

[Mission](#)[Editorial Committee](#)[Process and Structure](#)[Code4Lib](#) **Search**

Issue 54, 2022-08-29

Archiving an Early Web-Based Journal: Addressing Issues of Workflow, Authenticity, and Bibliodiversity

SWITCH is a journal of new media art that has been published in an online-only format since 1995 by the CADRE Laboratory for New Media at San José State University (SJSU). The journal is distinctive in its commitment to presenting scholarship and criticism on new media art in a visual format that reflects and enhances its engagement with the subject. This approach, which includes the practice of redesigning the journal's platform and visual presentation for each issue, raises significant challenges for the long-term preservation of the journal, as well as immediate issues related to indexing and discovery. This article describes the initial stages of a collaboration between the Martin Luther King, Jr. Library and the CADRE Laboratory at SJSU to archive and index SWITCH and to host a copy of the journal on SJSU's institutional repository, SJSU ScholarWorks. It will describe the process of harvesting the journal, share scripts used to extract metadata and modify files to address accessibility and encoding issues, and discuss an ongoing curricular project that engages CADRE students in the process of augmenting metadata for SWITCH articles. The process reflects the challenges of creating an authentic version of this journal that is also discoverable and citable within the broader scholarly communication environment. This effort is part of a growing multi-institutional project to archive the new media art community in the Bay Area in a 3D web exhibition format.

By Nick Szydłowski, Rhonda Holberton, Erika Johnson

Introduction

SWITCH, a journal of new media art, has been published in an online-only format since 1995 by the CADRE Laboratory for New Media (CADRE) at San José State University (SJSU). In 2021, the Martin Luther King, Jr. Library at SJSU began to work with CADRE to archive the journal in SJSU ScholarWorks (<https://scholarworks.sjsu.edu/>), the university's institutional repository, which is hosted on the Digital Commons platform. SWITCH was one of the first online academic journals to embrace the visual possibilities of the World Wide Web (WWW). The first electronic-only scholarly journal, Postmodern Culture, was founded five years earlier, in 1990 (Amiran, 1991), but, like most of the earliest electronic journals, it was initially published in a text format that did not allow for graphics or specify the layout or design of the page. The year SWITCH was published, the Association of Research Libraries (ARL) compiled a print directory of all online serials, which was not limited to academic journals or publications in English, listing only 675 journals and newsletters of any type (King and Kovacs, 1995). The majority of these were published via email or gopher – less than half used the WWW. Going back a year to 1994, the directory lists only 36 journals with URLs (Mogge, 1999). SWITCH was truly a pioneering publication.

The process of archiving SWITCH presents a particular set of challenges due to the journal's status as an early electronic journal coupled with its emphasis on design and visual presentation. One unique feature, referenced in the journal's title, is the fact that each issue of the journal showcases a new visual design and layout. Nearly all of the journal's issues are still live on the web in their original formats, and the evolution of the journal represents a capsule history of web design strategies and site structures. However, these original versions are not visible to researchers as part of the broader scholarly literature, or easy to find using search engines or library discovery tools. Because of the unique features of each issue, the journal can be challenging to browse and cite. For these reasons, it was very desirable to produce a version of the journal that would be easy to discover, browse, and cite, while maintaining as much of the authentic experience of the journal as possible.

This article describes the development of a process to archive the journal in a PDF format and to create metadata for each article. We have automated as much of this process as possible using a modular, multi-step workflow. Each step of the workflow is described separately, and while the entire workflow might be most applicable to journal-based projects, several of the modules can and have been used in targeted web archiving and in the archiving of digital humanities and other web-based scholarship projects. Moreover, the collaborative nature of the project is an example of how libraries and other cultural heritage institutions can work more directly with communities of practice in a digital environment.

The workflow also includes stages which are very much not automated, but instead use the archiving process to create unique curricular opportunities for students in the CADRE program. Students who are involved in the creation of current issues of SWITCH have worked in class to write abstracts and create additional metadata for specific articles in the archived journal. These assignments have been incorporated into classes, allowing students to gain information literacy skills and understand the role of metadata in scholarly research, to encounter challenging scholarship in their field, and to bring their own experiences as scholars and artists to bear in making that scholarship discoverable. The students' work is incorporated into the presentation of the archived journals, and students are publicly credited with their work on the collection's metadata.

