How do users discover digital Preservation tools?

Report on a survey of professionals

|  |  |  |
| --- | --- | --- |
| **Amber L. Cushing** | **Tess Burchmore** | **Sarah Conroy** |
| *University College Dublin*  *Ireland*  *Amber.cushing@ucd.ie*  *ORCID:* 0000-0003-0186-0689 | *University College Dublin*  *Ireland* | *University College Dublin*  *Ireland* |
| **Phoebe Doyle** | **Niamh Hegarty** | **Rebecca Kelly** |
| *University College Dublin*  *Ireland* | *University College Dublin*  *Ireland* | *University College Dublin*  *Ireland* |
| **Payton Kufeldt** | **Morgan McGann** | **Cian Ormond** |
| *University College Dublin*  *Ireland* | *University College Dublin*  *Ireland* | *University College Dublin*  *Ireland* |
| **Gerard Quine** | **Martina Reba** | **Ronan Woods** |
| *University College Dublin*  *Ireland* | *University College Dublin*  *Ireland* | *University College Dublin*  *Ireland* |
|  |  |  |

**Abstract – While information is available that charts the development of digital preservation tools and their use via case studies, less is known about *how* users discover and adopt digital preservation tools in different contexts. This study reports on a short survey of 68 professionals who utilize digital preservation tools, and how they discover and adopt those tools. Findings suggest that the role of community is important when discovering and adopting tools. Findings were then applied to digital preservation education to inform the ways in which tools are taught in formal digital preservation education programs.**

**Keywords – Digital preservation tools, surveys.**

**Conference Topics – Community, innovation.**

# Introduction

Part and parcel of digital preservation is the use of tools to complete to complete preservation activities. As registries demonstrate, there are a variety of proprietary and open source tools available for use. Previous research has explored digital curation competencies [1], [2] educating digital curators [3], [4] case studies of use of certain tools and services for example [5], [6], and the consideration of tools in digital preservation workflows [7].

Yet, there is limited understanding as to *how* the professionals who engage in digital preservation discover and select these tools for use. For example, if someone was just starting out, where would they start? How would they find and select the most useful, relevant tools for their situation? While this question can help address the how and why of tool selection, it also has implications for digital preservation education.

Recently, Yoon et al. [8] surveyed digital preservation syllabi in the US and Canada and found the role of teaching digital preservation tools via class activities to be an important part of formal digital preservation education. However, how can educators best go about teaching use of these tools? File formats will change over time and digital objects may become increasingly complex, so how can educators provide students with lasting skills, beyond just providing practice with the tools that are currently favored?

This paper reports on a short survey of digital preservation professionals and how they select tools for use. The survey was completed by students enrolled in the Digital Curation: Core Concepts module at University College Dublin between September – December 2021. While the findings of the survey are limited based on the sample size, they can elaborate an understanding of the context of digital preservation work practices, by shedding some light on how professionals seek out resources to support tool discovery and selection. The survey results can also be viewed in the context of digital preservation education. While the survey has the potential to inform how tools emerge and may grow a user base, it can also inform the ways in which educators approach tool use and education in the context of professional digital preservation education.

## A. Selecting and using digital preservation tools

Overall, most advice related to selecting tools for digital preservation is organized via the two most well-known models governing digital preservation practice: the DCC Digital Curation Lifecycle and the OAIS Reference model. While these models organize information about tools and assist the user in discovery of tools information, little is known about how individual users discover newly created digital preservation tools and software solutions. Peer reviewed journals often struggle to keep up with the ever-changing open-source technology landscape. Conferences, blogs, websites, message boards, and social media platforms often function as the most up to date information available about digital preservation tools, advising on practical aspects of various tools and systems associated with digital preservation [9]. The Digital Preservation Coalition [10] exists “to secure digital legacy” worldwide. Their website provides a comprehensive guide to the use of open source and proprietary tools as part of a digital curation workflow, making clear that the user must consider the institutional setting when working with combinations of tools to reduce costs.

