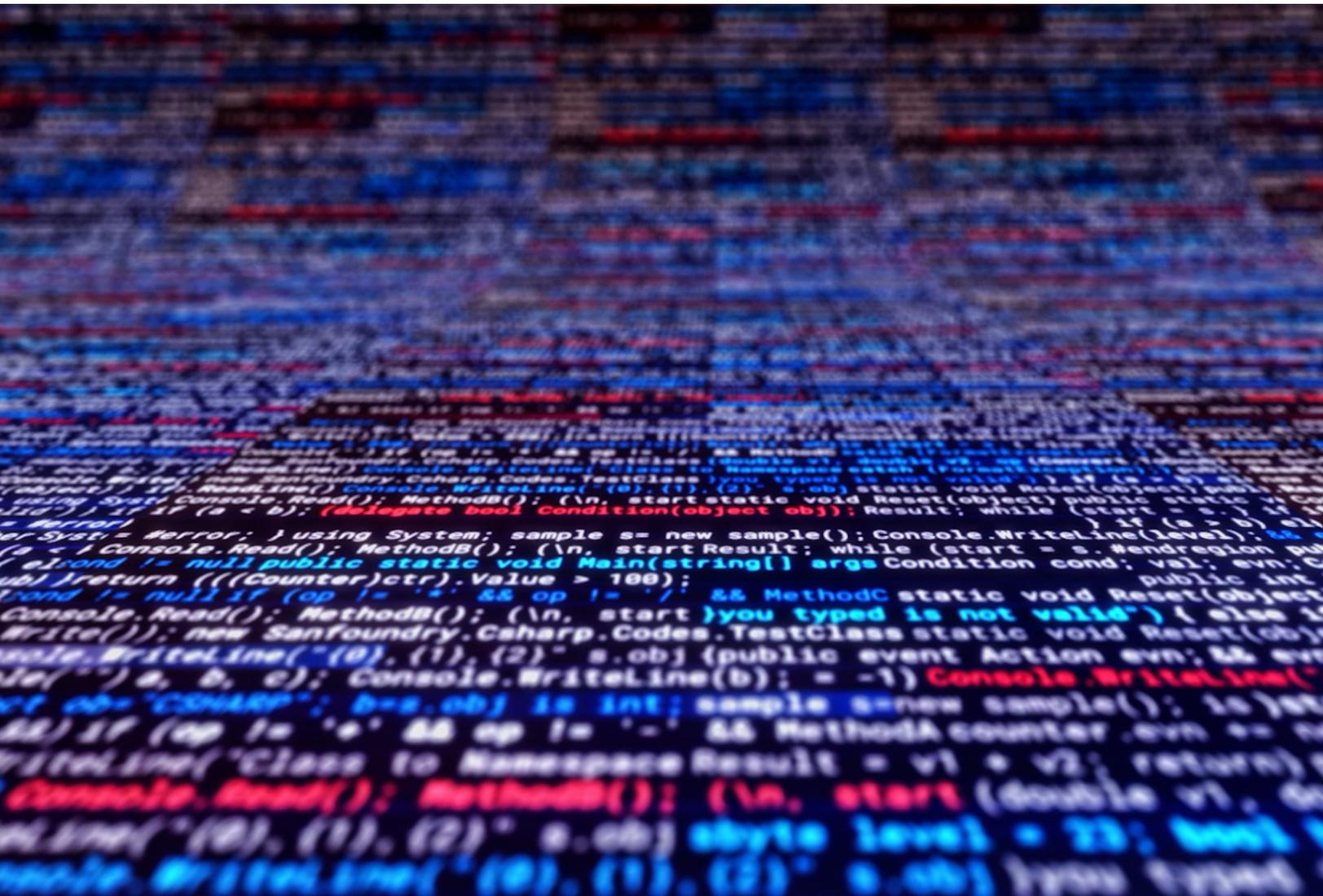


# Supporting Software Preservation Services in Research and Memory Organizations

A White Paper from the Software Preservation Network's Research-in-Practice Working Group

Jessica G. Benner, Seth Erickson, Wendy Hagenmaier, Monique Lassere, Christa Williford, Lauren Work





# Supporting Software Preservation Services in Research and Memory Organizations

A White Paper from the Software Preservation Network's Research-in-Practice Working Group

Jessica G. Benner, Carnegie Mellon University

Seth Erickson, Penn State University

Wendy Hagenmaier, Georgia Tech

Monique Lassere, Harvard University

Christa Williford, Council on Library and Information Resources

Lauren Work, University of Virginia

November 2022



Copyright © 2022 by Software Preservation Network and  
Council on Library and Information Resources

This work is licensed under a Creative Commons  
Attribution 4.0 International (CC BY 4.0) license



Co-published by

Software Preservation Network and  
Council on Library and Information Resources

This report and supplementary materials—including aggregated and anonymized survey results, the codebook used for analyzing interview transcripts, concept maps, and other materials—are available on Zenodo: <https://doi.org/10.5281/zenodo.7086618>.

**About the Software Preservation Network:**

The Software Preservation Network (SPN) is an independent, community-led membership association focused on ensuring long-term access to software through community engagement, infrastructure support, and knowledge generation. Guided by the belief that software is critical information infrastructure, SPN works to create a space in which organizations and individuals from industry, academia, government, cultural heritage, and the public sphere can contribute their myriad skills and capabilities toward collaborative solutions that will provide persistent access to software and software-dependent data.

**About CLIR:**

The Council on Library and Information Resources (CLIR) is an independent, nonprofit organization that forges strategies to enhance research, teaching, and learning environments in collaboration with libraries, cultural institutions, and communities of higher learning.

---

# Contents

<b>Executive Summary</b> .....	<b>1</b>
<b>1. Introduction</b> .....	<b>4</b>
<b>2. Study Overview</b> .....	<b>5</b>
Survey Design .....	5
Demographics of Respondents .....	6
Initiating and Conducting Follow-up Interviews .....	7
Coding and Analysis of the Data .....	8
Limitations of the Study .....	8
<b>3. Software Preservation as a “Field”</b> .....	<b>10</b>
Demographics of the Software Preservation Community .....	10
Four Personas of Software Preservation .....	13
Conceptions of Software in Software Preservation Work .....	16
<b>4. Barriers to Preserving and Providing Access to Software</b> .....	<b>18</b>
Professional Level .....	18
Institutional Level .....	19
Individual Level .....	21
<b>5. Moving Forward</b> .....	<b>24</b>
Theme 1: Representation in the Field .....	26
Theme 2: Defining the Field, for the Field .....	28
Theme 3: Networking and Community Building .....	30
Theme 4: Informal and Formal Learning .....	33
Theme 5: Shared Infrastructures and Model Practices .....	35
<b>Conclusion</b> .....	<b>39</b>
<b>References</b> .....	<b>40</b>
<b>Appendix A: Definitions</b> .....	<b>44</b>
<b>Appendix B: Survey Instrument</b> .....	<b>45</b>
<b>Appendix C: Interview Guides</b> .....	<b>50</b>
<b>Acknowledgments</b> .....	<b>58</b>
<b>About the Authors</b> .....	<b>59</b>

## Executive Summary

Researchers and practitioners in information science and allied fields have long discussed the need for software preservation: preserving software is a prerequisite for keeping and providing access to digital cultural heritage and research, where software is increasingly considered a research product or artifact in itself. This need has only grown more acute as research and memory organizations amass vast and rapidly growing collections of software. Many of these organizations struggle to manage these collections alongside other materials, even when keeping software is critical for continuing to provide access to those materials. Recognizing software preservation as a widely shared, complex problem that individual organizations cannot solve alone, the [Software Preservation Network \(SPN\)](#) was launched in 2016 to raise awareness, build capacity, and foster collective actions that engage a broad range of stakeholders who have an interest in this work.

This white paper outlines findings from a survey- and interview-based study of *software preservation service providers*—an expansive category that includes archivists, librarians, preservation specialists, technologists, and other professional roles. The survey and interviews were conducted in 2019 by the SPN Research-in-Practice Working Group. An analysis of the survey and interview data suggests broad support for inter-organizational collaboration in software preservation, despite challenges that surface because of the malleable concepts of “software” that can vary across preservation contexts. Results presented in this paper include a set of service provider personas, an overview of the conceptions of software that exist in research and memory organizations, a delineation of multilevel barriers to software preservation, and a related set of themes and targeted recommendations.

The themes and recommendations resulting from the study are listed below and further discussed in Section 5 of the paper.

### Representation in the Field

Future collaborative efforts in the field of software preservation should be grounded in an awareness of the need to broaden demographic representation within software collections and to expand diversity within the community of software preservation service providers.

**Recommendation 1:** Invest in collaborative collection development and research documenting and promoting public awareness of the participation of women and individuals with diverse racial, ethnic, linguistic, and cultural backgrounds in the creation and creative use

of software. Highlight the significance of the contributions of individual developers and preservationists in ways that everyone can identify with and appreciate.

**Recommendation 2:** Increase mentorship opportunities in the software preservation community, prioritizing quality mentorship training that can better support individuals from groups currently underrepresented among software preservation service providers.

**Recommendation 3:** Openly reflect upon and raise awareness of how structures within the software preservation community, research, and cultural memory organizations reinforce and reward whiteness; interrogate how software design and development have arisen in a culture of whiteness.

**Recommendation 4:** Create venues for those not currently providing software preservation services to share why they are unable to do so in an open, collaborative environment.

### Defining the Field, for the Field

Shared conceptions of “software” and “software preservation” are needed to advance the development of good practices and other social and technical infrastructures.

**Recommendation 5:** (a) Identify and refine a set of shared, capacious definitions of “software” and “software preservation,” perhaps building upon the documentation from the SPN Fostering a Community of Practice (FCoP) project. (b) Codify “good practices” for different software preservation use cases and ground those practices in the shared definitions developed in Recommendation 5a.

**Recommendation 6:** Help software preservation service providers assess and connect the aims, benefits, and costs of software preservation to the missions, goals, and financial landscapes of their departments or organizations.

### Networking and Community Building

Opportunities to connect software preservation service providers are needed. Intentional community building could help compensate for the isolation of service providers who have sole responsibility for software collections at their organizations, without ready access to peers who have experience dealing with similar responsibilities.

**Recommendation 7:** Establish networking and community building, both formal and informal, as the community’s top strategic priority and the scaffold upon which all other software preservation work is built. Co-create a community culture that is known for modeling vulnerability, respect, and trust.

**Recommendation 8:** Identify opportunities for translating and deepening connections among software preservationists across domains and disciplines—from hobbyists to collectors, industry to academia, cultural heritage to government, and beyond.

### Informal and Formal Learning

Software preservation activities draw upon a wide-ranging and eclectic skillset, and these skills require ongoing refreshment through formal and informal learning.

**Recommendation 9:** Pilot informal peer-to-peer, birds-of-a-feather learning groups focused on topics of shared interest that are open to the software community at large. Create incentives for these groups to develop open learning.

**Recommendation 10:** Develop or request proposals for a workshop on cultivating individual and organizational resilience in the face of complexity.

**Recommendation 11:** Recruit a team of organizational leaders and software preservation service providers to present at a series of conferences for administrators on the nature of software preservation work, including day-to-day activities, costs, and benefits.

### Shared Infrastructures and Model Practices

Advancements in software preservation will require that research and memory organizations collectively invest in shared “infrastructures” that benefit the international software preservation community. Shared infrastructures can include social groups or activities, technologies, research products, documentations of practice, or some combination of any of these resources.

**Recommendation 12:** Create a “levels of software preservation” matrix that empowers organizations and service providers to identify and prioritize tangible software preservation actions from a shared baseline of understanding.

**Recommendation 13:** Pursue a pilot project for formal expertise-sharing and collaborative workflow development among organizations that are invested in software preservation. Explore sustainability models within the scope of the pilot.

## 1. Introduction

*We probably have an inadvertent software collection.*

*Software does not live by itself but is part of an ecosystem...that is also evolving and sometimes can't be preserved...What are the best practices to deal with this complexity?*

—Software preservation service providers interviewed for this study, 2019

For decades, researchers and practitioners in information science, digital preservation, and allied fields have discussed the necessity of software preservation: preserving software is a prerequisite for keeping and providing access to digital cultural heritage and research, where software is increasingly considered a research product or artifact in itself. The need for greater attention to software preservation has only grown more acute as research and memory organizations amass vast and rapidly growing collections of software. Many of these organizations struggle to manage these holdings alongside other digital and physical materials, even when keeping software is essential for continuing to provide access to other digital artifacts. Recognizing software preservation as a widely shared, complex problem that individual organizations cannot solve alone, the [Software Preservation Network \(SPN\)](#) was launched in 2016 to raise awareness, build capacity, and foster collective actions that engage a broad range of stakeholders who have an interest in this work.

As part of SPN, the Research-In-Practice Working Group was established in 2017 to identify information gaps and opportunities for conducting research with the potential for contributing to the field and informing the ongoing strategic development of the organization. This paper outlines findings from a survey- and interview-based study of *software preservation service providers*—an expansive category that includes archivists, librarians, preservation specialists, technologists, and other professional roles. The survey and interviews were administered in 2019; they aimed to identify trends, gaps, and opportunities for designing and providing software preservation services within and across the variety of organizations that collect software.

An analysis of these data suggests a broad consensus among research and memory organizations that continuing and extending inter-organizational collaboration remains a valued strategy for implementing effective software preservation. At the same time, the material and conceptual bounds of “software” are malleable, and they can vary enormously across preservation contexts. The study results presented here include a set of service provider personas, an overview of the conceptions of software that exist in preservation contexts, a delineation of multilevel barriers to software preservation activities in a range of contexts, and a related set of themes and targeted recommendations. Interrelated themes and recommendations emerging from the analysis are presented here for a general audience interested in the stewardship of research and cultural heritage and in the advancement of software preservation practices consistent with this stewardship. Appendix A provides a list of terms and definitions for readers less familiar with the domain of software preservation.

## 2. Study Overview

The initial goal of the study was to develop and test an approach for understanding the evolving landscape of software preservation across “research and memory organizations,” which encompass a wide range of institutions and groups concerned with the generation, dissemination, and preservation of knowledge and culture. To this end, the Research-In-Practice group (referred to as “research team” or “team” throughout this paper) designed a survey (Appendix B) and interview protocol (Appendix C) to collect data about practices that fall under the umbrella of “software preservation” and the software-related activities and services currently available or being planned in these organizations. The group agreed to use these data to address the following research objectives:

- To characterize how service providers understand software and software preservation and how this affects the design and implementation of services
- To identify what skillsets are needed to understand, collect, preserve, and provide access to software
- To identify barriers software preservation service providers face in preserving and providing access to software
- To identify activities and future directions for the software preservation community and, where appropriate, the Software Preservation Network

Data were collected through an online survey built using the Qualtrics software platform and a set of semi-structured follow-up interviews. The survey questions, interview protocols, data use agreements, and consent language were approved by Institutional Review Boards at three academic institutions. The rest of this section lays out the design of the survey and interviews in more detail and provides brief demographic trends from the survey.

