

Final Metadata Strategy Project

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 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/>