MLIS PORTFOLIO

by

Malu Arreguin

Submitted to the Department of Information Studies
As part of the Graduate School of Education and Information Studies
At the University of California, Los Angeles
In Partial Fulfillment of the Requirements
For the Master of Library and Information Science
Spring 2023

Table of Contents

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

Professional Development Statement

Upon graduating from UC Berkeley in the Spring of 2019 I found myself committed to the prospect of attending graduate school for a variety of reasons. To some extent, it was due to personal goals, as I always envisioned myself attending graduate school and possessing a graduate degree. In addition, I come from an extended family who mostly possess graduate degrees and felt the need to be on the same level as my cousins, aunts, and uncles. Lastly, while I adored getting my bachelors in English literature, I did not quite know where to go with only that degree. I didn't see myself as a writer in any corporate sense, and beyond that, I spent my time in undergrad frolicking really, and never thought too far beyond my small world of reading Victorian literature and contemporary poetry and writing extensive, argumentative papers on these topics. I lived a very silly life at UC Berkeley, having transferred after spending four years at community college beforehand, and I was just happy to see myself do anything different at this point in my life while feeling stupidly lucky to be able to have that privilege in that community. I felt both certain and uncertain of myself constantly, chalking things up to "I was going to become whatever I was going to become." This isn't a bad way to live, but it doesn't sound promising when you try to explain that concept to others. Thus, I had instilled the need to revert to graduate school to keep me afloat in the eyes of the world.

Having graduated with my degree in English, I initially planned to get my Masters in either English or Creative Writing. In my first round of applications, I only applied to overtly competitive programs that were far, far away from my hometown. I did not get in to any and while at the time it felt like a loss, this was all unfolding in March of 2020, to which I now see as such a blessing in disguise. Thus, in the midst Fall 2020, I set my sights elsewhere for my graduate education, as I am originally from Orange County and decided I did not want to be too

far from my hometown given these unforeseen circumstances. Having always carried a love of libraries but in conjunction an interest in electronic literature and poetry that I discovered in undergrad, I felt the realm of Library and Information Science sounded very intriguing and applied to the MLIS program at UCLA. In time, I found myself interested in the technology aspect of the humanities, and so I opted to do my concentration as a mash-up of sorts between Informatics and Archives, specifically in the digital sphere, as well as working towards the Digital Humanities Graduate Certificate UCLA offers as well.

After enrolling in Introduction to Programming and Data Management in my first quarter, I decided to follow this path further and enrolled in a separate Data Management course as well as Digital Preservation. For the Spring quarter I enrolled in Metadata and Human Computer Interaction to further my comprehension in these areas. As it stands, my desire to deepen my understanding of the digital humanities through information technologies is on the forefront of my efforts at UCLA. For my last quarter in the program, I feel as though I've created a course schedule that blends the heart of my concentration into one, between Informatics, Digital Asset Management, and Archival Description and Access.

Throughout the years I have harnessed the valuable skill of being able to work autonomously and further developed my aptitude to pay close attention to detail in my work. As an undergraduate student studying English at UC Berkeley, the duties of reading and examining texts to write and edit papers became my routine, further allowing me to engage in the necessary skills of responsible time management and intelligent direct communication. I have been able to utilize these skills in numerous settings over the years from my time spent as a staff member for the Berkeley Fiction Review to my internship with a global non-profit the summer following my graduation. As a Store Manager I utilized my organizational marketing skills, attention to detail,

and leadership to maintain the business's operations effectively.

During my time in the program, I have worked as a Graduate Student Assistant for UCLA Capital Programs whose responsibilities include organizing and processing internal documents for digital archiving and database management. This position helped me solidify my desire to engage in digital preservation in its various forms. Concurrently, I have also been working for the UCLA Division of Graduate Education as a Graduate Career Outcomes Project Assistant with responsibilities including conducting alumni research and providing further database management in support of various initiatives across UCLA. Within this position I have been able to hone in my skillsets with multiple software's including Microsoft Excel for data organization, as well as Tableau and Flourish for creating data visualizations. From the latter half of 2022 I interned with San Diego State University as a Wikidata Project Research Assistant, helping the Cataloging and Metadata Strategies Librarian at SDSU in building faculty profiles on Wikidata. Through this project we were able to create queries using SPARQL to better extend our efforts regarding the projects desire for establishing linked data. As of the Spring of 2023 I have begun working as an Archival Intern with the American Film Institute where, within their Louis B. Mayer Library, I have been processing and organizing archival collections in digital environments and been able to deliberately work with their digital asset management system. At AFI, I can contribute to both cataloguing and creating metadata for these digital assets within this position. My hope upon completing the program is to work in the areas of either digital preservation or data management as I have discovered throughout my masters' how valuable information equity and accessibility is, and further the ability to capitalize on the digital platforms and software we now have today to better assist in these culminating efforts.