### Survey Design

The survey contained 25 questions pertaining to collections, services, practices, and needs. Since one original motivation for the project was to begin a process of collecting longitudinal data over time, many of the survey questions were designed with this in mind. The survey was launched in January 2019 and was open for approximately one month. Communication and publicity for the survey were shared through the SPN newsletter, website, and Twitter. The survey was also promoted through listservs used by communities of research and memory organizations (the target population for the survey). In administering the survey, the research team explicitly discouraged group responses. Interviews with survey respondents were conducted during the summer of the same year.

### Demographics of Respondents

The survey received 124 responses from individuals representing a wide variety of professional roles and organization types.

- Types of institutions represented by respondents included academic research universities (57.3%), government entities (19.4%), and a variety of “other” organizations including public libraries, entertainment companies, museums, and commercial entities (23.4%).
- Types of job roles among respondents included librarian (24.2%), archivist (24.2%), information technologist (13.7%), curator (8.1%), administrator (8.1%), conservator (4.0%), and others (17.7%), including digital preservationist, software preservationist, professor, program manager, digital asset management specialist, archaeologist, scientist, data specialist, data preservationist, photographer, and collections manager.
- The two most common job titles, archivist and librarian, were present in multiple types of institutions (Figure 1). The job titles information technologist, administrator, curator, and conservator were common across multiple types of institutions; the remaining titles were specialized to one type of institution.
- Fifty-five respondents (44%) affirmed that they were currently providing software preservation services, while the remaining majority indicated that they did not currently provide specific services in this area. Within the survey instrument, software preservation services were defined broadly to encompass collecting and describing software media and related artifacts as well as activities such as migration, emulation, or providing support to researchers with software-related questions.

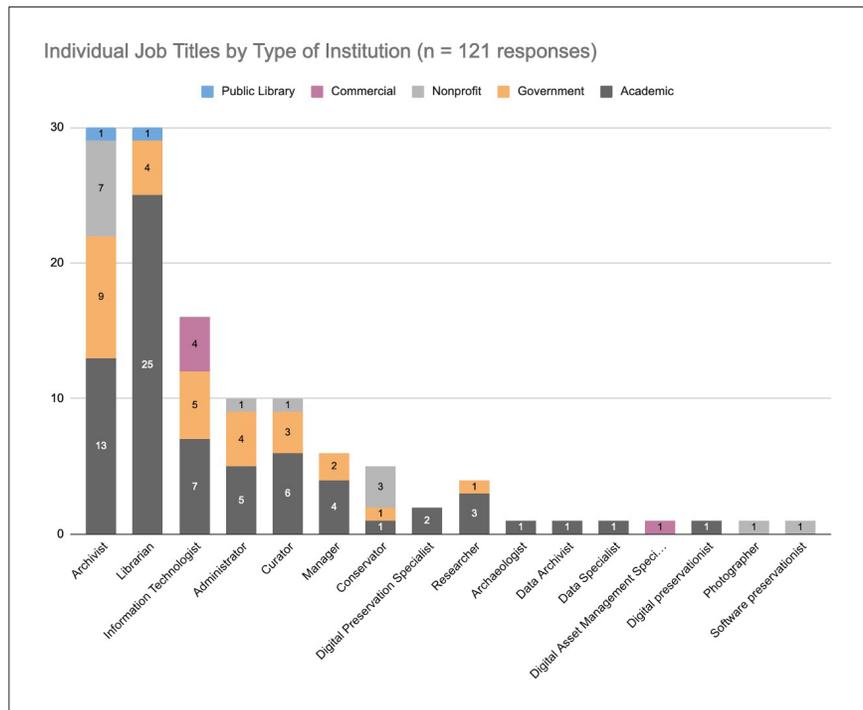


Fig. 1: Job titles of survey respondents, by type of institution

The 17 follow-up interviews engaged individuals based at a similarly broad range of institution types, including research universities, government entities, museums, and commercial organizations. However, all but one of those selected for the interview sample indicated that they were either currently providing software preservation services or aspired to do so in the future; one interviewee reported no intention to provide such services.

### Initiating and Conducting Follow-up Interviews

The last question in the survey invited respondents to volunteer to participate in a follow-up interview. Approximately 34 of 124 respondents indicated an interest in a follow-up conversation. The research team considered potential interviewees with the goal of selecting a sample that could represent a wide range of institution types (e.g., government units, academic libraries, nonprofits, museums, and private corporations), as well as a range of current or planned software preservation services activity—from active to aspirational. To be precise, the unit of analysis in this study is not the institution but the “service provider,” which the research team understands as an individual with a professional role, typically affiliated with an organization. Organizational variety is of interest to the extent that it may be a significant axis for differentiating individual professional roles and software preservation practices.

Based on the survey responses, the research team created three interview guides: One for “people currently providing software curation services,” one for “people not currently providing software curation services but planning to in future,” and a third for “people not currently providing software curation services and not planning to in future.” The questions in the guides were designed to allow interviewees to further elaborate on the motivations, practices, and resources they need to do their work. Interview guides were developed, tested, and refined by members of the research team in different institutional contexts in spring 2019.

Potential interviewees were contacted by email during the summer of 2019. A copy of the list of questions to be asked during the interview was provided for their review and convenience. Interviewees were asked to affirm their consent to the interview by replying to the email and stating that they agreed to participate in the study.

The team conducted 17 interviews in summer 2019 using the video conferencing software Zoom. Most of the time, research team members paired up to conduct the interviews, with one team member leading the interview using the guide that best fit the interviewee’s apparent working context given their survey responses, and the other taking notes in a shared document. Interviews were recorded when the verbal consent of the interviewee was given.

The semi-structured interviews ranged from thirty minutes to one hour and forty-five minutes. After the completion of an interview, the interviewers deposited copies of the interview notes and recordings in the group’s secure cloud storage space provided by Dropbox, labeling each with a unique identification number tied to the interviewee’s survey response.

Interview recordings were transcribed using Temi, an automated transcription service. Transcripts were then reviewed and further ed-

ited for clarity by research team members listening to the recordings in real time. All research team members listened to interviews and edited transcripts for interviews in which they did not participate to further their understanding of the data. Research team members edited transcripts to ensure speaker roles (interviewee and interviewer) were clearly identified, and that specific references to organizations, people, or places were redacted.

The research team sent copies of the edited transcripts to interviewees for review and approval. If interviewees requested clarifications or redactions to the transcript, a corrected copy of the transcript was noted as “clarified” and deposited in the team’s shared cloud storage.

### Coding and Analysis of the Data

The research team adopted an inductive, “bottom-up” methodology oriented toward theory building rather than hypothesis testing. In this approach, categories relevant to the research objectives emerged through the iterative coding of interview transcripts and survey responses. Each team member coded one interview they did not conduct, and one they did. This step focused on generating codes via a lightweight process of creating themes and categories corresponding to interview quotations in transcripts, writing research memos, and discussing codes and memos as a group. Once the group had each coded at least one interview and discussed and refined the emerging codes, the first draft of the codebook was created.

The research team identified Dedoose, a web-based software application used for analyzing qualitative and mixed-methods research data, as a useful tool for further coding. Dedoose allows users to work collaboratively in a single project space integrating the draft codebook and all transcripts. Codes could then be selected and applied to transcripts within this project space and to memos, allowing for more detailed analysis.

After creating a Dedoose project space, the research team tested the draft codebook by applying existing codes to additional interviews, engaging in co-coding to gain consensus: team members separately reviewed and coded the same interviews, and then met in pairs to discuss any code application differences or clarifications needed. The research team also coded open text answers from survey responses. The codebook was updated with new, revised, or retired codes until the group had coded all transcripts and arrived at a final iteration of the codebook.<sup>1</sup> To provide further context for each code, the group furnished a definition as well as an example quotation pulled from related interviews.

### Limitations of the Study

The research team acknowledges the potential influence of SPN’s priorities and perspectives and how these both consciously and unconsciously affected this study’s design, reach, and findings.

---

<sup>1</sup>Supplementary materials from this study—including aggregated and anonymized survey results, the codebook used for analyzing interview transcripts, concept maps, and other materials—are available on Zenodo under open licenses that permit adaptation and reuse: <https://doi.org/10.5281/zenodo.7086618>.

Additionally, apart from the relatively small numbers of survey respondents and interview participants, the study had several limitations that might be improved upon in future related projects. First, some logic pathways in the survey failed to perform as expected. For example, clarifying comments for the question, “Are you considering providing services related to software preservation in the future?” were only solicited from those who answered “No” or “Yes” to the question. Since most respondents selected “Maybe,” the research team missed an opportunity to explore in greater depth the variety of factors affecting respondents’ future expectations.

It is also important to acknowledge that the research team chose to do follow-up interviews only with those willing to opt in. Unsurprisingly, those indicating an interest in discussing their work in an interview were more likely to already be providing services or were aiming to do so in the future. Only one opt-in interviewee was not currently providing software curation services and not planning to in future, and it is hard to draw conclusions from this single instance. The team only interviewed individuals who had responded to the survey and did not reach out beyond this group. Survey responses skewed heavily to academic settings (over 50%), national libraries, and government entities, rather than to hobbyists or smaller, community-based groups interested in legacy software. By focusing on people who work in academic and government settings, many of whom hold library or archival degrees or related professional certifications, the research team missed opportunities to connect with active hobbyists, “tinker-ers,” or other volunteers who provide a rich set of services related to software preservation from outside these settings.

The survey questions were designed with strictly professional, rather than personal, demographics in mind: the lack of attention to participants’ personal backgrounds and identities constrained the team’s ability to take participants’ personal backgrounds and life experiences into consideration in the analysis of the barriers and challenges facing software preservation service providers in their work. Adding questions about participants’ race, ethnicity, ability, age, gender, and educational backgrounds would have enabled a more nuanced assessment of the diversity of the survey population; informed efforts to improve the diversity of the sample population to be interviewed; and made possible additional observations and recommendations about how social factors affect the allocation of attention and resources to the work of software preservation.

### 3. Software Preservation as a “Field”

Survey respondents, all of whom self-identified as “software preservation service providers,” worked in a variety of institutional contexts, including academic libraries, government agencies, museums, and technology companies. Respondents held a range of professional titles, including librarians, archivists, conservators, curators, technology specialists, administrators, academic researchers, and programmers. Most respondents indicated that their home organizations do not provide dedicated software preservation services; where dedicated services are available, the majority of respondents ( $n = 28$ ) indicated that they provide software preservation services infrequently.

These responses suggest that software preservation is a diffuse and loosely defined field of practice that intersects a wide variety of professional categories and institutional contexts. In addition, software preservation often is not afforded a dedicated institutional role, nor is it the primary responsibility for those who provide software preservation services; this point was particularly underscored in the interviews. In the context of university libraries and archives, which represented the bulk of the survey responses, software preservation frequently occurs as a secondary goal supporting more central missions such as providing access to digital records and resources (i.e., “software-dependent content”) and facilitating reproducible research (i.e., “computational reproducibility”). Software preservation services are often achieved by assembling and coordinating “pieces of people,” as one interviewee put it, with various professional training and skill-sets.

Certainly, there are exceptions. Our interviews revealed a handful of institutions committing significant resources and dedicated personnel to software preservation. These included government offices in the European Union and United Kingdom with legal mandates to collect software, cultural heritage organizations with an explicit focus on historical software and computing, and a software company with a commercial interest in preserving software. In addition, initiatives such as the software curation fellowships created by the Council on Library and Information Resources (CLIR), the Software Sustainability Institute (SSI), Research Software Engineering, and others have helped raise the profile of software preservation and software curation as dedicated professional domains, at least in the context of scientific software. At least one recently advertised position in an academic library emphasized software preservation as a primary responsibility.

#### Demographics of the Software Preservation Community

While the survey did not ask respondents to provide demographic information, a look at the varying communities that intersect with this field provides some indications of current diversity and representation within the software preservation community. There is no prevailing categorization of who exactly composes the “software pres-

ervation community.” Software is so embedded in the fabric of everyday life and work that many people and organizations are (or should be) invested in preserving it, including archives, museums, research organizations, commercial corporations, stewards, hobbyists, software historians, and many more. The best available view of the people making up this field are the members and target populations of SPN, the SSI, and Software Heritage—the three largest groups working toward large-scale preservation of software.

Unfortunately, only SSI has done surveys that explore the demographics of its constituency. A 2014 survey of researchers working in research-intensive universities in the United Kingdom (Hettrick 2014) highlighted that the same proportions of men and women *use* research software (92% of all researchers surveyed), but fewer women *develop* research software (only 30%). Further, 63% of the men had received formal software development training while only 39% of women had. A further survey in 2018 (Philippe 2018) showed that research software engineers (RSEs) were mainly men (73% or more) and dominantly working in the fields of computer science, physics and astronomy, and biology. These surveys suggest that many women do not come into the field with the same levels of preparation and support related to software development as men do. Neither survey tracked demographics beyond gender identity.

SPN includes both researchers and research software engineers, but according to their website, the Network supports a wider community of “colleagues from design firms, public libraries, history of computing museums, research data archives, university libraries and the open source software community” than SSI and Software Heritage. SPN members include “legal scholar-practitioners, digital preservation specialists, metadata specialists, data curators, digital conservators, knowledge managers, archivists, software developers and data journalists.” The next few paragraphs describe the demographics of a few of these groups including software developers, the open source community, libraries, and archives.

In 2020, software developers in the United States were mostly men (81%). The two most common races/ethnicities for all developers were White non-Hispanic (53.4%) or Asian (33.6%). Less than 5% of software developers were Black or African American (4.5%) or White Hispanic (4%), with all others reporting less than 1% (Data USA 2020). In a similar study, GitHub did a survey of attitudes, experiences, and backgrounds of those who use, build, and maintain open source software (GitHub 2017, n = 5,500). The study revealed a profound gender gap of 95% men, 3% women, and 1% non-binary; it also surfaced indications that women were more likely to experience harmful language, stereotyping, and harassment but less likely to leverage a main advantage of open source software development, which is getting help from people they do not already know.

Archives, museums, and libraries make up a large portion of SPN’s membership. While women are better represented, there is a persistent lack of racial and ethnic diversity across the archives, museum, and library professions. In 2019, 55% of archivists, curators, and museum technicians were women; 82% were White non-Hispanic, 5% were White Hispanic, 4% were Asian, 3.5% were Black or African American; all other categories were 2% or less (Data USA 2019). The most recent survey conducted by the archives profession,

the A\*CENSUS II All Archivists Survey, reports the population of archivists and memory workers are 71% women, 23% men, and 3% non-binary. Although there are more women overall and the number of women increased since the last survey in 2004 (Banks 2006), women continue to be underrepresented in senior and executive positions. For race and ethnicity, while the percentage of Black, Indigenous, people of Color (BIPOC) doubled since 2004, the profession remains excessively white (84%) with all other categories reporting 5% or less (Skinner and Hulbert 2022).

