## Issue Statement

The need for long term digital preservation is only growing more prominent in our tech-centric society and information scientists are in a unique position to formulate this into quantifiable techniques for data management. This proposal aims to discuss the efforts necessary to broaden information equity access via implementation of digital preservation strategies in their various forms.

## Issue Paper

The academic field of literary studies is vast as it focuses on the study of literature in its various forms and further encompasses the analysis of literary texts and the historical and cultural context in which they were produced, and the ways in which they reflect and shape societal values, beliefs, and ideologies. Interestingly, what makes the field of literary studies so unique is that literature is what provides the opportunity to experience perspectives and worlds different from our own. The value of this aspect of literature is foundational to the advancement of our societies. While the study of literature is broad, how we study it and utilize it is much more abstract than other disciplines. Further, the notion of utilizing a dataset in the context of literature is inherently overlooked and less understood. While an individual scholar could research a single author for an article, the utilization of data management software for literary research can approach the same author or body of text from millions of different angles and perspectives (Clement). One example of a powerful approach to reimagining literary analysis and its impact on scholarship would be the use of the application of text analysis tools. At length, the utilization of literacy technologies is a tool that is long overdue, and with this turn to a technological approach to analysis, also establishes the need for long-term digital preservation methods to enable information equity in its various forms.

Prior to the twenty-first century, the tools available to process large amounts of data were few and far between, and further were relatively inaccessible to the public. However, in this day and age, information is both readily accessible and able to be reimagined and manipulated through means that were previously impossible. The development of text analysis software is a key example of this change in the study of literature, and further, marks the shift from the traditional notion of the humanities that academics have long studied to the modern emphasis on the digital humanities and how these tools can reimagine and analyze the branches of knowledge that concern humans and both analytical and critical methods of inquiry (Rockwell, 209). Text analysis software uses natural language processing (NLP) algorithms to analyze text data and extract useful insights from it. From here, these types of software can provide many additional data insights including topic modeling, named entity recognition, and text modeling to name a few (Michel, 176). Some examples of these various tools include the Stanford Topic Modeling Toolbox, which, as stated on the website, "brings topic modeling tools to social scientists and others who wish to perform analysis on datasets that have a substantial textual component." (SNLP) In regard to text modeling and an array of other features, the Google Cloud Natural Language API, "[…] provides natural language understanding technologies to developers." (Google) In terms of user accessibility, shifting the focus to the text-analysis tool Voyant is a great starting point for those that are new to the world of utilizing technological tools in their digital humanities endeavors and research. Voyant is an open-source, web-based application that supports scholarly reading and interpretation of texts. Further, once data has been input it can produce an array of data visualizations that in turn allow for ease of accessibility to complex information systems and topics. Once these technologies have been used and their products have been extracted for widespread usage, it is critical that these conclusions do not become obsolete

(Bode, 77), and as such it becomes vital to turn our attention to the necessity of digital preservation software.

Information creation and access prior to the technological revolution was generally only available in a physical format. However today, information is created and accessed through many forms both physical and importantly, digitally born materials. Due to this advancement in literary creation, museums, libraries, and archives for example, must engage in a uniquely different task for preserving information than required of past preservation efforts. Having systems and methods in place to keep such collections from obsolescence are critical to their survival. Digital preservation software, therefore, is software specifically designed for this task (Reiger). One example could be the service Preservica which can provide the required means to be able to store files and automatically generate backup copies, metadata to make files easily discoverable, and, most importantly, it ensures the longevity of the data. Using software like this can help manage and sort through hundreds of different files that may need to be preserved.

While digital preservation software can act on its own, it is also closely tied to digital asset management (DAM). The primary function of digital asset management is to make it so that assets are easily findable. Once files have been uploaded, they automatically become digital assets. Digital asset management systems are extremely useful for version control since they can store older versions of a digital asset. Different organizations have different use cases for these systems, but broadly, they are all used in the same manner—to store digital assets and make digital assets manageable. Businesses use DAM systems for brand management, as it can make, among other things, company-specific fonts, and logos readily available. Museums use DAM systems to store information about digital assets such as digitized objects and exhibits. Libraries and archives also use DAM systems to store large amounts of videos and images. In today's age,

the ability to store digital material is crucial to a collection, especially if the collection is not born digitally (Lindemann).

