Reaping the Harvest: Developing the Idaho Harvester

Robert Perret
University of Idaho, rperret@uidaho.edu

Michelle Shannon
University of Nevada, Las Vegas, michelle.shannon@unlv.edu

Amy Thompson
University of Idaho, amyjo@uidaho.edu

Courtney Berge
Augusta University, berg9737@gmail.com

Devin Becker
University of Idaho, dbecker@uidaho.edu

Follow this and additional works at: https://digitalcommons.usu.edu/westernarchives

Part of the Archival Science Commons

Utah State University
MERRILL-CAZIER LIBRARY

Reaping the Harvest: Developing the *Idaho Harvester*

Robert Perret
Michelle Shannon
Amy Thompson
Courtney Berge
Devin Becker

ABSTRACT

Social media platforms like Tumblr offer a quick, easy, and popular way for archives to conduct outreach and engage users with collections without requiring technical expertise like computer programming and web design. However, as the University of Idaho Library, Special Collections and Archives Department discovered after years of using Tumblr for online outreach, there are significant disadvantages to using a third-party social media platform. Unable to control the discoverability, display, and preservation of blog posts, it became clear that Tumblr was no longer serving the department's evolving needs, necessitating an alternative solution moving forward. Special Collections & Archives partnered with the library's Data and Digital Services Department to develop a custom self-hosted platform for sharing archival collections and staff institutional knowledge. This platform, the *Idaho Harvester,* combines the useful functions of a blog with those of a digital collection to ensure the digitization and research conducted for outreach is easily discoverable and effectively preserved. This article describes the disadvantages of a third-party social media platform for archival outreach and the benefits of a self-hosted platform like the *Idaho Harvester* to demonstrate the need for a multi-platform approach to online archival outreach. Although working with a self-hosted platform like the *Idaho Harvester* requires more technical skills than working with a social media platform like Tumblr, we argue that it meets the goals of discoverability, display, and preservation better than Tumblr or other social media platforms. As archival work becomes increasingly digital, this case study serves as an example of how archival outreach demands have evolved over time and how to assess online outreach tools and develop a multi-platform approach that works for archives' unique and evolving needs.

Introduction

Since the dawn of the Internet, archival repositories have utilized the web to engage with users, provide access to unique content, and create an open and inviting environment for researchers. While providing a quick, easy, and popular way to conduct outreach, social media platforms also present a unique set of challenges. Archives must determine if they want to maintain discoverability of posted content, as social media often displays content related to what is happening "now," burying past posts without an effective way to search or find them. Search tools that do exist

are often rudimentary and do not always present relevant results; when results are relevant, they are often stripped of context. Archives must also navigate sudden and irretrievable loss of content as social media companies change hands or implement sweeping "content take-downs" for user protection. Staff turnover and changes in institutional file management systems often mean that previous efforts are lost. This poses a preservation challenge for archives that pour a lot of time into scanning and conducting research to create online content.

To address the discoverability and preservation concerns with third-party social media platforms, the University of Idaho Library's Special Collections & Archives (Spec) partnered with Data & Digital Services (DDS) to develop the *Idaho Harvester*, a platform that combines functions of a blog with those of a digital collection.[1] This combination allows blog posts to draw from the base digital collection as well as other digital collections to publish curated posts or miniature digital exhibits, while maintaining discoverability and preservation in a way that meets archival needs. Before developing the *Idaho Harvester*, Spec relied solely on Tumblr to promote collections and engage with patrons online. Tumblr was built around GIFs, images, and tags, which made it a useful tool to share historical images, institutional knowledge, and archival research. Tumblr served Spec's purposes for many years, facilitating easy promotion and interpretation of Special Collections & Archives content along with staff members' work experiences. However, over time the issues of poor indexing, poor discoverability, and lack of control over the preservation of content became more problematic until it was clear that a new approach was needed. The *Idaho Harvester* allows both casual readers and serious researchers to find and use Spec's archival collections in a new, engaging, and intuitive way while maintaining the discoverability and preservation that the department grew to need.

The *Idaho Harvester*'s technical infrastructure relies on a customized instance of CollectionBuilder, which is an open-source framework for digital collections developed by librarians at the University of Idaho Library. The platform requires contributors (limited to library and archives staff) to work with a more complex process to create blog posts and add items to the built-in digital collection. While the process for posting content to the *Idaho Harvester* is more involved than the process for posting content to social media platforms like Tumblr, these workflows enable UI Library to gain control over its blogs' content. This helps ensure better short- and long-term preservation by giving contributors full access to the code and text creating the site's content. It also helps enable UI Library's digital librarians to control how blog posts and digital collection items are indexed by search engines like Google and preserved via inclusion in the library's digital archive.

