

Presentation for workshop “Curating web archives: who cares for content?” session at 2014 IIPC GA

Michael Neubert, Library of Congress

The purpose of my talk is to describe the evolution of the role of curators in selection of websites at the Library of Congress, and perhaps more importantly, the evolution of the way that the curators understand what their role is.

I should clarify the terminology I will use - by “curator” I mean a staff member who has responsibilities for selection of particular subject or format for either our print or special collections – there are about 200 such individuals at LC, some with very narrow responsibilities, some with very broad ones, and some overlap. Web archiving is a format of material that has been added to the responsibilities of these curators. I will be talking about “curators” who “select” websites for acquisition through web archiving.

Another term - “collection” – at LC a web archiving “collection” has a defined scope that provides guidance for what should be selected for that collection. Today the Library of Congress has a “collection-based” approach for 99 percent of newly selected sites that we harvest – that is, almost all websites selected are associated with a particular collection.

The Library of Congress began its web archiving program almost 15 years ago. At first, it was heavily focused on national-level elections and events-as-themes, such as a collection of websites after the 9/11/2001 attacks related to that tragedy. Very few staff members were involved in selection of websites for archiving. Over time more curators from the Humanities and Social Sciences Division became involved in web archiving, in part because they worked outside their curatorial subject expertise on the labor-intensive identification of sites for national election collections. Even though relatively few curators were nominating significant numbers of sites in their subject areas, it became common to say that web archiving represented a new “format” and that curators had a responsibility to select websites in the areas of their responsibility for web archiving.

To support curators’ ability to nominate sites for archiving more freely, a new collection was created that was scoped to include any site that didn’t fit into an already existing collection – a kind of miscellaneous collection. The number of curators trained to do web archiving selection was increased – most of these newly trained curators nominated sites on subjects for which they were responsible, although not always, in this new collection.

The thinking was that by providing general support to curators to do website selections in their respective subject areas that the result overall would be a balanced and good selection of archived sites for the overall program and LC collections more broadly. Unfortunately as more and more curators were trained, it became clear that this was not true – that curators were inclined to do some minimal number of selections to satisfy a management mandate. It did seem that the results corresponded to LC collecting priorities.

For curators, a serious problem was that web archiving selection is an intensely hands-on activity while most of our acquisitions of print general collections materials is done filtering fire hose flow of materials, such as Copyright deposit of monographs. And some curators had very large areas of subject responsibility for which they couldn't realistically be expected to make balanced or sufficient selections.

Several years ago the Library reestablished its Collection Development Office and hired a new Collection Development Officer. I worked closely with this individual to refocus the selection of websites for web archiving to more closely represent institutional priorities and to align the program more closely with how curators are able to work to support the program, and to disseminate these changes to all the curators.

To this point we had not trained or involved all 200 curators in web archiving. There were two extremes among the curators with different versions in between – on the one end were curators who knew little or nothing about web archiving and on the other end were curators who knew a lot and had done many selections but who mostly felt guilty about not doing more. Coincidentally it was determined that a new general training program for curators was necessary and I was asked to include one 30 minute segment - an overview of web archiving. I used this opportunity to “reset” all the curators’ understanding of the web archiving program. I answered the following five questions:

- What is web archiving?
- What are LC’s web archiving goals?
- What responsibilities does a curator have?
- Where is there web archiving program documentation? And
- Who can answer my questions?

30 minutes isn't much to cover these questions – although you'll notice there is no discussion or presentation of web archiving tools as such. The heart of the presentation was to describe LC's web archiving program goals and to link that to an explanation of what a curator's responsibilities are – and what they aren't. What this training did was to clarify that outside of scoped collections, a curator did not have some undefined responsibility to select websites for archiving. As we shift to long-running collections that complement our traditional collecting priorities with print, curators end up with short periods of more intense activity to get these collections up-and-running and can spend less effort over time after they are established. For example, one new collection will harvest the websites of international organizations that deposit print collections at the Library of Congress. Several relevant curators worked with me to develop the collection proposal and parameters and the curators are selecting the organization sites, but once it is up and running it will take (hopefully) much less effort on their part. This is a model that both syncs with the institution's collecting more generally and also fits realistic expectations we can have for what the curators can do.

I would make two comments that are somewhat tangential to this discussion. One often hears about the ephemeral nature of websites and that a central goal is to preserve these ephemeral sites that are at risk of disappearing. It is not clear that acquiring as a priority sites at highest risk of disappearance necessarily reflects the acquisitions priorities for a national library. The LC program is now more focused

on collections of websites reflecting collection development priorities – much of which may also be at some lesser risk in a link rot or other sense, but that is not the main thing. Explaining this was part of the “reset” process with curators.

Another issue is the lack of meaningful use cases on what people do with archived sites that can be presented to curators. Most if not all curators think primarily in terms of how users will make use of collection materials that are acquired (even if they don’t think in terms of the phrase “use cases”). My strong feeling is that if as a community we had more compelling use cases to provide curators that they would identify with this work more strongly and they would be more enthusiastic.

Thank you.