Online Art Ephemera:

Web Archiving at the National Museum of Women in the Arts

Heather Slania, National Museum of Women in the Arts

Abstract—Artist and art subject file collections contain important primary source ephemera for art historical research—but what happens when the ephemera are online? The National Museum of Women in the Arts has been web archiving art-related online ephemera using the Internet Archive’s Archive-It since November 2011. This case study presents the considerations and challenges of archiving such types of material and provides a foundation for arts institutions to begin more collaborative web archiving.

[Int this article is a revised version of a presentation given at the “New Voices in the Profession” session at the ARLIS/NA Conference held in Toronto, Ontario, in April 2012.]

INTRODUCTION

Artist and art subject file collections contain important primary source ephemera for art historical research. However, a large percentage of contemporary ephemera is not being collected. Art-related materials are constantly uploaded to the web by museums, galleries, and artists, but they have not received much attention thus far from art librarians. The phrase “online art ephemera” refers to websites as well as the information contained within a website in any format, including video and audio. Art librarians are certainly interested in collecting online art ephemera, but questions regarding ideal methods and appropriate funding may stop many from doing so. Despite the popular notion that everything on the Internet lasts forever, websites disappear from existence on a daily basis. Studies have found that there is a half-life of four to five years for web citations in peer-reviewed journals.1 Even websites that are

captured by the Internet Archive\(^2\) may not be rendered properly, often revealing inconsistencies in format or appearance from their original designs. If institutional missions prompt librarians to collect primary source information for future art historical research, there is a gap when web material is not prioritized in current collection development plans.

Although some researchers have paid attention to the challenges of web archiving in general, few have investigated web archiving in an art-specific context. The use of dynamic content, for example, is of critical importance to art librarians. Websites with dynamic content utilize scripting languages (such as Flash and JavaScript) to create interactive experiences for each user accessing the page. This is in contrast to basic HTML websites which do not require additional code to render the page properly. Websites that employ dynamic content are difficult or impossible to archive by standard methods, either automatically by the Internet Archive or manually through the use of services such as the Internet Archive’s Archive-It.\(^3\) Capturing the way that a page is rendered visually is one of the most important—and difficult—components in saving online art ephemera. Of the eighty-three URLs captured for this study, 21 percent had an extreme loss of functionality or were completely nonfunctional. Since visual information is as important to capture as textual information, this failure rate is unacceptable.

An additional web archiving issue that art librarians must address is the sheer volume of material. There is duplication between vertical file collections, and some duplication may be warranted in online web archives, but it is not an ideal practice in a centralized online research environment. Web archiving is challenging even for well-funded institutions due to the amount of material—this is too large a project for any one institution. Duplications of effort are not efficient from either a time or financial perspective, especially in the age of declining budgets. Unprecedented collaboration between institutions in collection development will be necessary so that expenditures are kept to a minimum while building comprehensive ephemera collections.

Previous authors have addressed two components of this topic: web archiving and artist files. The literature on web archiving generally covers the broader issues facing the web archivist, but often specializes in the copyright, technical, and selection problems that web archivists face. The largest question on copyright is whether or not web archivists should allow rights holders to opt in or out of web archiving, either through active or passive means. Web content is under copyright, and there is currently no specific exemption for web archiving in US law. The technical problems focus on the difficulty of capturing websites that utilize dynamic content. Additional technological problems include the eventual storage, migration, and emulation issues once current technology becomes outdated. The selection issues are also not trivial—proper scoping is vital to achieve the best results. Of these three challenges, the copyright problem has the most potential for a solution, and has, in fact, been

\(^2\) Internet Archive, http://www.archive.org, is a non-profit digital library that has archived over 150 billion websites.

\(^3\) Archive-It, http://www.archive-it.org, and other web archiving services allow institutions to build, manage, and search their own web archives without needing to host the web archiving technology themselves.
solved in many countries outside the United States. The literature on art-specific vertical files is slim and barely recognizes the need to expand into online ephemera. However, websites do fit the definition of ephemera in each article. Despite the omission of websites from the literature, collecting online art ephemera is a natural extension of the original vertical file collection’s purpose.