I joined this program to give myself guidance that I once did not have and I am happy to say this program has pulled me in very specific directions and has allowed me to build a solid foundation for myself and my future career. I have been able to experiment in multiple areas of the library and information science sphere and now feel a strong connection to this world that I am excited to continue forth. Further, I am appreciative that I have been able to work in real world environments within the program to expand my career development efforts and am happy to say I have gained several prominent mentors through my various internship and workplace supervisors throughout this program as well.

List of Courses

Fall 2021

Education 260A—Introduction to Programming and Data Management
Information Studies 211—Artifacts and Cultures
Information Studies 212—Values and Communities in Information Professions

Winter 2021

Information Studies 241—Digital Preservation
Information Studies 262A—Data Management and Practice
Information Studies 270—Systems and Infrastructures

Spring 2022

Information Studies 260—Description and Access
Information Studies 272—Human Computer Interaction
Information Studies 439—Seminar: Special Collections
Information Studies 464—Metadata

Fall 2022

Digital Humanities 250—Special Topics in Digital Humanities
English 203—Digital Theories and Methods
Information Studies 289—Seminar: Special Issues in Information Studies: Readers' Advisory

Winter 2023

Digital Humanities 201—Introduction to Digital Humanities
Digital Humanities 299—Special Projects in Digital Humanities
Information Studies 202—History of Books and Literacy Technologies
Information Studies 400—Professional Development and Portfolio Design

Spring 2023

Information Studies 214—Informatics
Information Studies 289—Seminar: Special Issues in Information Studies—Digital Asset Management

Information Studies 438B—Seminar: Advanced Issues in Archival Science—Archival Description and Access Systems
Information Studies 498—Internship

Examples of Coursework

Major Paper
Final Metadata Strategy Project
Written for Information Studies 464: Metadata
Spring 2022

For the context of this assignment, I have been hired on as a Metadata Librarian at the UCLA William Andrews Clark Memorial Library to organize their initiative of transforming their collection of Oscar Wilde assets into a digital repository for their forthcoming online archives website. The intention for this is to establish greater ease of accessibility for scholars around the world and the public at large. The collection of Oscar Wilde assets at The Clark is the most prolific in the world and the initiative to integrate these items into an online source will greatly contribute to Oscar Wilde's influence and the preservation of these rare materials. The collection incorporates rare books, manuscripts, and iconographies. The intended users of this soon to be digital archive are scholars of Wilde, 19th century literature and history, and patrons that are interested in discovering a greater context of information regarding this prominent literary and historical figure. Further, while the digital archive will be accessible for all regardless of circumstances, we presume that given our reputation as a Special Collections Research Library that principal users will be undergraduate and graduate students conducting primary research through our source database. As an example of this, we will create our own case. For this, we will introduce Celeste, a PhD student at the University of California, Berkeley who is focusing on 19th century literature and is writing her dissertation on the philosophies of the Aesthetic Movement in the Victorian Era and its historical impact. As Wilde is a significant figure in these corresponding movements, she believes that conducting research that incorporates iconographies of Wilde alongside his original manuscript notes would be useful in the

development of her dissertation. Now, as this scholar lives in Berkeley which is approximately an eight-hour drive away from The Clark, by creating this digital repository, we are allowing her to conduct her research—or at least begin to—without having to travel a substantial distance. Further, due to the influx of the COVID-19 pandemic, we have all learned that circumstances can arise without warning that can dictate users' abilities to physically interact with original source materials. Through establishing this online digital archive, we are creating both a long-term digital preservation plan for The Clark's collections and furthering the efforts of the dissemination of information to all.

Here at The Clark, it is our hope that through collaborative efforts this digital archive will culminate in an effectively constructed site that showcases both electronic editing and digital representation. Through the implementation of these design principles the archive will be able to advance the development of other editions, catalogues, databases, and scholarly tools into one prominent electronic archival resource. It is our ambition to extend these efforts to create faithful digitized reproductions of rare and original materials in both detail and scale. For this, we intend to have pieces published through both a 'preview mode' and a 'standard mode' on the website. For books that have been fully published we will strive to provide a wealth of contextual information including full and accurate bibliographical details and meticulous descriptions of the content found within each digitized image. Users of the archive will be able to attain a new degree of access to these pieces through the combination of both advanced text and image-searching tools that are made possible by our editors-controlled vocabulary, detailed image descriptions, and innovative software. As previously stated, for this long-term project we will be implementing controlled vocabularies to allow for ease of text and image searching when users interact with the digital collection. For this project our intention is to establish a local vocabulary

to thoroughly detail the specific efforts of this endeavor. Moreover, the archive will be built under the presumption of a considerable allowance for future enhancements in both hardware and software.

