United States Newspaper Digitization in an Age of Uncertainty

William Schlaack
Digital Reformatting Coordinator, University of Illinois at Urbana-Champaign, Champaign, Illinois, United States of America.

Kyle Rimkus
Preservation Librarian, University of Illinois at Urbana-Champaign, Champaign, Illinois, United States of America.

Copyright © 2017 by William Schlaack and Kyle Rimkus. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: http://creativecommons.org/licenses/by/4.0

Abstract:
This paper investigates the sustainability plans of large-scale historical newspaper digitization programs at public university libraries in the United States of America. Since 2005, twenty-two public universities have gained federal support via the National Digital Newspaper Program (NDNP) to contribute digitized newspapers to Library of Congress's Chronicling America web platform, which currently hosts over eleven million freely-available indexed and searchable newspaper pages. While the NDNP has catalyzed much newspaper digitization in US libraries, the future of federal funding to support cultural memory efforts is uncertain, leading academic libraries to question the long-term sustainability of their own newspaper reformatting programs. This paper summarizes interviews with nine program coordinators at NDNP partner institutions as well as the authors’ own institution. Through this process, a snapshot of the present state of newspaper digitization in US academic libraries emerges, reflecting the challenges in shifting away from grant-funded projects in favor of locally supported digitization programs. Based on these interviews, the authors propose potential future strategic directions such as seeking alternative sources of funding through donors, as well as increased shared effort between academic institutions. By taking stock of the current state of American newspaper digitization in academic libraries, this paper seeks to identify challenges and opportunities for libraries that wish to sustain historical newspaper digitization programs in the long-term.

Keywords:
newspaper digitization, National Digital Newspaper Program (NDNP), grant funding, sustainability

Background
The United States National Digital Newspaper Project (NDNP) is a collaborative effort between the Library of Congress, the Division of Preservation and Access of the National Endowment for the Humanities (NEH), and various awardee institutions including historical societies, university libraries, and state libraries across a current total of 43 states, the District of Columbia, and one U.S. territory, with the intent to reach one institution in all 50 states as well as additional US territories such as Guam and the Virgin Islands. The program began in 2005 as a means to continue the success of the United States Newspaper Program (USNP, 1982-2011), which enabled university and state libraries as well as historical societies to...
compile extensive bibliographic records on and to conduct largescale microfilm reformatting of their print newspaper holdings. Subsequently, the NDNP distributed grants to institutions to digitize USNP microfilm, and to make the digital surrogates available through the Chronicling America web-platform.

Grants for NDNP have been distributed annually since 2007 to one representative institution per state through grant applications. Each grant takes place over the course of a two-year cycle during which each awardee makes title selections and then delivers up to 100,000 pages of newspapers digitized to Library of Congress technical guidelines and digitization standards. Prior to the 2016 grant cycle each institution was required by the Library of Congress to restrict its selection to newspapers published between 1836-1922, as all copyrightable works published in the United States before January 1, 1923 are presently in the public domain. On July 7, 2016 the NDNP announced that it would expand its date horizon, allowing institutions to digitize and ingest copyright-free content published from 1690 to 1963. Special attention was also given in the most recent grant cycle to expanding non-English publications in languages such as Cherokee, Icelandic, and Finnish, and, in the case of our own efforts at Illinois, Polish, Czech, and Lithuanian.

Upon its creation the NDNP sought to utilize standards, distributed organizational structures, tools, and shared workflow procedures that would in turn lead to support the goal of becoming "...a long-term effort to provide permanent access to a national digital collection of newspaper bibliographic information and selected [digitized] historical newspapers." Indeed, the Library of Congress has established a robust open source technical infrastructure to support its newspaper program. Through its open-source structure and availability NDNP awardees are encouraged, as many indeed have, to create their own locally managed websites built around the Chronicling America architecture.

The Library of Congress’s excellent technical framework is, however, subject to certain limitations. Namely, Chronicling America only accepts content digitized via NDNP digitization, meaning that institutions that digitize newspapers from other sources must seek an alternate platform for the preservation and dissemination of their page images.