The archiving of the journal and the integration of the process into the CADRE curriculum are both part of a larger project to archive the new media art community in the Bay Area in a 3D web exhibition format. While the larger project involves a number of additional institutions and partners as well as a different technical approach, this initial project is an important step in creating and modeling active collaborations between the community of new media artists and cultural heritage institutions, and in building capacity to preserve complex digital artifacts that represent the early days of art on the WWW.

A Very Brief History of SWITCH

SWITCH was founded in 1995, only two years after the WWW made it feasible to create and share web pages using a graphical layout. The journal was edited by CADRE students under the supervision of artist and CADRE faculty member Joel Slayton. Figure 1 shows the homepage for the first issue of SWITCH, on the theme of Information on the Internet.



the electronic art journal of the C.A.D.R.E. Institute
in the School of Art and Design at San Jose State University



CADRE

Figure 1. Issue #1 of SWITCH as it was originally published.

Screenshot from http://switch.sjsu.edu/archive/switch_v1n1/switch.html.

The first 16 issues of the journal each feature unique visual presentations and appear to consist of hand-coded HTML and CSS. With the 17th issue, published in 2001, the journal was relaunched on a platform which presented all of the previous issues in a standard visual format. That platform was only used for one additional issue (issue 18, on the topic Interface: Software as Cultural Production) before the journal returned to the tradition of introducing a new visual design for each issue. However, the re-launched versions of issues 1-16 came to be the most accessible and discoverable versions of those issues. The earlier versions were so difficult to find that it was only through the archiving process that our team became aware that the original versions of those issues were still live on the SWITCH server. Figure 2 shows version 2 of the first issue of SWITCH. Version 2 appears to be based on a Linux, Apache, MySQL, and PHP (LAMP) architecture. While in version 1, each article consisted of one or more HTML files stored directly on the SWITCH server, for version

2 the SWITCH editors stored the contents of each article in a MySQL database, and the HTML of the page was generated via PHP when the user requested the page. While each issue has its own header image in version 2, the format of each issue and article is much more consistent than in the original version.



Figure 2. Version 2 of Issue #1.

Screenshot from http://switch.sjsu.edu/archive/nextswitch/switch_engine/front/front.php%3Fcat=5.html

The initial decision to begin the project by archiving the 18 issues on the version 2 platform was made before we became aware that version 1 of the journal was still live on CADRE's server, and it offered the primary advantage of providing a large set of structurally similar files on which to work out our automated workflow. We will discuss some of the other advantages and disadvantages of focussing on version 2 later in this article.

Process and Workflow

The process used to archive issues 1-18 and ingest them into SJSU ScholarWorks is as follows:

- First, HTTrack software (<https://www.httrack.com/>) was used to create a localized archive of the publicly available SWITCH server
- Second, scripts were used to address accessibility and encoding issues, simplify HTML structure, and extract basic metadata from each article
- The resulting simplified HTML files were batch-converted to PDF using Adobe Acrobat
- After the PDF files and basic metadata were published in SJSU ScholarWorks, CADRE students augmented the metadata with abstracts and additional description, added alt text to images using Adobe Acrobat, and audited the extent to which version 2 articles faithfully represent the original versions

We will discuss each of these steps below, with particular attention to methods that may be applicable to other attempts to perform targeted web archiving.

Harvesting SWITCH to a local copy

The project team created a local copy of SWITCH using HTTrack, a GPL-licensed offline browser created by Xavier Roche and other contributors. While there are many methods available for crawling the web, HTTrack may be attractive to library practitioners due to its straightforward and stable graphical user interface and its options for controlling the behavior of the web crawler. The tool is largely absent from the library literature, but one member of our team has been using it in library practice for over a decade.

We began our crawl on the SWITCH archive page (<http://switch.sjsu.edu/archive/archives/index.html>) and allowed HTTrack to follow links on the site to archive the contents of the SWITCH server. This approach yielded a local archive of HTML and other files with a folder structure that reflects the page URLs. Under this approach the articles from version 2 of issues 1-18 could all be found in a single local folder.

For this project, the local copy was of particular use due to the complex and sometimes ambiguous URL structure of the SWITCH site. Put simply, the server contains over 25 years of web development by CADRE students and faculty. Creating an offline copy allowed team members to browse the site through the file browser of their local operating system, providing a different view which clarified questions about the journal's history and content. This method also effectively transforms the output of content management systems and similar database-backed sites into a set of HTML files, allowing for subsequent transformation and archiving.

