IJDC | *Conference Paper*

Building a Culture of Reuse: An Analysis of Reusable Software and Policies for Institutional Libraries

Alex Kinnaman
Virginia Tech University Libraries

Yinlin Chen
Virginia Tech University Libraries

Abstract

This paper will present findings from a multi-case study on the need for and valuable assets of reusable software and policies for digital library infrastructures. This paper supports the conference theme of reusability. Curating for reuse is a strategy that should not be limited to digital assets, and can extend to digital library software, policy, infrastructure, and implementation. Specifically, we seek to understand how data curators utilize reusable digital library software and policies and how we at Virginia Tech University Libraries can improve the reusability of our resources in order to promote openness, transparency, and reusability.

Submitted date 2021 ~ *Accepted* date 2021

Correspondence should be addressed to Alex Kinnaman, 560 Drillfield Drive Blacksburg, VA 24061. Email: alexk93@vt.edu

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

At Virginia Tech University Libraries (VTUL), we strive to ensure that our University's digital assets are accessible, discoverable, and sustainable. We also strive to ensure that our digital library policies and software are open and reusable to members of our field to promote information sharing and promote good practice. When we think about curating for reuse, we aren't limited to our digital assets, but to our governance structure, policies, software, and technical infrastructure. As an organization we publish our policies and software for our Digital Libraries Platform (DLP), the service foundation for our digital library, for public use. This paper will present findings from a multi-part study on the reuse and impact of digital library infrastructure policy and software at VTUL. We also use this paper as a platform to advocate and share the invisible labor of backend digital libraries work. Specifically, we seek to learn at what level our policies and software are reusable and if we can improve them for greater use and reuse through the following research questions: Research Question 1: Are VTUL digital library resources understandable and reusable to VTUL stakeholders? Research Question 2: Are VTUL digital library resources understandable and reusable to organizations external to VTUL, and how can we improve our resources for reusability?

Literature Review

As digital library practitioners, we often rely on others' resources to explore and build out our own, but what makes certain resources more appealing to use than others?

Software Reusability

The goal of open-source and open-access software is to encourage users to reuse existing software and not reinvent the wheel (Krueger, 1992). Open-source software may be developed in a collaborative public manner and gather crowd wisens to better and extend the existing software. While the techniques advanced dramatically through these years, new development concepts, cloud computing, artificial intelligence, and architectures are working together to encourage software reusability that can create better services for users' needs. Software reuse has been used in various domains, such as in education (Lakhan & Jhunjhunwala, 2008), geospatial (Wang et al., 2016) and biology (Gremme et al., 2013). A software's ability to be reusable is considered a valuable feature and lends to a better quality of software (Frakes & Kang, 2005).

Policy Reuse

Where software reuse is recognized as a highly valuable component to the field, digital library policy reuse is a generally unexplored area of traditional research. Guidance in digital library policy standards can be

found in digital repository certification criteria. ISO 16363:2012 (ISO 16363, 2011) requires that a trustworthy repository have an Access Policy, and Preservation Policy, and a Collection Policy and also requires non-policy documentation like a Mission Statement and Designated Community to improve transparency. Another example is the CoreTrustSeal, which recommends similar policies as ISO 16363 including a collection development policy and a preservation policy, as well as a public policy for addressing issues in non-compliance (CoreTrustSeal, 2019). These examples and other repository certifications provide some guidance on identifying digital library policies, but how do we know what policies we actually need or how to develop them?

There are many policy frameworks and guides available for digital preservation and collection development policies specifically, such as the MetaArchive's Preservation Policy Template,[1] NEDCC's Digital Preservation Policy Framework,[2] the Sustainable Heritage Network's Collection Development Policy Example,[3] and International Federation of Library Associations and Institutions' Guidelines for a Collection Development Policy,[4] to name a few. These account for the necessary policies according to TDR criteria, but there is limited information on what other policies might be relevant, like a fixity or audit policy. Research into digital preservation policies specifically has offered deeper insight, such as Noonan's (2014) digital preservation policy framework development at the Ohio State University Libraries case study that guides readers through the discussion of values, section development, and policy approval process. Beagrie, Rettberg, and Williams (2008) also performed a digital preservation policy study, providing a deep dive into digipres policy customization.

However, similar to Krueger's (1992) note above about software reuse, M. Farkas (2015) writes, "We spend a lot of energy trying to create things from scratch when, frequently, another library may have already done something very similar." Professionals tasked with digital library policy development know that we search for and reuse others' public resources as we develop our own, and this is entirely dependent on discoverability and accessibility.

