Joyce Wang
Spring 2019
MLIS Portfolio

Table of Contents

Issue Paper

In Pursuit of the Archival Commons: Designing Finding Aids for Participatory Archives

An "archival commons" involves user participation in the archive's construction of meaning and knowledge. As an archive's main entry point, finding aids should embody this ethos. I will argue for archivists to build a more participatory finding aid framework by drawing from Web 2.0 functionalities and scholarship regarding archival transparency.

The usability and affordances of online finding aids has been the subject of much debate since their introduction in the late 1990s. This paper is about rethinking the finding aid within the framework of current discourse surrounding the creation of the "archival commons" – a vision of the archive that seeks to realize a more user-centric and peer-based approach to description and access, most commonly through participatory Web 2.0 technologies such as user-generated content and folksonomies. As has been argued by numerous scholars such as Lisa Coats, Dennis Meisner, Elizabeth Yakel, and Joy Palmer over the past decade, current methods of archival processing remain more committed to dealing with the perceived needs of the collection than those of their users. Some of the reasons for this involve existential ruminations regarding the role of the archivist and the nature of archival authority, while others deal with the formidable conundrum of adapting to new technological and cultural shifts in an increasingly digital world. I will argue that the online finding aid should serve as a platform for the making of the "archival commons", engaging the public in the eternally iterative operation of archival processing and knowledge management. As most of my work experience has been within academic archives and creating finding aids for the OAC, the examples that I draw from and the proposals that I make will likely be most relevant to finding aids created in such settings. I will consider questions of how findings aids are produced and maintained, and what is to be gained by making them participatory sites for the production and preservation of memory and meaning. I will draw upon Elizabeth Yakel's Polar Bear Expedition project as a launch point for discussion into utilizing Web 2.0 functionalities for enabling user participation in archival knowledge making. I will reflect on the implications such changes will hold for the profession as a whole, and propose some next steps forward.

Beyond "finding": Finding aids as mediated sites of knowledge

The term "finding aid" is a rather insular one, being used commonly only within the archival profession. As stated by the website of the US National Archives and Records Administration (NARA), finding aids are "tools that help a user find information in a specific record group, collection, or series of archival materials".[1] The Society of American Archivists (SAA) defines a finding aid in two ways. The first one echoes the sentiments of NARA almost verbatim: "a tool that facilitates discovery of information within a collection of records".[2] The second definition is a bit more elaborate, defining finding aids as "a description of records that gives the repository physical and intellectual control over the materials and that assists users to gain access to and understand the materials".[3] Both sources then provide a few definitions-by-examples, stating that finding aids could be in a variety of formats including card catalogs, calendars, guides, inventories, and registers.[4,5] However, I find that such definitions obscure the fundamentally epistemological problems that underlie the task of archival processing. While it is certainly true that the finding aid's primary function is to provide access points to the materials it describes, it is also necessary to acknowledge its implicit involvement in the act of knowledge making and of representation.

As discussed by sociologist Erving Goffman, objects and artifacts change meaning and are re-contextualized over time.[6] This is especially the case for archival collections, which serve

---

1 "Finding Aid Type," National Archives and Records Administration, https://www.archives.gov/research/catalog/lcdrg/elements/findingtype.html.
2 "Finding Aid," Society of American Archivists, https://www2.archivists.org/glossary/terms/f/finding-aid.
3 Ibid.
4 "Finding Aid Type," National Archives and Records Administration, https://www.archives.gov/research/catalog/lcdrg/elements/findingtype.html.
5 "Finding Aid", SAA.
6 Erving Goffman, *Frame Analysis: An Essay on the Organization of Experience* (Boston: North Eastern University Press, 1991), 19.

as the means for forming conceptualizations and re-conceptualizations of the past, present, and future. In the process of their creation, finding aids give rise to a certain archival narrative through the ways in which they describe and structure representations of materials in a physical or digital collection. They are themselves mediated documents, created within a certain temporal, cultural, technological, and institutional context that shapes the ways in which they are interpreted and used. Since archival materials are oftentimes unique, ephemeral, and possess contextual value beyond their immediate content or appearance, the task of describing them is a complicated, murky issue filled with anxieties about what gets left out or is misrepresented. It is essential to account for these factors in order to create a finding aid that does not present itself as an objective list of item descriptions and locations. Beyond being a tool for researchers to know what items are where, the finding aid should be a platform where these "archival anxieties" can be brought to light and scrutinized, one that is reflective of the changes in values, meaning, and knowledge production of the individuals and communities that use them.

