established, traditional media elsewhere in your larger collecting scope:

> This collection of blogs, social media sites, video, and organizational websites documents the international art exhibition, La Biennale di Venezia, in 2013, 2015, and 2017 on the web. The crawls began on April 28th of 2013, May 1st of 2015, and May 10th of 2017 and continued through to the end of the exhibitions in November...[T]he Venice Biennale web project complements the Clark's Venice Biennale Collection of exhibition catalogues, press kits, and ephemera beginning with the 52nd Biennale in 2007. [See more from the Clark Art Institute Library]

To keep your collecting manageably scoped and consistent, be as specific as possible about both what *does* and *does not* belong:

> This web archive contains the websites of the Center for Jewish History; its partner institutions: American Jewish Historical Society, American Sephardi Federation, Leo

Baeck Institute, Yeshiva University Museum, YIVO Institute for Jewish Research, and other organizations that occupy office space at CJH. Websites include event listings, calendars, online exhibitions, news and press releases, videos and audio from events, blogs highlighting collections at the partner institutions, digital publications, research guides, encyclopedias, conference materials and other information documenting the organizations…The web collection documents the publicly available content of the web page and does not archive material that is password protected. [See more from the Center for Jewish History]

Before you archive though, be sure to decide on a policy approach to creators' rights. Especially if you collect material from outside of your own organization or institution, consider whether a letter of permission, deed of gift, or other "opt-in" or "opt-out" steps are necessary to begin:

> …The Web Resources Collection Program follows principles and techniques of non-intrusive harvesting. We attempt to notify all organizations and/or individuals whose websites are selected for archiving. We refrain from archiving websites that do not wish to be included in this project and will remove harvested content from the archive upon request by website owner(s)." [See more from Columbia University]

Collecting statement worksheet

Using your POWRR Plan and your answers to the key questions above as context, answer the following questions/prompts in order to outline the scope of your web archiving program.

The goal of this collecting effort is to preserve born-digital materials that pertain to:

The institution(s) and/or stakeholders of this web archiving program/project are:

This program aligns with the institutional/stakeholder mission in this way:

The target audience for the use of preserved material is:

Web material will be selected/appraised for preservation by (designated staff, faculty, etc.):

This collecting effort will respect creator/owners' intellectual property rights by (select one):

- ❏ Obtaining written consent prior to acquisition
- ❏ Providing "opt-out" recourse prior to and/or after acquisition

These materials will be made accessible (publicly? on-site only?) via:

What is specifically <u>not</u> within the scope of this collecting mission (an example being social media):