Open Access at Virginia Tech

Virginia Tech values and promotes open access on various levels in the University. In March 2021,the Virginia Tech Board of Visitors passed a new version of Policy 13000, Policy on Intellectual Property with the language "Authors grant to the university a nonexclusive license to their scholarly articles in order to provide open access (free, public, online access) to them via the university repository." This provides authors the right to publish their work in an open access format in addition to any journals or platforms that accept their publication behind a paywall. The University

[1] https://metaarchive.org/wp-content/uploads/2017/03/ma_dp_policy_template.pdf
[2] https://www.nedcc.org/assets/media/documents/SoDAExerciseToolkit.pdf
[3] https://www.sustainableheritagenetwork.org/digital-heritage/collection-development-policy-example
[4] https://www.ifla.org/wp-content/uploads/2019/05/assets/acquisition-collection-development/publications/gcdp-en.pdf

Libraries also maintains an Open Knowledge Committee that developed our Open Values, established the Open Access Subvention Fund, and is dedicated to the ongoing support and education of open access strategies.

VTUL is dedicated to open source work and has continuously contributed to the open-source community. All the software in the Virginia Tech Digital Library Platform is open to the general public. Currently, there are seventeen software and libraries published on GitHub. We also contribute our work to several open-source projects, such as Fedora and Samvera. Through open-source software, it allows anyone to reuse or redistribute the code and further leads to collaboration and communication and better software could be created.

Invisible Labor

The success of libraries is often determined by metrics on the number of patrons, the number of digital items made available, the amount of money saved to patrons by using the library, and other quantifiable numbers (Clark, 2022). Measuring this impact is vital to the health and growth of a library, but may not explicitly include the labor to appraise content, digitize content, create metadata, develop digital services, implement digital services, or preserve content. At VTUL for example, we made an organizational shift in our strategic plan to focus on advancing impact, university goals, diversity, accessibility, and physical space usage. None of these categories include digital library development and improvement, indicating that our shift in priorities does not include highlighting backend work in systems, software development, digital preservation, and other invisible labor on a public-facing platform. Sholler (2019) states, "it is difficult to quantify the research value of any technology in science," rendering this work invisible. Internally, libraries operate in silos that separate departments to their individual work (Kowalski, 2017), making communication and collaboration difficult to achieve. IT units and other backend work can go unnoticed. We hope to improve the transparency, understandability, and reusability of related resources at VTUL.

Methodologies

We used a multi-methods approach through informal interviews with VTUL stakeholders, and a web survey distributed to the wider communities of IT advisors, digital library managers, and digital preservationists to explore the following research questions.

Informal Survey

Research Question 1: Are VTUL digital library resources understandable and reusable to VTUL stakeholders?

We first want to measure the usefulness and reusability of policy within our own University Libraries. From 2019-2021 we developed a series of publicly accessible policies for each primary service provided by the DLP designed to convey transparent practices and decision-making. VTUL

has an IT Advisory Group[5] consisting of representatives from multiple departments in the University Libraries that "provide a platform for communication and promote transparency and inclusion between Library IT and other University Library departments." This group was the governing body to review and approve all of the policies developed for our DLP. We will engage this group with a short survey to gauge collective agreement, feedback on the process, and their interpretation of the reusability of our digital library policies. Survey questions will include:

1. As an IT Advisory Group member, are you familiar with the process for reviewing and approving policies for the digital library?
2. What was most understandable about this process?
3. What was the least understandable about this process?
4. Do you feel that you would reuse these policies in the future for your own work?
    a. If yes, how so?
5. Do you feel that a colleague in other University Libraries would be able to understand and reuse the digital library policies for similar purposes?
6. Do you have any other comments you would like to include?

Web-based Survey

Research Question 2: Are VTUL digital library resources understandable and reusable to organizations external to VTUL, and how can we improve our resources for reusability?

This web-based survey (Virginia Tech IRB #22-430) consists of eight questions designed to explore the impact of our open policies and open software outside of our University Libraries and collect feedback on how to make our resources more reusable based on community needs. This survey was distributed through known listservs in the communities of IT advisors, digital libraries, and digital preservation.