Against this background, the purpose of this case study is to explore the considerations and challenges in archiving online art ephemera. More specifically, this article has two objectives:

1. To explore the extent of current web archiving limitations on art-related websites.
2. To create a foundation for future collaboration among art librarians in evaluating, selecting, and describing online art ephemera.

This case study from the National Museum of Women in the Arts (NMWA), a small, private non-profit art museum in Washington, DC, is an example of how arts institutions large and small can approach web archiving. NMWA archived eighty-three websites relating to contemporary women artists. In order to fit within budget, time, and document limits, this website collection was further refined into three components: individual websites created by artists working in conceptual or new media art; profiles of women artists represented by contemporary art galleries who have interest in new media and conceptual art; and organizations of women artists.

The process of scoping, selecting, crawling, and describing this collection uncovered a variety of problems and factors that must be addressed.

**LITERATURE REVIEW**

Web archiving art-specific information is a topic still in its infancy. However, the component parts, web archiving and art ephemera (vertical files), have been written about extensively. The issues discussed in earlier works are still relevant today, despite technological advancements.

**WEB ARCHIVING: MAIN CHALLENGES AND CONSIDERATIONS**

Writing in 2002, Peter Lyman outlined four challenges that are found in web archiving: cultural problems (the public is unconcerned with archiving while in the present), technical problems (how to archive and properly render the digital object), economic problems (who is responsible for doing the archiving and with what money), and legal problems (copyright).4 Stemming from these problems, Lyman offered three questions about collections, authenticity, and preservation that further explicate the issues that web archivists face (Table 1).5 Lyman’s problems and ques-

---

5. Ibid., 41.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognizing need</td>
<td>Problem: Cultural</td>
<td>Organizational challenges</td>
<td>Current climate</td>
</tr>
<tr>
<td>Cost</td>
<td>Problem: Economic</td>
<td>Costs</td>
<td>Organizational issues</td>
</tr>
<tr>
<td>Copyright</td>
<td>Problem: Legal</td>
<td>Legal challenges</td>
<td></td>
</tr>
<tr>
<td>Technical</td>
<td>Problem: Technical</td>
<td>Technical challenges</td>
<td></td>
</tr>
<tr>
<td>Selection</td>
<td>Question: What should be collected?</td>
<td>Approaches to the collection of websites (automatic harvesting, selective capture, deposit, combined)</td>
<td>Collection development concerns (selection, unit of selection: content versus context, acquisition)</td>
</tr>
<tr>
<td>Authenticity and Description</td>
<td>Question: How do we preserve its authenticity?</td>
<td></td>
<td>Collection development concerns (authenticity, metadata, presentation: authenticity indication)</td>
</tr>
<tr>
<td>Long-term Preservation</td>
<td>Question: How do we preserve or build the technology needed to access and preserve it?</td>
<td>Collection policies and coverage, Software, Long-term preservation</td>
<td>Collection development concerns (organization, presentation: look-and-feel, preservation)</td>
</tr>
<tr>
<td>Access</td>
<td>Access policies</td>
<td></td>
<td>Needs addressed by Web Archives (persistent access to scholarly materials, provision of value-added services, persistent access to institutional repositories, additional user needs)</td>
</tr>
</tbody>
</table>
tions form the basis of current discussions on the main challenges and considerations found in web archiving projects.

Michael Day’s 2003 presentation closely follows Lyman’s analysis, while adding the importance of developing proper access measures for online web archives. Nearly five years later, Kathleen R. Murray and Inga K. Hsieh outlined the Web-at-Risk project—a project initiated in 2005 to begin a new web archiving service. Through the assessment of this project, they discovered many of the challenges and considerations Lyman discussed. Their findings were drawn from surveys with curators, focus groups with librarians/archivists, and interviews with end users/content providers. The challenges and considerations discussed in each article are compared in Table 1. No one article touches on every issue; however, together they provide a complete framework for beginning web archivists.

While the three articles focus on the considerations and challenges found in web archiving, Adrian Brown’s Archiving Websites provides the most comprehensive guide to creating and managing a web archiving program. This 2006 book is essential reading for anyone attempting a web archiving program, even in 2013. While Brown covers every topic the three articles mention, he is the most in-depth in regard to selection and preservation. The section on copyright is specifically geared towards UK law and is several years out of date. The collection methods chapter might be useful for those needing more education on how websites are built and captured, but it does not thoroughly explain the different options organizations have when choosing an archiving tool. Overall, the book is a practical how-to guide with case studies that illustrate each step of the web archiving process.