Library workers encompass both librarians and library technicians or assistants: data collected in 2020 indicated that both groups dominantly identify as White (83% and 79%, respectively) (DPE Research Department 2021). Black or African American and Hispanic or Latinx workers are each almost 10% of the library workforce; Asian American and Pacific Islanders represent almost 4%. Women remain over-represented in the library profession overall (83% for librarians, 77.5% for technicians and assistants). The only large study exploring disability within the library workforce is a 2012 extension of the Diversity Counts study run by the American Library Association in 2006. In these extended data tables, the reported number of people with disabilities working in libraries is very low at 3.7% of librarians (4,439 of 118,666) and 5.5% of library assistants (6,796 of 122,768) (Davis and Hall 2012). In the Association of Research Libraries (ARL)<sup>2</sup> 2019–2020 Annual Salary Survey<sup>3</sup> academic libraries report similar racial and ethnic trends (Mian 2022). An earlier study of ARL libraries included specific types of work within academic libraries. In each category women and White people are over-represented, except work with technology, which is still dominated by White men: *preservation* (n = 225) is 70% women, 76% White; *cataloging, metadata, resource description* (n = 998) is 69% women, 70% White; *special collections, rare books, archives* (n = 719) is 67% women, 77% White; and *technology* (n = 914) is 30% women, 75% White (Schonfeld and Sweeney 2017). Schonfeld and Sweeney also provide one of the few intersectional views of academic librarianship, reporting: White women 43%, White men 28%, Asian women 6%, Asian men 3%, Black women 5%, Black men 3%, Hispanic women 3%, and Hispanic men 2%. Archives, museum, and library data presented here are for the United States only.

Library and Information schools in the US are training future library and archival professionals. The most recent statistical report from the Association for Library and Information Science Education (ALISE 2022, Table II-2) reported the following gender and race/ethnicity trends for the 18,298 students enrolled in the 2021 ALA-accredited master's programs across the US and Canada: 80% women, 19% men, 0.7% non-binary, and 64.5% white, 10% Hispanic, 5% Black or African American, 4% Asian, and 4% international with the remaining 12% reporting more than one race, a category reporting

<sup>2</sup>ARL data is used because the Association of College and Research Libraries, a division of the American Library Association, did not include questions about demographics in their surveys until 2020; at the time of this publication, the data was not available.

<sup>3</sup>The data table was accessed via University of Virginia ARL Library access.

less than 1%, or cases in which the race/ethnicity was unknown. Women in general continue to be over-represented in library school programs; however, minority women continue to be under-represented both in library schools and in the library workforce.

This overview describes demographic trends affecting the communities of people who are or may be working in software preservation. In the computing world, many women do not come into the field with the same levels of preparation and support related to software development as men do. Further, many women are not present in or taking advantage of the benefits of the open source community, perhaps because they do not feel comfortable or welcomed. In archives, museums, and libraries, while women are often over-represented, minority women and men continue to be under-represented, and a significant gap in racial and ethnic representation persists in these fields. Further, even in libraries where there are large numbers of women, they are under-represented in technology-focused roles within the profession. These insights highlight the need for supports that help everyone engaging in software or technology work feel welcomed and able to make new connections within the network; it also highlights the need for opportunities for women and others who may not have the same levels of preparation with respect to software development and digital preservation to learn the skills they need to do the work. To address the imbalances in representation, collecting institutions and professional associations must continue to combat the cultural factors that discourage non-White participation in preservation-related careers, including open acknowledgement of current diversity gaps and previous failures of recruitment and retention strategies within research and memory organizations and information schools.

#### Four Personas of Software Preservation

To further characterize software preservation as a field, the research team developed a set of personas that reflect activities, challenges, and insights expressed by the 17 interviewees who participated in the study. The personas do not correspond in a one-to-one fashion with interviewees. Instead, they are meant to represent a range of professional contexts, skills, and responsibilities that emerged in the interviews. They help to illustrate findings related to this study's first and second research objectives: (1) to characterize how service providers understand software and software preservation and how this affects the design and implementation of services, and (2) to identify what skillsets are needed to understand, collect, preserve, and provide access to software. Note that, beyond these professional roles, an even greater number of individuals also make important contributions to advancing software preservation as a field. Current and former software developers, historians of technology, preservation hobbyists, and legacy software enthusiasts bring critical knowledge, experience, and skills to problem-solving within the complex and rapidly evolving field.

### Persona 1: Research Data Librarian at an Academic Library

Alicia is a research data librarian at a university library in the United States. One of her primary responsibilities is to manage research data (including software) deposited in an institutional repository managed by the library. Curation, in this context, involves documenting and improving submissions with an eye toward preservation, reuse, and the “reproducibility” of research findings. Recently, the number of submissions to the repository involving software has increased: “A lot of students are building software suddenly,” she explains.<sup>4</sup> Alicia describes a recent dissertation from a graduate student in English involving an interactive web application. To process the student’s submission, Alicia gathered the project source code, documented its software dependencies, downloaded the student’s website with a web crawler, and generated checksums for all the digital assets. This workflow was not prescribed; it was formulated in an ad hoc manner, and Alicia is considering ways to handle future software submissions in a consistent manner. At the same time, she is concerned that her team cannot realistically curate all the software produced by researchers at the university, much of which is shared on sites like GitHub and Zenodo, not the institutional repository. Alicia suggests that the best use of her time might be to educate students and faculty on documentation, metadata, and citation standards for research software: “I think that coming up with more of a ‘how to deposit your [software] anywhere’ guide is needed.” Alicia looks to the broader software preservation community for recommended practices that can inform her own curation workflows and for resources that researchers can use to better document and preserve the software they create.

### Persona 2: Digital Preservation Librarian

Jean is a digital preservation librarian at a university library in the United States. With its recently updated strategic plan, the library has set a goal to improve and expand its digital collections. To make progress toward this goal, Jean is considering a project to assemble a complete inventory of software in the library’s collections. Software, Jean notes, is very hard to find, and it is often overlooked: She adds, “We probably have an inadvertent software collection.” It lives in unexpected places like CD-ROMs in the back of books, amid research data sets in the institutional repository, or on unprocessed and obsolete physical media in special collections. Software dependencies (needed to render digital assets) are yet another category of software to consider. Jean realizes that building a complete software inventory will require coordination and “buy-in” among many different units in the library. Jean is unsure whether anyone in the library has the expertise and time to coordinate such a project. While the library has increased hiring in many areas related to data and digital curation, no positions are dedicated to software preservation; instead, “We have pieces of people.” Jean admits that making progress in the project will likely require focusing on one type of software, such as research software, at the expense of others. Jean sees the sheer *variety* of software and its uses as a major conceptual and organizational hurdle.

---

<sup>4</sup> Quotations used in the persona descriptions are sourced from interview transcripts.

### Persona 3: Commercial Software Preservationist

Steve is a preservation specialist in a large video game production company. The company devotes resources to software preservation for the purposes of “franchise resurrection” in which old (sometimes decades-old) digital assets are used to re-release games. Steve works with development teams to gather and preserve game assets and development environments necessary for future republishing efforts. Steve mentions several challenges in his work. Cloud-based games are particularly challenging because, “You’re not only preserving the object itself, but all these ancillary [network] services that exist around the object in order to get it to work.” Similarly, some game development tools can only be activated by connecting to the vendor’s license server; if these servers go offline in the future, it may become difficult or impossible to re-create game assets. Fortunately, the company can devote considerable resources to these challenges by, for example, negotiating permanent licenses with vendors. Other challenges are not readily addressed through resource allocation—some are cultural. Steve notes that the “start-up” mentality that predominates in the game industry often fails to acknowledge the deep cultural legacy of video games. He explains, “Our department’s mission is to preserve, but the institution’s mission is not. . . [S]ometimes there are conflicts of interests where teams are too busy trying to ship a title or finish some work that is mission critical.” Steve looks forward to a day when the values and concerns of software preservation are embraced broadly in the game development industry—until then, “We have to educate and evangelize.”

### Persona 4: Digital Preservationist in a National Library

Adrian is a digital preservation specialist at a European national library whose department is responsible for collecting and preserving all kinds of digital materials. Adrian says, “We collect certain kinds of software, the ones that we consider publications in a cultural heritage sense. That’s mostly video games, but also educational material and some interactive artworks.” Compared to preservationists in government agencies in the United States, Adrian’s department receives sufficient government support to carry out its mission of collecting and preserving culturally significant software. In a department of about two dozen people, two or three employees focus primarily on software preservation. In addition, government funding and the legal structure support the work: “We also have a set of rules and regulations revolving around copyright, and actually the work of libraries and archives in general, which allows us to do things that other libraries and particularly private collections’ libraries aren’t allowed to do, such as break copy protection, make copies of things for preservation purposes.” Among the country’s game developers, there is a high level of awareness that they must provide the library with a copy of each game title they release, which greatly facilitates Adrian’s work. To advance the field of software preservation, Adrian suggests strengthening international collaboration and fostering communities where difficult software preservation challenges are given due consideration.

## Conceptions of Software in Software Preservation Work

Conceptions of Software in Software Preservation Work individuals bring very distinct sets of priorities and challenges to the work of software preservation, which are shaped substantially by the contexts within which they are working. The breadth of the types of software and software-related materials they must manage can further complicate these priorities and challenges. In survey responses to a question about the software components that service providers collect, the most frequently mentioned category was documentation (65%), followed by executables (58%). Additional choices included source code (50%) and “other” (18%). Respondents described a range of “other” components: vintage media and obsolete carriers, virtual machines and software “containers” used in computational research, screenshots of the software in use, secondary software and other dependencies, and “stories” (i.e., narrative accounts or oral histories) about the software.

These survey responses underscore how malleable and varied the material bounds of software are across preservation contexts. In interviews, service providers often described an interest in preserving “everything” or “the whole thing”; however, what the *whole software* consists of can shift significantly from one context to another. For a preservationist focused on commercial software, the “whole” may include physical media, packaging, documentation, and advertisements; for a data curator concerned with software-based scholarship, the whole might include a complete inventory of digital assets and dependencies needed to re-run or reproduce the work (as illustrated in the persona of the research data librarian above), or a snapshot of a Git repository (Steeves and Nguyen, 2020). Museums dedicated to the history of computing and media archaeology labs might push these boundaries toward the hardware and peripherals on which the software was created or used. The persona of the commercial software preservationist (above) illustrates how network services and institutional relationships might also enter the fold.

Given these varied aspects and manifestations, the research team aimed to identify perceptions of software active in software preservation work. In the analysis of interview transcripts, three significant “conceptions of software and software preservation” emerged; these are themes that describe how service providers value software and how this animates their work:

- **Software as inherently valuable:** Software preservation is often motivated by the belief that software is valuable “in itself” as, for example, a cultural expression (similar to a book or film), a product of research (similar to an academic journal article), or a commercial product. Software preservation practices may be quite varied and extend to *non-digital aspects* of software (e.g., documentation, first-person accounts, physical packaging).
- **Software as support:** Software’s role in digital preservation is often secondary. It is a necessary condition for accessing, rendering, or using other (more primary) digital resources such as obsolete file formats. In this perspective, software’s value is purely instrumental or supplemental.

- **Software as part of “the digital”:** Software is often treated as another category of digital information; it may be stored alongside, managed, and accessed in more or less the same way as other kinds of digital assets. Here, software preservation is attenuated as a distinct practice, giving way to more general digital preservation practices.

Although these terms suggest a tiered structure, they do not represent mutually exclusive categories. In interviews, service providers routinely shifted among perspectives, suggesting that multiple conceptions of software may simultaneously animate preservation work. Collecting culturally significant software, for example, may entail a “supplementary collection” of software dependencies. Similarly, preservationists may recognize and work to enhance the inherent value of research software, but they may lack institutional resources to provide services beyond “bitstream” preservation.

In addition to these categories, the research team identified the difficulty of pinning down “software” as an important theme in itself. In the following section, the challenges associated with conceptualizing software are presented as a source of significant barriers in software preservation work. First, however, it is important to recognize that the malleability and variety within software and software preservation activities present certain opportunities for the software preservation community—opportunities to affect broader cultural perceptions of what software is, who it includes, and who it excludes. Rather than blindly accepting prevailing attitudes and values associated with software, service providers should approach software preservation practices creatively, reparatively, and with an openness to interpreting or redefining the significance that software materials hold for future generations. To realize these possibilities and opportunities to engage in software preservation as an active, ever-evolving pursuit, it is necessary first to grapple with the obstacles that stand in the way.

## 4. Barriers to Preserving and Providing Access to Software

With the study's third research objective, "to identify barriers software preservation service providers face in preserving and providing access to software," the research team sought to identify the obstacles or challenges mentioned by study participants as hindrances to their software preservation work. The team examined trends in responses to the survey questions related to barriers and analyzed co-occurrences between the "barriers" code and other codes across the interview and survey free text data. This analysis indicated that barriers appear at three main levels of the software preservation ecosystem: (1) at the level of "the profession," which could be defined as a grouping that brings together multiple institutions and individuals; (2) the level of "the institution," which could be defined as a grouping that brings together multiple individuals working for one employer; and (3) the level of "the individual," which could be defined as a person who self-identifies as a software preservation service provider.

Throughout these levels, barriers take the form of gaps—the lack of a component, characteristic, or resource needed for preserving and providing access to software. Collectively, these gaps impede progress in the software preservation space. Although the barriers are usually understood at all three levels (profession, institution, individual), each barrier has a gravitational connection to one level in particular—a place, according to the data, where the barrier is most acutely experienced. The team categorized the barriers into levels accordingly. These professional, institutional, and individual barriers often reinforce each other in a sort of vicious cycle, and barriers that appear at one level may be more effectively addressed at other levels.

### Professional Level

*The profession lacks shared conceptualization(s) of "software" and "software preservation"*

At the level of the profession, the main barrier evidenced in the data stems from the multifaceted contexts characterized in the previous section of this paper. There was a clear co-occurrence between the code "barriers" and the code "software as something challenging to conceptualize." In other words, the key barrier at this level seems to be a lack of shared conceptualization(s) and articulation(s) of software. When gathered together at the level of the profession, institutions and individuals may not be speaking the same language about what *software* and *software preservation* mean and may lack terminology for referring to different use cases or critical nuances within software preservation work. Depending on the overall motivation or goal, *software* and *software preservation* can carry numerous meanings, resulting in challenges to common understanding, repeatable workflows, or generalizable resource estimation. As one participant

reflected, conversations at the professional level “are so wanting to jump and move ahead . . . community wide. I think that’s really short-sighted. . . . It could be so powerful to come up, as a community, with a model of software that we understand, that has different parts and understandings, that is not monolithic sounding or looking. That would allow us to ‘right size’ preservation strategies and figure out how we collaborate on things.” Another participant echoed the lack of shared conceptualizations when they asked: “What needs to be preserved when I am preserving ‘a software?’” To a certain extent, this conceptual barrier (and vocabulary gap), which appears so frequently at the professional level, may be the cumulative result of tensions and disconnects—other barriers—found at the level of the institution.

### Institutional Level

#### *Institutions lack internal strategic alignment*

Within the context of this analysis, it is useful to consider “institutions” to be employers that bring individuals (and departments) together to accomplish a mission. According to the data, overarching institutional missions, strategies, and business processes are often uncomfortably aligned or even at odds with a department’s or individual’s software preservation goals, as in the persona of the commercial software preservationist described in the previous section. This lack of strategic alignment appears in the data from respondents working across contexts, including corporations, governments, museums, and others.