---

2 Ibid.
3 Chronicling America (http://chroniclingamerica.loc.gov/) relies on a robust storage and delivery infrastructure managed by the Library of Congress, with a public access system that allows users to conduct full-text searches across its vast corpus of newspapers, to zoom into individual page images, and to download PDF versions of those page images. All newspapers in Chronicling America are freely offered and fully searchable through optical character recognition (OCR) and page-level indexing.

4 These guidelines and standards have evolved throughout the various grant cycles. Current and past guidelines can be found here: https://www.loc.gov/ndnp/guidelines/
7 The Chronicling America suite of tools consists of a Preservation Microfilm Scanner Target (PMT), a Digital Viewer and Validator (DVV), and the Chronicling America: Historic American Newspapers user interface, with its own robust suite of open source toolkits ranging from the Django Web Framework, Apache HTTPD Web Server, a MySQL database, and various Python libraries (the Django application that the Library of Congress used to make the Chronicling America website is often colloquially referred to by awardees as “ChronAm”, https://github.com/LibraryOfCongress/chronam).
Furthermore, recent political developments suggest an even sharper need for institutions to become self-sufficient when it comes to newspaper digitization. For example, President Donald Trump’s budget blueprint for Fiscal Year 2018 proposes to completely eliminate funding for the National Endowment for the Humanities (NEH), along with many other independent federal agencies that support projects and development for libraries, museums, and archives across the country. For many public academic institutions, this news came as a shock and in turn raised serious questions of programmatic sustainability.

**Peer Institution Survey: Planning and Results**

The University of Illinois at Urbana-Champaign Library has received four rounds of NDNP funding since 2009. In addition, our Library strives to digitize newspapers with funding from other grant and donor sources when possible, and to provide access to newspapers that do not fall under the purview of federally-funded IDNP collection-building in our Illinois Digital Newspaper Collections portals. As we consider the future of our program in newspaper digitization, we cannot expect to receive the same level of federal funding we have received for the past eight years. We know that many of our peer institutions face similar challenges. In fact, awardees at NDNP partner meetings frequently pose questions about how best to continue the success of the NDNP program when grant funding ends. To consider our own future as well as the general direction of post-grant newspaper digitization projects and programs, we conducted an informal survey about how other public academic library NDNP awardees were engaging with this challenge, if at all. The questionnaire consisted of eight questions focusing on programmatic planning, staff and administrative support, and sources of funding and future planning (See Fig. 1).

---


9 The Illinois Digital Newspaper Collections currently contain 110 newspaper titles with a total of 132,668 issues comprising 1,440,499 pages. URL: http://idnc.library.illinois.edu/. UIUC also continues to host the outcome of the Illinois Newspaper Project which features newspapers and microfilm that have been identified and cataloged by the USNP. URL: http://www.library.illinois.edu/ inp/database.php
We queried twenty-two peer institutions (in our case, other public universities) awarded NDNP grants, nine of which participated in our interviews, for a 50% response rate. Eleven institutions initially accepted our interview request. We gave them the questionnaire in advance and allowed them to answer the questions in writing or discuss the questions via telephone. Of these 11 institutions, seven went on to accept a telephone conversation, three institutions elected to fill out the questionnaire in writing, and one institution reviewed the questionnaire and noted that they simply have ceased their digitization work and have no plans to reestablish their program\textsuperscript{10}. Participants were informed that their answers would be used to inform this paper, but would not be linked to their names or institutions, this ensure candid responses on delicate topics of staffing and administrative planning.

Question 1: What challenges, if any, have you faced with regard to funding your newspaper digitization efforts?

Of those ten interviewed, seven (70%) reported issues with funding as the primary challenge facing their digitization efforts. Two (20%) others mentioned issues with institutional memory loss, the lack of concrete newspaper collection direction and selection process, and lack of clear goals and models for programmatic approaches to digitization. One (10%) reported modest success with no significant foreseeable challenges. This institution pointed to their

\textsuperscript{10} The nine public universities surveyed here are (in no order): University of Oregon, Central Michigan University, University of North Carolina at Chapel Hill, University of Kentucky at Lexington, University of California – Riverside, Pennsylvania State University, University of South Carolina, University of Florida at Gainesville, University of Hawaii at Manoa, and West Virginia University.
early adoption and implementation of open-source software that helped avoid significant costs associated with upkeep.