Modifying harvested content and extracting metadata

Browsing the offline copy revealed 237 articles that had been published on the version 2 platform. The team chose to focus first on those articles first in order to establish a workflow that could be customized further to address the later issues. Examining those 237 articles suggested three goals for the next phase of the workflow: addressing encoding problems in the version 2 articles, taking steps to enable the creation of accessible PDF files, and extracting metadata. All of these goals were addressed primarily using the Python library Beautiful Soup 4 (<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>) using a working copy of the archived site.

Encoding issues

Visual inspection of both the archived and live versions of the version 2 files revealed a large number of missing special characters when the files were displayed in a web browser. After troubleshooting, it became apparent that while the header of the version 2 pages indicate text encoded according to the ISO-8859-1 standard, the text that represents the articles is actually encoded using the Windows-1252 standard. This was likely an artifact introduced during the process of creating version 2 of the journal. This was addressed by rewriting the files while replacing the encoding text. The very simple code below was used to address these issues. [1]

```
1 | filedata = filedata.replace('UTF-8', 'windows-1252').replace('iso-8859-1', 'windows-1252')
```

When the files specified the correct encoding, web browsers were able to render the special characters correctly.

Accessibility issues

The version 2 site presented two categories of accessibility issues which, unless addressed, would be carried over to any subsequent versions. Firstly, the layout features many areas with insufficient contrast between the text and background. Secondly, all images lack alt text.

The contrast issues could be addressed in an automated fashion by lightening the background of the content area and darkening text for certain HTML elements and classes. The initial plan for image alt text was for CADRE students to add it to the PDF file for each article at the same time that they wrote the abstract for the article. Adobe Acrobat provides a reasonable set of tools for adding alt text to PDF files, but testing that process on initial batches of SWITCH PDF files revealed that the complex HTML of the original files resulted in a PDF file that was incompatible with Acrobat's accessibility tools. SWITCH's version 2 platform, launched in 2001, uses nested tables to execute page layout, as was typical at the time. The article content is contained in a single cell of a table, which is the child element of another table, which is itself the child of a third table. A simplified and schematized version suggests why these pages might present challenges for tools like Adobe's that are expecting modern web page structures:

```
1 | <table><! -- page layout table -->
2 |   <tbody>...<tr><td>
3 | <table><! -- content area table -->
4 |   <tbody>...<tr><td>
5 |     <table><! -- v2 database output table -->
6 |       <tbody>...<tr><td>
7 |         <! -- article content html-->
8 |       </td></tr></tbody></td></tr></tbody></td></tr></tbody></table>
```

In order to produce PDF files that could be made accessible using the available tools, the team chose to rewrite the HTML files to produce files that are visually identical to the original version 2 pages, but which use CSS and inline styles rather than tables to produce the layout. The contrast and page structure adjustments were combined into a single script which rewrites the files to a new folder. A simplified version is included below – the full script is available at <https://github.com/NickSzydowski/switch/blob/main/rewrite-accessible-colors.py>.

Because the version 2 articles follow a consistent structure, it is possible to extract the unique content from the page programmatically. Article content and issue-level navigation were displayed in table cells with the classes "DBOutput" for the article "specialObjs" for the navigation. The first three images on the page are also retrieved, as they make up the header image which changes from issue to issue. Outside of these five elements, the remainder of the page is the same from article to article and can be replicated with greatly simplified static HTML.

```
1 | article = soup.find(class_='DBOutput')
2 | special = soup.find(class_='specialObjs')
3 | images = soup.find_all("img")
4 | image1 = images[0]
5 | image2 = images[1]
6 | image3 = images[2]
```

As an example, here is the code used to add the header images to the new simplified file:

```
1 | <div id="issueImage" style="background-color:#444444; clear:both; height:200px;">
2 | {secondImage}
3 | {thirdImage}
4 | </div>
```

The original HTML is too complex to easily reproduce. In it, these adjacent images are at the top two separate cells of a single table row. The special navigation is below the first image, in the same cell, and the article itself is below the second image, again in the same cell. In this case, simplifying the HTML structure of the page had accessibility benefits even when converting from HTML to PDF. A similar process could be used to extract the article content from the page in order to create a new HTML or XML version of a web-based publication.