As stated in Miller's piece 'Introduction to Metadata for Digital Collections', the efforts of this endeavor are to, in essence, publish a digital collection, which Miller refers to as, "a collection of digital resources, along with the metadata about those resources, made available online through an interface that allows users to search and or browse the contents of that collection." (Miller) For this undertaking within the digital archive, our intention is for each asset to contain textual metadata that includes the entire image information record. In turn, these textual records are designed to be a part of the archive's image files. From here, the hope is for these information records to combine the technical data recorded alongside additional bibliographic documentation, as well as information pertaining to provenance, present location, and the institution. It is our intention for these image files to contain several different kinds of information beyond the image themselves and that inserting these image information records into the portion of the image file reserved for textual metadata that this integration will allow the record to move with the image, even if it were to be downloaded and separated from the digital archive.

Further, within this project the metadata standard we will be using throughout this collection will be TEI which will be encoded using XML. We chose to utilize TEI—the 'text encoding imitative'—for this because we feel that it is both the most prominent and powerful tool for creating metadata for humanities documents. In addition, as this project is formulated through a digitizing initiative corresponding to a physical special collection, this allows TEI to be an excellent method to represent the textual features of these objects for research purposes in

both a direct and succinct format. TEI will also provide a simple foundation for researchers to make use of without the necessity of special-purpose software and permits for a rigorous definition and processing of texts. TEI can also provide for user-defined extensions and further conform to both existing and emergent standards. As TEI exclusively uses XML to encode textual data, this will in turn allow data to remain usable even as platforms and file formats develop with time. Our efforts stand on the premise of looking beyond what we are currently trying to accomplish with this initiative and lie in the desire to create a digital archive for the present moment that will be able to evolve alongside the passage of time—similar to the literature, celebrity and renown of Wilde. To contextualize XML, we would indicate italics with a plain ASCII tag. Further, XML does not have to be purely descriptive and can permit us to both recognize and encode the structure of documents through a title or heading that can be tagged and described.

A set of XML tags designed for a specific purpose will be referred to as a Document Type Definition, or DTD, which will establish a hierarchical system of contexts and constraints to allow for its tags to be used in the creation of consistent document structures. The hope is for our institution to utilize several Document Type Definitions for the project to further encode each asset at both the object and collection level as well as to incorporate and encode the textual metadata that constitutes the image information record as we have previously mentioned.

Upon completion of the website, our intention is for users to alternate between two modes, namely ‘gallery mode’ and ‘reading mode’. Both modes will provide a separate viewing experience for the user with ‘gallery mode’ being labeled as the default site setting that will showcase numerous digitized copies without the inclusion of text on the ‘object view’ page. Within the ‘object view’ page, a set of tools will be incorporated for each image; these

instructions will include rotate, zoom, enlargement, transcription, and citation in our initial publishing of the site. Within the ‘gallery’ window, we plan to provide an ‘information box’ which we aim to be a critical apparatus for the copy where a user can view the textual transcription, image description, and editors' notes for the object being observed. At the bottom half of the ‘object view’ page, we plan to provide a subset of the following tabs pertinent to what is being viewed. These could include ‘copy and group’, ‘same matrix’, ‘textually referenced’, and ‘letter information’ to name a few. Seeing as the digital assets in this collection are inherently related to one another due to the nature of this special collection, we plan to incorporate a tab for user comparison of these relationships where an individual can select objects for comparison that will then place these digitized assets side-by-side, similar in practice to that of ‘gallery mode’. Within this ‘compare’ window, information will be highlighted to correspond with the original object in the ‘gallery’ window. In ‘reading mode’, copies of works will be displayed as a horizontally scrollable window that will contain the objects of the copy in sequence with their respective textual transcriptions and without the use of a supporting instrument for website navigation.

Regarding the enhancement of metadata, during the initial procedures of the project the intention is for the metadata enrichment to be a closed effort confined within the staff of The Clark Library. However, upon the completion of this initial web-publishing, the implementation of either metadata enhancement through such sources as crowdsourcing or machine learning could potentially be implemented. When envisioning these future options, we at The Clark have imagined the pros and cons to both options and see the inherent value each could offer our program. As it stands currently, it seems that the efforts of crowdsourcing would be the more inclusive option, especially as the collection grows and the introduction of others to offer new

methods that were previously untapped begin to emerge. Further, the continuous development of this digital repository will be motivated in part by responding to the needs of a dispersed and varied audience of readers and viewers. Yet, due to the nature of the original publishing of assets being regarded under our special collections category, it is possible a machine learning enhancement tool could potentially accomplish tasks quicker and more thoroughly than a human in future years. However, we do not have an answer to that question yet, but we are certain that with time, and as the collection and digital archive evolve, we will identify the answers to these matters.