WEB ARCHIVING: COPYRIGHT PROBLEMS

Among the writers mentioned above, Lyman discussed the importance of the copyright/legal problem most thoroughly. It is a continued concern among web archivists. Web archivists in the national libraries of Australia and the Netherlands have been at the forefront of this issue and serve as good examples of the different styles of copyright permission that may be sought from rights holders (Table 2). In 2010, Lachlan Glanville compared the copyright measures taken by those countries in their web archiving. Australia uses the opt-in style, which is popular with many large research institutions. With the opt-in style, rights holders must give their permission to archive. The sites that do not give permission are either not archived or are saved but are not available to the public. The Netherlands uses an opt-out (with deadline) style of archives selection: archivists send rights holders a message explaining their intention.
tion to harvest and give a deadline for response in order for the rights holder to avoid harvesting. Finally, there is the full opt-out style that does not provide rights holders with any notification. Here, rights holders may be excluded only by contacting the body in charge and requesting that their sites be removed from the archive. This style is most notably used by the Internet Archive.

While many institutions in the United States and elsewhere have been wary of following Internet Archive’s bold opt-out strategy because of its potential risk, recent court rulings suggest that this is not as risky a behavior as was originally believed. In 2012 Jonathan Band, distinguished intellectual property lawyer and Georgetown University law professor, summarized two rulings (Perfect 10 v. Amazon.com (2007) and A.V. v. iParadigm (2009)) and two opinions (United States Patent and Trademark Office Opinion (2012) and Association of Research Libraries Code of Fair Use Best Practices (2012)). The two court cases reaffirmed that web archiving is transformative and therefore does not violate copyright. In light of these cases, the USPTO and ARL recommend using the opt-out strategy for archiving web materials. While US institutions that do not use the opt-out style may be waiting for a change to the US Copyright Act to specifically cover web archiving, it does not seem necessary to delay in light of these recent decisions and opinions.

WEB ARCHIVING: TECHNICAL PROBLEMS

Technical problems have also been targeted in the web archiving literature. Two 2011 articles point to the technological difficulties that occur with web archiving. Abbie Grotke, web archiving team lead at the Library of Congress (LC), admits that the library does not have the capacity to do all of its own web archiving and instead relies on a contractor on the West Coast to perform the crawling. LC collects four to five

---

Table 2. Copyright permission options for web archiving

<table>
<thead>
<tr>
<th>Style of Copyright Permission</th>
<th>Description</th>
<th>Example Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opt-in</td>
<td>Rights holders must give permission to archive.</td>
<td>National Library of Australia</td>
</tr>
<tr>
<td>Opt-out (with deadline)</td>
<td>Rights holders must respond that they do not want their site archived before an assigned date.</td>
<td>Koninklijke Bibliotheek (Netherlands)</td>
</tr>
<tr>
<td>Opt-out</td>
<td>Rights holders may contact archive to take out websites from archive.</td>
<td>Internet Archive</td>
</tr>
</tbody>
</table>

---

13. Ibid., 1-8.
terabytes of data per month, requiring around-the-clock technical expertise. Additionally, she recognizes that web technologies are constantly being revised and created, which means the crawler is always playing catch-up. In light of this, LC seems pragmatic—the institution is not trying to keep up with the technology problems itself, but instead is focusing on selection and access issues.

While there is consensus in the general web archiving literature that certain types of web technologies are difficult to crawl, only one article thus far has specifically probed the issue of dynamic content and web archiving. Megan Sapnar Ankerson compares the difficulties found by broadcast historians to those that future new media historians will find. Flash is her example of problematic software because the source code of Flash sites is not fully available to be archived like the coding of HTML pages. It is difficult to archive anything from these sites, meaning that the only documentation left might be a website’s mere existence. Not only will historians not know the content of the websites, but the inability to document the use of new web technologies from their inception is a loss as well. Ankerson’s main solution is to advocate for crowdsourcing preservation efforts and for web historians to work with cultural institutions to save web history. She does not, however, offer any suggestions to interrupt web technologies’ creation cycle by trying to influence web developers to consider building archiving functionality into their sites.