The questions in the survey included:

1. What is your field?
2. Please indicate all of the areas your position interacts with in your organization [check from list]
   Please navigate to https://apps.es.vt.edu/confluence/display/LIBDPLD/Digital+Library+Policy+and+Documentation
3. How likely are you to reuse or adapt other digital library policy for your own work? [scale 1 - 5]
4. Regarding policy, what is the most important factor to you when you consider reusing resources? [check from list]
   Please navigate to https://github.com/vt-digital-libraries-platform
5. How likely are you to reuse or adapt other digital library code for your own work? [scale 1 - 5]
6. Regarding digital library code bases, what is the most important factor to you when you consider reusing resources? [check from list]

[5] https://apps.es.vt.edu/confluence/display/LIBDPLD/Advisory+Council+on+Digital+Collections

7. If you wanted to make your digital library resources more reusable and discoverable, how might you go about doing so? [free text]
8. Do you have any other comments you would like to share regarding digital library resources? [free text]

After the survey closed, survey results were anonymized and each question's responses were condensed into broader topics for analysis.

a. Limitations

Our informal survey was distributed only to a single group within VTUL and was designed to gauge familiarity and understanding rather than to gather formal data. The IRB web-based survey combines policy and software and may have deterred participation due to lack of experience in either field. Combining both areas also prevents us from distinguishing which responses might be indicative of the respondent's expertise. This paper is limited to a scope of VTUL digital library policies and software and does not necessarily apply to other universities. We hope to be a guide for other universities developing or assessing their own digital library software and policies.

We had intended to include usage statistics in this report but previously unknown obstacles deterred this. Our Confluence is managed by the University Division of IT and denied a request for acquiring usage statistics using either the built-in Analytics feature or embedding Google Analytics tags into our page and does not maintain statistics themselves.

Results

Informal Survey
Research Question 1: Are VTUL digital library resources understandable and reusable to VTUL stakeholders?

The IT Advisory Group has 25 members, including 6 individuals that have recently rolled on as others rolled off and do not have experience with the policy development process. The internal survey distributed to this group revealed several gaps in familiarity with digital library policies and documentation available at VTUL. Respondents familiar with this process were evenly split with half having some experience and half being very unfamiliar. We have not published a new policy since 2021 so this was not surprising. For those who were familiar, respondents indicated that the purpose of the policies and the timelines were very clear, but that the least clear component was providing feedback. 75% of respondents indicated that they would reuse these policies for their own use, but only 25% feel that a colleague in another University Library would reuse these policies.

This brief overview reveals that even internally there is a lack of communication about the existence and purpose of our digital library

policies that can be improved. The silos in libraries that Kowalski (2017) mentions apply to VTUL in that digital library infrastructure labor is not as widely understood as patron-based services. While many of the technical policies and raw software code may not be directly relevant to our colleagues' work, we want to increase understanding and ensure that if there is a need for reuse, it is reasonably accomplished. We have experienced the need in the development of Memorandums of Agreement across departments, all of which are formatted consistently and approved by our Legal Department. The DLP is not the only digital library platform we utilize at VTUL. Similar policy and software needs may be shared across departments and would benefit from a decreased expenditure of time researching and developing these resources.

Web-based Survey

The survey results are based on 49 completed responses. Below is a brief overview of the response to each question. The full and anonymized dataset is publicly available in the Virginia Tech Data Repository (Kinnaman & Chen, 2022).

1. Field of study
   The top four fields were Digital Libraries/Archives (25%), Libraries (14%), Digital Preservation (11%), and Information Science (11%). Other fields ranged from Physical Archives, Digital Humanities, Library Systems, Computer Science, and Research Support. This was expected and indicated we received feedback from our desired communities.

2. Interacting departments
   Participants were asked to indicate all of the areas your position interacts with in your organization. The provided list included Digital Libraries, Information Technology, Digital Imaging, Metadata, Data Management, Digital Preservation, Special Collections, and a write in option.

   Results indicated a wide range of interdepartmental collaboration with 16% interacting with Digital Libraries, Information Technology and Digital Preservation; 14% interacting with Metadata; 12% interacting with Special Collections; 10% interacting with Digital Imaging; and an additional 3% for write in options. Write-in responses included physical preservation and conversation, scholarly communications, leadership and administration, and access/discovery services.

3. Policy use/reuse
   Respondents were asked to navigate to the Virginia Tech Digital Library Policy and Documentation page and rate on a scale of 1 - 5 on their likelihood to reuse or adapt the policies Virginia Tech has available, with 1 being not likely and 5 being highly likely. The breakdown of responses is the following:

1 (very unlikely): 10.64%
2 (not likely): 8.51%
3 (neutral): 21.28%
4 (likely): 27.66%
5 (highly likely): 31.91%

These results were expected. Over half of the responses indicate that people would be likely to highly likely reuse and adapt the VTUL digital library policies on various levels, indicating a noted desire to have digital library policies available and accessible for reuse. The other percentage of respondents that would be less likely to reuse or adapt may not have found a resource relevant to their work or do not work in a field where they would need to interact with similar policies. Additional feedback is discussed in the results for Question 8 on aspects that would allow for more effective reuse and adaptation.