Current conditions: The static online finding aid

The creation of the online finding aid was enabled by the development of encoding standards such as Encoded Archival Description (EAD), which was developed in the mid-1990s to provide a XML-based standard for making online finding aids searchable and indexable by machines.[7] Since its development, EAD has become the defacto standard for encoding online finding aids. However, EAD has since been treated almost as an end in itself, with many scholars wondering whether EAD was truly in sync with the information retrieval behavior patterns of

---

7 Daniel V. Pitti, "Encoded Archival Description," *D-Lib Magazine* 5, no. 11 (1999): 2.

their users.[8] Overall, there has been little study into the usability of current finding aids, and the metrics by which to quantify this have yet to be established. In tandem with the development of the EAD was also the creation of one of the largest existing union databases for archival finding aids: the Online Archive of California (OAC). However, a brief look at almost any finding aid on the OAC shows that the affordances of finding aids have not really progressed beyond those of their paper-based days. Finding aids on the OAC exist as a long block of text that forces users to find materials using one of either two ways: the browser keyword search function Ctrl+F or by scrolling through the hierarchical ordering of series and subseries that was established by the archivist throughout the course of processing. Oftentimes, there are no clickable hyperlinks that serve to connect a collection to others of its kind, and no way to allow for user feedback. In some respects, digital finding aids provide even less affordances for their users than do their paper counterparts. In a 2007 study, University of Michigan's School of Information professor Elizabeth Yakel writes that many scholars preferred paper finding aids over electronic ones because they showed physical signs of use – such as dog-eared pages and annotations – that helped to inform researcher's information retrieval process.[9] In other words, the finding aid became more valuable with evidence of user participation.

Experiments in participatory finding aids

The rise of the internet and the era of decentralized authority has led to a number of experiments into the formation of participatory sites of knowledge and the "archival commons". In discussing the notion of the "archival commons", I will borrow a definition provided by Scott

---

8 Lisa R. Coats, "Users of EAD Finding Aids: Who Are They and Are They Satisfied?" *Journal of Archival Organization* 2, no. 3 (2004): 31.

9 Elizabeth Yakel, Seth Shaw, and Polly Reynolds, "Creating the Next Generation of Archival Finding Aids," *D-Lib Magazine* 13, no. 5/6 (2007).

R. Anderson and Robert B. Allen, who described it as "a space where cultural professionals, researchers, and interested members of the general public could contribute narrative and links among objects of interest held by archives, libraries, and/or museums and *systematically reflect those activities within the primary repository itself*".[10] Since finding aids serve as the main entry points to a collection, they are fundamentally crucial to realizing such a "commons". One of the most commonly referenced examples of a finding aid that has successfully embedded participatory engagement into its very design has been the Polar Bear Expedition Digital Collection. Spearheaded by Elizabeth Yakel and the Finding Aids Next Generation (FANG) research group, the project experimented with new approaches for engaging communities in a participatory archive. The collection refers to an assemblage of sub-collections composed of journals, correspondence, photographs, and other materials that document the experience of US soldiers who were sent to fight the Bolsheviks during the Russian Revolution. The site was launched in 2006 and featured various means for promoting "social navigation".

In her 2007 performance report for the site, Yakel describes the four primary tools that her research team built into the Polar Bear website to facilitate user engagement: commenting, collaborative filtering, bookmarking, and visitor awareness.[11] Participants in the experiment were allowed to create user accounts where they could contribute to metadata about a given item, as well as asynchronously communicate with an archivist "at any point" – presumably within the standard working day hours. They would also be able to view the contributions of other users and know who was currently active on the site. Their site also provided users with a series of

---

[10] Scott Anderson and Robert Allen, "Envisioning the Archival Commons," *The American Archivist* 72, no. 2 (2009): 383.