---

1. University of Idaho Special Collections and Archives, "Idaho Harvester," University of Idaho, last modified 2003, https://harvester.lib.uidaho.edu/.

Literature Review

As the Internet rose in popularity in the 1990s, cultural heritage institutions began thinking about and adopting digital media as an avenue for outreach and engagement. Heyliger, McLoone, and Thomas note that while archival blogs began as early as 1996, widespread adoption of digital outreach did not begin until 2009.[2] They examined 2,850 repositories listed in Terry Abraham's *Repositories of Primary Sources* and found that in 2013, just 8% of repositories had a social media presence. They found that universities were most likely to post about university history, write for student and faculty audiences, and use students and support staff as blog authors, as has been the case at University of Idaho Library, Special Collections & Archives. They also found that blogs across institution types often had an irregular posting schedule, with weekly and monthly schedules the next most common, as the frequency of posts was directly correlated to staff size. All institutions identified the benefit of social media outreach as: increasing awareness of collections, highlighting specific materials, and promoting events, respectively.

Archivists, however, have been thinking about how to engage users with digital content beyond social media posts well before widespread adoption of social media in 2009. In 2002, Margaret Hedstrom was already thinking about the digital archives of the future. She argued that as human-mediated archives give way to computer-mediated archives, archivists must re-examine their role in connecting users to collections.[3] Her suggestions included creating online tools to allow patrons to make their own interpretation of archival holdings. In this way, Hedstrom advocated for online tools that did not just tell users about collections, like in a social media post, but that facilitated better engagement between users and archival materials as a result of how the tools were designed. Max J. Evans made a similar suggestion in 2007, arguing that by sharing a common online environment, holdings would be more easily discoverable and usable by archivists and users alike, resulting in a community of "highly intelligent men and women who will come to appreciate archives."[4]

Kate Theimer similarly examined how online platforms could connect users to collections in an engaging, rather than a passive, way. In 2008, Theimer began interrogating the meaning of the phrase "Archives 2.0," a popular conception that archives were evolving away from physical connection and increasingly towards online connection. Theimer argues that the idea of Archives 2.0 is "an approach to

---

2. Sean Heyliger, Juli McLoone, and Nikki Thomas, "Making Connections: A Survey of Special Collections' Social Media Outreach," *American Archivist*, 76 no. 2 (Fall/Winter 2013), 374-414. https://doi.org/10.17723/aarc.76.2.t820u33100443q55.

3. Margaret Hedstrom, "Archives, Memory, and Interfaces with The Past," *Archival Science* 2, no. 1-2 (March 2002), 21-43. https://doi.org/10.1007/BF02435629.

4. Max J. Evans, "Archives of the People, by the People, for the People," *American Archivist* 70, no. 2 (Fall/Winter 2007): 400. https://doi.org/10.17723/aarc.70.2.d157t6667g54536g.

archival practice that promotes openness and flexibility," posing the notion "that archivists must be user-centered and embrace opportunities to use technology to share collections, interact with users, and improve internal efficiency."[5] She notes that the role of archivists has moved away from controlling or limiting access to collections and towards facilitating the use of, and even building, collaborative archives where there is a relationship between staff and users. She encourages archivists to "go where your users are" by making digital collections available in online spaces that may be outside of their complete control.[6]

Mary Samouelian had similar ideas about how to utilize technology to actively connect users to collections. In 2009, Samouelian surveyed the "Web 2.0" landscape, where there is increased discussion and engagement between users and content producers in the form of comments, likes, and sharing. She found that the degree of online presence varied greatly among archival institutions. While most archives had a webpage "announcing their existence," she believed this was not enough, lamenting that little had resulted from the 2008 SAA meeting where Web 2.0 had been a topic of discussion.[7] Encouraging archives to do more to engage online users, she pointed out that anecdotal evidence showed increased use of collections when there was an active digital component.

By the time Katie Elson Anderson wrote in 2015, libraries and archives had noticeably increased online presence, specifically with microblogging platforms like Tumblr, which bridged the format of traditional blogs with the social media features of Twitter. Anderson examined the relationship between libraries and Tumblr and noted that special collections departments had adopted Tumblr at notably high rates. She attributed this growth to features including unlimited characters allowed in posts, a robust "reblogging" system that strongly maintained a chain of attributions when content was shared and reshared by users, and the fact that Tumblr was indexed by search engines in a way that was comparable to a traditional blog.[8] Anderson surveyed all types of libraries and found that special collections departments posted images most often, with 88% of their posts including images.[9]

By 2018, archivists were thinking about how digital preservation played a role in online outreach and engagement with users. Angela Fritz lauded digital project management as a new service model that represented a "professional reawakening."

5. Kate Theimer, "What Is the Meaning of Archives 2.0?" *American Archivist* 74, no. 1 (Spring/Summer 2011): 60. https://doi.org/10.17723/aarc.74.1.h7tn4m4027407666.