---

**Fig 2. – Summary of Response to “What challenges, if any, have you faced with regard to funding your newspaper digitization efforts?”**

---

All respondents underscored lack of administrative support as an important barrier to maintaining adequate staffing (i.e. hiring new, dedicated digitization staff or ensuring that current staff have enough dedicated time for newspaper reformatting efforts), paying for vendored digitization work costs, or supporting continued digital preservation and access management costs (e.g. digital storage space, costs associated with software digital library upkeep and maintenance). Four institutions (40%) made mention that they have actively sought creative new budget models. These included partnering with a non-profit digital library for scanning efforts, setting up a “premium features” model and soliciting donations/patron subscription fees, and utilizing an on-site microfilm scanners to enable ongoing in-house digitization, albeit often at a reduced digitization standard.11 While almost all institutions surveyed identified funding challenges, many have found success by seeking programmatic approaches either in collaboration or through in-house means.

**Question 2: What other grants (apart from NDNP) or funding sources has your institution utilized for newspaper digitization projects (local, state, university, federal, or private sources)?**

The most popular source of non-NDNP funding to emerge from our interviews was the Library Services and Technology Act (LSTA) grant, with seven universities (70%) identifying this as the primary or sole source of funding aside from the NDNP. The LSTA grant program

---

11 The University of North Carolina's North Carolina Digital Heritage Center has documented and made their in-house digitization process freely available on GitHub: https://github.com/ncdhc/ndnp-local-batch-process/wiki/An-approach-to-producing-NDNP-compatible-batch-ingest-packages-locally
is administered by the federal Institute of Museum and Library Services (IMLS) and acts as the only federal program in the United States exclusively for libraries. The LSTA grant is dispersed to individual state and territory libraries who provide matching funds and then in turn award libraries, archives, and museums through an in-state grant process. The Grants to States program is the main source of federal funding support for libraries, archives, and museum services in the United States, with over $150 million awarded every year. It is worth noting that, if passed, the same budget proposal that seeks to eliminate funding for NEH would also eliminate funding for IMLS and thus cut off this source of funding as well.

In addition to the LSTA/IMLS funding source, one institution mentioned that they garner funding from the Center for Research Libraries (CRL). The institution targeted a specific collection of Latin American newspapers and thus could receive funds from CRL’s Latin American Materials Project (LAMP) and Latin Americanist Research Resources Project (LARRP). This funding route presents a unique solution for libraries facing funding crises: target unique collections with specific funding sources in mind. Three libraries (30%) stated that beyond NDNP they have no alternative sources of funding, with one interviewee stating that they frankly need NDNP to maintain any semblance of continued newspaper digitization. One library (10%) stated that their institution has an endowment dedicated to digitization, newspaper preservation, and digital access. Despite state-wide budget shortfalls, this institution cited the endowment as a useful source of ongoing funds to support newspaper digitization without dependence on tuition dollars or federal grants.

**Question 3: How completely have you digitized your state’s public domain newspaper collections? Is there still a lot of work remaining?**

When asking about what percentage of each state’s public domain newspaper collection is digitized, the results overwhelmingly suggest that there is a lot of work remaining for each state. Nine institutions (90%) report that while they do not have an exact figure, they know the bulk of the state’s content is not in accessible digital format. In conversation, many project managers spoke to the need for a better overview of currently digitized content in contrast to what is remaining both on microfilm and on paper. Four institutions (40%) directly mentioned that while they have considerable coverage of their so-called establishment newspapers, i.e. papers that are from urban areas or are papers of record, they lag behind on rural as well as non-English content. This coincides with the NDNP’s current drive to expand non-English newspaper holdings and demonstrates the demand from collection and project managers to digitize such content. A third of institutions surveyed mentioned that they are the primary source for microfilm in their state, placing them in a strong position to drive digitization from film, should funding materialize.