In many cases, administrators and institutional resource allocators may not yet see compelling, direct connections between software preservation activities and high-level institutional priorities such as increasing usage, fundraising, tuition revenue, or student retention. In other situations, the connections between software preservation activities and institutional priorities may be more visible. For example, a resource allocator at an academic institution may have a strong interest in preserving and promoting research software that scholars create through grant funding. Even where the connections are visible, however, resource allocators face a heavy lift. Software preservation activities require the expertise of individuals and departments across an organization. As a result, resource allocators must perform elaborate strategic planning to understand use cases and to enable a software preservation program that supports cross-functional collaboration.

Where it is only implicitly tied to the institutional mission, software preservation can look like a resource-intensive rabbit hole compared to better-understood activities with known costs that meet collecting, preservation, and access goals in the shorter-term: “You have to have a really strong business case, and most of the time you don’t because only one in a thousand people maybe every four years [needs the software]. That’s just not enough to invest in.” The unknown (but assumed unfavorable) cost-benefit ratio of software preservation means it lands further down on the institutional priority list, if it makes the list in the first place. It is unsurprising, then, that survey respondents who said they were not considering providing software preservation services in the future selected these two reasons most frequently: lack of need and lack of staff.

Before touching on staffing barriers, it is necessary to acknowledge an aspect typically characterized (and perhaps overemphasized) as one of, if not the most significant barrier to software preservation: the legal landscape.

*Institutions lack strategies to evaluate intellectual property questions*

Although intellectual property questions arise in the contexts of profession-wide conversations and individuals' daily work, they loom largest at the level of the institution. Employing institutions are most likely to face legal risk in the form of copyright lawsuits if their workers collect and provide access to software in which third parties have financial or other interests. Study participants raised a variety of questions related to the legal landscape, such as how their institutions can legally provide remote access to software and software-dependent content, how to understand legal guidelines for collecting and sharing software internationally, and how organizations should preserve software that relies on proprietary dependencies or stacked licenses. These questions sometimes become paralyzing distractions to progress.

Ironically, interest in intellectual property considerations can, in some institutional contexts, work to reduce the strategic alignment barrier discussed above. For institutions such as corporations or research agencies with mandates to steward software created by their employees, interest in optimizing intellectual property, encouraging attribution, and facilitating reuse can serve as points of strategic alignment among software creators, software preservation service providers, and institutional leadership. In these cases, intellectual property concerns may serve as an incentive for investing in software preservation—a motivator rather than a paralyzing barrier. As one participant said of advocating for software preservation with institutional leadership, “If you’re having trouble getting in the room or getting someone to listen to you, you can bring these things up.” Even when intellectual property considerations can be a boon, though, it may not be sufficient for software preservation service providers themselves to bring up the legal incentives; the message may only resonate with leadership if delivered by an attorney.

It is important to note that the most resounding theme related to the legal landscape was this: although legal questions feel important and garner a lot of attention, investing energy in them at an institutional level can seem premature or even counterproductive if support and dedicated labor are not available to allow for practical work on software preservation. Participants recognize the complex legal issues that affect an institution's ability to preserve software, but those issues may not be the most urgent. The practical problem of needing labor capacity to even ascertain what software is in their collections (let alone to preserve or provide access to it) is a barrier they must get past before being able to appreciate the most relevant legal issues. “Maybe I would complain more about legalities and rights,” one participant stated, “if I felt the operational side was actually something we could do.”

*Institutions lack dedicated labor and talent*

In this light, lack of human time and attention becomes perhaps the most urgent barrier to software preservation at the level of the institution—it is an institutional issue because employers are ultimately responsible for their staffing and training strategies. As a result of the scarcity of individuals with expertise in this area and time to apply it, engagement and representation in software preservation activities can look like a matter of institutional privilege. That is, institutions with workers attending to their software collections have capacity to take practical action; consider legal questions; and build technical, experimental, and collaborative skills in more than one employee (crucial to knowledge management and sustaining institutional progress). This issue exacerbates the profession-level barrier of conceptualization. If one employee from a small organization has time to participate in profession-wide conversations about software preservation, they see a very different problem space and set of shared action items than does a team of participants from an institution comparatively rich in staff time and expertise. The conceptualization barrier becomes not just a lack of common understanding of software and software preservation, but an empathy and empowerment gap between a wide range of workers, each operating within a unique set of limitations and privileges. As one participant explained, their institution is not considering providing software preservation services in the future “because the requirements are beyond the scope of what we have the resources for. And that’s less about the technological hardware resources and more about the human resources.”

**Individual Level***Individuals lack control over resource allocation*

For this analysis, “the individual” is defined as a person who self-identifies as a software preservation service provider. The vast majority of these individuals, according to the data, are practitioners at lower levels of institutional administrative hierarchies rather than administrators. Because of their position within their organization, these service providers may have limited influence or institutional perspective to best align their software preservation aims with the mission and goals of their employers. In other words, these individuals generally lack authority to correct institutional barriers related to strategic alignment and staffing and may lack explicit responsibility to work on software preservation at all. Further complicating matters, at higher levels of organizations, where authority is often situated, leaders frequently “have a fundamental misunderstanding of what’s needed for preservation,” in the words of one participant. Individuals may see themselves as embodiments of expertise and personally identify as mission-driven members of a professional alliance, but their employers may see them fundamentally as workers whose priorities must slot into the institutional mission.

The result is that individuals can experience competing messages and motivations coming from the profession and from the institution and remain stuck in the middle, balancing allegiances. The institution pays the bills (salaries and memberships), but the profession provides

the support on which the individual depends to accomplish their work. Within software preservation conversations at the level of the profession, an individual may be asked to represent the interests of their employing institution. Within their employing institution, the individual may feel pressure to represent the interests of the profession. In effect, individuals may carry the burden of driving both local and professional strategy without the authority, time, or compensation needed for such work. “We don’t really have a stick,” remarked another participant. “We act more as a service. And so, we have to educate and evangelize.” This barrier aligns with findings from several other recent studies of digital stewardship work in memory institutions, underscoring the pervasiveness of the issue (Blumenthal et al. 2020, Atkins et al. 2017, Handel and Matienzo 2021, NDSA 2022). To compensate for their lack of control over resource allocation, service providers engage in burnout-inducing advocacy, an effort that competes for energy with one area in particular that individuals urgently need for software preservation: good practices.

#### *Individuals lack “good practice” workflows and skills*

The most popular answer selection in response to a survey question about what participants hoped an international software preservation consortium could provide to them that they could not otherwise accomplish on their own was “best practices.” One participant emphasized that access to best practices was their selling point to convince administrators to sponsor a membership in the Software Preservation Network: “Best practices was one of the big things that we kept hitting upon when we were advocating for joining SPN.” The research team used the nebulous term *best practices* in the aforementioned survey question without defining its meaning. When discussions of *best practices* showed up in the interview data and survey free responses, though, study participants used the term to mean documentation of what concrete, everyday (even imperfect) software preservation workflows and skills look like over time. Multiple participants echoed the sentiment of one respondent who asked for “examples and models of good practice (rather than best practice).” According to the data, “good practices” for software preservation means having shared understandings about (1) repeatable workflows for collecting, preserving, and providing access to software and software-dependent content for different use cases and (2) sustainable techniques for cultivating or sharing the skills required to do this work over time. The lack of documented, shared understanding about the workflows and skills required to carry out practices is a barrier that limits individual service providers daily.

In the context of working with software donors, for example, service providers often struggle with workflows that require donor-creators to take extra steps to make their software preservable or reusable. Once the service provider has custody of the software, workflows for one software preservation use case do not necessarily help with (or scale for) other use cases, and bit-preservation workflows (treating software as just another digital archival record) may be insufficient in cases where reuse is the goal. Survey results and interview responses reflect enthusiasm for emulation but also indicate that emulation is seen as being beyond the reach or capacity of many

practitioners. Even when supported by hosted tools, emulation workflows still require service providers to invest significant effort, which may channel limited resources into technical solutions where higher priority problems may be more about organizational or professional culture than technology. One participant reflected on the importance of understanding their donor and user community's immediate needs related to software preservation, asking, "What are your priorities? . . . Emulation . . . [is] such a high bar." Another participant remarked that emulation is "not the [only] thing that we . . . need to talk about. People . . . want to throw technology at problems."

When it comes to workflows for providing access to preserved software, increased clarity about legal guidance over the past several years shows great promise for empowering service providers, even if it means disrupting existing *ad hoc* practices that some institutions have historically performed under the cover of ambiguity. One participant mentioned that explicit limitations in the 2018 Digital Millennium Copyright Act (DMCA) exemption, which states that institutions should provide access to software "on-premises" (i.e., in a reading room or similar physical location), disrupted their institution's *ad hoc* access practices: "Once the new exemptions came out, it actually kind of curtailed how much we were sharing, especially with the provision that it had to be on premises" (Albert 2018). (The on-premises limitation was loosened for software other than video games in 2021) (US Copyright Office 2021).

Service providers are also concerned about workflows that "get the collection material to talk to the software in a meaningful way that isn't a huge technical burden for our reference staff." And when designing workflows to deliver software and software-dependent content to end-users, service providers need practices for understanding who the (often heterogeneous) users are and measuring how services can be improved.

Regarding the team's second research objective (to identify what skillsets are needed to understand, collect, preserve, and provide access to software), the skills required for preserving a particular piece of software can be highly specialized and technical; the skillset for one use case can also differ significantly from what is required for another piece of software to be preserved for a different reason. As one participant put it, each software preservation project is "its own Mount Everest." If the need for these skills arises only occasionally, service providers may not deem it worthwhile to cultivate or maintain them—especially when many employees must teach themselves or train short-term colleagues. When service providers invest in their technical skills and do not see corresponding market-rate compensation increases, they may be lured away from memory institutions by higher-paying jobs.

The data illuminate the diverse array of activities classified under the umbrella of software preservation services. "Software does not live by itself but is part of an ecosystem (data, API, etc.) that is also evolving and sometimes can't be preserved," noted one participant who then asked, "What are the best practices to deal with this complexity?" In the context of such varied practices, the dream of universal workflows that can be documented and easily repurposed or skills that can be cultivated in any individual situation may be unattainable, but individual service providers can join together to support their practices in other ways that propel the community forward.

## 5. Moving Forward

When the research team administered the survey, the team asked all respondents to prioritize a list of possible resources that a professional organization like the Software Preservation Network could provide: best practices, peer-to-peer support, avenues to funding, help with advocacy, and infrastructure (Figure 2). Although some respondents identified each of the possibilities as important, the results indicated strong preferences for best practices, infrastructure, and peer-to-peer support.

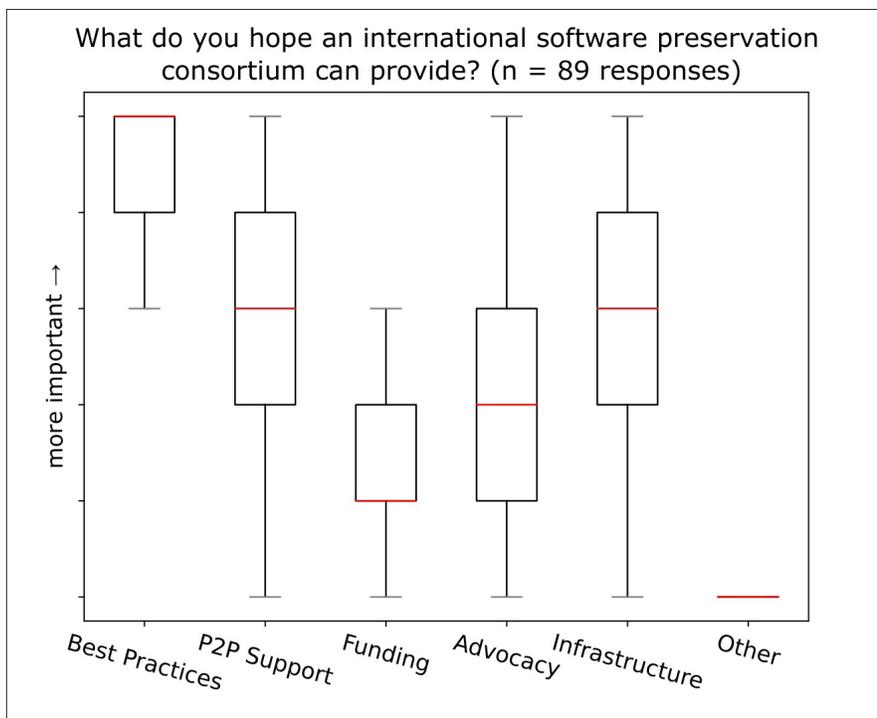


Fig. 2: How survey respondents prioritized potential resources that an international software preservation consortium could provide

These preferences are perhaps indicative of the most keenly felt frustrations of the survey respondents. Knowing what others view as good practice, who to ask for more information, and what is needed to make software preservation happen are everyday struggles, whereas advocating for additional time, funds, staffing, or technology happens on rarer occasions, given the limits of service providers' influence within their organizations. Our analysis of the barriers to implementing software preservation services exposed the complex interrelationships among the varied needs that study participants identified. Insufficient funding, the absence of clear mandates to preserve software, and the variety and wide distribution of practitioners currently engaged in software preservation present challenges for the

individual, the institution, and the field, leaving those who would advocate for deeper investments in staffing, training, and technical infrastructures without a clear place to start.

The in-depth interviews provided excellent opportunities to further explore practitioners' thinking about setting and pursuing field-wide goals. Starting with the study's final objective, which is "to identify activities and future directions for the software preservation community and, where appropriate, the Software Preservation Network," the team asked each interviewee a direct question: "What professional development opportunities, documentation, or other resources could SPN provide to help you meet your goals for software curation and preservation?" An analysis of the interview transcripts identified suggestions for collective action not just in the responses to this question, but throughout the conversations. Five major themes emerged through the analysis, and the research team identified specific recommendations for collective action to address each of these areas.

Following the conclusion of the research team's interviews, SPN's Steering Committee released a framework for research and innovation (Dietrich and SPN Steering Committee 2020). The framework describes broad roles and organization categories for facilitating the germination, refinement, and execution of community projects. The recommendations below could be situated within the framework: they represent a range of possible initiatives and projects that the community might pursue. SPN's role in these projects may vary, leading some while supporting or simply promoting others.