In her review of the state of the art of digital preservation in 2018, Rieger [11] listed “the availability of new preservation systems and tools” as part of “what’s working well” in the field. Yet, availability of tools does not always translate to a general understanding of how to discover which tool is going to best fit specific work contexts. The *Digital Preservation Handbook’s* [10] “Technical Solutions and Tools” section advises that “before selecting digital preservation tools it is important to consider carefully the technical workflow and institutional setting in which they are embedded.” The *Handbook* references the Northumberland Estates Case Study, which recommends evaluating new tools using a “product analysis scorecard” which helps the user map the product/tool to the OAIS reference model compliance requirements. The guide also recommends assessing how the tool will deliver preservation actions discussed in an existing preservation plan.

The relationship between tools and workflows has enjoyed recent attention in the OSSArcFlow project. According to the project’s *Guide to Documenting Born-Digital Archival Workflows,* the purpose of the project was to document born digital archiving activity and to offer advice in the selection and implementation of workflows. According to the project guide, institutions report having “to manage significant gaps and overlaps between different tools and environments” [12]. The guide lists tools that are commonly utilized at different steps of a digital preservation workflow and emphasizes the importance of assessing how tools may be compatible within a current or ideal workflow. The guide focuses on how practitioners utilize tools on the context of a workflow, but tool selection is not discussed specifically in detail.

The *Handbook* [10] also highlights the role of Digital Preservation Registries, particularly the Community Owned Digital Preservation Registry (COPTR), a wiki based registry of digital preservation tools [13]. According to the webpage, COPTR’s “main aim is to help practitioners discover preservation tools that will help them tackle particular preservation challenges.” The Registry allows the user to search for tools according to DCC Digital Curation Lifecycle stages which was developed to assist in research data management, function, content type, and file format. COPTR’s strengths lie in it’s ability to act as a clearing house for digital preservation tools information, but providing links to further, more detailed information about digital preservation tools. However, while COPTR is well positioned to provide access to information about tools and software solutions, there is less information available about how to select a tool to fit a specific context.

One of the most common formats that offers evaluation of digital preservation tools is the practitioner case study. In this format, the user describes their use of a tool/software solution in the context of their own individual case. These case studies can be found in formal venues such as the *International Journal of Digital Curation*, as well as conference series, such as *IPRES,* and informal venues, such as blogs*.* While practitioner case studies can provide insight into how a tool might work in a specific situation, the cases typically focus on implementation and review, with little detail about how one learned of the tool/software in the first place. For example, Trujillo et al. [5] describe how the five college library team came together to assess a need to work as a team to address digital preservation, including readiness for such a program. Consultation of the POWRR Registry of tools is mentioned, but there is less detail about *how* the team came to adopt Archivematica.

Another venue for the discussion of digital preservation tools is the communities that develop around the use of these tools. Tools typically fall into the categories of proprietary or open source, and both categories are accompanied by user communities that share advice and reflection about use of the digital preservation tool. Informal discussion occurs via social media, message boards, and email lists. An active community can be central to the development of an open source tool in particular, as it allows for up-to-date discussion on the which to complete tasks and to understand what needs to be done for these tools to be improved upon. These communities are worldwide, but some are targeted to a smaller, local community, such as the Dutch Digital Heritage Network “Erfgoedkit,” which provides Dutch language support for archivists selecting digital preservation tools to enhance digital heritage efforts [14].

## B. Tools and digital preservation education

In the last two decades significant work has been published about digital preservation education, providing snapshots of how competencies may remain stable or change over time. This work has explored digital preservation competencies, the teaching of technology skills in digital preservation, and the content of digital preservation courses and modules [1]. [2], [3], [8], [15], [16]. Much of this work cements the need to provide students with the ability to understand technology but does not offer substantial detail about how best to teach students to find, discover, and use technical tools.

Starting in 2007, Lee et al. [1] list “understanding technology” as a necessary competency. In their discussion of digital stewardship pedagogy, Bastian et al. [3] describe technology as a necessary competency because technology is necessary in the oversight of collections, but stress that this understanding of technology must be grounded in context. The authors also discuss their development of a Digital Curriculum Laboratory, where “users of the laboratory can experiment with and evaluate tools and standards for their relevance to the kinds of content” via teaching scenarios designed by educators (p. 617). Like Bastian et al. [3], Feng and Richards [2] utilized literature review analysis and found that “hands on” technical practice in digital curation education is vital.