6. Ibid., 62.

7. Mary Samouelian, "Embracing Web 2.0: Archives and the Newest Generation of Web Applications," *American Archivist* 72, no. 1 (Spring/Summer 2009): 43. https://doi.org/10.17723/aarc.72.1.k73112x7n0773111.

8. Katie Elson Anderson, "Libraries and Tumblr: A Quantitative Analysis," *Reference Services Review* 43, no. 2 (2015): 157. https://doi.org/10.1108/RSR-12-2014-0060.

9. Ibid., 167.

Discussing the proliferation of online outreach and access to collections, she noted that "a variety of library units share the work of surfacing special collections and university archives, operationally integrated through cross-functional teams throughout the organization working together to pursue common goals and outcomes to ensure interoperability, sustainability, and broad accessibility of digital collections."[10] To manage the cost of this increased work, she recommended incorporating digital projects into daily operations and argued that cross-training between departments should be viewed as a necessity. She argues that sustainability should be considered as well during interdepartmental collaboration and the creation of new processes. Ultimately, Fritz recommended a shift in focus from the preservation of paper-based collections to a focus on "virtual space management," arguing that "expanding access to digital content is firmly grounded as a foundational core service of any archives of special collections."

Case Study

Between 1998 and 2009, Spec occasionally posted digital content to Digital Memories, a simple HTML website hosted by the Library.[11] Spec was not engaged with online outreach on any platform between 2009 and 2014. In 2014, Spec began posting digital content again, this time using Tumblr as the primary platform for online outreach. Initially, one staff member created weekly posts supporting university events or sharing unique items found in the collections. There were several benefits to using Tumblr for outreach to support the Web 2.0 ideals of online interaction. Tumblr made it easy to engage with other special collections and archival institutions by following each other's feeds and liking, commenting on, and sharing each other's posts. Posts could be grouped by subject by "tagging" posts with keywords so that they are displayed with other similar content, reaching a broader audience. Tumblr also allowed Spec to cross-post on general library social media channels (including Twitter, Instagram, and Facebook) through automated services like IfThisThenThat (IFTTT), expanding its reach to audiences on other platforms. Finally, the process of creating a Tumblr post was easy, intuitive, and not time-intensive, even for the novice user.

In the years following Spec's adoption of Tumblr in 2014, however, departmental workflows evolved; all Spec staff started creating longer in-depth posts to increase outreach and engagement with Library collections, along with the history of the University of Idaho and North Idaho. The work spent on researching and writing longer posts surfaced several disadvantages in Tumblr for discoverability and preservation. For example, Spec decided that each blog series, or groupings of

10. Angela Fritz, "From Collection Silos to Digital Content Hubs: Digital Projects Management in Special Collections and University Archives," *Advances in Library Administration and Organization* 38 (2018), 187-198. https://doi.org/10.1108/S0732-067120180000038014.190.

11. University of Idaho Library, "Digital Memories," University of Idaho, last modified 2009, https://www.lib.uidaho.edu/special-collections/dm/index.html.

thematically-related posts, should be displayed prominently and serially. However, Tumblr was not designed to organize content in that way, limiting the cumulative, contextual, and organizational benefits of presenting material in this fashion. Spec was also unable to control how content was indexed on Tumblr, as the platform only offered searching across the entire platform rather than allowing users to search a specific feed. Searches were also limited to whole posts or tags, excluding the digital objects (e.g., images or videos) included in a post. Accessing older posts was also challenging since posts were organized chronologically with the newest posts shown first, and Tumblr required users to be logged in to browse the list of old posts, limiting access to users with accounts. Finally, Spec could not ensure the longevity and preservation of content posted on Tumblr. For example, when a new policy was deployed by Tumblr that banned and automatically removed content that was deemed offensive or inappropriate by an algorithm, several posts were removed suddenly and without warning, even though the content did not meet the established criteria. This resulted in the loss of digitized images, staff institutional knowledge, and archival research.

By 2020, the frustrations with Tumblr's limitations led Spec to investigate other platforms for writing about and sharing archival collections. This led to a departmental collaboration with the University of Idaho Library's Data and Digital Services (DDS). Both departments shared similar concerns about the discoverability and preservation of the content created on Spec's Tumblr account, as DDS was responsible for UI Library's website, digital collections, and other digital content. Several DDS librarians were also looking for ways to expand on the multi-modal authoring features they developed for their digital collection software tool, CollectionBuilder, and building a customized CollectionBuilder-based blog platform seemed like a way to develop some of these capacities. With all these reasons prodding them forward, Spec and DDS partnered to create a new platform that would address some of the problems Tumblr presented while allowing for more flexibility and control of the outreach content themselves.