[11] Magia Krause and Elizabeth Yakel, "Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid," *The American Archivist* 70, no. 2 (2007): 285.

recommended links that "sought to provide alternate and unexpected interrelations between subjects and collections that will enable researchers to make unanticipated connections between records".[12] The ability for users to participate in making changes to the archive occurred without overturning the central authority of "The Archivist" (the pseudonym under which the Bentley Historical Society's group of archivists collectively presented their online presence).[13] However, the success this archive has garnered as an experiment in collaborative memory and knowledge building was largely dependent on the fact that it had a size-able, engaged user group and the fact that a significant portion of its collections had already been digitized. Furthermore, since 2011, the project appears to have lost maintenance and now only remnants of it are accessible through the Internet Archive's Wayback Machine. As with most digital projects in general, sustainability remains a perennial issue. One important observation to note in my research into participatory online finding aids is that, much like the Polar Bear Expedition Project, all deal with digital collections. However, it is important to emphasize that digitization need not set the basis for establishing a participatory finding aid framework. Participatory finding aids should not be conflated with the notion of digital archives. Rather, they should speak to the ways in which issues of finding aid description and access are mediated between the archivists and their research community.

Creating participatory finding aid platforms for the "archival commons"

Suggestion 1: Identify the needs of the commons and decentralizing curation

---

12 Elizabeth Yakel, Seth Shaw, and Polly Reynolds, "Creating the Next Generation of Archival Finding Aids," *D-Lib Magazine* 13, no. 5/6 (2007).

13 Ibid.

One of the first and most challenging steps towards creating finding aids for the realization of the "archival commons" would be to identify and bring together the researchers who utilize the finding aids and to access their needs. This would require communication between the archivist and their various user communities regarding the types of research they are conducting and the types of materials they would be interested in studying. Archivists should seek to understand the uses to which finding aids are employed and identify areas where their users are encountering difficulties or left wondering about the ambiguities of an item's description or arrangement within the finding aid. They should also seek to understand what types of materials users are interested in to make determinations about whether an item should be digitized for greater access. This can be done through focus groups if a collection is particularly popular and has a large user group. For collections that do not receive as much attention, information on researchers' experience of a finding aid can be collected at the individual level, such as through voluntary entry and exit surveys that users can fill out so that they can share their expectations upon coming into an online finding aid, and upon leaving, specify whether or not their expectations had been met or changed in interacting with the finding aid.

Suggestion 2: Allow users to contribute information through comment fields and tagging

Finding aids do not have to be transformed into digital archives Finding aids such as the ones on the OAC currently lack any features for finding aid users to respond to the how materials are described or arranged within a collection. One approach may be to attach to every entry in a finding aid an expandable comments field where individuals can impart their specialized knowledge regarding the materials in question, or to allow the appending of a keyword tag. Initiatives such as the Hypothes.is Project are already seeking to "annotate all knowledge" by

providing an open-source software that will allow for an annotation layer to overlay the online site of publishing companies and digital libraries such as JSTOR, arXiv, and Wiley, allowing scholars to annotate documents and contribute to a growing body of knowledge.[14] Furthermore, in the spirit of the commons, where there is no centralized authority dictating how information can be used, the comments need not be concerned with the materials directly. For example, if a researcher knows that one archive contains the correspondence of Person A to Person B and another archive contains the correspondence of Person B to Person A, they should be encouraged to provide links establishing the connection between the two resources. To allow such knowledge to disappear after it has served the needs of a single researcher, as things are structured currently where individuals utilize the archive in silo, would be a waste of the intellectual labor and other resources put in to making that discovery and lead to redundant work. Helping to foster signs of use and engagement with the materials will have a positive feedback effect and compel other individuals to share their views as well.

Implications of a participatory archive on the role of an archivist

While numerous archival institutions are aware that many of their collections are in need of reprocessing or that their finding aids are severely outdated, the task of updating pales in comparison to dealing with the deluge of materials that have yet to be processed at all. As such, the finding aid is quite often presented – willingly or not – as a "final version" when in reality it should reflect the iterative nature of archival processing. It is important that archivists develop the mindset for producing, as Kate Theimer puts it, "iterative products, not 'perfect' products".[15] Recently, the Center for Primary Research and Training at UCLA's Young Research Library has

---

[14] "About Us", Hypothes.is, April 10, 2019, https://web.hypothes.is/about/

[15] Kate Theimer, "What is the Meaning of Archives 2.0?" *The American* Archivist 74, no. 1 (2011): 63.

been piloting a graduate student position for a "Redescription Scholar" tasked with "identifying finding aid descriptions that contain outdated, culturally insensitive or oppressive language in order to redescribe collections that better reflect self-description by communities".[16] It would be absurd to pretend that the meaning relayed by archival finding aids were representative of an absolute truth; as such, the capacity to demonstrate that archivists are not, in-fact, "all knowing" is a crucial aspect that needs to be designed into a finding aid. Such "fallibility" could be demonstrated by documenting changes made to finding aids through version control, allowing user to see the ways in which a finding aid has developed over time. Changes resulting from user input could also be denoted in a different color in order to preserve the original description. In doing so, the finding aid becomes a dialogic platform, a document of use and interpretation that records shifts in discourse regarding the materials over time.