The same script addresses the accessibility issues present due to inadequate contrast on the version 2 site. Because the original site uses a mix of inline styles and CSS, it was simplest to use a multi-pronged approach to addressing styles. Here, as a last resort, background colors that were repeatedly specified in inline styles are replaced directly.

```
1 | # accessibility changes - make changes here where inline styles and other issues make CSS difficult
2 | output=output.replace('bgcolor="#666666"', 'bgcolor="#888888'')
```

These changes maintain the look and feel of the version 2 site while facilitating improved accessibility. Figure 3 shows screenshots of each version.

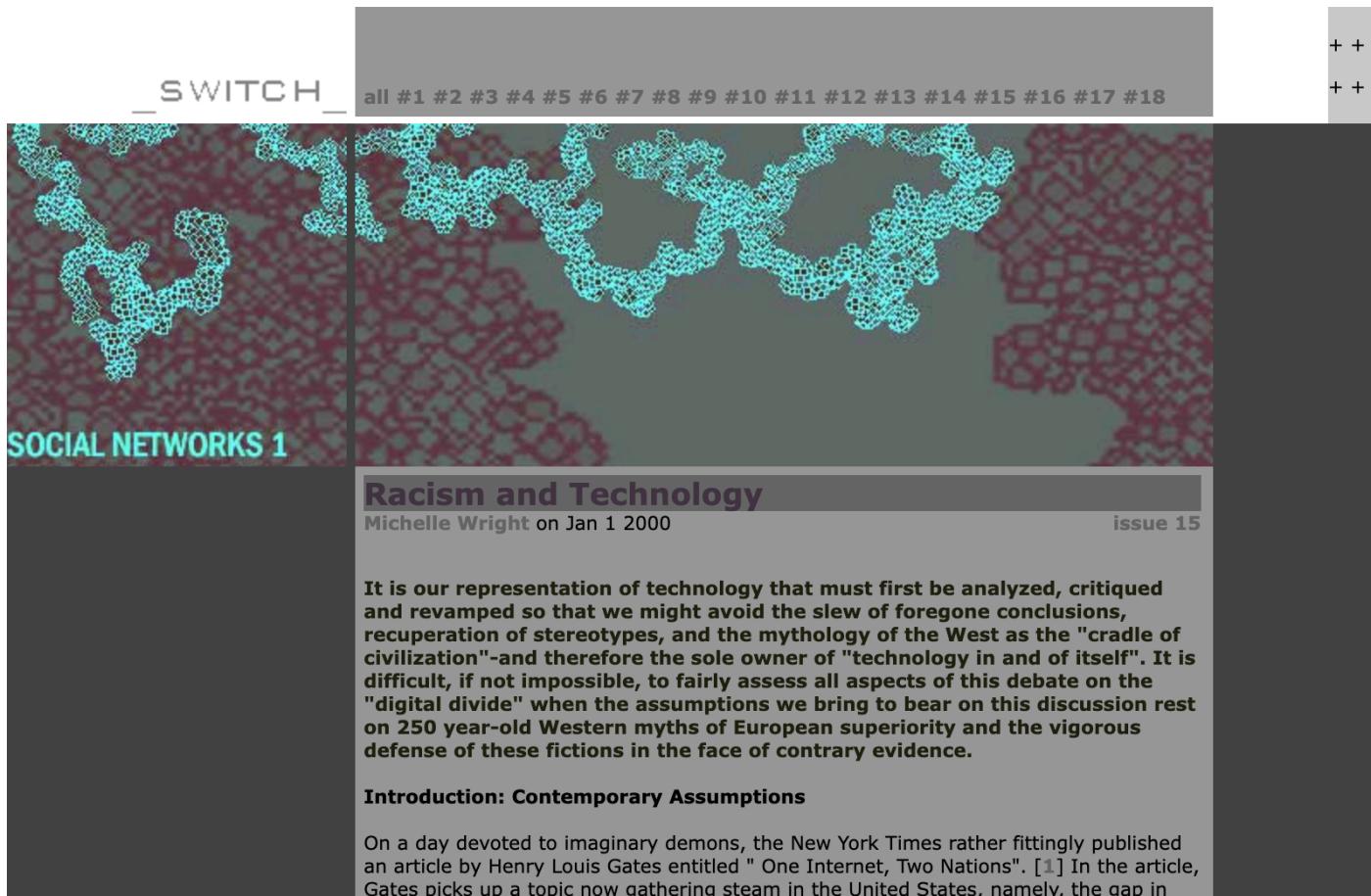


Figure 3. Comparison of original and archived article.