After roughly four months of collaborative work between Spec and DDS, the *Idaho Harvester* launched in December 2020 and currently contains over 30 years of staff research and digitized archival content, including materials from Tumblr, various legacy webpages and databases, and photographs that Spec faculty and staff have taken for events and exhibits. Fittingly, content from Tumblr, as well as legacy websites (proto-blog posts from the 1990s), were "harvested" to seed the new platform. Using Tumblr's built-in data export option, Spec and DDS used Tumblr's administrative tools to harvest the text of existing posts, which were written in HTML with image links embedded. However, the original images used in blog posts were not included and had to be harvested using the data cleaning software OpenRefine and the web file retrieving software GNU Wget on the exported data to download the images stored on the platform. This pointed to another shortcoming in the Tumblr workflow: Tumblr posts did not systematically save images, PDFs, or files attached to a blog post. By going through the process of harvesting data from Tumblr, it became clear that the new *Idaho Harvester* platform should be designed to ensure that the

digital items themselves, along with their metadata, were preserved, accessible, and connected to their associated blog posts while also browsable independently from their associated blog posts. With this in mind, the *Idaho Harvester* was built on top of the digital collection framework CollectionBuilder, allowing both migrated and new blog posts to draw on a built-in, ever-growing digital collection of over 1700 items that grows with new posts. The *Idaho Harvester* content is in turn digitally archived with established preservation standards used for all of the UI Library's digital collections.

Technical Apparatus

The *Idaho Harvester* is a heavily-customized version of CollectionBuilder, an open-source framework for creating digital collection and exhibit websites that are driven by metadata and powered by modern static web technology. CollectionBuilder is built on Jekyll, a static site generator that transforms plain text into static websites and blogs. Because CollectionBuilder is built on Jekyll, staff were able to infuse code from Beautiful Jekyll, a ready-to-use template to help create beautiful websites quickly, as the platform's stylistic theme to create a unique look and feel for blog posts and the homepage. Although there are many static site generators available for use online, Jekyll is one of the most established and allowed staff to heavily-customize the styles, features, and functionalities of the *Idaho Harvester* in a user-friendly way that best presents and connects users with the content.

By combining CollectionBuilder with Beautiful Jekyll's blog features, staff were able to create several customized discovery options and local standards for display and publication. There are several ways that the *Idaho Harvester* is customized beyond the base CollectionBuilder template. First, the base CollectionBuilder template has a simple Browse page intended for users to browse the items in a digital collection or exhibit. The *Idaho Harvester* is customized beyond that—in the navigation bar, users can browse Posts, Items (from the Collection dropdown menu), or Series (by choosing from the dropdown menu).[12] Second, the base CollectionBuilder template has simple features for staff to write interpretive "About" pages for exhibits while including digital objects from the exhibit to accompany the interpretive writing on the "About" page. Staff can not only embed all major formats of digital objects (image, text, audio, video), but can actually include digital objects from any digital collection or exhibit currently hosted by the University of Idaho Library, not just digital objects from the *Idaho Harvester*. This was incorporated into the *Idaho Harvester* code by drawing on JSON data from each digital collection that is stored in the site's repository, and contributors can use specific commands in a blog post to pull on any other existing digital collection. This feature expands the opportunity for interpretive writing to approximately 90,000 items that have already been digitized.

---

12. University of Idaho Special Collections and Archives, "Blog Posts | Idaho Harvester," University of Idaho, last modified 2003, https://harvester.lib.uidaho.edu/posts/.

Several software programs are required to work with the code behind the *Idaho Harvester*, including Visual Studio Code (a text editor), Git (a version control system), GitHub Desktop (a computer application that makes it easy to work with the GitHub repository), Ruby (a programming language), and Jekyll (the static site generator). The CollectionBuilder's Software for Working with CollectionBuilder instructional webpage is a detailed guide on how to install and use all of these software programs.[13]

Since the *Idaho Harvester*'s technical apparatus is quite complex, it requires Spec staff to engage with a fairly technically-demanding process. Not only do Spec staff have to download the software required to edit and run the *Idaho Harvester*, they must also write blog posts in Markdown (a lightweight formatting language which allows plain text to be both human readable and transformable into HTML webpages). Additionally, Spec staff must be able to navigate code in a text editor and even use certain preset code to include images, PDFs, video, or audio into blog posts.

The complexity of this system, coupled with the acute attention to detail required for a static site generator to work without breaking, proved to be much more difficult than initially thought at the outset of the project. Some of the challenges are described below, but initially the thought was that the benefits of a self-hosted platform that ensured the digitization and research conducted for online archival outreach was easily discoverable and effectively preserved outweighed the cost of learning complex technical programs and workflows. Staff would not only gain increased control of the presentation and preservation of digitized content, but staff also gained web development and data management skills that benefited the University of Idaho Library as a whole. The same tools and workflows used to create blog posts and add digital objects to the *Idaho Harvester* can also be used to build standalone digital exhibits, digital collections, or to edit and update the Library's website.