**Question 4: Do you have any future newspaper digitization projects planned for after your NDNP grant cycle is completed? How do you plan to fund those?**

Next, we questioned if the institution had any plans for newspaper digitization projects after the completion of their NDNP grant cycle or. If the institution anticipated a future project or was currently working on one we subsequently asked what plan there was to fund such a

---

12 For an overview of this program please see: [https://www.imls.gov/grants/grants-states](https://www.imls.gov/grants/grants-states).
project. Eight institutions (80%) commented that they do have plans for future newspaper digitization projects. Of those responding in the affirmative, four cited funding opportunities provided by LSTA. The same number of institutions also mentioned turning to local or regional grant opportunities. Two institutions mentioned supplementing their digitization costs with a dedicated budget line for continued digitization in their library budget, as their library has designated newspaper digitization as a component of the library’s mission. One library mentioned their use of endowment funding to cover future digitization expenditures. Additionally, and perhaps most intriguing to hear, was one institution’s mention of working with a commercial newspaper digitization website, and viewing their relationship with this company as reciprocal and symbiotic. Nonetheless, many these founding sources, aside from a continued library budget line and endowment funding provide only short-term project success, rather than long-term programmatic funding.

Two (20%) institutions simply stated they have no further plans for digitization beyond NDNP and subsequently remarked that they were interested in continuing these efforts but were curious as to how they could “downgrade” standards or strip down workflows to be more economical for future budget lines or donor-driven funding. One institution commented that they are investigating how to get content online quickly and efficiently by circumventing NDNP’s high technical and quality control standards, which they cite as an impediment to working at scale. Even so, all institutions surveyed mentioned that they would like to continue newspaper digitization on a programmatic level, regardless of their ability to fund or manage such an endeavor.

Question 5: What other platforms and digital repositories, if any, have you used for your digital newspaper collections? How do you plan to maintain these?

One essential piece of any newspaper digitization project or program is the digital repository where the content resides and is subsequently made accessible to patrons, therefore we asked which platforms, if any, each institution is using and how they are financially supporting their platform of choice. Four institutions (40%) reported utilizing the open-source ChronAm digital repository software as their primary local platform, often with cosmetic branding variants. Two institutions (20%) reported using the Veridian platform. Internet Archive, DSpace, Olive ActivePaper Archive, and SobekCM all garnished one response from the remaining institutions. Half of the institutions reported the desire to have an open-source based digital repository, with Open Online Newspaper Initiative (Open ONI), a fork of ChronAm being mentioned by two of those institutions. Two institutions expressed dissatisfaction with their current repositories, but mentioned that it is difficult to argue for developer and staff time to support advancements in this area, given a general lack of personnel resources and specialized knowledge of technical standards for newspaper search and display. Nonetheless these institutions often spoke to the newspaper collection being a part of library and campus digital storage infrastructure. Funding for these options came mostly (60%) in the form of institutional support for digital storage fees. One institution reported utilizing donor funding and charging three years of storage fees to external institutions who deposited content in their repository, in order to support ongoing maintenance.

---

14 Veridian is Digital Asset Management Software (DAMS) develop by DL Consulting (http://www.veridiansoftware.com/knowledge-base/veridian/).
15 SobekCM is an open-source developed digital repository developed by Mark Sullivan at the University of Florida (http://sobekrepository.org/).
costs. Another institution reported using LSTA grant funding to cover the associated costs of their digital repository and storage. One institution was unaware of local plans for maintaining local digital storage.

Fig. 3 – Summary of Response to “What other platforms and digital repositories, if any, have you used for your digital newspaper collections?”

Digital Repository and Access Platform Used

<table>
<thead>
<tr>
<th>Platform</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChronAm</td>
<td>40%</td>
</tr>
<tr>
<td>Veridian</td>
<td>20%</td>
</tr>
<tr>
<td>Internet Archive</td>
<td>10%</td>
</tr>
<tr>
<td>Dspace</td>
<td>10%</td>
</tr>
<tr>
<td>Olive ActivePaper</td>
<td>10%</td>
</tr>
<tr>
<td>SobekCM</td>
<td>10%</td>
</tr>
</tbody>
</table>

**Question 6:** Have you worked with, or do you plan to work with, private donors to support newspaper digitization efforts? How did you/do you plan to go about this? What successes and/or challenges did you face as a result?