In the years ahead, the profession will have to face the influx of increasingly large and more heterogeneous collections, both in terms of their historical contexts and the types of media or materials they comprise. Already, many archival institutions are reporting that their backlogs take up close to a third of their entire collection holdings.[17] In their proposal for a "more product and less process" framework, Green and Meissner argue that archivists should focus on minimizing the time they spend on description and instead focus on providing just enough descriptive context to allow researchers to work with material.[18] This will require revising existing workflows and forms of practice, including a more reciprocal relationship between archival professionals and their users. Furthermore, in order for a finding aid to truly embody the

---

16 Courtney Dean, 'Winter 2019 CFPRT positions', email, 2018.

17 Mark Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, no. 2 (2005): 211.

18 Ibid, 237.

spirit of participatory engagement and the "archival commons", it is important that archivists themselves strive to be transparent about their role as active agents in the processing of an archival collection. UNLV archivist Michelle Light and Harvard librarian Tom Hyry have suggested the addition of a colophon that would come at the end of the finding aid and would "record what they [the processing archivists] know about the history and provenance of a collection and to reveal appraisal, arrangement, description, preservation, and other decisions they made while working on a collection".[19]

Conclusion

A call for a more participatory finding aid that embodies the ethos of the "archival commons" means providing ways for users to contribute to and shape the way knowledge is represented in the archive. It also means re-evaluating the role of the archivist as gatekeepers in the preservation of memory and knowledge. It will raise questions about the nature of archival authority, and will run counter to the opinions of those who believe in the sanctity of the archival record. However, a more participatory framework can help open archives to new users and ensure that the past contributions of researchers or other users can be integrated within the archival narrative. Naturally, it would be important to mention the issues that will arise within the framework of a participatory archive, such as the inevitable need for content moderation. However, the design and affordances of finding aids should reflect the notion that knowledge is created through a collaborative process, and that it is a "living document" reflective of the changes in meaning and significance that it acquires through the course of its use.

---

[19] Michelle Light and Tom Hyry, "Colophons and Annotations: New Directions for the Finding Aid," *The American Archivist* 65, no. 2 (2002), 218.

Works Cited

Anderson, Scott, and Robert Allen. "Envisioning the Archival Commons." *The American Archivist* 72, no. 2 (2009): 383-400. doi:10.17723/aarc.72.2.g54085061q586416.

Coats, Lisa R. "Users of EAD Finding Aids: Who Are They and Are They Satisfied?" *Journal of Archival Organization* 2, no. 3 (2004): 25-39. doi:10.1300/j201v02n03_03.

Dean, Courtney. 'Winter 2019 CFPRT positions'. Email, 2018.

"Finding Aid." Society of American Archivists. Accessed March 31, 2019. https://www2.archivists.org/glossary/terms/f/finding-aid.

"Finding Aid Type." National Archives and Records Administration. Accessed March 31, 2019. https://www.archives.gov/research/catalog/lcdrg/elements/findingtype.html.

Goffman, Erving. Frame Analysis an Essay on the Organization of Experience. Boston: Northeastern University Press, 1991.

Greene, Mark, and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." The American Archivist 68, no. 2 (Winter 2005): 208-63. doi:10.17723/aarc.68.2.c741823776k65863.

Krause, Magia, and Elizabeth Yakel. "Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid." The American Archivist 70, no. 2 (2007): 282-314. doi:10.17723/aarc.70.2.lpq6124788lt10kv.

Light, Michelle, and Tom Hyry. "Colophons and Annotations: New Directions for the Finding Aid." *The American Archivist* 65, no. 2 (2002): 216-30. doi:10.17723/aarc.65.2.l3h27j5x8716586q.