WEB ARCHIVING: SELECTION

The copyright and technical problems will likely find their solutions outside of libraries; however, selection will certainly be one of the main issues that art librarians will face. Each cited article mentioned that selection is necessary for web archiving, whether by subject, creator, genre, or domain. Even the Internet Archive, the only archive that does not engage in some form of selection, does not actually archive everything.

Brown’s Archiving Websites is again the most thorough work on selection and collection considerations. He discusses the processes, methods, and criteria used in selection. In addition to outlining the types of collections web archivists can choose to build, he stresses the importance of how much, when, and why to archive websites. For example, websites featuring rapidly changing news stories should be captured at a different frequency than rarely updated websites on less at-risk topics.

Also of note is Tracy Seneca’s 2009 update on the Web-at-Risk program highlighted earlier. She describes the program’s development of collection plans based on more traditional collection development. The specific problem not typical in traditional collection development is that it is impossible to know what will be collected until after a website has been archived. This is a particular concern for those organizations using web archiving services that need to keep within data and document budgets.

18. Ibid.
Within her article on the problems presented by Flash, Ankerson provides an important note on selection. Specifically, she cautions against taking documents, audio, video, etc., out of their home websites, in case there will not be enough evidence left to properly contextualize digital objects. Additionally, she reminds the reader that since the web is ephemeral in nature, “decisions concerning what to archive are also decisions concerning what not to archive,” which will only serve to highlight the viewpoints and motivations of decision-makers. Written from a perspective of a historian rather than web archivist, she provides a useful viewpoint that is not found in most of the web archiving literature.

**VERTICAL FILES**

Art-related vertical files, currently still bastions of paper-based material, have thus far not crossed paths with born-digital content, except perhaps when CDs and DVDs are placed directly into files. However, now that art ephemera, like exhibition brochures and announcements, are created for the web, often without any corresponding print documentation, it seems only a matter of time before virtual artist and art subject files are created. Clive Phillpot’s 1995 article does not touch on the web as a medium of ephemera production since most people obviously were not thinking about this in 1995. However, he argues for the collecting of ephemera for its primary importance and as a starting place for art historical research. While paper art ephemera still exists, online art ephemera is increasingly taking its place and should be saved in its original form—electronically.

Some of the more recent literature on artist files still neglects to recognize the web as an ephemeral medium. Even while discussing how to digitize artist file material, Terrie L. Wilson and Erika Dowell’s 2003 article does not consider born-digital material. Despite overlooking the web, their only defining criterion for artist files is that they contain ephemeral material. Even when web ephemera are mentioned directly, the issue is still not given full consideration. Javier DoCampo and Rosario López de Prado’s 2001 article on exhibition ephemera lists websites as one of the formats that are ephemeral. While recognizing the need to collect this material, they do not discuss its potential treatment, focusing instead on more traditional material types. The art library community is beginning to discuss web archiving but has not yet considered the issue as thoroughly as other research communities.

**WEB ARCHIVING AT THE NMWA**

The web archiving program at the National Museum of Women in the Arts (NMWA) was not meticulously planned. A website redesign was well underway when the author, the director of the Library and Research Center, was hired (the library had been

---

21. Ibid.
closed for four years prior to her arrival). There were no plans at the time to archive the previous version of the museum’s website, making it necessary to quickly establish a web archiving program to save the institution’s own digital history. This led to an opportunity for the library to begin a small study on archiving women artists on the web.

Internet Archive’s Archive-It was the only serious web archiving service considered because the author was already familiar with this product through its use at the University of Maryland.25 The website selection process was much more difficult. Determining the right websites to archive out of thousands of potential sites was, and still is, challenging. Given the document and data limits dictated by the subscription, three points of selection criteria were chosen within an overarching theme.

The broad theme for the collection is contemporary women artists. While there are many websites about other women artists throughout history, the web is more likely to host documentary evidence for contemporary artists. Among contemporary women artists, the most likely to be using the web are new media and conceptual artists. Additionally, new media and conceptual art are difficult to collect and are therefore not strengths of NMWA’s collection. While the library collection closely mirrors the artists, topics, and works from the permanent collection and exhibitions, this was an appropriate opportunity to further the library’s mission in facilitating knowledge creation about the history and achievements of inspirational women artists worldwide. Since these works are not a strong point of NMWA’s collection, the artists selected were those featured in the books Internet Art26 by Rachel Greene and Digital Visions: Computers and Art by Cynthia Goodman.27 In addition, a few conceptual artists whose work is in NMWA’s collection were chosen for their use of the web.