In conference with Whalen's piece regarding rights metadata, there are several key factors to note as we create this digital archive. First, although these articles in this collection confer with the rights of Oscar Wilde, his works are in fact in the public domain as he died at the at the end of the nineteenth century. However, while it would be easy to look past this area due to this key fact, we at The Clark believe that the establishment of this digital archive should promote the efforts of information dissemination and descriptive metadata standards, and thus, are planning on incorporating rights metadata in this effort through the same practice implemented at the Getty Research Institution, as Whalen's piece notes. This effort will involve the enactment of a specific rights metadata dictionary for our special collection as we are confident these actions will promote further developments in the foreseeable future. As stated in the piece, Whalen discusses that, "[…] two important improvements came from this […] clarification of which work was being described in the record, and the addition of terms to the drop-down menus that allow users to better understand some of the ambiguities or unknowns about the rights information provided." (Whalen)

Upholding the data quality of our assets has been and will continue to be a longstanding goal in our efforts of providing reliable description and access to our special collections' materials. While numerous variables can contribute to data quality issues, we at The Clark feel confident that we will be able to appropriately address these issues in a manageable format if and when they were to arise. As our task stands of transforming and enriching data from our current records into a machine-readable format for the purposes of this digital archive, we understand that these efforts will need to be thoroughly scrutinized to ensure consistencies across databases. With the onset of our plans, we hope to reestablish core metadata fields in our data, reestablish content rules across our data models, provide metadata enrichment, and integrate preservation metadata during digitization. All of these efforts have the potential to incur metadata quality issues, and the impacts of these matters can affect the overall usability of our online digital archive. From inconsistent recall and precision to data ambiguity, small miscalculations have the potential to lead to a culminative loss of trust from users in our digital archive and our institution as whole. In establishing and upholding quality metrics in our project we will be able to create the digital archive we believe is needed for the 21st century. Further, we will be following the quality metric values presented in 'The Continuum of Metadata Quality: Defining, Expressing, Exploiting', which includes completeness, accuracy, consistency, appropriateness, provenance, timeliness, and accessibility. (Bruce and Hillmann) At length, the effort of foreseeing potential data quality issues alongside resource description issues are matters we have been considering throughout this process and will continue to do so as our needs and obligations transform over time.

Sources

1. Bruce, Thomas R., and Diane I. Hillmann. “The Continuum of Metadata Quality: Defining, Expressing, Exploiting.” *Metadata in Practice,* American Library Association, 2004, pp. 239-256.

2. Miller, Stephen J. “Introduction to Metadata for Digital Collections.” *Metadata for Digital Collections,* Facet Publishing, 2015, pp. 1-24.

3. Whalen, Maureen. “Rights Metadata Made Simple.” *Getty,* The Getty Research Institute, Los Angeles, https://www.getty.edu/publications/intrometadata/rights-metadata/

Elective Work
Design Manifesto
Created for Information Studies 272: Human Computer Interaction
Spring 2022

https://ocmalu5.wixsite.com/maluarreguin/about-3

Design Manifesto

My sentiment on design is rooted in ethical sustainability, user accessibility, and creativity. The merging of these three areas to establish a piece that is distinct, environmentally conscientious, and simple to operate is the primary intention of my design pursuits.

Sustainable: I feel that creating artistic products without ease of use integrated in the design process perpetuates the standard of disposability, which in turn, leads to further environmental degradation. Upholding this awareness is a perspective I believe all creators must prioritize to some extent, as the loss of resources is a legitimate threat to future generations that should not be ignored.

Trust: I believe that the need for clarity of concept and ease of use should follow and, ideally, go on to utilize an artistic aesthetic that allows for a designer's inner personality to be showcased through the product of their work. Through my design efforts, a user should be able to trust that my design was tested to fulfill its objective for the long-term, so as an individual grows and changes, the product they once purchased remains both effective and synonymous with the stylistic glee that I once imparted into it, regardless of the passage of time.

Vision: I intend for the aesthetic of my designs to be unique and whimsical statement pieces in and of itself, with the hope for them to be able to stand alone and hold a conversation with a stranger waiting in line for their Sunday morning pain-au-chocolat if need be.

