Note: All images (that are not screenshots) are Creative Commons images acceptable for noncommercial use with attribution/links back to the original work.

Introduction

My name is Heather Slania and I am the Director of the Library and Research Center at the National Museum of Women in the Arts (NMWA). Today, April 1st, 2012, is my one year anniversary at the museum. Immediately upon starting I faced a number of challenges: hiring new staff, managing the Mellor Prize series of books (book publishing is an area where I previously had no experience), and migrating our ILS off a server that was not backed up and maintained as the library had been closed for three years prior to my arrival. But while I was hitting the ground running with these projects, I was also presented an opportunity: the website of our organization was being redesigned for our 25th anniversary.
What was so great about a website redesign besides rewriting all of the library pages for the new website? In my function I am not only librarian, but archivist, and before I was hired at NMWA no one was considering archiving our old website for posterity. In addition, I realized that if we were able to archive our own website I could also create a digital archive of women artists on the web. The National Museum of Women in the Arts is a small museum and we aren’t usually able to collect new media art as we don’t have many opportunities to exhibit it. Often artists will put up documentary evidence of their art, if not the art itself, on their web page.

So, I presented my case to the director of the museum that we should not only archive our own digital presence, but begin archiving contemporary women artists on the web. Clearly my case was successful because I am speaking with you all today. So with that, we purchased a subscription to Archive-It.
I first became aware of Archive-It when I was working as a cataloging librarian in the University of Maryland’s Special Collections. The University of Maryland they started archiving pages two years ago focusing on the university, State of Maryland, and historical preservation. Archive-It is a project of the Internet Archive, most famous for it’s Wayback Machine which provides access to archived websites that were crawled by the Internet Archive’s web crawlers. While the Internet Archive casts across the entire net to archive webpages, Archive-It creates distinct collections of webpages that you chose.

So now you probably have the question:
Why should art librarians collect web pages?

Introduction

If the Internet Archive is already crawling the web and archiving it, why would I care to create my own collection of archived websites? In answering this question let us see if pictures speak louder than words.
Here is Jenny Holzer’s website captured by the regular web crawlers at the Internet Archive. If we try to go to the next page, what happens?
Hrm.
Wayback Machine doesn’t have that page archived. It doesn’t seem to be available on the live web, either.
Want to search for all archived pages under
http://www.jennyholzer.com/?


There was no other image captured. So let’s look at what I was able to capture.

So this page was captured by me using Archive-It. Did it perform better?
Yes, it did perform much better. I was able to get the City Index...

... photos from each city in the index...
We can ask a lot of questions about this website to gain insight into the artist: Why did Jenny Holzer organize this series on her website? She didn’t have to put it online. However if she is working in projections, viewing the artifacts of those installations through a computer monitor and on video seem like a more interesting choice than the images being published in a book. She chose these specific images. She displays edited video of different cities. What happens if she does away with this and does something entirely different with her website? What does that say about her and this project? All this information on the Projections series might be missing, or at the very least very hard to find all together, if we were not archiving it.
If this site wasn’t captured, I think we’d be missing something about Jenny Holzer. Artists, galleries, and museums use the World Wide Web extensively, yet art librarians have not focused much attention on webpages as art or artifacts. Art webpages document artistic trends as well as relationships between artists and viewers/consumers. So I want to posit that:
Art librarians should collect web pages.

We care about electronic articles and books where the images have not been reproduced, but we haven’t been thinking about all the information that might be going missing online. If the Internet Archive isn’t able to grab all the information we consider important, then I believe that we should be working to help archive the web. So how do we go about doing that?
In the rest of my presentation I am going to cover some of the considerations and challenges that one has to recognize and deal with in order to take on this project. We will go over the more conceptual considerations first by looking at the Curatorial Issues of using Archive-It. We will then look at some of the technical and other challenges in using Archive-It.
Curatorial Issues: Who Should Archive?