In order to collect the online art ephemera of new media and conceptual women artists, NMWA targeted the galleries of these artists as well as their personal websites for archiving. In addition to the two components of gallery and artist websites, a third component—women artists’ organizations—was added under the broader theme. Because there are not many organizations that focus directly on women artists, the websites of organizations and artist groups such as the Guerrilla Girls provide important documentary evidence in tracing women artists’ movements.

In total, eighty-three websites were originally marked to be archived in the Contemporary Women Artists on the Web Collection (http://www.archive-it.org/collections/2973). Crawls were originally archived on an annual basis, as most of the websites seemed as though they did not change on a day-to-day basis, and there was uncertainty over how much data would ultimately be archived. Like the Internet Archive, NMWA did not ask for permission to archive websites. The opt-out style of copyright permissions was used, and robots.txt files were respected.28

25. For a comprehensive list of on-demand and web harvesting crawlers and initiatives, Wikipedia’s entry on web archiving is the most comprehensive: http://en.wikipedia.org/wiki/Web_archiving. Besides Archive-It, the Web Archiving Service from the California Digital Library is one of the more popular on-demand services.
28. Robots.txt files allow website owners to inform web crawlers that their site is not open to be crawled.
Analysis of Findings

The project of web archiving was much more challenging than originally anticipated; however, the end result was extremely worthwhile. Test crawling, a method for viewing the URLs that would be archived given the current crawler parameters, was an integral part of the process. A test crawl can show whether web archivists are archiving too much or too little content and provide insight into the means of strengthening a crawl—all without affecting document or data budgets. After running the test crawl, a report is generated listing exactly which URLs would be archived, what type of digital objects would be included (video, PDFs, etc.), and whether any of the site was blocked by robots.txt files.

Test crawls were run dozens of times on the collection. Problems discovered in the test crawl stage were often solved to satisfaction, and the majority of information was able to be captured with most of the context remaining. Some problems encountered in the creation of this collection were solved by looking at the test crawls and changing the scope of the crawl before the collection was archived. Two examples of this were crawler traps and databases.

The NMWA website itself had a crawler trap. A crawler trap occurs when a webpage generates an infinite number of links, in this case a dynamic events calendar that generates the next month of the calendar by clicking the “next” button on the webpage, whether or not any events are scheduled. Because the crawler is unable to tell that there is no real content on the webpage, it will try to go to the next page for as long as the program is running. Without performing a test crawl, a web archivist will not know that a web crawler is capturing the organization’s events all the way through year 4567 and beyond. In this case, constraints were added to the collection so that the calendar was skipped over by the web crawler (events information was captured elsewhere on the website).

Databases provide a different dynamic content problem for web archiving. This problem was discovered when trying to archive the WomenArts Network, a database of women artists from around the world. The crawler is unable to query the database when trying to archive a website. However, a stable URL was generated when using the web form to search the database. This made it possible to capture the content inside the database that was missed in the test crawl. Additionally, the technology behind the databases is not archived, so as of this writing, it is not possible to use an archived search form to access the captured information. There are, however, potential ways around this problem. Unlike the Internet Archive’s Wayback Machine, where users can only search a known URL, Archive-It is full-text searchable on a collection level. To find the material in the database, a user can search the collection using controlled vocabulary terms from the form. Using this method, we had success replicating results from the live database with the archived database.

Despite these successful instances of solving problems using test crawls, it may not be possible to fix some dynamic content problems. Even with test crawls, helpful suggestions, and work by the Archive-It team to solve the problems discovered, technology problems remained significant throughout the archiving of this collection (Table 3).
Problems with web archiving became apparent after running crawls on the websites—many used dynamic content that could not be archived fully. Of the eighty-three websites, 21 percent had an extreme loss of functionality or were completely nonfunctional when archived. The numbers are even worse when considering the loss of video and images—a full 42 percent are missing important documentary evidence of art. While there are currently no comparison figures for similar collections, based on conversations the author has had with other web archivists, these seem like extraordinarily high numbers and may be specifically related to archiving contemporary art content.