Screenshot from version 2 http://switch.sjsu.edu/archive/nextswitch/switch_engine/front/front.php%3Fartc=30.html

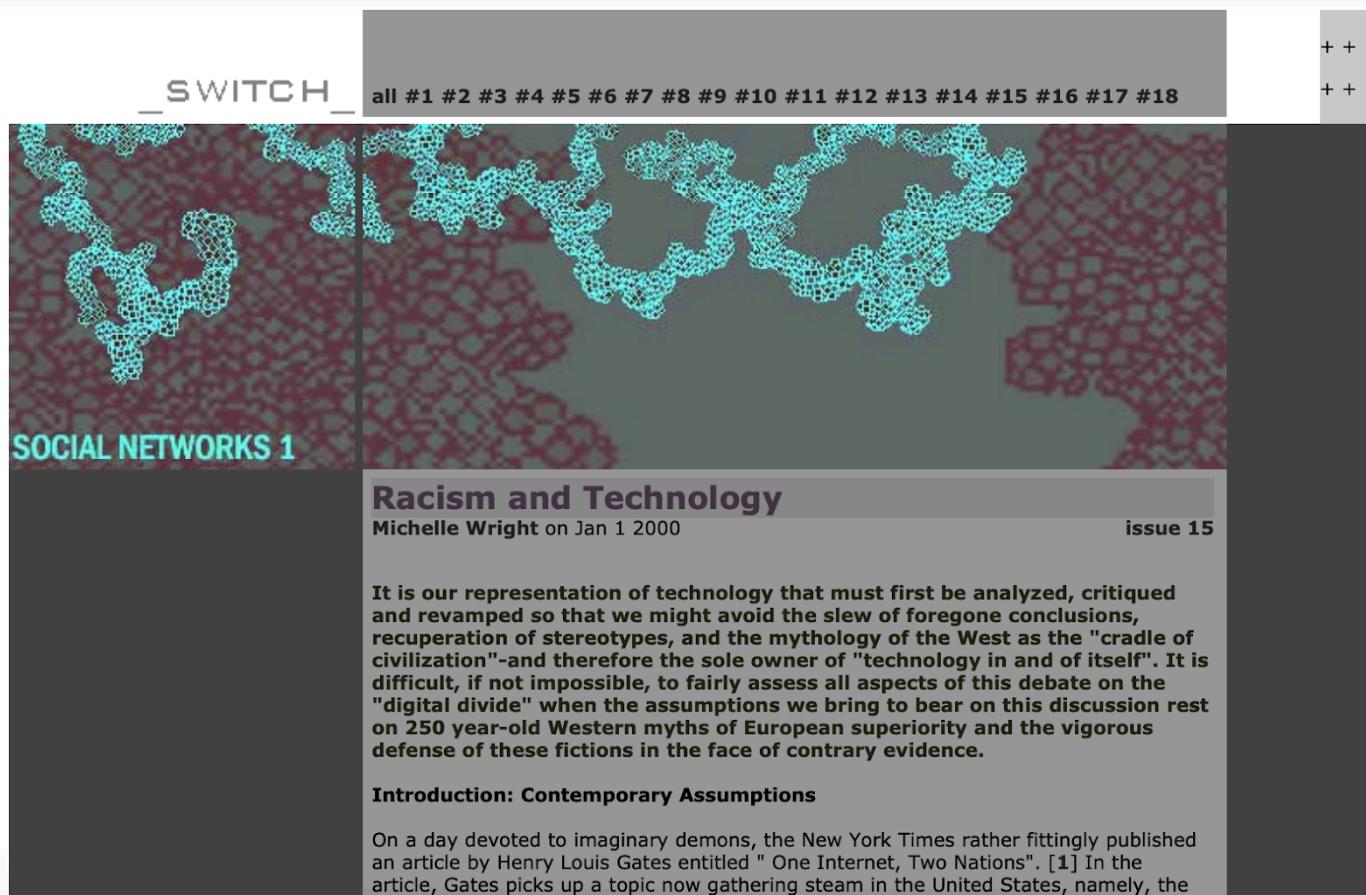


Figure 3a. Comparison of original and archived article.