Curating and collecting digital archives has ties to curator, librarian, and archivist so almost anyone could be involved with this project. If you are in a type of institution which has all three positions and all of you are interested in this type of project: consider yourself lucky. I asked our modern and contemporary art curator if she was interested in helping me select digital artworks/artists available online but she was not interested. However, the first people interested in preserving art and art history on the web were curators, so you might find some interest with curatorial staff. If you don’t, don’t let it discourage you from collecting web pages as it is important information that isn’t being saved.

So let’s say you realized you are interested in collecting web pages but then how do you decide what to collect? I don’t want to worry you...
Curatorial Issues: What to Archive?

But trying to archive the web is like looking at a donated collection of billions of items that you need to select a few items out of to fit in your collection development policy. You must do some serious conceptual thinking about what you would like to collect before you begin curating your collection.

So what should you collect?
The sky’s the limit in what you can collect. We collect women throughout history all over the world, so I could be collecting pages dealing with 19th century French women artists or First Nations artists. But because of limited funds I focused on artists who are most likely to use the digital medium in their art or who is creating art that is hard to capture in books. If your focus is Russian Architecture, perhaps you want to collect YouTube videos of people walking through famous buildings in St. Petersburg. Or if your institution is interested in miniature art, there are many societies and associations of miniature art, some of whom have online exhibitions. At the very least you should be thinking of archiving your institution’s website and social media, but I am positive that there are useful art resources on the web that you should begin collecting now.
For my project I have concentrated my collection on women conceptual and new media artists. Since there are so many, I primarily focused on women included in the books *Internet Art* by Rachel Greene and *Digital Visions: Computers and Art* by Cynthia Goodman. In addition, I selected a few conceptual artists as Roni Horn. I am interested not only in the art they create, but how they represent themselves, and how they are treated in the market place. With that I selected three collecting areas within women conceptual and new media artists.
Collecting: Artist Webpages

First I collected artists’ webpages. Artist webpages provide insight into their work, even when the webpage medium seemingly constrains artistic statements. For those artists working in digital or video art, the web provides an essential display of rarely shown works.
Collecting: Net Art

As a subset of artist webpages I collected some Net Art. One can argue that most Net Art is not very good, and that artists, especially those outside of digital/electronic or conceptual art, are not utilizing the web to its fullest extent. As Steve Dietz said in his essay “Why have there been no great net artists?” historical net art “[i.e.] art that is older than five minutes- more often seems outdated than historic.” But if the Internet Archive isn’t able to grab all the information we want from net art without human intervention, we should collect despite its seeming outdatedness. First it is useful to capture artists in the beginning stages of coming to grips with new tools and challenges. These artworks, even the early crude ones, illustrate different ways artists began thinking of this new medium and incorporating interactivity into their art. Many artists working in net art were playing with interactivity in their art before they started working
online, and capturing the digital versions goes part way to show how the interactivity works as an artistic concept.
Collecting: Gallery Webpages

Second I collected gallery webpages. An artist's webpage is often extremely different from the profile pages created by representing galleries. The dialogue between gallery webpages and artists' personal webpages highlights tensions between art and market, and additionally showcases trends in marketing. If you are a museum focusing on a particular time period, it might be useful to archive the sites that specialize in that kind of art. If you are in an education institution, it would be interesting to document how different galleries are, or are not, using the web to sell art.
Finally I focused on women artist organizations and associations. The internet is a powerful tool for organizing and showcasing political messages. I wanted to track how different women artist related organizations utilize the internet to illustrate and promote their agendas.

Again, this is how we chose to scope our collection: selected artist web pages including art, gallery pages, and women artist organizations. But if this kind of project really interests you, I suggest you take time and scope really carefully before beginning. There are enough technical issues to think about without worrying “Should I really archive it?”
There are a number of challenges to using Archive-It. I will cover some of the more technical issues first including crawler traps, databases, and flash/javascript.
Crawler Traps