INF STD 272

Design Manifesto

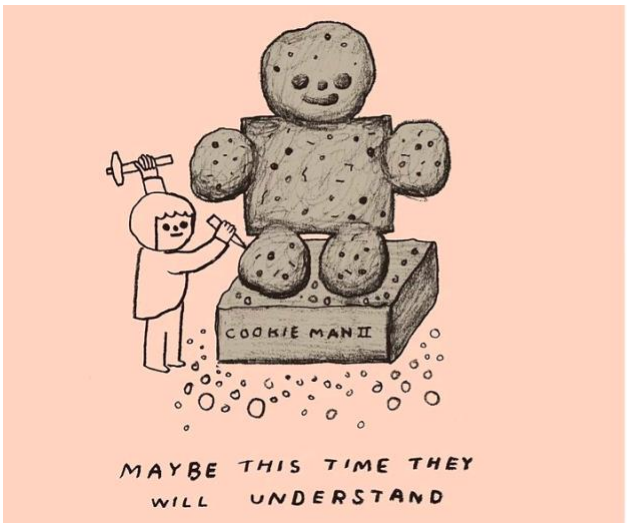
My sentiment on design is rooted in ethical sustainability, user accessibility, and creativity. The merging of these three areas to establish a piece that is distinct, environmentally conscientious, and simple to operate is the primary intention of my design pursuits.

Sustainable

I feel that creating artistic products without ease of use integrated in the design process perpetuates the standard of disposability, which in turn, leads to further environmental degradation. Upholding this awareness is a perspective I believe all creators must prioritize to some extent, as the loss of resources is a legitimate threat to future generations that should not be ignored.

Trust

I believe that the need for clarity of concept and ease of use should follow and, ideally, go on to utilize an artistic aesthetic that allows for a designer's inner personality to be showcased through the product of their work. Through my design efforts, a user should be able to trust that my design was tested to fulfill its objective for the long-term, so as an individual grows and changes, the product they once purchased remains both effective and synonymous with the stylistic glee that I once imparted into it, regardless of the passage of time.



Vision

I intend for the aesthetic of my designs to be unique and whimsical statement pieces in and of itself, with the hope for them to be able to stand alone and hold a conversation with a stranger waiting in line for their Sunday morning pain-au-chocolat if need be.

Core Work
Final Paper
Written for Information Studies 260: Description and Access
Spring 2022

Metadata is built on the concept of analyzing and organizing data about data and this perspective helps to set the standard for what users are intending to accomplish through the utilization of this practice. While metadata is valuable for means of description and access in all information systems, it can also be very convoluted and lacking integration as there are numerous metadata standards that often overlap in their intentions to better organize and structure the data systems and objects that they represent. One of the most common and prevalent metadata schemas due to its ease of use, is Dublin Core. This schema is popular and recognized as user-friendly because of its standardized vocabulary language which only includes fifteen elements. However, this ease of use does not make it unsusceptible to issues and long-term drawbacks to its greater utilization in organization of metadata. As Pomerantz describes in the text, *Metadata*, the elements in Dublin Core are made to be as "unambiguous as possible" which corresponds to, as he states, "[the] subtleties of meaning are flattened out". This perspective in turn inspires us to ask the questions: 1) what we are losing through this schema and 2) how does that impact our data? Ironically, it can also be said, that in essence, the strength of this schema also carries its weaknesses. The significance of these questions is a common issue and great challenge for all metadata standards, as the dream of standardization being universally implemented across all domains may very well be impossible to achieve. However, by looking at this issue through the lens of the utilization of one standard, and in turn contrasting that analysis to another, users can identify how and when certain standards are needed and why the distinction between them is ultimately useful depending upon the subject material being recorded.

The necessity of metadata is only growing in our current technological age. As the advancement and dissemination of information further becomes an inherent and necessary right in modern societies, it is integral that we maintain systems that can promote the longevity of items, objects, artifacts, documents, images, and histories for the interminable future. Thus, the need for standardization of metadata is an ever-growing field as scholars continue to search for tactics that best incorporate information for future users from both a back and front-end perspective. In its conception, Dublin Core originated to describe and recognize varying items from both digital to physical resources. Today, this schema can be utilized for simple resource description to combining metadata vocabularies of different metadata standards. Further, Dublin Core can provide metadata interoperability for other vocabularies in both linked data cloud and Semantic Web applications. ("Metadata Basics") This schema is built on the utilization of fifteen standardized elements in its description of data. As taken from the Dublin Core Metadata Initiative website, these include 'contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, and type'. These areas are the foundational elements for this descriptive metadata schema.