Screenshot from archived PDF. <https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1118&context=switch>. Note increased contrast for accessibility.

Web archiving approaches that are suitable at scale may struggle to address accessibility issues that are common in older websites. Additionally, users of older sites may agree that character encoding problems leading to missing characters or the garbled text, sometimes called mojibake, are a particularly common issue for older sites. (McNulty, Alvarez, and Langmayr, 2021) The practices described here show some of the benefits of a more hands-on approach to archiving early web content, in particular when that content is of importance to a community of practice or an institution. There are clear parallels here to the preservation and conservation practices that libraries apply to print materials: materials of local and community importance warrant hands-on conservation treatment as needed, while the much larger set of “general collection” materials can be treated with practices that are capable of scaling up to larger bodies of material.(e.g. Boice, Draper, Lunde, and Schwalm, 2017).

Metadata extraction

The final script that was executed against the set of files used BeautifulSoup to parse out the article-level metadata available for each article and write that metadata to a CSV file. Available fields were title, author, issue number or issue description, date published, and filename. This module is an example of how close attention to the structure of these older files makes it possible to parse out data that is otherwise difficult to acquire programmatically. The pages do not include metadata in the header, and few elements have unique class names or IDs. The code used to acquire the issue and date information is included below.

```
1 | issue =soup.find(class_='DBoutput').find('table').find(align='right').find('a').get_text(":", strip=True)
2 |
3 | date =soup.find(class_='DBoutput').tr.next_sibling.next_sibling.a.next_sibling.replace('\n', '').replace('\r', ''')
```

Even this minimal metadata, when added to a modern journal hosting platform, dramatically improved the discoverability of individual articles, as well as the journal as a whole. Upon uploading to Digital Commons with minimal metadata, articles were almost immediately indexed in Google Scholar and easier to find on Google.

This information also provides a structure for the remaining steps of the project, in particular the curricular component which gives CADRE students an opportunity to augment this metadata and grapple with the challenges of archiving this content.

Curricular component and metadata augmentation

Students in the CADRE class ART 104 – Interdisciplinary Seminar in Digital Media Art are involved in the creation of new issues of SWITCH and engaged academically and artistically with the themes addressed in the journal. Being directly involved in archiving past issues of the journal is an opportunity for students to gain information literacy skills by learning more about how scholarship is created, shared, and archived. In particular, the challenge of writing abstracts for individual articles helped students understand how metadata influences the discoverability of scholarly articles.

Students each chose three articles to work on over the course of the semester. For each article, students wrote abstracts and assigned subjects and keywords, added alt text to images, and researched the author or authors of the piece. Students also located the article on the original SWITCH site and

compared the archived version to the original. In many cases, students identified missing content or other issues that were introduced during the creation of version 2 in 2001. Of the 98 articles students examined in Spring 2022, 36 had missing content from version 1 to version 2, with the most common issue being articles where only the first few paragraphs of the original text were reproduced in version 2. This occurred when an article was published over multiple URLs in the original version – many articles were laid out with a few paragraphs on one HTML page, with a next button or other navigation that brought the reader to a new URL for the next part of the text. Even in 2001, this pattern seems to have caused problems for the editors of SWITCH as they migrated to a new platform.

We expect to address this problem by adding the additional text to the downloaded HTML files and regenerating the PDFs of these articles. This is an example of the flexibility afforded by taking a more hands-on approach to digital preservation – while this stage of the process is likely to be labor intensive, it will restore the complete text of the articles that has been missing from the most visible version of the journal for two decades.

Conclusions

As part of a pilot project for a much larger preservation effort focused on the digital art community in the Bay Area, we hoped that this project would help to establish ways of collaborating that could carry over to the broader project. Additionally we believe some of the lessons from this project may also be applicable to similar contexts, especially those where relatively small research libraries, or other smaller cultural heritage institutions, aim to work with creative communities whose work involves technology. At our own institution, we believe the library's work with the digital arts community provides a template for efforts to support digital scholarship in other fields.

Challenges

There were some unanticipated challenges that we think will be instructive to those who archive web-based art and scholarship. To be straightforward, the closer we looked at either the original SWITCH site, the 2nd version of the journal, or our archived surrogates, the more problems we found – problems with authenticity and missing content, problems with functionality and discoverability, problems with accessibility, and so on. This is not unique to SWITCH by any means. Both our previous experience with web archiving and existing studies (Ayala, 2020) suggest that these challenges persist across web archiving projects.