Of the artist websites, 67 percent were not able to be fully rendered when archived, with only 33 percent of websites completely functional. Lynn Hershman Leeson’s website is a good example of what a fully dynamic website looks like after archiving (Figure 1).29 The live site currently has moving animations, video, and images; however, nothing could be archived from a normal crawl.30 While some information was

---


---

Table 3. Dynamic content archiving problems by severity of problem

<table>
<thead>
<tr>
<th></th>
<th>Artist (33 artists, 36 websites total)</th>
<th>Gallery (6 Galleries, 42 websites total)</th>
<th>Organizations (4 Organizations, 5 websites total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completely functional</td>
<td>12 websites</td>
<td>9 websites (1 gallery)</td>
<td>1 website</td>
</tr>
<tr>
<td>Some loss of functionality—broken links</td>
<td>2 websites</td>
<td>23 websites (3 galleries)</td>
<td>1 website</td>
</tr>
<tr>
<td>Moderate loss of functionality—video/pictures not loading</td>
<td>14 websites</td>
<td>–</td>
<td>3 websites</td>
</tr>
<tr>
<td>Extreme loss of functionality—video/pictures not loading plus broken navigation/links</td>
<td>4 websites</td>
<td>10 websites (2 galleries)</td>
<td>–</td>
</tr>
<tr>
<td>Completely nonfunctional</td>
<td>4 websites</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total problem websites</td>
<td>24 websites</td>
<td>33 websites</td>
<td>5 websites</td>
</tr>
<tr>
<td>Percentage of archived websites not completely functional</td>
<td>67%</td>
<td>79%</td>
<td>80%</td>
</tr>
<tr>
<td>Percentage of significant problems (moderate to non-functioning)</td>
<td>61%</td>
<td>24%</td>
<td>60%</td>
</tr>
</tbody>
</table>

---
eventually captured by using a directory, the images are completely decontextualized.31

Gallery pages had several problems as well. Because NMWA focuses specifically on women artists, at first it seemed most logical to archive only those portions of gallery websites specifically pertaining to women artists. Unfortunately, only one gallery website was constructed in such a way that the crawler could find all relevant information on the artist. Three other galleries needed to be archived to a fuller extent in order to archive all relevant information. Ankerson’s warning about being very selective about content is relevant here: it may be that next time the NMWA collection plan is reviewed, gallery websites will be either archived in full or not at all.32 Even with that issue, two galleries could not be properly rendered even when selecting the entire website due to their dynamic navigation.

Organizational websites were the least likely to have significant problems; however, 60 percent of the archived websites still had a moderate loss of functionality. While there is not enough data to analyze with certainty why this is the case, organizations may be more hierarchical and conservative on a structural level which might cause them to use less cutting-edge technology when creating their websites than artists or galleries. It should be noted that private, password-protected sections of

websites were not counted as problematic in these findings—private webpages currently cannot be archived.

The problems found are significant, and the reality of archiving changed some of the original collection parameters. The website capture frequency changed after monitoring how much space each archived website filled. Websites originally intended to be captured on an annual basis were moved to monthly, and the only one captured on a monthly basis, jodi.org, was switched to daily archiving. It will be an ongoing process to determine the best frequency to capture websites in the collection.

Maintaining a small collection such as this one requires only a few hours a week after the collection is created. The largest problem in initiating the project was finding the time to get it started, especially in an institution where staff time is stretched. Crawls often take days to complete, at which point each crawled website must be searched for problems. For test crawls, this can involve going over hundreds of thousands of URLs to determine if the crawler is catching the right material, and parsing URLs for problem patterns. If a robots.txt file is found, additional time is needed to contact the site administrator to request that the archive crawler be exempt from the site’s rule. In many cases, website owners are not thinking of web archiving as something they intend to exclude by using the robots.txt file and are happy to allow web archivists to save the material.

Still in process for NMWA is the metadata for the collection. Archive-It utilizes the Dublin Core metadata element set to describe records at both a collection level and at a website level. Since this is mostly an artist-based collection and full-text searchable, the URLs currently do not have much individual metadata attached beyond the artists, organizations, and medium represented. As the collection is used more and standards are developed, it is certain that more metadata elements will be added to increase access.