Having established what the Dublin Core schema is, it is necessary to investigate the first argument of Pomerantz's original statement regarding how the semantics being shared across domains is one of this standard's inherent strengths. This can be recognized through the lens of the simplification of data and material sources. In this regard, it would be important to acknowledge that the purpose of metadata and the systems and infrastructures for preserving data is to make an object, item, and so forth accessible for the future. Naturally, one way to ensure that something (object, item, etc.) is accessible is to make it more simplistic. This position is something Dublin Core abides by, through having less categories and rules around the

description of an object the standard can enhance the user's comprehension when approaching objects and implementing this standard. At length, this foundation is precisely where Dublin Core's strengths prevail: through the process of establishing a limited vocabulary for the utilization of the standard, the barrier to entry is lower. This allows the standard to be user-friendly for both current and future users of an object, item, etc. for years to come as it is both consistent and more easily able to be classified and read for future analysis and investigation. (Ball)

Now, regarding the latter perspective that Pomerantz expressed in their initial statement, it is important to acknowledge that the simplification that is at the root of the Dublin Core schema is also its greatest weakness. While Dublin Core is valuable for promoting user accessibility, the negative cornerstone to that is the cost that comes from simplifying things down to a smaller pool of options. Here, the simplification results in the loss of material information and resources due to the minimizing of certain elements and options. Further, it is very possible to misinterpret data through this schema because the need to categorize in a specific way could create user confusion due to the limited categories that may not feel accurately applicable depending on what they are trying to describe. As stated in "Dublin Core: An Obituary", this "loss of specificity that occurs makes it difficult to then convert it into other systems or to transfer other data from different systems." (Beall) Beyond this sentiment, one could argue that metadata should be flexible to account for the various objects and artifacts it will come to be utilized for, and thus, through implementing a rigid system such as Dublin Core, we are losing space for broader identities and classifications. Therefore, it is easy to envision how this weakness in Dublin Core could potentially lead to difficulties with both retrieval and interoperability between standards.

Through further investigation of these constraints presented in Dublin Core, it is necessary to endeavor to foresee a disaster scenario due to the weaknesses found in this metadata schema. As this standard only allows for fifteen elements, more obscure and articulate information is at risk for long term obsoletion due to these potentially restricting standards. Take for example if a museum were needing to create a Dublin Core record for a photograph, due to the elements being so generalized we could in turn create a record that is inherently unspecific. Due to this, the prospect of losing data about the photograph is one potential disaster scenario. Beyond this, a second issue stems from the lack of consistency that could exist within a record, which, unfortunately, could revert the preservation efforts of this archived photograph null.

While the weaknesses in Dublin Core are prominent and should necessarily be considered, the Dublin Core Metadata Initiative (DCMI) has already begun to foresee a progressive future in this schema. According to the DCMI, the current focus at the organization is furthering the implementation of linked open data and the role that Dublin Core will enact in this future. With this initiative, the goal is to connect data from a variety of open sources—including libraries, health organizations, and government data—by utilizing this common standard. The hope through this process of establishing linked open data, is to develop the semantic web to where all data is both linked and easily analyzed by computers. The Dublin Core Metadata Initiative believes that through implementing Dublin Core metadata in conjunction with web content that a business or website can then become a part of the semantic web. At length, the data that is being stored in the semantic web is only increasing with the progression of time due to the rapid developments in cloud computing. Therefore, the implementation of a standardized metadata schema will allow further maintenance of this data and create ease of accessibility for all.

After examining Dublin Core, it is useful to compare it with another schema to analyze what other metadata practices can offer for long term preservation, description, and access. While Dublin Core can be utilized in various areas for resource management, certain other standards can prove just as (if not more) useful depending on the circumstances of what is being standardized. As an example, regarding the management of literature metadata, Dublin Core in conjunction with TEI are the two standards that are most implemented in this subject area, which in turn makes them a natural comparison as each provides effective tools in the formation and standardization of metadata. The Text Encoding Initiative, or TEI, was designed for marking up humanities documents and focuses on structure using coding. Further, it can be customized to include any variation of metadata. This ease of use is what has garnered TEI the reputation as a popular standard, and, as stated from the Introduction to the TEI Guidelines, "[it] provide[s] a means of making explicit certain features of a text in such a way as to aid the processing of that text by computer software running on different machines. This process of making explicit we call markup or encoding." In comparing the two standards, TEI and Dublin Core overlap in their respective intentions towards accessibility. However, TEI is constructed with more flexibility than what is found in Dublin Core due to its limited set of fifteen elements. In TEI, the following, broader elements must be followed, which include "common core identifiers, and textual features are easily shared and accessible; easy modification features for additions and removals, multiple textual encodings of the same textual resource should be possible, mark-ups should be user defined, following certain small requirements, and the sufficient need for documentation to be written alongside what is encoding." These features are the core module for TEI and highlight the differences in flexibility and fluidity it presents as a metadata standard in comparison to the Dublin Core Metadata Initiative.