The challenges that unfolded as the *Idaho Harvester* was developed quickly demonstrated the need for additional documentation and training to ensure both current and future Spec staff could use the platform as intended. Detailed documentation written by Spec staff provides instructions and screenshots to support completing a large number of tasks, such as creating a blog post from scratch, adding digital objects to the digital collection, troubleshooting common errors, and creating a new series. These instructions can be found on the Wiki portion of the GitHub repository and are available for current and future staff to reference while working with the *Idaho Harvester*.[14]

To further facilitate informed contribution to the *Idaho Harvester*, the Harvester Helper website, also on GitHub, was created to alleviate several persistent issues

13. CollectionBuilder, "Software for Working with CollectionBuilder," CollectionBuilder, accessed August 2023, https://collectionbuilder.github.io/cb-docs/docs/software/.

14. University of Idaho Library, "Idaho Harvester Blog & Working Collection Wiki," Github, last modified November 3, 2021, https://github.com/uidaholib/spec-lumber/wiki.

encountered by Spec staff.[15] The site includes interactive forms that automatically generate code required to insert images, PDFs, videos, and audio files, as well as including a basic template for creating a blog post. The site includes links to the Wiki, as well as a QR generator for library staff to use more broadly. By creating documentation that is detailed, widely accessible, and written from the perspective of Spec staff that engage with the platform regularly, the hope is that the *Idaho Harvester* will continue to be useful to future staff and patrons.

It took Spec and DDS roughly four months between the first meetings initiating the building a new platform in August 2020 and launching the *Idaho Harvester* in December 2020. It began with a weeklong "sprint" between the two departments to discuss, brainstorm, and map out what the new platform should look like and how it should function. After that, members of DDS began adapting CollectionBuilder to create the platform and harvest legacy content into the new platform. Spec members began populating the platform with contextual information while also cleaning up metadata and adding organizational features like tags. This project was a top priority for both departments during this time.

Functionality

While developing the *Idaho Harvester*, Spec and DDS focused on the features and functions necessary to benefit users, specifically prioritizing the discoverability, display, and preservation of content.

Discoverability and Display

It was important to Spec and DDS to create a platform that had a simple, user-friendly display to facilitate easier navigability and discoverability of content. An important way to ensure discoverability was by clearly displaying tags in each blog post and digital object so that users could search for established keywords. Select tags are listed on the homepage while the Subject word cloud includes a complete list of all tags (see Fig. 1 and 2). Additionally, related tags are listed within a blog post or digital object (see Fig. 3 and 4).

Grouping blog posts and digital objects by tag allows users to see relevant materials across collections based on a keyword subject, such as "women", "recreation", "mining", or "wilderness". Since the Idaho Harvester is not limited to a single archival collection, users have the ability to view and engage in dozens of collections at once.

---

15. University of Idaho Library, "Harvester Helper," Github, last modified 2022, https://uidaholib.github.io/harvester-helper/.

Figure 1. Subject tags on the homepage
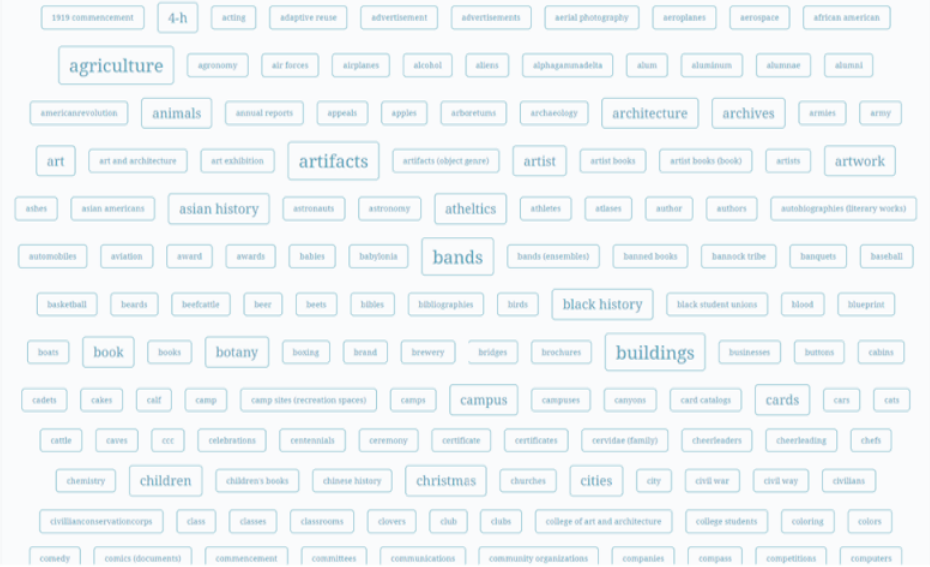
Figure 2. Subject tags in the word cloud

Figure 3. Subject tags on the homepage

Figure 4. Subject tags in an item record

Another way that the *Idaho Harvester* enhances discoverability and display is by including links to associated blog posts, finding aids, and/or catalog records in a digital object item page. This better connects users to the context surrounding a digital object since they can easily find other resources that the object is associated with (see Fig. 5).