The most important discovery overall was that this project could take place at a small institution. Having a strong reason to begin web archiving is imperative. In general, most institutions respond favorably to allocating money to preserve their own histories. Adding an additional project to an already-existent one to maximize the expenditure was seen as cost-effective, even if it did mean that more funds were needed to accommodate the additional project.

**FUTURE DIRECTIONS**

The use of dynamic content and cutting-edge web technologies will be a problem that will plague art web archivists as long as there are web developers and artists attempting to create new and groundbreaking platforms. While the Internet Archive is working constantly at improving the crawler’s ability, it would benefit from engaging in massive outreach efforts to web developers and programmers. By reviewing examples of web technology history that have gone missing, perhaps web developers will be more inclined to build the concept of archiving into their new technologies.33 Whether or not artists even want their art web ephemera saved (and may be choosing

33. Ibid.
it as a medium because it is difficult to preserve) is another question entirely, and one that is worthy of future discussion.

While this project did not begin as a means to encourage collaboration among institutions, it quickly became apparent that such collaboration was vital to success. The most challenging aspect once more arts institutions begin the project of web archiving is making all art web archiving initiatives interoperable. The first step is to select an appropriate space to store the information. It may be tempting to create a web archiving service specifically for online art ephemera; however, for a small art library, it would be impossible to do this kind of work if any advanced technology were needed in-house. Besides, if it is difficult for the Library of Congress to keep up with its own needs for web archiving, it is unlikely that many other institutions would be able to shoulder the technology issues. As an institution whose main mission is to preserve the web, the Internet Archive has the best resources and is the web archiving service most likely to solve the important technology problems. These are not trivial matters—as technology shifts, one must keep in mind the mistakes made with floppy disks, videotapes, and DVDs and not repeat them. Once a web archiving service is selected, interested art librarians will need to convince their parent institutions to engage in web archiving. Provided that web archiving is in accordance with the organizational mission, the proposal will likely be well met. If the proposal includes archiving the institutional web history of the parent institution, including social media if relevant, it is even more likely to succeed. In large institutions it might be challenging to determine exactly who in the institution should be in charge of web archiving. Librarians who manage artists’ files may be the best individuals to manage online art ephemera as well due to their long-standing experience collecting the breadcrumbs of history.

Interested art librarians and archivists will also need to engage in collaborative collection development planning. There is too much material needing to be collected to create highly duplicative collections. Brown discusses several types of themes that might be worthwhile: Subject matter (e.g., folk art, textile art, miniature art), Creator (e.g., galleries, New York artists), Genre (e.g., blogs, net art), or Domain. Currently, domain is not a useful selection category for art; however, if the .art domain is widely adopted when it is available it might be of great help to art web archivists. Some duplication will occur between theme types, but the more specific and localized these collections can be, the better.

Along with an efficient way to discover which websites are being captured across arts institutions, access tools for users to easily find the archived information are equally important. Currently, only individual collections are available for full-text searching in Archive-It. In order to make them most accessible, all collections, regardless of web archiving service, should be searchable together. In one search, users should be able to determine which artists, galleries, museums, and art topics have been archived and by which institution. Librarians and archivists also need to adopt a

uniform way of describing the collections for access. Whether traditional library or archival methods of description and access are used, or whether new methods are created—it is imperative that institutions work together.

CONCLUSION

The recent past is quickly disappearing, difficult to capture, and of enormous size. To save this increasingly important documentation of the art world, art librarians and archivists, web developers, arts institutions, and artists need to work together. While the proliferation of dynamic content brings up vexing issues for those wishing to archive the art web, there are many people and institutions outside of art librarianship who are concerned with this problem. Selecting the online art ephemera to save and determining the best way to access it are perhaps the gravest concerns facing art librarians on this topic.

The literature has proven that web archiving is transformative and fits under fair use. Copyright issues should not stop art librarians and others from web archiving. While copyright problems persist with digitizing modern and contemporary paper material, public born-digital material is safe for US librarians to archive without securing permission from individual site administrators. In order to complete this work with available resources, processes must be as streamlined as possible. While the ARLIS/NA Artist Files Working Group is making steps towards dealing with online art ephemera, the author hopes this article will support the process by providing a foundation and a vision towards the future.