TEI offers useful benefits for users as a metadata standard. As the TEI website states, the first strength TEI offers is that it is easily verifiable via an online checked validator which establishes ease of use for all current and future users of this system and enhances accessibility. In addition, TEI is sustainable and open source and is regularly updated. The schema also supports text analyses efforts and adheres to a common standard that can operate on a variety of systems. Further it addresses the problem of varying texts in being designed to be adaptable to these different texts. These factors allow the initiative to be continuously progressive rather than stagnant and abiding by a predetermined catalyst which makes it vulnerable to becoming outdated and insufficient for future user needs in the face of the rapid ascension of technology in the current day and age. In comparison to this perspective, TEI does, as all standards inevitably do, come with its' own inherent disadvantages. In this case, these weaknesses include: 1) the process of encoding can be time-consuming due to the efforts of marking up documents. Additionally, 2) the way documents can be marked up also has the potential to be inconsistent as these markups are dependent on the contributor, and thus could make descriptions different not only across but also within varying projects.

Drawing from these weaknesses, it is important to imagine what a potential disaster scenario could encompass due to the flaws in this specific standard. As discussed, the inherent likelihood of experiencing inconsistencies so easily within the process of this standard is a major issue as fluctuations and differences could lead to long-standing difficulties. However, while TEI can substantiate these differences, it can also, through creating well-formed XML documents, be easily validated against standard TEI schema which allows modifications to be correctly documented across platforms quite easily due to the use of software and online validator. In essence, while any metadata standard would inherently possess its own positives and negatives,

and although the barrier to entry for the user to interface with TEI is greater than that of Dublin Core, the positives in TEI inherently solve the negatives. Further, the weaknesses present in Dublin Core are more significant, as the threat of loss of specific information that exists in Dublin Core is something TEI is wholly qualified and designed to prevent. In the end, TEI utilizes expansive space to encourage proper description for long term accessibility needs that can be implemented as successful digital preservation efforts at museums, libraries, and other archival institutions.

As we've discussed**,** all metadata standards are susceptible to the loss of information. However, TEI does prioritize the improvement of its guidelines in the future to address these concerns. In a study conducted on TEI's user base, it was discovered that while TEI addresses many of the needs for its users, there are some areas and users that TEI does not address properly and these individuals' concerns are now being recognized and acknowledged for the improvement of the standard in whole. In "The Present and Future of the TEI Community for Manuscript Encoding" the authors address how it appears that the issue most users addressed was the amount of effort it took to simply use TEI, due to the technicalities inherent in the standard such as TEI schemas and further personalization and customization that is available. In the article, it is hypothesized that this issue could initiate from the directions in the guidelines provided on the use of TEI rather than the negatives of TEI itself. However, it is still critical to have these concerns addressed properly for the further ease of description and accessibility for all. Further, it appears some drawbacks come from the specificity utilized in the schema and the desire for more permissible guidelines that allow users to better handle uncertainties. In turn, the Text Encoding Initiative through the configuration of a broad and diverse range of scholars' efforts is reconfiguring its priorities to strengthen the guidelines to address these grievances

within the standard in the hopes of implementing long-term solutions in the future of this initiative. (Burghart and Rehbein)

At length, both Dublin Core and TEI offer simplistic and useful ways to organize resource materials into metadata schemas which in turn is necessary for further preservation, description, and access. The foundation of this comparison between standards stems from the notion of Dublin Core's inherent strengths and weaknesses as acknowledged in Pomerantz's original statement. However, after thoroughly comparing the Dublin Core Metadata Initiative to TEI, it is made clear that despite its weaknesses the standard will remain for its ease of entry and implementation as a metadata standard. While Pomerantz's argument was right to be addressed as standards should be critically and conscientiously analyzed rather than accepted as they are, wholly unchangeable and immutable. Yet, this challenge and analysis is vital and necessary for greater understanding, development, and continuous improvement and most keenly a representation of the evolution inherent in our current world. As our technologic and preservation needs progress with the influx of data and the necessity for long term accessibility, the potential loss of information only increases, which is in turn why these systems must be reexamined to ensure longevity. Although Dublin Core does in fact permit for a level of flattening out of information that could very well continue to occur, the long-term positives outweigh the arguable negatives that are inherent in this metadata standard. In the end, both Dublin Core and TEI offer positive and negative aspects and these must be acknowledged and accepted before the utilization of one over another on any given classification that could arise in the future.

Sources

1. Ball, A. “Review of the State of the Art of the Digital Curation of Research Data (version 1.1).” *ERIM Project Document erim1rep091103ab11*. University of Bath, 2010., Accessed June 5, 2022. http://opus.bath.ac.uk/18774/2/erim1rep091103ab11.pdf

2. Beall, J. “Dublin Core: An Obituary.” *Library Hi Tech News*, 21(8), 2004, pp. 40-41., Accessed June 1, 2022., http://taddeo.emeraldinsight.com/

3. Burghart, Marjorie, and Malte Rehbein. “The Present and Future of the TEI Community for Manuscript Encoding.” *Journal of the Text Encoding Initiative*, no. 2, Feb. 2012, pp. 1–27., doi:https://journals.openedition.org/jtei/372#tocto2n5.