Using literature analysis, Kim et al. [16] list “understanding software” as a necessary knowledge and skill for digital curators. Yet in their analysis of job postings, the authors found that knowledge, skills, and abilities associated with specific tools and applications was requested in 45% of job postings analyzed in 2013. “Working in an information technology-intensive environment” was listed in 50% of the job postings. Kim et al. [16] summarized these findings as a “curation technologies competency” that included “competency required to identify, use, and develop tools and applications to support digital curation activities” (p. 79).

Yoon [8] analyzed 59 digital preservation syllabi (US and Canada only) to develop their findings and found “a need to integrate digital preservation tools and technologies into course content through class activities” (p. 1). In their case study, Cushing and Shankar [15] found that practitioners desired continuing professional development (CPD) education about how to use digital curation tools.

Existing research clearly demonstrates the need for digital preservation education to include how to select and use digital preservation tools, but this research lacks specific detail about how knowledge related to how to select tools can be delivered to students effectively.

# Methods and procedure

The current study was designed by the primary author to answer the question, “how do digital preservation practitioners discover and select digital preservation tools?” In addition to responding to the research question, a secondary goal of the project was to improve student knowledge about digital preservation practice, and also to develop student research skills in relation to digital preservation work and work practices.

The study was completed by eleven students enrolled in the 2021-2022 Term 1 Digital Curation: Core Concepts graduate level module at the UCD School of Information Studies. The students worked in teams of 2-3 students. After one team of students completing a preparatory literature review, another team designed the data collection methods for the study. A pair of students then developed the questionnaire and used it to collect the data, and a final team of three students conducted data analysis. The primary author completed the final analysis and discussion.

The online survey was targeted towards members of the digital curation field. As such, no specific group was pre-selected to participate in answering the survey questions. To reach working digital preservation specialists, the questionnaire was created and then subsequently shared by the University College Dublin School of Information and Communications Twitter channel because this platform allows for fast communication to a large group of people. To further our reach, people already working in the field were also asked to share a link to the online survey with their contacts through word-of-mouth recruitment. A well-connected digital preservation manager also distributed the call for participants via several listservs and on their own Twitter account.

SurveyMonkey was used to administer the 13-question survey which contained questions relating to the use and implementation of digital preservation tools. The survey was open for a two-week period in late Autumn, 2021. Permission for researchers to collect and analyse the answers given by participants was obtained using an Information Packet. Any free text that identified places, names, or other personal details was deidentified during data cleaning and analysis.

The survey included a mixture of question types to allow for numerical data to be extracted as well as to allow participants to express their opinions on current digital preservation practices. Questions were created following the consultation of several sources on the state of current digital preservation practices, such as the Community Owned digital Preservation Tool Registry (COPTR) which provided a list of current tools used in the field. Question types included yes/no questions, open-ended multiple-choice questions, and a singular short answer question which asked participants to express any outstanding opinions on the implementation and use of digital preservation tools that were not covered by previous survey questions.

# Results

In total, 68 participants completed the survey. Numerical data from the yes/no and multiple-choice questions was analysed using descriptive statistics produced in Excel to highlight any trends in the data.

The participant who respondent to the questionnaire were frequent users of digital preservation tools. In figure 1, most reported using tools daily.

Table 1 providesmultiple choice answers to the question: how do you find information about useful tools? Among the participants, 70.59% answered that it depends on a project by other institutions/organizations. The most popular responses were conferences and online communities. Next, 48.53% chose word-of-mouth and 45.59% chose blogs as their source of information. Finally, 30.88% of those who responded to the survey used social media, while 27.94% used other sources.

Figure 1. How often do you use digital preservation tools?

|  |  |  |
| --- | --- | --- |
| Format | Quantity | Percentage |
| Following projects | 48 | 70.59 |
| Conferences | 46 | 67.65 |
| Online forums | 45 | 66.18 |
| Word of mouth | 33 | 48.53 |
| Blogs | 31 | 45.59 |
| Social media | 21 | 30.88 |
| Other | 19 | 27.94 |

Table 1. Responses to the question “how do you find information about useful tools?”