Pitti, Daniel V. "Encoded Archival Description." D-Lib Magazine 5, no. 11 (1999).

doi:10.1045/november99-pitti.

Theimer, Kate. "What Is the Meaning of Archives 2.0?" *The American Archivist* 74, no. 1 (2011):

58-68. doi:10.17723/aarc.74.1.h7tn4m4027407666.

Yakel, Elizabeth, Seth Shaw, and Polly Reynolds. "Creating the Next Generation of Archival Finding Aids." D-Lib Magazine 13, no. 5/6 (May/June 2007).

doi:10.1045/may2007-yakel.

Major Paper:

DNA Databases as Archives

IS 431: Archives, Records, and Memory

Fall 2017

Professor Anne Gilliland

This paper was written for Professor Gilliland's introductory archives course in response to an open-ended prompt to consider the recordkeeping principles and archival concerns of a digital database. I decided to draw from my undergraduate studies in microbiology and molecular genetics by applying an archival lens onto DNA databases. The term "archive" is rarely used within the scientific community, so I thought it would be interesting to interpret the rise of DNA databases and the sociopolitical issues they've garnered through the framework of archival principles, and to understand their placement in the greater history of records management. The majority of the paper was written in Fall 2017 and has been edited and extended since. The original paper can be found on the electronic portfolio.

As genetic sequencing technologies become increasingly efficient and affordable, the proliferation and use of DNA records and the implications of maintaining DNA databases or repositories become an increasingly pressing matter. Despite being comprised of the same components – records that must be managed, arranged, and preserved – that characterize an archive. DNA databases do not tend to be the subject of attention within the archival discipline. They do not fall under traditional notions of "the archive" as a physical or digital space where items of historical and contemporary importance, evidence, or narratives are stored and interpreted. However, trying to argue that DNA records are stored in a "database" and not an "archive" – evoking the perennial debate on the semantics between "repository", "databank", "archive", and "database" – and that archival studies has no application to this field would be a significant oversight. Doing so obscures the archival concepts and issues attending the use and creation of DNA databases, such as those of memory, evidence, and transparency. This paper will first contextualize DNA databases within a historical context of scientific recordkeeping, then proceed to frame the DNA record as both a material and digital document. The paper will go on to discuss some of the controversies surrounding the establishment of modern-day DNA databases, putting them in the context of archival theory and ethics particularly with regards to issues of record's privacy, confidentiality, accessibility, and acquisition.

DNA Records: Origins in the recordkeeping of a "universal" classification

The discovery of DNA in the mid-20th century as the molecule responsible for generating almost all life (with the exception of RNA-based organisms) on Earth has fundamentally redefined scientific efforts towards classifying and understanding life forms. Prior to the development of tools for detecting genetic variations, biology taxonomies were largely based on

morphological and behavioral features such as an organism's number of limbs or their reproductive patterns. To draw from a classical Western example, Aristotle was one of the earliest thinkers to come up with a classification system for animals. His classification was based on such aspects as "those with and without blood" and viviparity versus oviparity.[1] Classification is deeply rooted in the act of recordkeeping and the construction of collections and archival "microcosms" for making more generalized conjectures about natural phenomenon at large. Early modern European cabinets of curiosities were developed largely in part out of efforts to classify, describe, and thus make intelligible findings from Western European explorations and colonial expansions into the New World. Carefully curated and arranged, cabinets of curiosities served as a means of representing the boundaries of the known and unknown world.[2] Given the impossibility of grasping a totality of the world, the creation of a discrete collection, archive, or repository becomes the prerequisite for the production of any form of knowledge.

However, it is also important to note that there is a reciprocal relationship between classification and the creation of archives. While collections often serve as the source materials for creating taxonomies and certain knowledge schemas, the process of aggregating materials for an archive is always already based in certain preconceptions about the social and natural world. For example, one of the fathers of physiognomy Giambattista della Porta, wrote a treatise in the 1570s relating the morphology of a person's hands to their temperament after visiting a prison in Naples, Italy.[3] His observations and documentations of hand morphology and other bodily

---

[1] James Lennox, "Aristotle's Biology." (*Stanford Encyclopedia of Philosophy*, 2017).

[2] Isabel Yaya, "Wonders of America: The curiosity cabinet as a site of representation and knowledge," *Journal of the History of Collections* 20, no. 2 (2008), 173.