### Taking Action

A non-exhaustive list of organizations, groups, and individuals who might contribute to implementing, championing, or funding the recommendations related to the five themes described in this section:

- Administrators and leaders of individual libraries, archives, museums, and companies
- Administrators and leaders of research and memory organization consortia and collaboratives
- Leadership of software preservation-oriented organizations such as Software Preservation Network, Software Heritage, and Software Sustainability Institute
- Organizational members of Software Preservation Network
- Groups of people who are more experienced with providing software preservation services or members of the 2019–2020 Fostering a Community of Practice (FCoP) Cohort
- Funders such as the Institute of Museum and Library Services (IMLS) or CLIR
- Organizers of common gathering spaces within the software preservation community such as iPRES International Conference on Digital Preservation, National Digital Stewardship Alliance, Maintainers, and BitCurator Consortium
- SPN Working Groups: Law and Policy Working Group, Metadata Working Group, Research-in-Practice Working Group, Training and Education Working Group, Technological Infrastructure Working Group, and Community Engagement Collaborative
- Working groups focused on specific topics within software preservation-oriented organizations or more broadly, e.g., Levels of Digital Preservation Working Group of the National Digital Stewardship Alliance
- Groups within the software preservation community who are actively developing and disseminating software, such as the Emulation-as-a-Service-Infrastructure (EaaS) team
- Organizers of existing digital preservation professional development efforts such as the Digital Preservation Outreach and Engagement Network (DPOE-N)
- Everyone with any interest or connection to software or software preservation

## Theme 1: Representation in the Field

In previous sections, the team identified limitations in the study and outlined some of the demographics related to the software preservation community. While continuing to reflect on the survey's findings and interview conversations, the team agreed that any findings from the study should first be grounded in an awareness of the need to broaden representation and participation within the community of software preservation service providers. Any and all collective actions that SPN or others may take will become markedly more effective if designed to maximize inclusivity at the outset.

Just as demographic questions were not included in the study's survey design, the study's interview protocols did not prompt interviewees for their thoughts on diversity and representation within the software preservation community. Even so, the topic did come up in a few of the conversations. The topic of access to resources and over-representation by highly resourced institutions arose in at least two conversations. This imbalance was discussed in detail in Section 4 as it plays a role in who gets to be represented at the table when decisions are made about standards and practices within the "field." One interviewee commented on the impact this has on what practices service providers can implement, "Office-to-office people are differently resourced. So, someone might do it this way because they have a bunch of money and then someone else might not." In terms of who is represented at the table, one interviewee observed that "A lot of the people who are participating in ... digital preservation broadly, ... especially software preservation now, are big institutions who have a lot of funding and a lot of resources and have the ability to put one specialist on this kind of project."

One interviewee called attention to the fact that two of the team members interviewing her were women:

People power. Yeah, because we're three women on the call. ...If our goal is to have 20% women in computing by 2020, which would be up from where we are right now, about 14% [which is] down from where we were in the '80s, which could have been as high as 40%. Those people doing that work are not going to be women in very high numbers unless we do something about it. People power's important.

Continuing the conversation, the same interviewee further emphasized the importance of looking at these realities as a profession, "So it's really important that we as women figure out what the barriers are that are preventing women from being active as research software engineers or software preservationists. ...And I think as a profession it's something that we need to pay attention to." In Section 3, we discussed the varied representation of women in technology, libraries, and archives. The importance of building a network of colleagues cannot be overstated, especially for underrepresented groups within the community; as one interviewee noted, "I think it's important for us [women] to know each other in the field and to keep those relationships."

The topic of race and ethnic representation was not a significant emphasis in the interviews the team conducted, but the team's analysis and review of current literature suggests that engaging a fully

representative constituency in establishing a shared understanding of “software” and “software preservation” is essential for informing the development of practices and other infrastructure that can ensure the widest possible access to legacy software in the future. It is also important to examine racial and ethnic representation in the broader context of understanding precisely who has been considered as part of the software preservation field—or purposely excluded from the field—despite contributions to the software community over time.

During an Archiving the Black Web panel in 2021, André Brock succinctly described this “difficulty with finding Black technologists in the archives.” Brock further explained that who is represented, and why, often includes outright erasure of the contributions of marginalized groups by institutions in relation to whose work is deemed worthy of archival stewardship and preservation (Brock et al. 2021). In his 2016 keynote for the National Digital Stewardship Alliance (NDSA) conference and essay in Medium, Bergis Jules also ties this reality to the related fields of digital preservation and digital archives, noting that “who gets represented is closely tied to who writes the software, who builds to tools, who produces the technical standards, and who provides funding for the work” (Jules 2016). The Software Preservation Network and others have a clear opportunity to undertake actions related to broadening representation both by acknowledging the historical realities in the field and by empowering new workers with an interest in preserving software.

### Recommendations

1. Invest in collaborative collection development and research documenting and promoting public awareness of the participation of women and individuals with diverse racial, ethnic, linguistic, and cultural backgrounds in the creation and creative use of software, building upon the work of scholars such as Charlton D. McIlwain (2021).<sup>5</sup> Highlight the significance of the contributions of individual developers and preservationists in ways that everyone can identify with and appreciate. Examples of actions that could be taken toward this recommendation:
  - Develop collaborative collection development and documentation strategies, oral history programs, and research initiatives focused on representation of underrepresented developers and software users.
  - Use zines or podcasts to highlight the history and contributions of a wide range of software developers, hobbyists, and preservationists.
2. Increase mentorship opportunities in the software preservation community, prioritizing quality mentorship training that can better support individuals from groups currently underrepresented among software preservation service providers. Examples of ac-

<sup>5</sup> An opportunity to hear McIlwain speak at a Software Preservation Network meeting in September 2020 was instrumental in shaping the research team’s thinking about grounding the recommendations in this report around addressing the profound gaps in representation in narratives about software history and in the information and cultural workforces.

tions that could be taken toward this recommendation:

- Organize a peer mentoring cohort for individuals who are interested in software preservation.
  - Partner with existing mentoring programs in digital library/archives contexts to engage mentor-mentee pairs in explorations of software preservation.
3. Openly reflect upon and raise awareness of how structures within the software preservation community, research, and cultural memory organizations reinforce and reward whiteness; interrogate how software design and development have arisen in a culture of whiteness. Examples of actions that could be taken toward this recommendation:
- Build inclusive practices into the routine work of organizations, similar to the practices outlined in BitCurator Consortium’s Inclusivity Statement (BitCurator Consortium n.d.), a living document of pragmatic actions taken within the community to address some of these long-standing issues.
  - Amplify and support the work of scholars and organizations who are highlighting underrepresented software designers, developers, and preservationists.
4. Create venues for those not currently providing software preservation services to share why they are unable to do so in an open, collaborative environment. Examples of actions that could be taken toward this recommendation:
- Create spaces that allow for open, honest discussion with trusted colleagues.
  - Create interactive posters that call for feedback about barriers to engagement in software preservation and present the posters at a tour of conferences, with the goal of reaching communities that are active in software preservation as well as communities that are not yet active in software preservation.

## Theme 2: Defining the Field, for the Field

The second theme echoes a principal barrier explored in Section 4. Establishing shared conceptualizations of “software” and “software preservation” holds the potential for empowering individuals to come together to build shared capacity and to advance the development of good practices and other social and technical infrastructures. The first clue to the need for attention to this theme arose from confusion among those recruited to take the survey: some individuals reached out to inquire about whether they should consider themselves sufficiently qualified to respond. As shown in Section 2, people who maintain software collections reflect a broad range of professional and institutional contexts, and they hold a wide array of other roles and responsibilities (Figure 1). Such responsibilities may include research, digital preservation, collection management, data curation, or other administrative or managerial functions. Because stewarding software is rarely the sole focus of their roles, many of these individuals do not readily self-identify as “software preservation service providers.” There is as-yet no common understanding of how software

preservation overlaps with and is distinct from other modes of stewardship, and, as emerged in the team's interviews, there is a hunger for engaging in any activity that would help establish this understanding. One participant asked, "What are some core tasks in software preservation? Or core concepts in software preservation where we have some projects that you can contribute to? These are things that need to be done."

One interviewee observed that, in their experience, discussions of the thorny legal and technical challenges related to software preservation tended to obscure the need to establish basic concepts useful for a theoretical mapping of the software preservation landscape: "We need to talk about more. Like infrastructure is the human and the technological and the resources. It's the wherewithal piece." Given the varieties of software and software-related materials, motivations for keeping these materials, organizations engaged in software preservation, and individuals doing the work, the interviewee explained that a broad array of adaptable and interconnected strategies will be necessary for keeping and contextualizing legacy software for the benefit of future generations. The success of these strategies would rely on a common understanding of the scope and variety of software, the history of software development, and how people are working with and using software collections.

The need for a greater foundation for research to further develop the field of software preservation has been pointed out by others, such as Amelia Acker, who concludes that:

There is an urgent need to develop comprehensive resources that describe existing approaches and known preservation standards specific to software technology and software emulation in U.S. cultural heritage organizations. This includes a conception of long-term access to digital cultural memory and a general vocabulary for understanding digital preservation from software development to software emulation perspectives with many types of users (Acker 2021, 1158).

The complexity of software materials makes software curation and preservation uniquely difficult, and at current resourcing levels impossible, to scale. A common understanding of the problems software preservation service providers face and good current strategies for addressing these problems have the potential to advance the field in three ways: by empowering those entering the field to get up to speed, by making it possible to share responsibilities across organizations, and by distinguishing the problems that some members of the field are already engaged with from those that have not yet been addressed. One interviewee described it as a "GIS" (geographic information system) of the field:

[We need a] GIS of information that we would need to do ... [software] preservation, [with] layers of stuff that you're able to explore: ... people, places, things, ... scenarios. ... [And] how could we think about ways ... [to] use software to capture some of what we need? ... [A] lot of times we need to leverage the technologies that we're trying to preserve, to preserve them. ... [T]he preservation technology should be able to [be]

self-looping, ... so we're not kind of going from the legacy and never getting into the future, and we're not starting with the future and never getting to the past: some way of parsing it out, so that people can participate in various kinds of projects that have outcomes to help people prove a model and then adapt it for more general use.

Notably, the Software Preservation Network's Fostering a Community of Practice (FCoP) initiative, which was active during the time the research team conducted interviews, has compiled and generated documentation suitable for beginning to establish a shared understanding of the concerns, challenges, and approaches relevant to preserving software in varied contexts.

### Recommendations

5. (a) Identify and refine a set of shared, capacious definitions of "software" and "software preservation," perhaps building upon the documentation from the SPN Fostering a Community of Practice (FCoP) project. (b) Codify "good practices" for different software preservation use cases and ground those practices in the shared definitions developed in Recommendation 5a. Examples of actions that could be taken toward this recommendation:
  - Bring several international software preservation-oriented organizations together to host a series of virtual working sessions to develop a set of shared definitions.
  - Create template job descriptions that help to define software preservation practices and illuminate the required skills and value of software preservation work.
6. (a) Identify and refine a set of shared, capacious definitions of "software" and "software preservation," perhaps building upon the documentation from the SPN Fostering a Community of Practice (FCoP) project. (b) Codify "good practices" for different software preservation use cases and ground those practices in the shared definitions developed in Recommendation 5a. Examples of actions that could be taken toward this recommendation:
  - Create a series of templates that empower software preservation service providers to articulate and analyze costs of stewardship activities and crosswalk those activities to strategic plans.
  - Convene a peer mentoring group to discuss, complete, and utilize the templates.

### Theme 3: Networking and Community Building

The next set of suggestions offered by interview participants involved creating opportunities to connect software preservation service providers together. All those interviewed had at least a passing familiarity with SPN's early activities, and many acknowledged progress in this area, but others took opportunities in the conversations to stress the value of continuing to build and sustain the network of stakeholders interested in software preservation. The rationale given for prioritizing networking and community building varied by participant. While some emphasized the value of building a foundation for formal professional development or collaborative research among community members, many emphasized the need for less formal engagement,

allowing time to become more familiar with work happening elsewhere and simply validating each individual's experiences of the complex challenges inherent in the work.

Several participants emphasized that intentional community building could help compensate for the isolation of service providers holding sole responsibility for software collections at their organizations, without ready access to peers with experience dealing with similar responsibilities. As one observed, "I find a lot of places have ... one person if they're lucky who might be thinking about this, and they have ... very little community or context. ... I'm interested to see ... where other people might be ... struggling." Having access to others who manage software collections can help providers more easily find different and creative ways to address the problems they are facing. Many stressed that software preservation is just one of their many duties; for those already overstretched, making advances in an emerging area without support from others who have complementary knowledge and experience is too ambitious to attempt.

A few interviewees elaborated on the importance of building networks of practice across institutional, professional, and national boundaries, noting especially the difficulties that those working outside academic research libraries face when seeking out peers. A participant based at a museum described feeling "a little bit siloed from broader academic contexts. ... I wasn't even aware of the software preservation initiative ... because it's more at a university level." While most study participants were based in the United States, those who were not emphasized the need to build and sustain "an international network ... where we [as service providers working outside the United States] could come as participants as well and feel at home discussing stuff." This participant explained that in other established international professional networks, "[Software preservation] is always ... the parallel discussion to what is the prime focus in all the other groups. ... It's like, 'The software thing? Yeah. That's difficult, too. Let's not talk about that now, let's talk about image preservation.' It would be ... really welcome to see an international forum with conferences and stuff like that, where this is the main focus."

Several participants underlined the need for a robust and active community focused exclusively on software preservation: "There are decades of software floating around on the internet. But it seems to be in a lot of other contexts. This is just like a new conversation that's just getting started. And I think that we're in this weird situation where we're kind of just getting started ... with the conversation, but there's already a whole bunch of things that can be done and people who care." Others acknowledged the substantial overlap among existing professional networks dedicated to preserving all varieties of digital artifacts, suggesting that active engagement with other related communities of practice would reduce duplication of effort while leading to more rapid advances in the field: "So much of what we're talking about [is] just strictly good digital practice for digital preservation [and] is not ... unique to software. And I feel like it would be so much stronger if that was broadly understood." The implication of this comment is that by working toward a shared understanding of the stewardship of digital collections and digital preservation across domains, collecting organizations like libraries, museums, and the public and commercial sectors could make more rapid advancements in software preservation.

Given the closely interrelated interests that software collection managers share with the stewards of other digital artifacts, any robust community of practice dedicated to advancing software preservation would require ongoing attention to monitoring the activities of other groups, identifying shared interests, and communicating about priorities and resources in ways that promote the work of other constituencies and interpret the significance of these efforts for keeping software. Similarly, monitoring and communicating expansively across academic, government, cultural heritage, and commercial sectors will require regular acknowledgment of the differences in the missions and priorities of these organizations and a consistent, thoughtful approach to stimulating participants' curiosity to learn about unfamiliar projects and working environments. Historians and theorists of software development, for example, work outside the professional networks inhabited by software collection stewards. Creating more opportunities for open exchange between these scholars and software stewards will raise awareness of barriers affecting software preservation practice while informing new potential approaches to documenting software collections. This approach to networking and community building will require that considerable energy be devoted to research, reflection, and translation.

Interview participants mainly emphasized the benefits of networking and community building for individuals, such as access to sources of expertise and support to help address each person's "struggles," potential sounding boards for new ideas, or demonstrations of workflows and practices that are adaptable to different contexts. A few, however, foregrounded the benefits of a robust network for institutions and the field at large. Institutions with limited resources could, suggested some, rely on a professional network to access specialists not available inside their organizations: "We shouldn't have to try to pretend we know patent law or any of these other things. There are people who would be willing [to assist]. If you ask people questions about the things that they do, ... then [they] are more than willing to share." Several participants expressed a wish for a directory of experts in the field: "We don't always have full access to [information, such as], 'Here's this expert on that thing,' but somebody knows. ... How do you even get access to the people who have that kind of knowledge?" One interviewee emphasized the value of maintaining healthy professional networks for easing new professionals' entry into the field: "I know folks ... in library land and outside of library land. ... who have a lot of expertise. ... I know the people I need to go to, and I know the questions I need to ask them, and I don't know how to distill that, and I don't know how to provide a roadmap to that [for newer colleagues]." Without an obvious way to build connections with others who have complementary expertise, this participant implies, the future of software preservation will remain uncertain.