There are numerous other ways that users can engage with materials. In the navigation menu at the top of each page, users can search over 800 blog posts, over 20 series, as well as browse image-thumbnailed posts by publishing year within a timeline. In the Collection tab of the navigation menu, users can search over 1,700 items from dozens of UI Library archival collections, searching by subject, geolocation, or post publishing date; users can even download collection metadata as a complete dataset for their own research purposes.

Additionally, users can read about the *Idaho Harvester* and its contributors in the About page, or subscribe to monthly email newsletters to stay current on the most recent blog posts and events in Special Collections & Archives. Finally, the bottom of each blog post includes a feedback form link for users to submit what they would like to see covered or improved with the platform. Users can also use the "Ask Us" link in

Figure 5. Blog posts and finding aids linked in an item record

the top right of any Spec department websites with multiple points of contact to reach Special Collections & Archives, or the greater UI Library.[16]

Preservation

The *Idaho Harvester* is built, stored, and managed directly by the University of Idaho Library's Special Collections & Archives and Data and Digital Services departments. All digital objects (images, PDFs, audio/visual files) are stored within UI Library's existing digital preservation infrastructure. Digital objects that are accessible on the *Idaho Harvester* are the "access copy" of the object, meaning it is a low-resolution version to make it easier for users to download and work with on their own. The Library keeps a "preservation copy" of the object, which is an archival quality high-resolution version that is stored in a secure digital environment with regular onsite and offsite backups and hash value checksums performed routinely.

The website itself is preserved through GitHub, which is built into the *Idaho Harvester* workflow. GitHub allows for precise version control so UI Library and its contributors can track changes and revert to previous versions. The digital objects, along with the research and metadata that goes into blog posts, are fully preserved and controlled by the University of Idaho Library, making sudden or unexpected loss of content very rare.

Other Features

The *Idaho Harvester* solves many of the problems that Spec had with Tumblr, specifically relating to discoverability, display, and preservation. However, there are pros and cons to each platform. For example, Tumblr allowed Spec staff to interact with users and other institutions directly by utilizing built-in interactive features such as likes, comments, shares, and follows. Wanting to ensure an ongoing connection with users, Spec and DDS established a monthly email newsletter that features blog posts from the previous month and information about projects and news from Special Collections & Archives. The newsletter allows Spec to stay in contact with interested patrons, including many prominent members of the local community, while giving them a steady flow of new content and information.

The *Idaho Harvester* also gives Spec the ability to digitize and make accessible materials from any collection on a case-by-case basis since the digital collection and the blog portions of the platform are separate (albeit intricately connected when necessary). Spec often digitizes materials for reference inquiries and can now easily make those reference scans publicly available to everyone, instead of just a single patron. The ability to add individual items from a wide variety of collections on an *ad hoc* basis increases access to the archival collection as a whole, allowing users to browse across a wider variety of topics and collections.

---

16. University of Idaho Library, "Special Collections and Archives," University of Idaho, accessed August 2023, https://www.lib.uidaho.edu/special-collections/.

Discussion

Positive Outcomes

Several positive outcomes have emerged since developing the *Idaho Harvester* as the primary means of online outreach for University of Idaho Library, Special Collections & Archives.

- The *Idaho Harvester* solves many of the problems that Spec had with Tumblr, specifically relating to discoverability, display, and preservation.

- The University of Idaho Library gained complete control over posts and content, eliminating the risk of content being lost suddenly or unexpectedly.

- New means of feedback and communication have resulted due to the email newsletter, including direct emails and chats from patrons inquiring about specific posts.

- The digital collection component of the platform allows Spec staff to easily add digitized materials on an *ad hoc* basis, improving access to collections overall.

The collaboration between Spec and DDS has also resulted in a symbiotic relationship between the interpretive content on the blog and the digitized materials that are available in the Library's Digital Collections. This relationship has provided more context for users, as well as more perspectives to consider when evaluating or using items from a digital collection. For example, the process for creating an *Idaho Harvester* blog post is similar enough that Spec staff can now build stand-alone digital exhibits and digital collections. This not only benefits users by increasing access to more materials but it also creates professional development opportunities and resume-building skills for Spec staff.