[3] "Giambattista della Porta." Science and Its Times: Understanding the Social Significance of Scientific Discovery. . *Encyclopedia.com.* (April 4, 2019). https://www.encyclopedia.com/science/encyclopedias-almanacs-transcripts-and/maps/giambattista-della-porta-0

features were already embedded in the social and legal norms of what constituted criminality in the 16rh century. As is being increasingly argued within the field of contemporary archival theory, no archive was ever created from a void to preserve a neutral narrative of the world. Such Jenkinsonian notions of the archive as the product of an objective reality completely disregard the various value systems, ideological constructs, and circumstances that defined their creation. As the basis for all multi-cellular life on Earth, the DNA sequence has acquired a certain level of empiricist, universalistic credo, overturning many paradigms within the field of biological taxonomy. At the same time, it is also important to treat DNA records with the same level of caution that should be accorded when interpreting all human artifacts. They are subject to re-interpretation, leading to unforeseen implications and generating unforeseen consequences. As new forms of material and digital (ultimately still material) records, DNA records and the databases or repositories in which they are stored invoke the very same issues of memory, accountability, evidence, identity, and responsibility that form the fundamental concerns of the archival practice.

DNA as a material record

The materiality of DNA is concealed in its microscopic size. The structure of DNA can only be rendered vaguely visible to the human eye under complex, stabilizing chemical conditions and the use of incredibly high-powered electron microscopes. For the majority of those who have taken any preliminary biology course, DNA takes the shape of a colorful, linear double helix connected by “rungs”. It can be replicated by protein complexes when the cell is ready to divide, and unravel at certain positions to change protein expression in order to modulate cellular functions. While the physical structure of DNA has been validated through

extensive and diverse methods of experimentation and modeling, the visualization that has become an abstraction no matter how chemically precise or structurally accurate the depiction. Even when represented using advanced 3-D modeling software that depict the physical proportions, dimensions, and interactions of the molecule, the physicality of DNA and its orchestration of biological processes does not lend itself to easy documentation. However, as a material record alone, DNA has very little use. For DNA to be studied and understood as the document of an individual, it almost always has to be interpreted in a way that resolves its genetic sequence (with the exception of certain biochemical research or medical studies where the structure of DNA might be more important than its sequence). DNA sequencing, starting from its early manifestation in the Sanger method developed in the 1980s to the high-throughput processes of today, is the primary tool used in the generation of the DNA sequence record. In this process, DNA is converted from a material phenomena into a "rational", understandable sequence, stored in a .fasta or .2bit file type to be processed by bioinformatic programs. By determining the linear sequence of adenine, cytosine, guanine, and thymine nucleotides that comprise a single gene or genome and performing large-scale comparisons with other samples, it becomes possible to identify certain "genetic markers" that could have identifying, categorical, comparative, or predictive purposes.

An individual's genetic sequence is a biochemical record that can be analyzed within certain parameters to describe certain aspects of its subject, whether it be the person's medical condition, ethnic background, or criminal status (in the case of forensics). Thus, DNA falls comfortably into the framework of a "document" as described by Suzanne Briet: "evidence in

support of a fact".[4] It should be qualified that the "fact" being referenced here is not absolute, but a belief widely held to be true within a given circumstance. As will be discussed later in the paper, the genetic record is subject to varying interpretations and use. Immutable (as far as the sequence is concerned) and difficult to conceal given the human body's penchant for shedding off dead cells, DNA is a ubiquitous and unique document that can paint a very intimate and perhaps novel portrait of the individual in question that reflects greater sociopolitical beliefs and concerns.

A myriad of DNA records and databases

Despite the materiality of DNA and the experimentally determinable nature of their nucleotide sequence, the generation of a DNA record is not free of bias. The terms "DNA record" or "DNA sequence" are very vague and do not reflect the variety of genetic documents that get deposited for different purposes by collecting institutions. Today, DNA databases are used in a variety of fields. They aid research in precision medicine, crime scene investigations, and genealogy mapping. As such, the types of genetic information stored in these databases can range from an individual's copy of a single genetic marker to, in rare cases, their entire genome sequence. Law enforcement agencies and forensic labs tend to emphasize the documentation of short tandem repeats (STRs), areas in the genome containing a highly variant (polymorphic) number of short nucleotide string repeats that are believed to meet the threshold of effectiveness for distinguishing between individuals. The genetic profiles of convicted individuals stored in the massive United State Combined DNA Indexing System (CODIS) are generated from an analysis of 20 select STR locis known as the CODIS Core.[5] On the other hand, popular consumer