### Recommendations

7. Establish networking and community building, both formal and informal, as the community's top strategic priority and the scaffold upon which all other software preservation work is built. Co-create a community culture that is known for modeling vul-

nerability, respect, and trust. Examples of actions that could be taken toward this recommendation:

- In organizational budgeting decisions, prioritize investment in community building and facilitation work.
  - Host informal meetups in which individuals can pose software preservation questions for group discussion, support, and troubleshooting.
8. Identify opportunities for translating and deepening connections among software preservationists across domains and disciplines, from hobbyists to collectors, industry to academia, cultural heritage to government, and beyond. Examples of actions that could be taken toward this recommendation:
- Produce a publication or video series that identifies and explores a diverse array of software preservation work and convene discussions to connect stakeholders.
  - Create personas that help people see themselves in software preservation.

#### Theme 4: Informal and Formal Learning

A fourth, closely related theme that emerged from interview conversations was the value of ongoing learning activities for service providers. Software preservation activities draw upon a wide-ranging and eclectic skillset, and these skills require ongoing refreshment. Participants most familiar with SPN mentioned past shared learning activities and expressed an appetite for more of the same, while at the same time broadening the range of offerings so that those with different levels of experience would be able to find a productive place to begin expanding their knowledge. Several individuals stressed that a robust network creates opportunities for continual learning, a necessity for both new and experienced practitioners as the field evolves and becomes better defined. Interviewees most frequently expressed a desire for informal interactions in venues suitable for exchanging ideas and solving problems: “I think that that’s a really useful way to also bring together people to talk about some of these issues and experiences that others have had so that we can learn from each other.”

Several participants mentioned that they came to the work of providing software preservation services without obtaining any formal training through degree-granting or certificate programs. As one interviewee described their experience, “These skills emerge at need. You know, sometimes ‘baptism under fire.’” Perhaps because of the dearth of formal training programs related to software preservation, more informal opportunities to learn are highly valued, even by those with substantial experience. One participant expressed particular interest in having “chance[s] to talk more about any kind of preservation or content. ...Not just saying like, here’s a bunch of slides and stuff to read, [but] an avenue for saying, ‘I have a project and I have questions and I need help.’” Another agreed, “I like that idea of the special interest group. You know, perhaps we could make a special interest group for the funny carriers.” A key feature of a robust network of individuals with identified areas of expertise would be facilitating learning opportunities, either through more deliberate “workshop-

ping” or through exploratory discussions focused on emerging areas of work.

Several interview subjects attested to the value of participating in more structured events and formal, planned programming similar to SPN’s past offerings such as the 2019 webinar series on the Code of Best Practices in Fair Use for Software Preservation (Meyerson et al. 2019): “I think the online webinars and training opportunities [are] great, and [I] definitely encourage SPN to continue on with that.” Another echoed this sentiment: “Training documents, ... webinar opportunities, bringing in special speakers—all of those things clearly need to happen, too. But I think that ... we’re already doing that. I think they’re [SPN] doing a good job with that.” Another expressed a wish for opportunities to advance learning through attending traditional conference-style reporting sessions: “I’ve got a project, and therefore I learned these things. How can we share more of that information?” Formal presentations, workshops, and courses can support any number of possible needs, but in the context of a rapidly evolving field, creating opportunities for those more advanced in the field to show others what is possible provides a valuable service, empowering others to make more informed decisions about directing attention and resources to their software collections.

Several participants emphasized the need for learning resources suited to a wide variety of experience levels. Accessing very basic, step-by-step documentation of a solution to a particular problem, for example, can save individuals significant time and frustration: “I just want to do the work. I don’t want to have to think about [questions such as], ‘Is this the best ... microservice, or is that the best microservice, [or] how come this doesn’t work?’ You know, I don’t want to think about any of that stuff.” Another interviewee stressed that since many individuals working in digital preservation encounter questions about software only occasionally, they need to be directed to resources that can help orient them, “You know what, I don’t deal with software, but I just got this request, and where do I go?” Since resourcing is a significant barrier to progress, providing learning opportunities for administrators more removed from the day-to-day of staff working directly with software collections could address a general “lack of understanding” about the levels of support and skill that work with software collections requires: “[Software-dependent artifacts are] much more expensive to acquire and preserve than your standard chair or teacup. ... I think that people misjudge how complex these works can be.”

## Recommendations

9. Pilot informal peer-to-peer, birds-of-a-feather learning groups focused on topics of shared interest that are open to the software community at large. Create incentives for these groups to develop open learning. Examples of actions that could be taken toward this recommendation:
  - Identify topics with strong shared interest.
  - Convene a variety of informal learning groups or formal, funded cohort groups to discuss and develop good practices related to the identified topics.

10. Develop or request proposals for a workshop on cultivating individual and organizational resilience in the face of complexity. Examples of actions that could be taken toward this recommendation:
  - Work with and pay for a consultant on organizational resilience to design a train-the-trainer workshop.
  - Invite workshop participants to share the material within their local contexts.
11. Recruit a team of organizational leaders and software preservation service providers to present at a series of conferences for administrators on the nature of software preservation work, including day-to-day activities, costs, and benefits. Examples of actions that could be taken toward this recommendation:
  - Run a campaign focused on administrators including social media, presentations, targeted 1:1 meetings, conference sessions, or symposia.
  - Incorporate the templates suggested as part of Recommendation 6.

### Theme 5: Shared Infrastructures and Model Practices

A fifth group of suggestions and hopes shared by the interview participants related to establishing and sustaining shared “infrastructures” that could benefit the international software preservation community at large. While a common set of concepts and definitions might be considered one kind of fundamental infrastructure and a sustainable community network engaged in ongoing learning activities is perhaps another, participants identified other more concrete ways that collective investment in infrastructures will be necessary to fuel advancements in the field. Participants’ notions of infrastructure encompassed observations closely aligned with work on cyberinfrastructures presented by Bowker, Baker, Millerand, and Ribes (2010) who described them as “value-added systems and services that can be widely shared across scientific domains, both supporting and enabling large increases in multi-disciplinary science while reducing duplication of effort and resources” (100). Though cyberinfrastructure focuses on scientific domains, this definition can also be usefully extended to nonscientific work (e.g., humanities research, cultural production, and museum exhibition). Shared infrastructures can include social groups or activities, technologies, research products, documentations of practice, or some combination of any of these resources.

Several emerging collaborative efforts have already provided shared infrastructures that serve the field of software preservation, and many of these were noted by study participants:

- **Emulation-As-A-Service Infrastructure (EaaSI)**: This is a software platform that enables the distribution and retrieval of software emulation environments, eliminating the need for organizations to maintain emulation environments that enable users to execute the software in their collections (SPN

n.d. [a]). Several interview participants expressed interest in and excitement about EaaS: “I feel like I’m just going to keep bringing up the EaaS project, but I think I am kind of really excited for it ... lowering the technological barrier to actually getting involved.”

- **Wikidata for Digital Preservation:** This project helps collaborators access and update information about computing relevant to identifying and describing software titles, file formats, and operating systems (Wikidata n.d.).
- **Fostering a Community of Practice:** This completed project brought together practitioners to describe and document current workflows and other practices implemented at six distinct participating organizations, describing in detail a diverse set of use cases that together illustrate some of the breadth and complexity of the field of software preservation (SPN n.d.[b]).
- Legal and policy resources such as the **Code of Best Practices for Fair Use in Software Preservation** (Aufderheide et. al. 2019), produced through a partnership between SPN and the Association of Research Libraries, and the **Preservationists’ Guide to the DMCA Exemption for Software Preservation**, created through a collaboration between SPN and the Cyberlaw Clinic at the Berkman Klein Center for Internet and Society at Harvard Law School (SPN 2018).
- **The National Software Reference Library:** Supported by the National Institute of Standards and Technology (NIST), this project collects digital signatures of known software applications, helping people identify and understand digital files in their collections (NIST 2016).
- **Software Preservation Network (SPN) Working Groups:** Volunteer-led groups of practitioners organized through SPN regularly identify shared needs and create resources that advance the field of software preservation. Besides Research-in-Practice, SPN’s other current groups focus on law and policy, metadata, technological infrastructure, training and education and community engagement.

These examples show a variety of “shared infrastructures” from time-bound, focused projects made possible through grants or community volunteers, to larger-scale repositories and portals designed to expand over time. The research team’s interviews revealed an appetite for more robust and varied resources such as these. Multiple interview participants, for example, mentioned the value of identifying and exploring more real-world use cases for software preservation, similar to those documented through the Fostering a Community of Practice project. Documented use cases hold value not only as models for implementing software preservation services but also for advocating for increased investments in labor, training, and technologies needed to maintain and create access to software materials.

While many of the interviewees mentioned the importance of software emulation for providing access and the need to devote resources to make emulation more widely available in the future, relatively few participants reported providing access to software through emulation themselves. Instead, their greater day-to-day concerns pertained to identifying and assessing software already on shelves or in digital repositories, developing software collections, creating metadata for software, conducting outreach to software developers at their institutions, and helping researchers find information related to legacy software. Similarly, participants noted the importance of legal and policy infrastructure while also expressing that they are not far enough along with software preservation to apply the guidance in day-to-day work. Although there are established practices for dealing with other types of digital materials that can be adapted for keeping software materials, it was clear to study participants that more refined and readily accessible model workflows that can be tailored to a research or memory organization's priorities and capacity would significantly benefit their daily working lives. The relative simplicity and adaptability of the National Digital Stewardship Alliance's (NDSA) Levels of Digital Preservation Matrix—now available in multiple languages—helps articulate a set of common needs and practices that allow practitioners working with different levels of experience and support to improve services over time, without imposing a “one size fits all” set of expectations for good practice. A similar or extended framework for levels of software preservation could prove a valuable resource for growing the international community of software stewardship and building consensus around the foundational concepts that inform the work.

A few interviewees suggested that SPN pursue the creation of a shared set of research and development infrastructures that could operate as a site for collaboration and learning: I think that one of the things that SPN could be [is] like a really good ... central place ... almost like a sandbox. ... [SPN could organize its activities around the questions,] ‘What are some core tasks in software preservation, or core concepts in software preservation, where we have some projects that you can contribute to, to learn the basics or [show that] these are things that need to be done?’ If you are an archive that has started getting Jupyter notebooks attached to papers, or you want to start being able to accession these, ... [you could] contribute to a task force or a working group that has this [activity as its focus].

Others emphasized the need for shared infrastructure to have the advantage of economies of scale in collection development and maintenance. One participant asked: “Are there shared infrastructure models that could be useful? Because we all don't need to be collecting the same software ... and really reinventing the wheel. So where can we look to the community to provide some of these solutions? [And] when do we need to do things institutionally based?” Shared infrastructure would help even out capacity and resources across the community of practice, since currently there are wide gaps, especially between larger and smaller organizations.

### Recommendations

12. Create a “levels of software preservation” matrix that empowers organizations and service providers to identify and prioritize tangible software preservation actions from a shared baseline of understanding. Examples of actions that could be taken toward this recommendation:
  - Confer with the working group that created the National Digital Stewardship Alliance’s Levels of Digital Preservation chart (NDSA 2019).
  - Query software preservation service providers about what their goals would be in using the matrix.
13. Pursue a pilot project for formal expertise-sharing and collaborative workflow development among organizations that are invested in software preservation. Explore sustainability models within the scope of the pilot. Examples of actions that could be taken toward this recommendation:
  - Consult with existing expertise-sharing collectives such as the Data Curation Network or the Memory Lab Network to inform the design of a pilot.
  - Host a gathering in which software preservation service providers sketch out workflows to show how software preservation activities are performed and where gaps or pain points exist.

## Conclusion

The research team set out to explore four research objectives: to identify conceptualizations, skillsets, barriers, and future directions related to software preservation work. Findings from this study suggest that, in its next phase, the software preservation community should embrace a constellation of actions to broaden representation, define the field, invest in community building, nurture continuous learning, and cultivate multifaceted infrastructures.

During the May 2022 Software Preservation Network community call, participants began reflecting on these findings. Research team members shared the themes and high-level recommendations presented in the Moving Forward section of this paper. Community members brainstormed specific activities in line with the recommendations, identified stakeholders who might contribute, and described how the outcomes might help individuals, institutions, and the field. Initial suggestions for potential activities ranged from creating template job descriptions that encompass software preservation work, to increasing involvement in hobbyist spaces, to offering preservation consultation assistance for underrepresented developers. The research team invites any and all readers—individuals or organizations, new or familiar—to engage with the recommendations and refine them into the software preservation futures they can collectively imagine.

## References

- Acker, Amelia. 2021. "Emulation Practices for Software Preservation in Libraries, Archives, and Museums." *Journal of the Association for Information Science and Technology* 72(9): 1148–60. <https://doi.org/10.1002/asi.24482>.
- Albert, Kendra. 2018. "A Victory for Software Preservation: DMCA Exemption Granted for SPN" (October 26). <https://clinic.cyber.harvard.edu/2018/10/26/a-victory-for-software-preservation-dmca-exemption-granted-for-spn/>.
- ALISE (Association for Library and Information Science Education). 2022 *Statistical Report: Trends and Key Indicators in Library and Information Science Education*. [https://ali.memberclicks.net/assets/documents/statistical\\_reports/2022/ALISE%202022%20Statistical%20Report%20final%20July%202022.pdf](https://ali.memberclicks.net/assets/documents/statistical_reports/2022/ALISE%202022%20Statistical%20Report%20final%20July%202022.pdf).
- Atkins, Winston, Carol Kussmann, Katherine Kim, and Aliya Reich. 2017. "Staffing for Effective Digital Preservation 2017: An NDSA Report" (September 13). <https://doi.org/10.17605/OSF.IO/3RCQK>.
- Aufderheide, Patricia, Brandon Butler, Peter Jaszi, and Krista Cox. 2019. *Code of Best Practices in Fair Use for Software Preservation*. Association of Research Libraries. <https://doi.org/10.18130/v3-kpcj-k283>.
- Banks, Brenda. 2006. "Archival Census and Education Needs Survey of the United States, Part 6. A\*CENSUS: Report on Diversity." *The American Archivist* 69(2): 482–93.
- BitCurator Consortium. n.d. "Inclusivity Statement." Accessed October 27, 2022. <https://bitcuratorconsortium.org/inclusivity-statement/>.
- Blumenthal, Karl-Rainer, Peggy Griesinger, Julia Kim, Shira Peltzman, and Vicky Steeves. 2020. "What's Wrong with Digital Stewardship: Evaluating the Organization of Digital Preservation Programs from Practitioners' Perspectives." *Journal of Contemporary Archival Studies* 7(1). <https://elischolar.library.yale.edu/jcas/vol7/iss1/13>.
- Bowker, Geoffrey C., Karen Baker, Florence Millerand, and David Ribes. 2010. "Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment." In *International Handbook of Internet Research*, edited by Jeremy Hunsinger, Lisbeth Klastrup, and Matthew Allen, 97–117. Dordrecht: Springer Netherlands. [https://doi.org/10.1007/978-1-4020-9789-8\\_5](https://doi.org/10.1007/978-1-4020-9789-8_5).