Many public institutions turn to private donors to fund newspaper digital efforts. In our study, six institutions (60%) reported not yet formally soliciting donor funds, saying that there were no concentrated efforts to solicit such contributions. Reasons for this lack included the absence of donor database or networking ability, and issues resulting from lack of administrative support for donor relations. Four of these institutions, however, remarked that they would appreciate being able to develop and eventually work with donor funds, with one mentioning that there are plans in the works with their advancement office. One university reported preliminary discussions with a major newspaper publisher interested in supporting mass digitization. Four of the universities (40%) stated they did indeed receive donor funding. However, the primary success came with working with other institutions, be they genealogy groups, historical societies, or museums.
Interviewees noted challenges in communicating the complexities and demands of newspaper digitization and digital preservation, particularly given staff turnover and the challenge to maintain ongoing contracts and effective communication to donors. Another institution concurred, and remarked on the difficulty of communicating the full cost of digitization to donors, with many failing to grasp the intricacies of digitization efforts – ignoring the need for preservation standards, and being unaware of the costs of “premium” services such as article-level segmentation. The cost of long-term digital preservation also came up. Donor support, even among those with more robust donor programs in place, often falls short with regard to funding the long-term storage and maintenance of digital collections. These are seen as difficult to market to potential donors, with two institutions characterizing this aspect of their newspaper digitization as “unsexy.” In addition, donor funding is sporadic, especially at a time of financial difficulties for many groups and individuals. While donor funding can often lead to successful project-based digitization efforts, caution should be taken when considering seeking out donor funds to support programmatic efforts.

**Question 7: What ongoing staff, if any, has your institution committed to supporting ongoing newspaper digitization in place of grant funding?**

For the last structured question, we investigated the allotment of non-grant funded ongoing staff time to support newspaper digitization. Most interviewees (seven in our survey: 70%) stated that they have some percentage of staff dedicated to newspaper digitization, however, none of these employees are solely dedicated to this process. The remaining institutions remarked that they have a skeleton crew that is still unable to cover the needed work – one institution mentioned that they have one full-time employee overseeing their efforts but that...
this is dependent on external funding, and this they need at least two full-time employees to reach a programmatic level. An additional university gave a similar response, stating that they do have one full-time employee dedicated to newspaper digitization supported by university funds, but that this individual is stretched thin and unable to maintain NDNP levels of digitization. The remaining institution noted that they have two part-time employees working 20 hours per week on digitization projects. Despite their assistance, they still see the need for increased staff support. The key theme of the answers is one of understaffed newspaper digitization programs, with many library administrations expecting such projects or developing programs to do less with more.

**Question 8: Do you have any other comments you would like to add regarding your sustainability planning for newspaper digitization?**

Finally, we asked interviewees to give any additional remarks they felt relevant to their newspaper digitization sustainability and planning. Many commented that staff supporting newspaper digitization need to be optimists and look at the positives of how far projects like NDNP and USNP have brought us. These programs have helped build efforts with clear technical standards and national best practices. There remains a marked need to publicize the importance of these projects and their successes. The desire to see an advocacy platform that would work across institutions to help promote digitized newspaper collections and provide awareness to administration, researchers, and donors alike is a critical component to the future of these programs. One institution described sustainability as a puzzle that needs to be tackled one piece at a time until the necessary aspects of a full-fledged program begin to take shape. In other words, we need to have small goals that fit into larger programmatic efforts.

---

**Possible strategic directions for the future**

**Cross-institutional repository development**

Approaches to sustaining newspaper digitization have been idiosyncratic due in large part to differences in institutional culture, rules, and scholarly need. In the landscape of library digitization, newspapers are often overlooked. Ambitious largescale digitization projects (Google Books, the Internet Archive) have been undertaken for books. One avenue of approach for newspaper digitization that might prove promising is an expansion of ingest policies that could accommodate more institutions, much on the model of the HathiTrust Digital Library. NDNP’s Chronicling America serves as a wonderful basis for such a platform, with the critical limitation that only institutions who have participated in NDNP and have digitized titles through the grant cycle can currently deposit page images. Using this or a similar infrastructure for future efforts and balancing it with economical workflows could prove beneficial to researchers and libraries alike. Indeed, the current lack of central vision beyond federal grant funding creates institutional paralysis. What form will our national program take if funding dries up? We need to see a continued evolution of our technical framework and management structure. The National Digital Information Infrastructure and Preservation Program (NDIIPP) provides one potential example of this. After three rounds of grant funding, the program morphed into the National Digital Stewardship Alliance (NDSA),