[4] Suzanne Briet, Ronald E Day, Laurent Martinet, and Hermina G B Anghelescu, *What is a Documentation?: English translation of the classic French Text* (Lanham: Scarecrow Press, 2006), 9.
[5] Douglas R. Hares, "Selection and Implementation of Expanded CODIS Core Loci in the United States," *Forensic*

genetics company 23andMe utilizes a different analytical procedure that genotypes individuals using markers known as single nucleotide polymorphisms (SNPs) – variations in a single nucleotide at a particular location across a genomic population.[6] In a 2016 blog post, 23andMe responded to concerns about law enforcement inquiries into the company's database, stating that "23andMe's test focuses on how you are *like* other people, while forensic tests focus on how you are *different* from other people".[7] The differences in the type of DNA record being created, they argue, is generally significant enough to render 23andMe's records incomparable with those in CODIS, thereby making them of little use to law enforcement. 23andMe further states that even if law enforcement were compelled to inquire into their records, the DNA records that the company creates do not satisfy the "chain of custody" requirements needed to establish a record as "evidence" since the services are offered online.[8]

DNA records are also rarely ever just a list of an individual's genetic sequence. They are oftentimes accompanied with various other types of data, such as those regarding an individual's socioeconomic status, ethnicity, lifestyle habits, and even political viewpoints.[9] Ultimately, it is important to recognize that a genetic record is generally not a literal read-out of an individual's genetic sequence, but the approximations of certain biochemical and statistical manipulations under the direction of fallible algorithms that satisfy (or give the illusion of satisfying) an institutional and consumer need.

Science International: Genetics 17 (2015).

[6] 23andMe, "23andPrivacy: Your Data and Law Enforcement," 23andMe Blog (2016).

[7] Ibid.

[8] Ibid.

[9] In a 2014 genome wide association study done on five populations (American, Danish, Swedish, Hungarian, and Australian), Peter K. Hatemi, a Professor of Political Science and Biochemistry at Pennsylvania State University, argued that genetic factors were involved in the formation of people's political ideologies.

Early DNA archives and efforts towards "empiricist" studies of difference

Though national DNA databases started to emerge in countries such as the UK, Brazil, and United States in the late 1990s – largely for forensic purposes – the widespread acquisition of genetic material from specific groups of people was embedded in a popular ideology of post-1960s biological research that DNA could empirically elucidate human evolutionary history. This fixation on the DNA sequence as a neutral document illustrated a high level of naivete on the part of the scientific community at large when it came to understanding the sociological and political impacts of the findings they made. Population geneticists and anthropologists alike were convinced that DNA analysis could produce an "objective" and "anti-racist" science free of the blatant prejudice that had defined eugenics and social Darwinism. They believed that the quantifiable materiality of DNA could somehow produce a non-ideological understanding of differences between groups of people. With such notions in mind, a group of population geneticists at Stanford University launched the Human Genome Diversity Project (HGDP) in the early 1990s.[10] Led by prominent geneticist Luca Cavalli-Sforza, the team sought to conduct an expansive survey of the diversity of the human genome by collecting DNA samples from more than 50 indigenous populations around the world.[11] While these samples were originally acquired with the promise that they would be used to help the medical needs of these communities, it became clear not long after that these documents were becoming a new site for colonialism.

---

[10] It should be noted that the HGDP is an entirely separate project from the Human Genome Project, which had also been launched around a similar time. The latter was an international collaborative project that sought to determine the nucleotide base pairs of the human genome and to profile their biological functions.

[11] Jun Z. Li et al., "Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation," Science, February 22, 2008, accessed April 13, 2019: 1.