Brock, André, Jr., Catherine Knight-Steele, Charlton McIlwain, and Raven Maragh-Lloyd. 2021. “Not new to dis, we true to dis – The Black Presence Online From Then to Now.” Panel presented at Archiving the Black Web, Online, April 29, 2021. Archiving the Black Web video, 1:39:39. <https://archivingtheblackweb.org/schedule/>.

Data USA. 2020. “Software Developers.” <https://datausa.io/profile/soc/software-developers>.

Data USA. 2019. “Archivists, Curators, & Museum Technicians.” <https://datausa.io/profile/soc/archivists-curators-museum-technicians>.

Davis, Denise, and Tracie Hall. 2012. “Diversity Counts, 2012 Update, Data Tables A-1 and A-2.” American Library Association. <https://www.ala.org/aboutala/offices/diversity/diversitycounts/divcounts>.

Dietrich, Dianne, and SPN Steering Committee, Research and Innovation Subcommittee. 2020. “SPN Research & Innovation Framework: Overview | Software Preservation Network (SPN),” (May 13). <https://www.softwarepreservationnetwork.org/spn-research-innovation-framework-overview/>.

Digital Preservation Coalition. 2015. *Digital Preservation Handbook*. <https://www.dpconline.org/handbook/glossary#C>.

DPE Research Department. 2021. “Library Professionals: Facts, Figures, and Union Membership.” Washington DC: Department for Professional Employees – AFL-CIO. <https://www.dpeaflcio.org/factsheets/library-professionals-facts-and-figures>.

GitHub. 2017. “Open Source Survey.” <https://opensourcesurvey.org/2017/>.

Handel, Dinah, and Mark A. Matienzo. 2021. “Facilitating and Illuminating Emergent Futures for Archival Discovery and Delivery: The Final Report of the Lighting the Way Project.” Stanford Digital Repository. <https://doi.org/10.25740/JM302FQ5311>.

Hettrick, Simon. 2014. “It’s Impossible to Conduct Research without Software, Say 7 out of 10 UK Researchers.” *Software and Research: The Institute’s Blog*, December 4, 2014. <https://www.software.ac.uk/blog/2014-12-04-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers>.

Higgins, Sarah. 2008. “The DCC Curation Lifecycle Model.” *International Journal of Digital Curation* 3(1): 134–40. <https://doi.org/10.2218/ijdc.v3i1.48>.

Jules, Bergis. 2016. “Confronting Our Failure of Care Around the Legacies of Marginalized People in the Archives.” *On Archivy* (blog), November 12, 2016. <https://medium.com/on-archivy/confronting-our-failure-of-care-around-the-legacies-of-marginalized-people-in-the-archives-dc4180397280>.

McIlwain, Charlton D. 2021. *Black Software: The Internet & Racial Justice, from the AfroNet to Black Lives Matter*. New York: Oxford University Press.

Meyerson, Jessica, Patricia Aufderheide, Peter Jazsi, Krista Cox, Brandon Butler. 2019. “Code of Best Practices for Fair Use in Software Preservation: Why and How?” Software Preservation Network and Association of Research Libraries video, 1:01:00. (February 25). <https://www.softwarepreservationnetwork.org/core-activities/law-policy/#resources>.

Mian, Anam. *ARL Annual Salary Survey 2020*. Washington, DC: Association of Research Libraries, 2022. <https://doi.org/10.29242/salary.2020>.

NDSA (National Digital Stewardship Alliance). 2022. “2021 Staffing Survey Report.” NDSA Staffing Survey Working Group. <https://osf.io/emwy4/>.

NDSA (National Digital Stewardship Alliance). 2019. “Levels of Digital Preservation Matrix V2.0.” NDSA Levels of Preservation Working Group. <https://osf.io/2mkwx/>.

NIST (National Institute of Standards and Technology). 2016. “National Software Reference Library (NSRL).” (April 19). <https://www.nist.gov/itl/ssd/software-quality-group/national-software-reference-library-nsrl>.

Philippe, Olivier. 2018. “What Do We Know about RSEs? Results from Our International Surveys.” *Software and Research: The Institute’s Blog*, March 12, 2018. <https://software.ac.uk/blog/2018-03-12-what-do-we-know-about-rses-results-our-international-surveys>.

Rios, Fernando. 2016. “Preserving and Sharing Software for Transparent and Reproducible Research: A Review.” <https://doi.org/10.17605/OSF.IO/3E69D>.

Schonfeld, Roger, and Liam Sweeney. 2017. *Inclusion, Diversity, and Equity: Members of the Association of Research Libraries*. Ithaka S+R. <https://doi.org/10.18665/sr.304524>.

Skinner, Makala and Ioana Hulbert. 2022. “A\*CENSUS II All Archivists Survey Report.” Ithaka S+R. Last Modified 22 August 2022. <https://doi.org/10.18665/sr.317224>.

SPN (Software Preservation Network). 2018. *A Preservationists Guide to the DMCA Exemption for Software Preservation*. Accessed August 15, 2022. <https://www.softwarepreservationnetwork.org/a-preservationists-guide-to-the-dmca-exemption-for-software-preservation/>.

SPN (Software Preservation Network). n.d.(a). “Emulation-as-a-Service-Infrastructure (EaaS).” Accessed August 15, 2022. <https://www.softwarepreservationnetwork.org/emulation-as-a-service-infrastructure/>.

SPN (Software Preservation Network). n.d.(b). “Fostering a Community of Practice (FCoP).” Accessed August 15, 2022. <https://www.softwarepreservationnetwork.org/fcop/>.

Steeves, Vicky, and Sarah Nguyen. 2020. “Investigating and Archiving the Scholarly Git Experience.” QDR Main Collection, December 14, 2020. <https://doi.org/10.5064/F6VOIB8H>.

US Copyright Office, Library of Congress. 2021. “Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies.” *Federal Register* 86, no. 206 (October 28, 2021): 59627. Accessed August 15, 2022. <https://www.govinfo.gov/content/pkg/FR-2021-10-28/pdf/2021-23311.pdf>.

Wikidata. n.d. “About Wikidata for Digital Preservation.” Accessed August 15, 2022. <https://wikidp.org/about>.

## Appendix A: Definitions

**Curation:** Digital curation includes a broad set of activities conducted along a lifecycle to conceptualize, appraise, ingest, preserve, and provide access to various digital materials. While preservation actions within the digital curation lifecycle are focused on the long term, curation actions are oriented toward the use of standards to describe and represent the digital content for current reuse or deal with materials as they migrate (Higgins 2008). We also consider reproducibility as an aim of curation activities.

**Preservation:** Digital preservation and its associated actions is a component of the digital curation lifecycle that takes a long-term view focused on ensuring “that data remains authentic, reliable and usable while maintaining its integrity” over time (Higgins 2008). There are short-, medium-, and long-term stages of preservation (Digital Preservation Coalition 2015). The differences between curation and preservation are difficult to untangle and can be dependent on the community doing the activities. In this study, we distinguish curation as the overarching set of activities that includes preservation actions.

**Research and memory organizations:** In using this term, the research team aims to encompass the wide range of organizations concerned with the generation, dissemination, and preservation of knowledge and culture—from museums to community archives to academic libraries and beyond. Both software and software preservation are critical to the missions of these organizations in two ways: (1) to understand works of software as cultural artifacts and (2) to maintain software tools required to provide access to other digital artifacts.

**Services:** With the term *services*, the research team aims to include a broad and expanding array of activities that relate to curation, preservation, and stewardship of software in different contexts. In a survey question about services, the research team used the following language to suggest possible activities: “Consultation about software preservation,” “Preservation of software,” “Metadata creation for preserved software,” “Access to preserved software,” and/or “Other (describe).”

**Software:** Findings from this study indicate that the term *software* carries a range of meanings for different stakeholders. The research team suggests the following definition of *software* as a useful baseline: “A list of commands that causes a computer to behave in a certain way” (Rios 2016).

## Appendix B: Survey Instrument

This survey instrument is provided as an appendix for the convenience of readers. A downloadable version is available on Zenodo under an open license that permits adaptation and reuse: <https://doi.org/10.5281/zenodo.7086618>.

**The Software Preservation Network (SPN) Research Working Group invites you to participate in a survey on practices, needs, and gaps related to software preservation.**

### Why Does This Matter?

For decades, researchers and practitioners in information science, digital preservation, and allied fields have discussed the necessity of software preservation: preserving software is a prerequisite for preserving and providing access to digital cultural heritage and research, and software is increasingly considered a research product or artifact in itself.

How are cultural heritage professionals working on preserving software? What are the obstacles to software preservation? Do best practices exist?

To help answer these important questions, we need your input. This survey is the first in what we hope will be a longitudinal study of software preservation services over time.

### Who Should Participate?

Any individual involved in activities that involve or rely on software preservation is encouraged to take the survey. For the purposes of this survey, software preservation encompasses a wide range of experimental or established services or actions at organizations such as collecting original software media and documentation, consultations with software producers or users of specialized or obsolete software, preservation of software code or executable files, metadata creation for preserved software, etc.

### How will the survey information be used?

Anonymized data from the study will be made available to the profession, along with analysis of current trends and possibilities for future research.

### The following 3 questions were shown to all respondents:

**Q1** (Required) Does your organization currently provide services (established or experimental) related to software preservation (e.g., consultation, preservation, creation of metadata for preserved software, etc.)?

- Yes
- No

**Q2 (Required)** Please indicate the type of organization you are affiliated with.

- Academic
- Commercial
- Government
- Nonprofit
- Other (please explain below):

**Q3 (Required)** Please select the option below that best describes your job position in your organization.

- Administrator
- Archivist
- Conservator
- Curator
- Information Technologist
- Librarian
- Other (please explain below):

The following set of questions was shown to respondents who answered Yes to Q1:

**Q4** How often do you provide the following services?

	Frequently	Rarely (a few times a year)	Never
Consultation about software preservation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Preservation of software	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadata creation for preserved software	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Access to preserved software	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Q5** If you provide a service other than those previously mentioned in question 4, please describe the service(s) and list associated frequencies.

**Q6** Which software components or related items do you preserve?

- Executables (e.g., .exe files, CD-ROMs, etc.)
- Source code
- Documentation about software
- Other (please explain below):

**Q7** What is the license status of the software you preserve?

- Open source license (e.g., MIT, Apache, etc.)
- Proprietary / commercial
- Don't know (🤷)
- Other (please explain below):

**Q8** What is the provenance of the software you preserve?

- Software created by or within your institution
- Software created by an entity or individual other than or outside of your institution
- Other (please explain below):

**Q9** What are your / your institution's motivations or use cases for preserving software?

- Preserve software because it is part of the cultural record
- Preserve software in order to preserve and / or provide access to other collections / data
- Preserve software for legal / regulatory compliance
- Preserve software so it can be reused
- Preserve software to aid in scientific reproducibility
- Other (please explain below):

**Q10** Do you have software-specific workflows in place for any of the following?

- Accessioning software
- Creating metadata for software
- Ingesting software
- Preserving software
- Providing access to software

**Q11** If you're comfortable sharing your workflows, please upload them here.

**Q12** Do you have any of the following policies related to software preservation?

- Collection development policy
- Deed of gift / donor agreement
- License agreement
- Preservation policy
- Access policy
- Other (please explain below):

**Q13** If you're comfortable sharing your policies, please upload them here.

**Q14** Have you received requests from researchers who want to access preserved software?

- Yes
- No

**Q15** How many staff are at all involved in software preservation at your institution? (for example, involvement might include technical maintenance of required systems, metadata work, selection / curation of software, providing access to preserved software, etc.)

- 0
- 1
- 2-4
- 5-10
- 11+

**Q16** What kinds of resources (financial, human resources, etc.) do you rely on to engage in software curation and preservation activities?

- Resources from your institution
- Resources from granting organizations
- Support and expertise from professional communities
- Other (please explain below):

The following set of questions was shown to respondents who answered No to Q1:

**Q17** Are you considering providing services related to software preservation in the future?

- Yes
- Maybe
- No

*Display The Following Question:*

*If “Are you considering providing services related to software preservation in the future?” = Yes*

*and “Are you considering providing services related to software preservation in the future?” = Maybe*

**Q18** What would be your / your institution’s motivations or use cases for preserving software?

- Preserve software because it is part of the cultural record
- Preserve software in order to preserve and / or provide access to other collections / data
- Preserve software for legal / regulatory compliance
- Preserve software so it can be reused
- Preserve software to aid in scientific reproducibility
- Other (please explain below):

*Display The Following Question:*

*If “Are you considering providing services related to software preservation in the future?” = Yes*

*and “Are you considering providing services related to software preservation in the future?” = Maybe*

**Q19** What information would you use and / or like to have to make an argument for obtaining the resources and support needed to engage in software preservation?

- Data about software preservation needs at your institution
- Data about software preservation activities at peer institutions
- Best practice guidelines, case studies, and / or instructions for engaging in software preservation activities
- Other (please explain below):

*Display The Following Question:*

*If “Are you considering providing services related to software preservation in the future?” = No*

**Q20** Why are you not considering providing services related to software preservation in the future? Please move these items to rank them in order from most important (1) to least important (6).

- \_\_\_\_\_ Lack of human resources
- \_\_\_\_\_ Lack of financial support
- \_\_\_\_\_ Lack of time
- \_\_\_\_\_ Lack of need
- \_\_\_\_\_ Other (please explain below):

The following six questions were shown to all respondents:

**Q21** What skill sets do you think are needed to collect, preserve, curate, and provide access to software?

- Knowledge of software development
- Knowledge of software uses
- IT skills
- Understanding of archival best practices Archives skills (e.g., accessioning, appraisal, provenance)
- Community engagement skills
- Reference skills
- Other (please explain below):

**Q22** What do you hope an international software preservation consortium could provide to you / your institution that you feel you could not otherwise accomplish on your own or with existing partnerships? Please move these items to rank them in order from most important (1) to least important (6).

- \_\_\_\_\_ Best practices
- \_\_\_\_\_ Peer-to-peer support
- \_\_\_\_\_ Pathways to funding
- \_\_\_\_\_ Support with advocacy within and / or beyond your institution
- \_\_\_\_\_ Technical infrastructure
- \_\_\_\_\_ Other (please explain below):

**Q23** What questions do you have about software preservation that you'd like to explore or learn more about?

**Q24** Do you have anything else you'd like to tell us?

**Q25** Please state your employer / institutional affiliation.

**Q26** Are you interested in being interviewed about these topics in more detail? If so, please provide your name and email:

## Appendix C. Interview Guides

These interview guides are provided as an appendix for the convenience of readers. A downloadable version of the guides is also available on Zenodo under an open license that permits adaptation and reuse: <https://doi.org/10.5281/zenodo.7086618>.