One technical issue you often come across is the Crawler Trap. When the crawler goes in to your site, some things, like calendars, send the crawler down a black hole. Here’s an example that we came across.
If you click on each page on the calendar, it’ll generate a new page: regardless if there is content. This begins at March 2012.
This illustrates July 2012 even though NMWA’s new website will be up by then.
We jump to April 2013 where ARLIS/NA will be meeting in Pasadena, but NMWA will be completely utilizing our new website by then.
And we can continue to April 2333 and beyond if we wanted. As you can see, these are pages that will generate new pages indefinitely. This is why before actually crawling a site you should complete a test crawl. This will make sure that you are actually grabbing what you really want to grab and not information that is not relevant. In test crawls a report is generated for the captured urls and you’ll be able to notice if something seems odd. You don’t want to blow your entire budget on capturing a calendar.
There are web pages, like crawler traps that you don’t want to capture, but then there are many web pages that you want to capture but are very hard to capture.
Databases

First, let us look at databases which are hard to capture. Let’s try to fill out this form on the archived site like we would have in the original.
We can try searching for a Canadian artist working in Body Image/Body Size.
But when we press search we can’t get at that information. However, there is a way around it. It doesn’t exactly mirror the original but most of the information can be saved.
If a database doesn’t have an ability to browse it through links, you can often create a search in the database and archive that search—which is what we had to do in this case.
You can then visit the collection page, and using the same search terms, search through the collection to find artists who would have the search form information on their page.
Even though the form itself didn’t work, we were able to get at the information by utilizing the full-text collection search Archive-It provides. Even though the website didn’t work the exact same way it does in its live version, we still retained most of the information in the form it would have eventually been presented to us. This is not always the case.
Lynn Hershman Leeson has a gorgeous site which has animations between each section. We can move to each section...
Image: lynn hershman.com

... look at interesting artwork...
... and view video, pictures, and other artifacts of the art if not the art itself. However, when I tried to archive her website:
This is all that I was able to archive. Because of the flash on her website, the web crawler doesn’t know how to grab the information. You can often tell the problem webpages because their web addresses don’t change even when you’re clicking and looking at different material.
We were able to go through the back end and capture some material.
But we only captured a small portion of Lynn Hersman Leeson’s website, and not any of the look and feel of the website—important components for future art historians.
Websites with interactive menus also have problems archiving. This is a well-designed website that would be wonderful to capture.

Image: http://www.cheimread.com/artists/jenny-holzer/
Challenges: Finding the Time

How do you find the time to archive? We all have busy schedules and most people who use Archive-It have similar workflow problems. The reason why Archive-It works better than the regular Internet Archive crawlers is because of human intervention. You must schedule regular hours each week to work on your Archive-It collection in order for it to be effective. It takes considerable time to scope your collection, to work through the problems that you find in your test crawls, and to add all the metadata. I spend a couple hours each week working on our collection.

If you feel organized and dedicated to creating your collection, your question is likely...
Challenges: Finding the Money

How do you find the money to create the collection? If you are at a large university, chances are your institution already subscribes to Archive-It. Look to your Special Collections and/or Archives department (the likely creators of your university’s collection) and ask them if you could archive websites related to your area of study. If you have art related archives and special collections, it would probably be best to find websites that complement your existing collections.

If you are at a smaller institution, I suggest offering to archive your organization’s web presence. Often times, saving your institution’s history is supported by administration. If your institution is interested in using Archive-It to save its own history, it will be easier to suggest archiving other items that fit your mission. So this is my mission for you.
Look at your mission.

Find web pages.

Save them!

Read through your mission at your institution and see if there is anything that should be saved on the web. Then go out and save them. By disregarding the utility and significance of collecting contemporary art websites, art institutions leave important art and artifacts to disappear without record. Your own institutional history might go missing as well.

Plus, wouldn’t you want to document an important website before it ceases to be?
Image: nmwa.org (old site)

Or before it changes from this...
Image: nmwa.org (new site)

... to this.
Additionally, the Artist Files Special Interest Group is now trying to work with the Internet Archive to determine how we can collectively archive artist information. It seems unrealistic for artists to stop using flexible tools such as flash/javascript to create their webpages. We must work with developers as well so that art is able to be archived. There will likely be technology changes along the way to allow archivists to archive web pages that are now hard to capture. However, partnering with developers in the first place so that they can help solve the problem before it begins seems to be a useful step forward.

Thank you! Email me to learn more.

hslania@nmwa.org