When asked to select the sources for information on useful tools, 70.59% of participants reported that they looked to another institution facing a similar situation for advice. After that, 48.54% of participants looked to conferences and online communities for information, followed by 30% of participants who relied on social media channels for this information. Lastly, 27% of participants selected “other sources.” These “other” sources included the Digital Preservation Coalition (DPC), Google/Web search, COPTR, Open Preservation Foundation (OPF), blogs, International Internet Preservation Consortium (IIPC), literature, peer recommendation, or sector training. The results of this question are not mutually exclusive, as many sources listed under “other” are considered to be digital preservation communities.

Next we asked which digital preservation tool features were most valued (see table 2). Open source was considered to be the most valued feature of a tool, with 76.47% of participants selecting this choice. Graphical User Interface was the second most valued feature of participants (60.29%), followed by API (30.88%), and Command-line (26.47%). Only 14.71% of participants’ most valued paid commercial support. Responses to “other” (13.17%) included an active online community, good documentation and sufficient functionality.

Next, participants were asked to give their views on the following statement: “It is extremely difficult to discover and choose suitable digital preservation tools” using a Likert scale. Figure 2 illustrates this, with 27 participants agreeing, 23 being neutral, and only 16 participants disagreeing and believing that it is easy to discover and choose digital preservation tool.

|  |  |  |
| --- | --- | --- |
| Feature | Quantity | Percentage |
| Open source | 52 | 76.47% |
| GUI | 41 | 60.29% |
| API | 21 | 30.88% |
| Command line | 18 | 26.47% |
| Other | 13 | 19.12% |
| Paid commercial support | 10 | 14.71% |

Table 2. The most valued features of a digital preservation tool

Figure 2. Responses to “It is extremely difficult to discover and choose suitable digital preservation tools”

When asked how much consideration is placed on file formats and standards when making the decision to use a digital preservation tool and/or software, most participants (51.47%) selected “a lot” followed by “some” (30.88%), “not applicable” (11.76%), “not much” (5.88%), and “none” (1.47%). In addition, when asked how much consideration is placed on information security when searching for a digital preservation tool, nearly half of participants (48.52%) selected “a lot”, followed by “some” (33.82%) “not applicable” (11.76%), “not much” (4.41%), and “none” (1.47%).

A Likert scale was used to gather responses to the statement “I place an extremely high priority on finding digital preservation tools that are sustainable with active community participation” (Figure 3).

We asked participants if they are frequent users of an online community for digital preservation tools. Slightly over half of participants (52%) stated that they were not part of an online community. Of those who were part of an online Community, the most used groups were the DPC community, Bitcurator community, and Preservica community.

Figure 3. Responses to the statement “I place an extremely high priority on finding digital preservation tools that are sustainable with active community participation.”

There was little overlap when asked about the use of specific tools: participants listed 48 different tools, with 34/48 tools listed being used by one person. The most common tools listed were JHOVE (15 participants), Bitcurator (7 participants) Archivematica (5 participants), and Exiftool (5 participants). Participants reported less autonomy in making decisions to utilize proprietary tools and software, with 48/68 (70.58%) reporting that the decision to adopt a proprietary tool often requires administrative, IT department, and budget approvals.

Lastly, we asked participants about the importance of Registries such as COPTR when finding digital preservation tools (Figure 4).

Figure 4. How important are registries such as COPTR and PRONOM in finding the digital preservation tools/software you may need?

# Discussion

While the survey data collected provides results concerning how digital preservation tools are discovered and selected, the survey results can also contribute to a discussion of how to approach the teaching of digital preservation tools in formal digital preservation education programs. The survey results will be discussed in both of these contexts.

The survey sample size was small (68 participants), and as such, results are not generalizable. In addition, future research that involves qualitative data collection on the issue, such as a focus group, could provide context to survey results.

*A. Digital preservation tool discovery and selection*

Overall, the results of the survey highlight different factors that influence tool discovery and selection. Participants seemed to prefer open source to proprietary tools, with most participants citing open source as an important feature of a tool. The greater digital preservation community also plays a role, which is not surprising. However, as demonstrated in the question: “how do you find information about useful tools?” and potentially the question about frequency of use of communities, participants may subscribe to different definitions of the term “community” in this context. For example, is a professional organization also a community? Must a community be aptly named, such as a space named an “online community forum”? Is social media a “community”? Results suggest that a “community” holds value for the discovery of digital preservation tools, but the definition of what constitutes a “community” lacks agreement. Whether this term needs defining in this context, and what benefits that might bring is unclear from the results. However, understandings of the term community in the context of digital preservation may prove a fruitful avenue for future research.