4. “DCMI Metadata Terms” n.d. Accessed June 5, 2022., https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

5. “Introducing the Guidelines – TEI: Text Encoding Initiative.” n.d. Accessed June 1, 2022., https://tei-c.org/support/learn/introducing-the-guidelines/.

6. “Metadata Basics” n.d. Accessed June 1, 2022. https://www.dublincore.org/resources/metadata-basics/

7. Pomerantz, Jeffrey. *Metadata*. MIT Press, 2015, pp. 1-64.

Advising History

Advisor: Johnathan Furner

September 21, 2021: Initial Advisory Meeting over Zoom.

February 27, 2022, through March 29, 2022: Email sent to Professor Furner regarding potentially finishing the program a quarter early. Unfortunately, Professor Furner was out of office due to a sickness during that time. Instead, I messaged Dee Winn on March 11, 2022, regarding who to reach out to regarding my questions. I was directed to either her, Michelle Maye, or Dr. Michelle Caswell depending on the nature of my questions. I subsequently did not message anyone, and later heard back from Furner when he returned from his sick leave.

September 7, 2022: Email sent to Professor Furner to get a signature on the internship form for that quarter, which I ultimately did not end up using my internship for course credit then due to a class conflict.

October 18, 2022: Email sent to Professor Furner regarding some confusion I had about the nature of the issue statement and paper portion of the portfolio.

September 22, 2022, to January 9, 2023: Email sent to Professor Furner asking if I could use my English 203 course to count towards my research methods requirements for my MLIS degree. Professor Furner approved of this change and the form was properly signed and sent to Michelle Maye in early January of 2023 after I had successfully completed the course.

January 19, 2023, to February 6, 2023: Professor Furner sent an email to me asking if I would like to set up a meeting soon to go over any potential questions regarding my progress towards my MLIS. I agreed and we set up a Zoom meeting for February 6, 2023. In the meeting Professor Furner went over my list of courses and approved of my plan for Spring quarter and that I was on the appropriate track for graduation.

February 2, 2023: Email sent to Professor Furner to get my intent to present form signed and submitted to Michelle Maye.

April 3, 2023: Email sent to Professor Furner to get a signature on the internship form for that quarter; I did enroll in the internship course.

April 4, 2023: Email sent to Professor Furner with my finalized spring course schedule to ensure everything looked appropriate for spring graduation, units wise. Professor Furner approved of it accordingly.

Professional Resume

Work Experience

Archival Intern **April 2023 – Present**
The American Film Institute
- Intern for the American Film Institute's Louis B. Mayer Library whose responsibilities include the processing and organizing of archival collections in digital environments.

Graduate Career Outcomes Project Assistant **April 2022 – Present**
The Division of Graduate Education at UCLA
- Responsibilities include conducting alumni research and providing further database management in support of various initiatives across UCLA.

Graduate Student Assistant **March 2022 – Present**
UCLA Capital Programs *Los Angeles, CA*
- Responsibilities include organizing and processing internal documents for digital archiving and database management.

Wikidata Project Research Assistant **June 2022 – December 2022**
San Diego State University *Remote*
- Internship with San Diego State University. Assisting the Cataloging and Metadata Strategies Librarian at SDSU in building faculty profiles on Wikidata.

Personal Assistant to Writer and Showrunner **September 2021 – February 2022**
- Job responsibilities included but not limited to maintaining private affairs, scheduling appointments, script proofreading, transcriptions, coordinating travel arrangements and overseeing properties.

Freelance Writer and Copy Editor **March 2021 – August 2021**
- Freelance writing work including assisting in non-profit and higher education grant writing projects, ghostwriting and copy editing for a business development e-book as well as various script proofreading assignments.

Education

University of California, Los Angeles **Expected Graduation in 2023**
Master of Library and Information Science Candidate *Los Angeles, CA*
- Degree in progress. Current concentrations include data management and digital preservation.

University of California, Berkeley **2019**
Bachelor of Arts in English *Berkeley, CA*

- Focused on Victorian, Modernism, and Post-Modernism literature and poetry.

Skills

- Well-versed in Microsoft Office Suite, Microsoft Teams, SharePoint, and Google Drive.
- Proficient in R Programming language for data analytics.
- Excellent verbal and written communication skills.
- Familiarity with data visualization software's such as Tableau and Flourish.
- Experience constructing SPARQL queries for linked data.