10
While the NDSA model is very different from the NDNP, it does demonstrate a community of academic libraries coming together to sustain what was initially a federal grant-funded program after said program’s mandate expired.

**Alternative sources for funding**

As institutions become increasingly unsure of future funding sources, many have started to look elsewhere. Several universities have reported considering partnerships with commercial ventures (in particular, newspapers.com) for low-cost digitization at the expense of forgoing open and accessible content during an embargo period. One institution interviewed accomplished such a “public-private” partnership with a great deal of success, while another reported difficulties. Working together to discuss contractual obligations and how to maximize benefits would be advantageous as many universities attempt to navigate the murky waters of commercial contracts. In this vein, the creation of a how-to guide for newspaper digitization funding would be beneficial for in situations beginning to take a programmatic approach. This document would assist program coordinators with navigating funding efforts. Such a document could also bring together commonly asked questions from donors and assist in the effort to demonstrate programmatic transparency. The development of models for donor tools (e.g. handouts, websites, etc.) to share between institutions would also be welcome, as the field lacks a straightforward way to articulate how donors can contribute to our efforts—what the true costs are, what money can go to support beyond digitization, the importance of digital preservation. With more informed and empowered newspaper digitization program managers, there will be a heightened ability to work with campus development staff to seek additional funding opportunities.

**Ideas for long-term programmatic development**

Several groups, most notably “Beyond NDNP,” have convened to discuss some of the issues mentioned in this paper, such as the need for institutions to look at standards, funding sources, economizing workflows, and sustainability. Ideally, a vision for continued collaboration between partner institutions would lead to devising and implementing future digitization plans. Just as the USNP informed institutions of their film holdings, there should be an effort for states to better understand what newspapers have been digitized and which remain to be digitized in each state. This would in turn help toward the development of better assessment and prioritization tools, with the goal of filling gaps in collections. In addition to this, our survey also highlighted the issue of institutional memory and turnover. Forward-thinking project managers need to ensure that documentation is adequate and updated regularly to allow for new staff to hit the ground running. Despite the idiosyncrasies shown throughout the survey there is a common thread of a desire to improve and expand upon current workflows and to transform newspaper projects into veritable newspaper programs.

**Conclusions**

While NEH and IMLS grant funding might very well persevere through federal administration difficulties, United States newspaper digitization programs seeking to transform their grant-based projects into ongoing programmatic efforts, should begin to take stock of their funding sources, digital preservation plans, staffing model, as well as regional and national partnership...
opportunities. With this in mind, we would like to underscore several key takeaways from our survey:

- Most institutions of higher-learning would welcome the continued digitization of their newspaper collections. These collections speak to the diversity of each state and are of high use for researchers.
- The NDNP and USNP have been invaluable boons for libraries to develop microfilm and digital newspaper collections, but in the absence of federal grant funding many partner institutions will be unable to continue digitization efforts.
- While organizations often collaborate well when receiving grant funding, this enthusiasm and cross-institutional collaboration would benefit the library community if it were carried over into post-grant discussion and networking.
- We should work more with advancement offices, campus-established donor networks, in addition to local and regional historical and genealogical societies to diversify sources of funding as well as cultivate potential areas of collaboration and promotion of collections of note.
- If possible, there should be further research into the possibility of a nation-wide digital newspaper repository and access effort similar to what was seen with the development of HathiTrust.
Acknowledgments

We would like to thank all of the staff who agreed to participate in the survey. In addition, we would also like to acknowledge NEH and Library of Congress for their support and assistance throughout NDNP and for providing the initial forum discussion on sustainability which became the inspiration for this paper.

Bibliography