The role of community is exemplified via upkeep of the community owned COPTR registry, which acts as a clearinghouse of information about digital preservation tools and how they can be utilized. This is supported by the finding that most participants selected “very important” in response to the question that asked about the importance of tools registries. The role of a community supporting a tool was also considered to be important, with the majority of participants’ placing a high priority on tools that are accompanied by an activity community of users. This is not surprising, as tools with an active community can lend support to someone just starting out with the tool and also provide guidance on situations where the tool may be most useful. Future research might dig further into these user communities, to discern the characteristics of the communities that are considered most successful and the relationship between community and tool success and longevity, especially since many open-source tools are developed and launched using temporary financial support. As such, understanding user communities (however they may be defined) may help to understand how to make digital preservation tools more efficient and sustainable and better support novice users.

While the role of communities and community owned resources like COPTR are important, it is concerning that even though these resources exist and many find them useful, most participants agreed with the statement that it was extremely difficult to discover new digital preservation tools. A natural follow up query when viewing this data is to ponder “why?”—what makes it so difficult to discover new digital preservation tools? Rieger’s [11] challenges offer some context to position this finding. Rieger [11] cites “lack of assessment metrics” as a challenge for digital preservation. Rieger’s statement is most likely meant to apply to the digital preservation field in general, as she points out the lack of “collaborative approaches to explore what constitutes success and how we identify it, and measure outcomes associated with digital preservation” (p. 12).

However, the same could be said for digital preservation tools: while COPTR provides information about tools and how they can be used, there is less information available that provides an assessment of tools, beyond case studies. These case studies can provide insight into how a tool was used, but the case study format is not easily compared and assessed. Further case studies can be difficult to discover, especially if they were not formally published in a conference proceeding or journal this is indexed for discoverability. The ability to easily compare and assess tools may be one of the vital missing pieces that allows a user to adequately discover *and then* compare different digital preservation tools. Perhaps, this can be addressed with the addition of an index of case studies of specific tools linked to the tools listed in the COPTR registry. This could ease discoverability of assessment information associated with tools and work toward empirical, generalizable assessment practices for digital preservation tools.

In addition, tool developers that host user communities may also create space to host case studies of tool use to ease discoverability. They could also go one step further and encourage and then highlight case studies of their tools. Finally, to build on Rieger’s [11] point, perhaps we should start to ask how best to measure success with a digital preservation tool and then formalize that measurement so that it can be accessible to the wider digital preservation community.

*B. Digital preservation education and digital preservation tools*

Previous research has made it clear that digital preservation education needs to address technology and the use of digital preservation tools, by providing students in-class experiences to engage with different digital preservation tools [8]. However, this in and of itself can be a challenge, an many academics do not use a variety of digital preservation tools on a regular basis and may need to rely on guest lectures from practitioners as the best resources to explain and demonstrate implementation to students.

Placed in the context of digital preservation education, the survey results suggest that in teaching digital preservation tools, educators may contextualize hands-on activities with tools with a greater discussion of how to go about assessing differences between digital preservation tools, when resources to perform this assessment may be lacking.

This is where the role of the digital preservation community in assisting with tool discovery, implementation, and use could be introduced to students. Use of the community could also be taught via different scenarios, as Batian et al. [3] suggests. These scenarios and teaching cases could allow students to assess a landscape and propose a new tool, which may include a requirement to explain how the tool may be integrated into the wider digital preservation community, as well as how the tool can be assessed. This aligns with the call for computational thinking in archival education, which includes the use and understanding of preservation tools [17].

Finally, it is worth noting that survey participants strongly preferred open source versus proprietary tools for digital preservation. This has implications for education, as students may need to ability to assess, compare, and contrast open source and propriety tool models. The best-case scenario would provide students experience with both formats, and the ability to make informed decisions of why to choose one model versus another model.