Because of the technological world we live in, it will become increasingly more likely that institutions decide to transition, if they haven't already, to some form of digital preservation. The benefits of using digital preservation software (DPS) create a whole new world for access and can save some artifacts that could be lost to obsolescence, but the extinction of the original is a loss that must also be dealt with. The constraints, on the other hand, are not true constraints but disadvantages of using DPS. Rather, these points should be seen more as inevitabilities or byproducts of digital advancement, and ultimately the community will have to seek to provide additional solutions to combat these issues. Regardless, the positive impacts include continued access, organization, sharing capabilities, and no literal decay (Shultz). For continued access, the simple fact of the technological world we live in, is that files are rapidly becoming obsolete. From vinyl to CDs to DVDs to completely digital-born artifacts, there is always something new to replace the last format we were certain would be here to stay. With DPS, institutions are able to transfer files from older formats into newer ones without permanently losing access. With organization, we can now have everything in one place, as once an item is digitized, you can link other objects to it easily. Now you are able to keep the description, metadata, and any linked data all in one place whether the system you are using allows that in the same record or all these items are in the same folder. Regarding sharing capability, the ability to share an artifact that is made digital is infinitely easier than a physical object. This creates opportunities for institutions to share an object as many times as they like—or as a license allows—without needing to concern themselves with safety, transportation, or damage during the process. Digitizing also opens the doors to sharing artifacts with all types of people across the world, and further promotes

information accessibility. For example, when an institution such as the United States Holocaust Memorial Museum creates a digital exhibition, this allows patrons from all over the world to view and access this space without having to physically travel to experience the exhibit. Lastly, no literal decay or physical preservation is required when utilizing a digital preservation software. While the original artifact is still irreplaceable and will require its own preservation depending on the format, you do not have to preserve the digital file in a physical sense, thus creating no decay. As long as the file format is transferred whenever necessary to remain accessible, the artifact will remain intact (Shultz).

In contrast, some drawbacks to consider include original obsolescence, decrease in quality, the frequency of technology evolution, automation, and cost (Blumenthal). For original obsolescence, when it comes to transferring files into a readable format, this process may require the sacrifice of the original. For example, with video games, the gaming console is just as essential as the game itself, and when the console is no longer repairable or usable that game falls into the ether if it is not preserved in another format. This can cause a decrease in quality due to the process of preservation in comparison to the quality of the original. Again, with video games, an essential component is the aesthetic of the original which can sometimes be lost in transference if the new version is not able to duplicate that same quality. This can also be seen with physical art. When a work of art is digitized, details like brush strokes and depth can be harder to recognize. Digital preservation software cannot always be a perfect match in quality, especially for items that are not digital born. With the advances that have become commonplace in our technological world, the need for transferring artifacts to more accessible files will only increase. Essentially, this means that transferring an item once in its life is not likely to be enough. Files will need to be maintained on a regular basis to ensure that nothing is lost to

obsolescence due to the frequency of evolving technologies. Another aspect to consider is automation, as some digital preservation software automatically generates descriptive data for digital assets. This can either prove to be not too bad, or horrendously wrong. For example, if a museum used a program to automatically generate descriptive metadata tags for their digital assets, including exhibits, these automatically generated tags could be wholly incorrect. For instance, only recognizing the shape of a displace case instead of what was inside of it or mislabeling an object as something completely different. While in most cases, using software to digitize assets is a good thing, for automated description it is not. Lastly, while there are many options for DPS that are open-source and therefore free, there may be some institutional reasons why a license must be purchased. This can be a serious deterrent for institutions that are already underfunded. Museums might also have to hire people to clean up badly automated descriptive metadata if software fails at generating it successfully (Blumenthal).