4. Important factors for policy reuse
Choosing one option from a provided list, respondents indicated the following were the most important factors:
Ease of reuse or adaptation: 27.37%
Relevance to your work: 25.26%
Usability/understandability: 24.21%
Accessibility: 11.58%
Discoverability: 9.47%
Other: 2.11%
The Other option received two write-in answers for licensing and thoroughness. Ease of reuse, relevance, and usability were comparable in terms of importance to the respondents.

5. Software use/reuse
Respondents were asked to navigate to the Virginia Tech Digital Library Platform GitHub and asked to rate on a scale of 1 - 5 on their likelihood to reuse or adapt the policies Virginia Tech has available, with 1 being not likely and 5 being highly likely. The breakdown of responses is the following:
1 (very unlikely): 8.7%
2 (not likely): 26.09%
3 (neutral): 26.09%
4 (likely): 28.26%
5 (highly likely): 10.87%
These results indicated that respondents are more hesitant to use and reuse and showed a varied range between not likely - likely with smaller responses for either very unlikely or highly likely.

6. Important factors for software reuse
Choosing one option from a provided list, respondents indicated the following were the most important factors:
Ease of reuse or adaptation: 27.84%

Usability/understandability: 26.8%
Relevance to your work: 24.74%
Accessibility: 8.25%
Discoverability: 6.19%
Other: 6.19%

The Other write-in options consisted of quality, community involved in coding, ability to cut and paste modules or snippets, current staffing of a library IT institution, and sustainability. These factors are much more specific than the provided options, which could indicate that software developers either have more guidance on their software requirements, or need more precise factors when considering reusing codes or code snippets. As with Question 3, additional feedback is discussed in the results for Question 8 on aspects that would allow for more effective reuse and adaptation.

7. Resource discoverability strategies
   Respondents provided a wide range of strategies and comments on how to increase reusability and discoverability in an open answer, write-in option. The top comments were storage on GitHub, providing clear documentation, providing clear licensing for reuse, and promotion at appropriate conferences and listservs. Additional comments included utilizing sitemaps for increased discoverability, improving the metadata quality of resources, and providing resources specifically on how to implement software. Other unique suggestions were to assign PIDs to policies, developing a consortia-managed wiki for resources, translation, and applying FAIR principles to the software.

8. Additional comments
   In addition to specific ideas on increasing usability and reusability, respondents left several thoughts that reflect general sentiments, such as “There is very little in my work that I have not borrowed from somewhere or learned from someone else” and “It'd be helpful if people didn't go into something with the mindset that they [...] need to reinvent the wheel.” Others provided broader ideas, such as “a Linktree or aggregator for such documentation might be cumbersome to maintain, but invaluable to the profession,” and “DOCUMENTATION.”

Through the case study interviews and test cases and the wider survey, we gained valuable information that has informed our decisions on what actions we will take to better prepare our digital library resources for reuse.

Interventions & Design Considerations

These results have encouraged us to continue with several of our current strategies, particularly on hosting policies and software in publicly accessible locations and always improving our documentation. We will also continue to share and ask colleagues for feedback for improvement, and maintain regular updates and reviews as needed.

In the short term we will be focusing on licensing, documentation, and further promotion. We have added a CC 1.0 Universal License to the main Digital Library Policy and Documentation page and will take steps to add them to each policy individually to clearly indicate that our digital library policies are fully reusable and adaptable. For policy and documentation, we will be adding other policies that we have referenced into the Resources section for each policy. We also plan to develop supplementary documentation that denotes sections specific to Virginia Tech. All of our software code is shared publicly on GitHub and is open for contribution and suggestions and provides some instruction on how to deploy our version of our digital library. However, the site lacks specific licensing on reuse that may deter users from reusing the software code. We plan to include licensing and increase current guidance on how to use the software and overcome common errors that may occur. We have promoted our policies and software at various conferences, but have not distributed via social media or relevant listservs. As we begin to actualize these short term goals, we will develop an intentional social media plan to share this work for the purposes of feedback and reuse.