For our context, the lesson here is that, for content that is critical to the history of a particular institution or community of practice, we should expect to expend significant resources addressing these types of issues during the archiving process, and we should expect that hands-on, bespoke interventions will be required to make that content accessible to new users. In the context of the digital humanities, this project is particularly instructive. In SWITCH, we have an example of web-based digital scholarship that is over 25 years old. The level of intervention that is required to archive this type of material is significant and is not likely to be reduced as the technical complexity of the scholarship increases.

Importantly, for this project, faithfulness to the published version of the journal was not sufficient to preserve and provide access to the content of the journal. This is not only because of the extensive missing content in version 2 of the published journal, but also because the published journal did not meet current accessibility standards or provide the necessary metadata or structure to make the publication visible as a journal in the modern scholarly communication environment. Web archiving at scale necessarily limits its goals to an accurate reproduction of the published site, but in this case that approach would not meet the goals of either the library or the community of practice that created the journal.

More broadly, this project points to the need to consider the value of bibliodiversity to the scholarly communication environment, and the need for libraries to adopt an approach to technology that supports bibliodiversity. (Shearer, Chan, Kuchma, and Mounier, 2020) SWITCH is distinctive among scholarly journals in ways that include its integration of subject matter and presentation and its close connection to a particular community of artistic practice, but it is challenging to retain the journal's distinctive features while also increasing its visibility in the comparatively rigid world of scholarly publishing. This project points to a general approach for preserving both the visual and structural characteristics that make a journal like this distinctive. In particular, the close involvement of practitioners and students in the process of archiving the journal has produced a more authentic version of the publication than the library could have achieved on its own, while making the journal more visible and accessible to users.

In response to these challenges, project participants adopted several specific practices that we hope to carry forward and apply to future projects:

- Flexible balance of hands-on and automated work
- Making archived work visible and legible to the broadest possible audience
- Close connection to communities of practice, including students

Flexible balance of hands-on and automated work

This project began with the intention of automating as much of the process as possible through web crawling and scripting. The initial aspiration of the project was to archive version 2 of SWITCH as it was published, but the input of various collaborators pointed the way toward a more ambitious project. Making the archived journal accessible required manual creation of alt text as well as more complex scripts to address contrast issues and simplify the HTML structure of the harvested pages. Student involvement created much richer metadata, grounded in the discipline, and also highlighted issues with version 2 which will require significant hand-work from the library, but will also make the archived SWITCH a much more complete record of the journal.

Hands-on digital preservation work is often addressed from a position of scarcity, with the implicit assumption that the volume of potential digital content is simply too great to allow labor-intensive intervention on individual objects or files. (Corrado, 2022) While library resources are necessarily limited, this project demonstrates how partnering with communities of practice can introduce a different dynamic, both by helping to clearly identify high-value material and by facilitating collaborative labor that can supplement work done within the library.

Making archived work visible and legible to the broadest possible audience

This project treats SWITCH as a journal, and in some cases it imposes a more conventional structure on the most experimental articles in the publication. While a traditional web archiving approach would capture the experience of the original publication more authentically – and would be an appropriate complement to this project – it would not make the journal's contents easily discoverable and browseable by a general audience.

This is another example where close collaboration facilitates a more interventionist approach. Migrating the journal to a very different context involved numerous small decisions. On this project library workers and CADRE faculty and students were able to discuss and work through those decisions, including

during in-class discussions and workshops. This process allowed members of the digital arts community to have consistent input, shaping the archiving process to fit the expectations of the journal's audience.

Close collaboration with communities of practice, including students

As the previous examples underscore, this project revealed both expected and unanticipated benefits of close collaboration between communities that create content and institutions that take on the responsibility of archiving that content. The Reference Model for an Open Archival System (OAIS) defines the designated community of a digital archive as:

An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time. (CCSDS, 2012, 1-11)

This case study suggests an alternate understanding of an archive's designated community – as a community of collaborators in the archiving process, as well as one of potential consumers of archived information.

For this project, the collaborative model influenced decisions large and small. This is an area where approaches and methods developed in the field of community archives might help inform the work of libraries who are active in archiving and facilitating digital scholarship created by faculty and students. (Caswell, 2020).