# Conclusion

This paper reports on a survey of 68 participants queried about the discovery and selection of digital preservation tools. Findings suggest that participants have difficulty discovering new tools and rely on tool user communities for support in selecting tools to meet their needs. However, there may be different understandings of what constitutes a community. Tool registries also play an important role in tool discovery, but information about how to assess and measure different tools is lacking and may aid future tool discovery and selection. In the context of education, digital preservation educators may be well placed to position hands on experience with digital preservation tools with learning about the vital role of digital preservation communities, as well as providing students with the skills to assess tools for use.

# REFERENCES

1. C.A. Lee, H. Tibbo & J. Schaffer, Lee, Christopher A., Helen R. Tibbo, and John C. Schaefer. "DigCCurr: Building an International Digital Curation Curriculum & the Carolina Digital Curation Fellowship Program," in *Proceedings of Archiving* May 21 - May 24, 2007, Arlington, VA, USA, Society for Imaging Science and Technology, 2007.
2. Y. Feng and L Richards, “A review of digital curation professional competencies: theory and current practices,” *Records Management Journal, vol*. 28, no. 1, pp. 62-78, 2018.
3. J. Bastian, M. Cloonan and R. Harvey, “From teacher to learner to user: developing a digital stewardship pedagogy,” *Library Trends* vol. 59, no. 4, pp. 607-622, 2011.
4. P. Botticelli, B. Fulton, R. Pearce-Moses, and C. Szuter, “Educating digital curators: challenges and opportunities,” *International Journal of Digital Curation* vol. 6, no. 2, 2011.
5. S. Trujillo, M. Bergin, M. Jessup, J. Radding, S.W. and McGowan, “Archivematica outside the box: piloting a common approach to digital preservation at the five college libraries,” *Digital Library Perspectives,* vol. 33, no. 2, pp. 117-127, 2017.
6. E. Iglesias and W. Meesangnil, “Using Amazon S3 in digital preservation in a mid sized academic library: a case study of CCSU ERIS digital archive system, *Code4Lib Journal,* vol. 12, 2010.
7. C. Post, A. Chassanoff, C. Lee, A. Rablin, Y. Zhang, K.. Skinner and S. Meister, “Digital curation at work: modeling workflows for digital archival materials,” *ACM/IEEE Joint Conference on Digital Libraries (JCDL)* 2019. Available: ACM Digital Library, [Accessed 7 March 2022].
8. A. Yoon, A. Murillo, and P.A. McNally, “Digital preservation in LIS education: a content analysis of course syllabi,” *Journal of Education for Library and Information Science,* v. 62, no. 1, pp. 61-86, 2021.
9. “Digital preservation tools and systems,” [no date]. Available: <https://www.nationalarchives.gov.uk/archives-sector/advice-and-guidance/managing-your-collection/preserving-digital-collections/digital-preservation-tools-systems>. [Accessed 7 March 2022].

[10] Digital Preservation Coalition, *Digital preservation coalition*

*handbook,* 2nd Ed. Available:

<https://www.dpconline.org/handbook>. [Accessed: 7 March 2022].

[11] O. Rieger, “The state of digital preservation in 2018: a

snapshot of challenges and gaps,” *Ithaka S+R,* [Online], 2018.

[12] A. Chassanoff, and C. Post, *OSSArcFlow guide to documenting*

*born digital archival workflows*, Available:

<https://educopia.org/ossarcflow/>, 2020. [Online]. [Accessed 7 March 2022].

[13] Community owned digital preservation tool registry (COPTR).

[no date]. Available:

<https://coptr.digipres.org/index.php/Main_Page>.

[14] “Toolbox for digital heritage” [no date]. Available:

<https://erfgoedkit.nl/>. [Accessed: 7 March 2022].

[15] A. Cushing and K. Shankar, “Digital curation on a small island,” *Archives and Records* vol. 40, no. 2, pp. 146-163, 2019.

[16] J. Kim, E. Warga, and W. Moen. “Competencies required for

digital curation: an analysis of job advertisements.

*International Journal of Digital Curation* vol. 8, 1 [Online].

Available: <http://www.ijdc.net/article/view/8.1.66>, 2013. [Accessed 7 March 2022].

[17] R. Marciano, “Toward a computational framework for library

and archival education: report on a preliminary literature and curriculum review” [1 April 2019]. [Online]. Available: <https://ai-collaboratory.net/wp-content/uploads/2020/04/imlsCASresearchGroupReport.pdf>. [Accessed: 7 March 2022].