In the medium to long-term, we hope to create policy templates for each policy we have that are more generalizable to a wider audience. We also hope to ingest all of the policy resources into our digital library platform to codify policy, assign identifiers, and embed quality metadata for greater discoverability. For our software, we hope to utilize sitemaps and other discovery tools for greater discoverability and implement more integration with other metadata harvesters like DPLA.

Although we at VTUL do not have bandwidth to lead larger movements alone, the results indicated that a larger digital library policy and software community would be beneficial. This could take the shape of a consortium that develops, shares, and maintains digital library resources in a single, aggregated location similar to the Linktree discussed above. This could also be limited to shorter events like policy edit-a-thons or software reviews to note where commenting would be especially helpful.

Conclusion

This paper explored the results of a web survey distributed to the communities of IT advisors, digital library managers, and digital preservationists that both rated VTUL digital library policies and software,

and provided feedback for improving resources at VTUL for enhanced reusability and discoverability. The results shaped specific interventions and design considerations we will apply in the short and medium-term to improve our digital library policy and documentation, and share these evolutions with relevant communities to continue supporting reusability across various communities involved in backend digital libraries work.

References

[report] Beagrie, N., Semple, N., Williams, P., and Wright, R. (October 2008). "Digital Preservation Policies Study." JISC. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.214.9056&rep=rep1&type=pdf

[journal article] Borgman, C. (2003). The invisible library: Paradox of the global information infrastructure. Library Trends (51). http://hdl.handle.net/2142/8487

[online magazine] Clark, R. (2022, January). Work made visible: What's lost when we obscure labor. American Libraries Magazine. Retrieved from https://americanlibrariesmagazine.org/2022/01/03/work-made-visible/

[report] Consultative Committee for Space Data Systems. (2012). Audit and certification of trustworthy digital repositories (Magenta Book CCSDS 650.0-B-1). Retrieved from http://public.ccsds.org/publications/archive/652x0m2.pdf

[report] CoreTrustSeal Standards and Certification Board. (2019). CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022 (v02.00-2020-2022). Zenodo. https://doi.org/10.5281/zenodo.3632533

[report] Electronic Resource Preservation and Access Network. (2003). "Digital Preservation Policy Tool." ERPAnet. https://www.erpanet.org/guidance/docs/ERPANETPo

[online magazine] Farkas, M. (2015). "Reuse, recycle, share: Let's stop reinventing the wheel." American Libraries Magazine. https://americanlibrariesmagazine.org/2015/02/12/reuse-recycle-share/

[journal article] Frakes WB, Kang K (2005) Software reuse research: status and future. IEEE Trans Softw Eng 31(7):529–536

[journal article] Gremme, G., Steinbiss, S., & Kurtz, S. (2013). GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM transactions on computational biology and bioinformatics, 10(3), 645-656.

[dataset] Kinnaman, A & Chen, Y (2022): Building a Culture of Reuse: An Analysis of Reusable Software and Policies for Institutional Libraries - Web Survey. University Libraries, Virginia Tech. Dataset. https://doi.org/10.7294/20234097

[journal article] Koulouris, A., Kapidakis, S., & Mekdon, F. (2005). Applying library policies for digital content access to the commercial sector. 1st Hellenic Society for Systemic Studies National Conference 2005 (HSSS 2005), in co-ordination with the University of Peloponnese, Tripolis, Greece, May 12-14, 2005 (pp. 719-727). Retrieved from https://www.researchgate.net/publication/28805345_Applying_library_policies_for_digital_content_access_to_the_commercial_sector

[journal article] Kowalski, M. (2017). Breaking down silo walls: Successful collaboration across library departments. Library Leadership & Management 31(2). doi.org/10.5860/llm.v31i2.7202

[journal article] Krueger, C. W. (1992). Software reuse. ACM Computing Surveys (CSUR), 24(2), 131-183.

[journal article] Lakhan, S. E., & Jhunjhunwala, K. (2008). Open source software in education. Educause Quarterly, 31(2), 32.

[journal article] Noonan, D. (July 2014). "Digital Preservation Policy Framework: A Case Study," EDUCAUSE Review online, 49(4). http://www.educause.edu/ero/article/digital-preservation-policy-framework-case-study

[conference proceedings] Sholler, D. (2019). "'Invisible Work' as a lens for understanding humanware's role in research: Cloud computing evidence from an interview-based study." Proceedings of the Humans in the Loop: Enabling and Facilitating Research on Cloud Computing. DOI: 10.1145/3355738.3355754

[journal article] Wang, S., Liu, Y., & Padmanabhan, A. (2016). Open cyberGIS software for geospatial research and education in the big data era. SoftwareX, 5, 1-5.