The implementation of these tools is an important consideration, as there is various software for different needs, each one can be navigated in a unique fashion between their counterparts. For example, there is the distinction when using digital preservation software between open source or commercial, and the service versus software aspect of these functionalities. A common distinction one will find when researching digital preservation software is between open source and commercial. These two systems can both provide functioning and valuable digital preservation software, but their distinction is important to note when first deciding between each version, and later operating the software itself. Once an institution has identified its needs and constraints accordingly, then it can properly choose and implement which form of digital preservation software would be the best solution. A commercial DPS could be a better option for a team that is looking for ease of use and accessibility and has

the monetary means to maintain the operating fees of a commercial software. In contrast, open source would be a better option for a team with a competent developer on board, as they can better manage the workflow of the DPS for their specific needs (Gilbert). Further, an open-source DPS allows for greater innovation in the field as well. Some examples of tools DPS uses include 3D, audio, binary data, container, database, disk image, document, e-book, email, geospatial, image, metadata, project management data, research data, software, spreadsheet, video, web, and non-content type specific (Skinner). Between these formats, a digital preservation software is then designed to implement certain actions to maintain these digital assets. A portrait of this life cycle could be as follows: create or receive the asset (acquisition), ingest, preservation planning, preservation action; access, use and reuse; store (storage); dispose (disposal), and cross-lifestyle functions. While this is an example of the utilization of a DPS, it is also important to note their management. As it stands, the management of a DPS could be through the software service provider or it could be through the persons who are employed to manage the assets at an institution. This can vary largely depending on the budget and multitude of assets being digitally preserved. For example, there could be an Information Manager who manages a multitude of data and asset management, thus, this individual can utilize DPS to ensure that all the vital material is stored in the system in the first place. From here, we could turn to the Digital Archive Manager at this institution, who would use DPS as their initial ingestion tool to then inspect the files and inform what operations would be needed for better long-term preservation, an example of this could be the need for metadata or updating file format (Skinner). These tools that a digital preservation software provide are a few examples of how this technology can ensure a significant contribution to information management within the infrastructures that use them.

At length, the need for long-term digital preservation standards regarding digital humanities data management are necessarily vital for the future of information accessibility and innovation in literacy technologies' long-term development. Through these practices we are able to offer information equity to all, and can further the process of historical archives and the commitment to formal upward mobility. Through data management standards for the infrastructure of digital humanities to the significance of digital preservation software, these tools are only the building blocks for what is to come and the foundation for where we have been in the field of literacy technologies and academic enhancement.

Sources


1. Bennett, Stephen, and Shultz, Matt. *Michigan Statewide Survey on Digital Preservation Readiness Report & Analysis*. 13 May 2019. https://mipres.org/wp-content/uploads/2022/01/gvsu_imls_stepping_stones_data_report1.pdf

2. Blumenthal, Karl, et al. "What's Wrong with Digital Stewardship: Evaluating the Organization of Digital Preservation Programs from Practitioners' Perspectives." *Journal of Contemporary Archival Studies* 7, no. 1 (2020): 13.

3. Bode, Katherine. "The Equivalence of 'Close' And 'Distant' Reading; Or, toward a New Object for Data-Rich Literary History." *Modern Language Quarterly*, vol. 78, no. 1, pp. 77. https://doi.org/10.1215/00267929-3699787

4. Clement, Tanya et al., "How Not to Read a Million Books." 2008, https://johnunsworth.name/hownot2read.html

5. Gilbert, Heather, and Mobley, Tyler. "Breaking Up with CONTENTdm: Why and How One Institution Took the Leap to Open Source." *Code4Lib Journal* 20 (2013).

6. Lindemann, Katie. "16 Tips for Selecting a Digital Asset Management System." *Canto*, 2 July 2021. https://www.canto.com/blog/16-tips-for-selecting-dam-system/

7. Michel, Jean-Baptiste, et al., “Quantitative Analysis of Culture Using Millions of Digitized Books.” *Sciencexpress*, vol. 331, pp. 176, https://www.librarian.net/wp-content/uploads/science-googlelabs.pdf

8. Rieger, Oya Y., et al. "The Effectiveness and Durability of Digital Preservation and Curation Systems." *Ithaka S+R,* https://doi.org/10.18665/sr.316990

9. Rockwell, Geoffrey. “What is Text Analysis, Really?” *Literature and Linguistic Computing*, vol. 18, no. 2, June 2003, pp. 209, https://doi.org/10.1093/llc/18.2.209

10. Skinner, Katherine, and Lippincott, Sarah. “Values and Principles Framework and Assessment Checklist.” *Commonplace*, 27 July 2020. https://commonplace.knowledgefutures.org/pub/5se1i1qy/release/4