Statistically, traffic to the *Idaho Harvester* has remained consistently strong since launching in December 2020. Below is a screenshot of the Matomo Analytics' Visits Overview page that summarizes the visits and interactions during the *Idaho Harvester*'s eighteen months of operation (December 15, 2020 to June 15, 2022) (see Fig. 6).[17] The web statistics for the past year compare favorably to UI Library's other digital scholarship projects. As content continues to grow, the Library expects traffic to the *Idaho Harvester* to grow as well, especially as traffic is driven increasingly by search engines and additional usage, and as the site gains more traction among the community.

---

17. Spec and DDS chose Matomo Analytics to track web statistics for the Idaho Harvester because the interface for analyzing visitors and trends is more user-friendly and intuitive, and because Matomo's open-source software allows for much better privacy protections for our users than, for instance, Google Analytics, due to the software being hosted and controlled on a University of Idaho Library server.

Visits Overview

 19,351 visits

 1 min 8s average visit duration

 75% visits have bounced (left the website after one page)

 1.9 actions (page views, downloads, outlinks and internal site searches) per visit

 296 max actions in one visit

 35,222 pageviews, **30,132** unique pageviews

 696 total searches on your website, **466** unique keywords

 580 downloads, **459** unique downloads

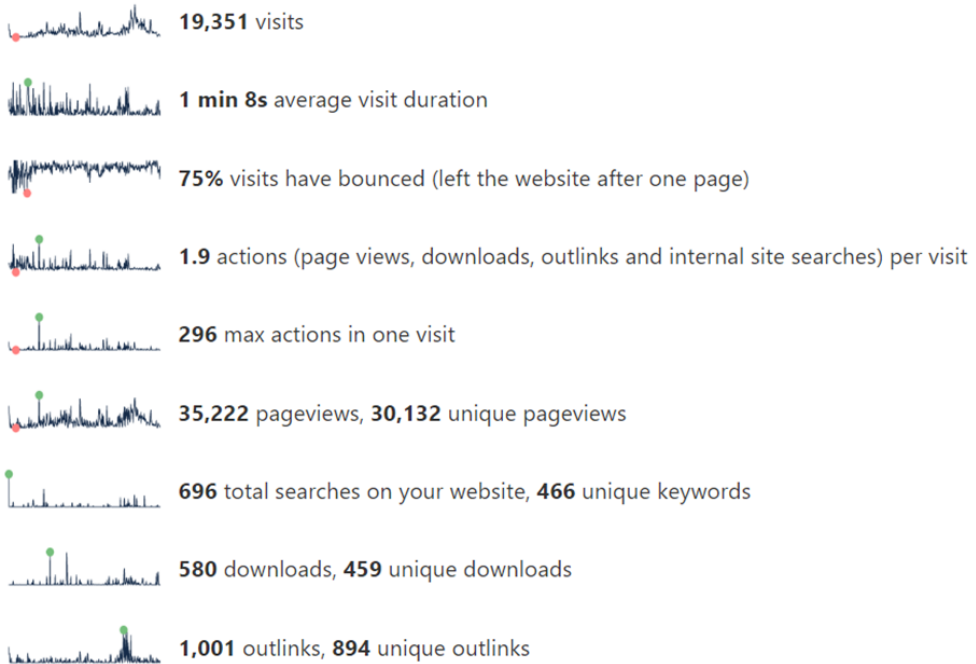 **1,001** outlinks, **894** unique outlinks

Figure 6. An overview of the Idaho Harvester site visits and activity from 2020-12-15 to 2022-06-15

Statistics show that most users do not enter the site through the *Idaho Harvester* homepage, but instead through a specific post or digital object (see Fig. 7). Since DDS librarians had seen for years that most users enter the library's digital collections through search engines, they developed the *Idaho Harvester* with this in mind and intentionally ensured that specialized content was more easily indexed by search engines through tags and keywords. The large number of unique entry pages (25% enter through search engines and over 50% enter through direct links) seemingly demonstrates that these efforts were worthwhile and work for the user (see Fig. 8).

Additionally, the Matomo Analytics' statistics for the *Idaho Harvester* show a strong contingent of return users that engage more deeply with the site and its content (see Fig. 9). On average, return users spend over three times as much time and perform more than twice as many actions during each visit. Although return users only account for 15% of visitors, Spec and DDS believe this group accounts for many of the users they hoped to reach with this project in the first place. To see that this group is engaged with the content provides further evidence that the effort was worth the time invested.

Figure 7. Entry page statistics from 2020-12-15 to 2022-06-15, organized by the directory that each page is stored in. Individual blog posts are stored in the "posts" directory and digital collection items are stored in the "collection" directory. This indicates where users first enter the *Idaho Harvester* website.