Finally, this case study suggests that deeply collaborative partnerships between communities of practice and cultural institutions may be a viable model for digital preservation practice. It is important to expand the number and types of institutions that can practice complex digital preservation (Blumenthal et al., 2020), and we believe this model is particularly valuable for smaller and less resourced libraries. This approach opens up opportunities to integrate digital preservation and digital scholarship practices into the curriculum, which are particularly valuable at teaching-intensive academic institutions. Finding sustainable digital preservation models for small institutions is critical to maintaining the diversity and representativeness of archived digital collections, including the bibliodiversity represented by early electronic journals like SWITCH. We hope that this case study helps to demonstrate the feasibility of this work in a variety of contexts.

About the Authors

Nick Szydłowski is the Scholarly Communications & Digital Scholarship Librarian at the Martin Luther King, Jr. Library at San José State University.

Rhonda Holberton is an Oakland-based interdisciplinary artist and Assistant Professor of Digital Media at San José State University.

Erika Johnson is the Institutional Repository and Digital Scholarship Coordinator at the Martin Luther King, Jr. Library at San José State University.

Endnotes

[1] Partial code is included for explanatory purposes. The complete code for the project is available at <https://github.com/NickSzydłowski/switch>.

Sources

Amiran E, Unsworth J. 1991. Postmodern culture: publishing in the electronic medium. *The Public-Access Computer Systems Review* 2(1):67-76. Available from: <https://www.ideals.illinois.edu/handle/2142/201>.

Ayala BR. 2020. Correspondence as the Primary Measure of Quality for Web Archives: A Grounded Theory Study. In: Hall M, Mer?un T, Risse T, Duchateau F, editors. *Digital Libraries for Open Knowledge*. Vol. 12246. Cham: Springer International Publishing. (Lecture Notes in Computer Science). p. 73–86. doi: [10.1007/978-3-030-54956-5_6](https://doi.org/10.1007/978-3-030-54956-5_6).

Blumenthal K-R, Griesinger P, Kim JY, Peltzman S, Steeves V. 2020. What's wrong with digital stewardship: evaluating the organization of digital preservation programs from practitioners' perspectives. *The Journal of Contemporary Archival Studies* 7. Available from: <https://elischolar.library.yale.edu/jcas/vol7/iss1/13>.

Boice J, C. Draper D, B. Lunde D, M. Schwalm A. 2017. Preservation for circulating monographs: Assessing and adapting practices for a changing information environment. *Technical Services Quarterly* 34(4):369–387. doi: [10.1080/07317131.2017.1360640](https://doi.org/10.1080/07317131.2017.1360640).

Caswell M. 2018. Affective bonds: what community archives can teach mainstream institutions. In: Flinn A, Bastian JA, editors. *Community Archives, Community Spaces: Heritage, Memory and Identity*. Facet. p. 21–40.

The Consultative Committee for Space Data Systems (CCDS). 2012. Reference Model for an Open Archival Information System. Washington, DC: CCDS. Available from: <https://public.ccsds.org/pubs/650x0m2.pdf>.

Corrado EM. 2022. Digital preservation is not just a technology problem. *Technical Services Quarterly*. 39(2):143–151. doi: [10.1080/07317131.2022.2045432](https://doi.org/10.1080/07317131.2022.2045432).

Directory of Electronic Journals, Newsletters, and Academic Discussion Lists. 1994. Association of Research Libraries, Office of Scientific and Academic Publishing.

McNulty J, Alvarez S, Langmayr M. 2021. Detecting Research from an Uncurated HTML Archive Using Semi-Supervised Machine Learning. In: 2021 Systems and Information Engineering Design Symposium (SIEDS). p. 1–6. doi: [10.1109/SIEDS52267.2021.9483725](https://doi.org/10.1109/SIEDS52267.2021.9483725).

Mogge D. 1999. Seven years of tracking electronic publishing: the ARL Directory of Electronic Journals, Newsletters and Academic Discussion Lists. *Library Hi Tech*. 17(1):17–25. doi: [10.1108/07378839910267154](https://doi.org/10.1108/07378839910267154).

Shearer K, Chan L, Kuchma I, Mounier P. 2020 Apr 1. Fostering bibliodiversity in scholarly communications: a call for action! doi: [10.5281/zenodo.3752923](https://doi.org/10.5281/zenodo.3752923).

Subscribe to comments: [For this article](#) | [For all articles](#)