### Service provider interview guide #1: for subjects currently providing software curation services

#### I. Interviewer Version (with follow-up questions)

##### Part One: Motivations, 4-5 questions + follow-ups, est 25 min

If you remember, how did you hear about the SPN Service Provider Survey?

(Confirm types of software preserved at the subject's institution, Q5-7) Can you tell me more about the kinds of software and software-related materials you have in your collections?

Follow-ups: What does/did the software do? Why did your institution acquire it? Where do you store it currently?

(Confirm whether researchers have requested access to software, Q13) If you have received requests from researchers to access and/or use software in your collections, how did you hear about them and what did the requesters want to do?

(Confirm motivations/use cases mentioned in the survey, Q8) Can you tell me more about why preserving software is relevant to your institution's mission or goals?

Follow-ups: Who are the people most likely to be interested in the software you're preserving, and why would they be interested in having access to it?

Follow-up, for those with compliance responsibilities: Tell me about the regulations or compliance issues that affect your approach to software preservation.

What information do you use to advocate for software preservation resources?

What are the biggest challenges you face in preserving software at your institution?

Follow-ups: What information or resources (a community forum, workshops, expertise, needs assessments, data, equipment, etc.) do you use to address these challenges?

**Part Two: Practices, Policies, and Resources, 7+ questions, est. 25 min**

(Confirm staffing levels for software preservation services indicated in the survey, Q14, or apparent from institutional website). Who are the primary people responsible for software curation and preservation at your institution and what do they do? What other responsibilities does each have apart from software preservation?

Follow-ups: How long have these positions been established at your institutions? About how long has each person on the team been in their current role?

(Confirm types of workflows mentioned/shared in the survey, Q9-10). How did you craft your workflows for software collecting, curation, preservation, and access?

Follow-ups: How long have they been established? Why did you choose to design your workflow in this way? What systems and standards do you use in the course of this work (e.g., accessioning, metadata, ingest, preservation, dissemination, discovery, etc.)? How do you document them? How often are they re-assessed and amended? What services or platforms do you use (emulation or repository services, version control tools, etc.)?

(Confirm types of policies mentioned/shared in the survey, Q11-12). Can you tell me more about your policies for software collecting, curation, and preservation (e.g., policies such as collection development policy, language in deeds of gift, etc.)?

Follow-ups: How do you document these policies and agreements? How widely do you share or publicize them? How long have they been established? How often are they re-assessed and amended?

What types of agreements for use and access do you provide?

What contextual documentation or other information (software installation manuals, etc.) would you have in an ideal world to preserve the software in your collections?

(Confirm resources mentioned in survey, Q15): Are there any other sorts of institutional resources (e.g., people, departments, servers, institutional grants) that you use for software curation and preservation activities?

Are there any resources beyond your institution that you use for software curation and preservation activities?

**Part Three: Skills and Professional Development, 3-5 questions + follow-ups, est. 10 min**

(Confirm skills mentioned in survey, Q20). Why did you pick [insert survey question response] as most important for collecting, preserving, curating, and providing access to software?

Follow-ups: Are there other skills you think are important to collect, preserve, curate, and provide access to software? If you were advising a new professional interested in software preservation, where would you tell them to go to develop these skills?

(Confirm hopes for SPN, Q21). What professional development opportunities, documentation, or other resources could SPN provide to help you meet your goals for software curation and preservation?

(Confirm subject's questions about software preservation, if any, Q22) What (other) questions do you have about software preservation that you'd like to explore or learn more about?

#### **Part Four: Wrapping Up, 1 question, est. 3 min**

Is there anything that I should have asked you today that we have not yet discussed?

[Before finally concluding, the interviewer should verify any details in the notes that weren't clear and thank the subject for volunteering the time to the study, reminding them of the group's plans for the interview data as needed.]

## **II. Version to share with interviewees**

### **Part One: Motivations**

If you remember, how did you hear about the SPN Service Provider Survey?

Can you tell me more about the kinds of software and software-related materials you have in your collections?

If you have received requests from researchers to access and/or use software in your collections, how did you hear about them and what did the requesters want to do?

Can you tell me more about why preserving software is relevant to your institution's mission and/or goals?

What are the biggest challenges you face in preserving software at your institution?

Follow-ups: What information or resources (a community forum, workshops, expertise, needs assessments, data, equipment, etc.) would you use to address these challenges?

### **Part Two: Practices, Policies, and Resources**

Who are the primary people responsible for software curation and preservation at your institution and what do they do? What other responsibilities does each have apart from software preservation?

How did you craft your workflows for software collecting, curation, preservation, and access?

Can you tell me more about your policies for software collecting, curation, and preservation (e.g., policies such as collection development policy, language in deeds of gift, etc.)?

What types of agreements for use and access do you provide?

What contextual documentation or other information (software installation manuals, etc.) would you have in an ideal world to preserve the software in your collections?

Are there any other sorts of institutional resources (e.g., people, departments, servers, institutional grants) that you use for software curation and preservation activities?

Are there any resources beyond your institution that you use for software curation and preservation activities?

### **Part Three: Skills and Professional Development, 3-5 questions + follow-ups, est. 10 min**

Which skills are most important for collecting, preserving, curating, and providing access to software?

What professional development opportunities, documentation, or other resources could SPN provide to help you meet your goals for software curation and preservation?

What (other) questions do you have about software preservation that you'd like to explore or learn more about?

## **Service provider interview guide #2: for subjects not currently providing software curation services)**

### **I. Interviewer Version (with follow-up questions)**

#### **Part One: Motivations, 4-5 questions + follow-ups, est 25 min**

If you remember, how did you hear about the SPN Service Provider Survey?

(Confirm types of software or software-related materials mentioned in the survey, if any, Q5-7) Can you tell me more about the kinds of software and software-related materials you have in your collections?

Follow-ups: What does/did the software do? Why did your institution acquire it? Where do you store it currently?

(Confirm whether researchers have requested access to software, Q13) If you have received requests from researchers to access and/or use software in your collections, how did you hear about them and what did the requesters want to do?

(Confirm motivations/use cases mentioned in the survey, Q16-17) Can you tell me more about why preserving software is relevant to your institution's mission or goals?

Follow-ups: Who are the people most likely to be interested in using software in your collections? Who are the people most likely to need access to software preservation services provided by your institution in the future?

Follow-up, for those who might expect to have compliance responsibilities: Tell me about the regulations or compliance issues that might affect your approach to software preservation in the future.

What contextual documentation or other information (software installation manuals, etc.) would you like to have in order to preserve the software in your collections?

What are the biggest challenges you face in getting started with preserving software at your institution?

Follow-up: What information or resources (a community forum, workshops, expertise, needs assessments, data, equipment, etc.) would you use to address these challenges?

### **Part Two: Skills and Professional Development, 3-5 questions + follow-ups, est. 10 min**

(Confirm skills mentioned in survey, Q20). Why did you pick [insert survey question response] as most important for collecting, preserving, curating, and providing access to software?

Follow-ups: Are there other skills you think are important to collect, preserve, curate, and provide access to software? If you were advising a new professional interested in software preservation, where would you tell them to go to develop these skills?

(Confirm hopes for SPN, Q21). What professional development opportunities, documentation, or other resources could SPN provide to help you meet your goals for software curation and preservation?

(Confirm subject's questions about software preservation, if any, Q22) What (other) questions do you have about software preservation that you'd like to explore or learn more about?

### **Part Three: Wrapping Up, 1 question, est. 3 min**

Is there anything that I should have asked you today that we have not yet discussed?

[Before finally concluding, the interviewer should verify any details in the notes that weren't clear and thank the subject for volunteering the time to the study, reminding them of the group's plans for the interview data as needed.]

## **II. Version to share with interviewees**

### **Part One: Motivations, 4-5 questions + follow-ups, est 25 min**

If you remember, how did you hear about the SPN Service Provider Survey?

Can you tell me about the kinds of software and software-related materials you have in your collections?

If you have received requests from researchers to access and/or use software in your collections, how did you hear about them and what did the requesters want to do?

Can you tell me about why preserving software is relevant to your institution's mission or goals?

What contextual documentation or other information (software installation manuals, etc.) would you like to have in order to preserve the software in your collections?

What are the biggest challenges you face in getting started with preserving software at your institution?

**Part Two: Skills and Professional Development, 3-5 questions + follow-ups, est. 10 min**

What skills are most important for collecting, preserving, curating, and providing access to software?

What professional development opportunities, documentation, or other resources could SPN provide to help you meet your goals for software curation and preservation?

What questions do you have about software preservation that you'd like to explore or learn more about?

What information would you use and/or like to have to advocate for resources to engage in software curation and preservation?

**Service provider interview guide #3: for subjects who do not intend to provide software curation services)**

**I. Interviewer Version (with follow-up questions)**

**Part One: Motivations, 4-5 questions + follow-ups, est 25 min**

If you remember, how did you hear about the SPN Service Provider Survey?

(Confirm types of software or software-related materials mentioned in the survey, if any, Q5-7) Can you tell me a bit about the nature of your collections and how software preservation relates to those collections, if at all?

Follow-ups: How and why did your institution acquire your software and software-related materials? Where do you store these materials currently? Who shares responsibility for caring for them? Do you anticipate acquiring (more/any) software and software-related materials in the future?

(Confirm whether researchers have requested access to software, Q13) If you have received requests from researchers to access and/or use software in your collections, how did you hear about them and what did the requesters want to do?

If a person came to your institution requesting access to software or software-related materials, where would you tell them to go for help?

(Confirm ranking of reasons for not planning to engage in software preservation, Q19) Can you tell me more about why you don't anticipate providing software preservation in the future?

Follow-up: If some of the biggest barriers were removed, do you think there would be interest in the development of software preservation services at your institution?

**Part Two: Skills and Professional Development, 3-5 questions + follow-ups, est. 10 min**

(Confirm skills mentioned in survey, Q20). Which of these skills are most important for collecting, preserving, curating, and providing access to software?

Follow-up: Are there other skills you think are important to collect, preserve, curate, and provide access to software? If you were advising a new professional interested in software preservation, where would you tell them to go to develop these skills?

(Confirm hopes for SPN, Q21). What professional development opportunities, documentation, or other resources could SPN provide to advance software curation and preservation?

(Confirm subject's questions about software preservation, if any, Q22) What (other) questions do you have about software preservation that you'd like to explore or learn more about?

**Part Three: Wrapping Up, 1 question, est. 3 min**

Is there anything that I should have asked you today that we have not yet discussed?

[Before finally concluding, the interviewer should verify any details in the notes that weren't clear and thank the subject for volunteering the time to the study, reminding them of the group's plans for the interview data as needed.]

## II. Version to share with interviewees

**Part One: Motivations, 4-5 questions + follow-ups, est 25 min**

If you remember, how did you hear about the SPN Service Provider Survey?

Can you tell me a bit about the nature of your collections and how software preservation relates to those collections, if at all?

If you have received requests from researchers to access and/or use software in your collections, how did you hear about them and what did the requesters want to do?

If a person came to your institution requesting access to software or software-related materials, where would you tell them to go for help?

Can you tell me more about why you don't anticipate providing software preservation in the future?

**Part Two: Skills and Professional Development, 3-5 questions + follow-ups, est. 10 min**

Which skills are most important for collecting, preserving, curating, and providing access to software?

What professional development opportunities, documentation, or other resources could SPN provide to advance software curation and preservation?

What (other) questions do you have about software preservation that you'd like to explore or learn more about?

## Acknowledgments

So many people have helped this project come to fruition. Jessica Meyerson's early encouragement helped the research team establish the need for the project, and she has continued to support us throughout. We would like to thank early members of the Research-in-Practice Working Group including Eric Kaltman who helped develop the survey and interview protocols. We received feedback on the survey and interview questions from Emily Davis, Sherry Lake, and Susan Parham, which improved the flow of each. We give the biggest thanks to the study participants who took time to complete the survey and especially to those who volunteered to talk with us in interviews.

We also thank the communications team at Educopia for helping to publicize the study, and the Software Preservation Network for paying for access to Temi and Dedoose, two tools used in the research. Thank you to SPN Community Facilitator Jessica Farrell who supported us to the finish line. Finally, this white paper has benefited from reviews by many people including Sherry Lake, Eric Kaltman, Jessica Farrell, and participants in SPN Community meetings during which the research team shared initial ideas.

Thank you to everyone who has been part of this project. We hope what we have shared is useful, and we hope to repeat the study again as the field moves forward.

## About the Authors

**Jessica G. Benner** is a librarian and GIS specialist at Carnegie Mellon University supporting the Computer Science, Machine Learning, and Philosophy Departments and anyone on campus using maps, GIS tools, or spatial data. Jessica's current interests include developing best practices for libraries teaching geography and GIS concepts, investigating how libraries and others are providing services for software preservation, and building capacities for digital accessibility. Jessica holds an MLIS and PhD in Library and Information Science and an MS in GIS and Remote Sensing, all from the University of Pittsburgh. <https://orcid.org/0000-0002-2427-1222>

**Seth Erickson** is a data services librarian in the DREAM Lab at the University of California, Santa Barbara. His research focuses on software curation, preservation, and reproducibility. Seth holds an MLIS and PhD in Information Studies from the University of California, Los Angeles. Between 2018 and 2022, while coordinating the SPN Research-In-Practice Working Group, he was a CLIR postdoctoral fellow and then research data librarian at Penn State University. <https://orcid.org/0000-0002-5570-7201>

**Wendy Hagenmaier** is the digital curation archivist at the Georgia Tech Library, where she leads the development of workflows for stewarding born-digital archives and special collections; she also manages the Library's **retroTECH** initiative. She holds an MS in Information Studies from the University of Texas at Austin. <https://orcid.org/0000-0003-4045-0068>

**Monique Lassere** is the digital archivist at the Houghton Library, Harvard University's rare books, manuscripts, and literary and performing arts archive. In her role, she stewards born-digital archival materials within the Manuscript Section. Her research interests include issues in software preservation, born-digital archives, and information maintenance. Prior to joining Harvard, Monique worked for the University Libraries at the University of Arizona, where she oversaw development of their digital preservation strategy and program. She also worked with a digital preservation startup company called Digital Bedrock, researching file formats and software obsolescence. Monique earned her MSLIS and BA in English Literature from the University of Illinois at Urbana-Champaign. <https://orcid.org/0000-0001-5395-8974>

**Christa Williford** is senior director of research and assessment at CLIR. She is responsible for designing and implementing documentation, evaluation, and development strategies for CLIR's programs, and for helping others shape and advance new initiatives related to the work of information organizations. A theater historian trained at Indiana University, she completed postdoctoral fellowships at the

University of Warwick and Bryn Mawr College and worked as user services librarian at Haverford College prior to joining CLIR.

 <https://orcid.org/0000-0001-6273-3793>

**Lauren Work** is the digital preservation librarian at the University of Virginia, where she is responsible for the implementation and maintenance of digital preservation strategy and systems for university digital resources. She also helps to lead digital stewardship collaboration within communities such as the Academic Preservation Trust, NDSA, the Software Preservation Network, and the BitCurator Consortium. She earned her MLIS from the University of Washington.

 <https://orcid.org/0000-0002-0941-6921>