Figure 8. Channel type statistics from 2020-12-15 to 2022-06-15. This indicates whether users enter the *Idaho Harvester* website directly or if they are routed from another site.
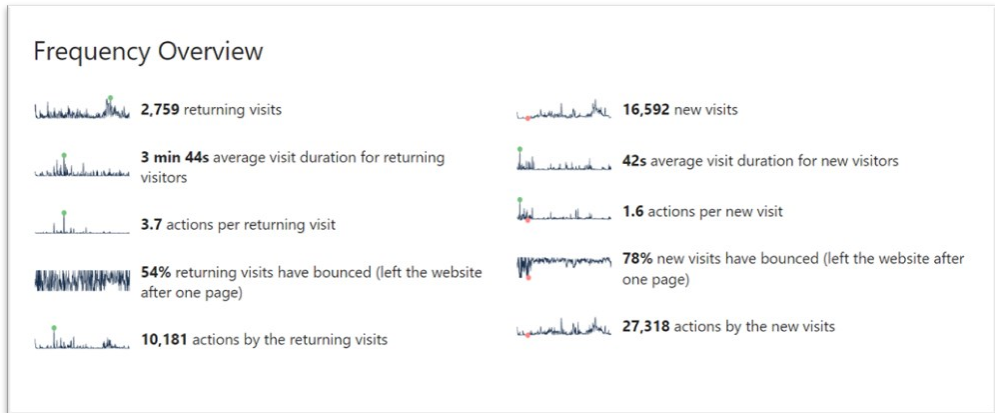
Figure 9. An overview of user frequency from 2020-12-15 to 2022-06-15. This shows the difference in engagement between new and returning visitors to the *Idaho Harvester*.

There are internal benefits as well. Since archivists spend countless hours working with collections and assisting researchers, they have a unique perspective of which items, topics, or collections are useful research tools. Traditionally, archivists have turned to social media to post images of interesting discoveries found in a collection to spark interest online, such as an old book, a weird object, a love letter, a land deed, etc. However, posting images of "interesting" items without including the broader context about that item strips the object of its richness and meaning, leading archivists to conduct research for social media posts. Contextualizing an item might mean finding information about the collection that the item resides in or how the item relates to other collections or other histories. Context allows researchers to learn more in-depth about a collection, topic, or region while also learning about related resources.

The research involved to contextualize archival material for online outreach is often time- and energy-intensive and not only serves external researchers, but internal institutional knowledge as well since research can help inform future reference interactions. It takes time to become familiar with archival collections and when institutions have a high turnover, it can be difficult for new staff to find the collections and resources that best fit a reference request. By preserving digitized content and contextual research, archival staff learn their collections more quickly and thoroughly and are able to better assist their patrons.

Challenges

During the first few months after launching the site, troubleshooting errors required considerable time from both the person adding items to the digital collection or creating blog posts, as well as staff who had the technical skills to help. Even to this day, staff run into strange errors or unexpected problems due sometimes to an extra space or misplaced comma being included in a blog post. This can be

extremely frustrating, so the addition of extensively revised documentation and the subsequent Harvester Helper website has been necessary to improve the success rate for staff who work with the *Idaho Harvester*. Further, if the *Idaho Harvester* is going to succeed going forward, Special Collections & Archives users will need timely and consistent support from technical staff, encouragement from positive user feedback, and consistent administrative support. These contributions will hopefully go a long way towards making any frustration tolerable, but they also require more attention and time from personnel across the organization than a social media site. Additionally, as updates or improvements made to the underlying system may disrupt established and documented procedures, DDS librarians in charge of the technical underpinning will need to be as attentive to the documentation for the platform as they are to the code underlying the platform itself. This is essential for continuing the trust between Spec and DDS and will hopefully ensure that any challenges can be resolved, and workflows can be revised as needed.

Conclusion

The *Idaho Harvester* was developed in collaboration with the University of Idaho Library's Special Collections & Archives and Data & Digital Services departments in response to concerns about discoverability, access, and preservation of content posted on Tumblr. Social media sites like Tumblr still play an active role for Spec to engage users in the Web 2.0 ideals of attempting informal online conversations or posting interesting or relevant material from Special Collections & Archives holdings. However, having used Tumblr as the main source of outreach for collections for approximately six years, several issues had presented themselves. Older content was difficult to find, similar content was difficult or impossible to organize, and changing policies led to a loss of posts that no longer fit Tumblr's Terms and Conditions. These issues increased over time as the department's staffing and priorities evolved, leading them to explore other options. The outcome was a collaborative process that utilized skillsets from both Spec and DDS to create the *Idaho Harvester*. This tool better serves Special Collections & Archives patrons and the wider UI Library by allowing greater freedom in content creation, content discoverability, and preservation of access.