In 1993, the HGDP was widely denounced by numerous indigenous communities and human rights organizations for patenting genetic materials without the awareness of the indigenous participants, as well as, according to anthropologist Jonathan Marks, for "creating artificial genetic distinctions" between groups.[12] Historian of science Cathy Gere termed the HGDP's collection of DNA and other biological materials as constituting a "biocolonial archive", a holding of plundered bio-records obtained through an imperialistic legacy.[13] The records contained in the database are neutral in the sense that they were obtained via a standardized experimental protocol for acquisition of a DNA sample. However, they came into existence within an institutional framework that failed to uphold values of transparency and confidentiality, and instead, treated people's DNA data as assets to be sold at the researchers discretion. Again, such studies were being conducted in the midst of theoretical trends in evolutionary biology which "rested on the conviction that indigenous genomes partook of a unique genetic purity" free of "the mongrelizing and displacing effects of cosmopolitan modernity". [14] In light of such views, the HGDP's collection of indigenous DNA samples recapitulated the actions of extractive, early 20th century anthropological methods, which were spurred by the urgency of needing to preserve the traces of a "vanishing" people. In separating the record and the human from which it was derived, the HGDP's "biocolonial archive" presents a classical case of the bureaucratic, extractive archive that failed to prioritize social good, particularly within historically marginalized populations.

---

[12] Jonathan Marks, *What It Means to Be 98% Chimpanzee: Apes, People, and Their Genes*. (Oakland: University of California Press, 2005), 202.

[13] Cathy Gere, "Evolutionary Genetics and the Politics of the Human Archive" in *Science in the Archives*. Lorraine Daston. (Chicago: University of Chicago Press, 2017), 208.

[14] Ibid, 185.

Surveillance and consent in the collection of DNA records

The methods by which DNA samples are acquired for a repository remain one of the most controversial aspects in the establishment of DNA databases. According to the Dutch archivist-scholar Eric Ketelaar, the agency of the individual from whom the documents were obtained should be one of the determinants of a documents' value and accessibility for posterity. [15] Records obtained under coercive circumstances and/or in the absence of informed consent should have restrictions in their use and access, or be eliminated all together. Unfortunately, that an archive may contain information about a person without their knowledge or consent is nothing particularly new. The covert collection of information has throughout history laid the foundations for maintaining state control, some of the most blatant examples being the Stasi files of East Germany and the FBI's counter-intelligence program (COINTELPRO), which had actively spied on and sought to discredit and disrupt civil rights, Black Power, feminist and other "politically subversive" activist groups during the 1960s. As mentioned in the previous section concerning the origins of the HGDP and its implementation as a biocolonial archive, issues of consent are inextricably tied to not only the collection of DNA but also to the context of their use.

The issues revolving around the collection of genetic information becomes particularly controversial in forensic and disciplinary settings. The United States Combined DNA Indexing System (CODIS) is a software platform that analyzes DNA records kept in forensic laboratories at the national (NDIS), state (SDIS), and local level (LDIS). All together, they make up the world's largest DNA database. As of April 2019, CODIS contains the STR profiles of more than

[15] Eric Ketelaar, "The Right to Know, the Right to Forget? Personal Information in Public Archives," *Archives and Manuscripts* 23 (1995): 14.

17 million individuals, 13 million of which originate from offenders.[16] The incorporation of DNA records into these three levels, each with their own standards of use and access, occurs in a hierarchical fashion. In most cases, samples are deposited first in the local level of the CODIS database. If the individual is convicted or arrested, the sample is then transferred into the state or national repositories. State repositories are regulated on a state-by-state basis, but many local and regional labs, while also officially subject to federal and state laws, operate with much greater autonomy. A 2013 *New York Times* article (notably published only a week after the Snowden NSA surveillance leaks) reported a rising number of DNA samples that were acquired covertly by law enforcement agencies across the country. Many records are obtained from low-level offenders in exchange for plea bargains, with the justification that data from the NDIS corresponded primarily to convicted criminals who were already going to prison. The *Times* article also reports on surreptitious feats of DNA collection, with investigators obtaining and retaining for unspecified amounts of time samples from discarded waste, crime scenes, and the victims themselves.[17] The blanket collection and storage of DNA without a clear retentions period makes it impossible for individuals to take control of their own genetic data. In another example, the Electronic Frontiers Foundation reports efforts to collect DNA from refugees and asylum seekers as well as their families.[18] The preemptive collection of DNA in such conditions can be seen as a breach of 4th amendment rights, with the mere implication of the person in the database being to some degree a document of guilt.

[16] "CODIS-NDIS Statistics," FBI, https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics.
[17] Joseph Goldstein, "Police Agencies Are Assembling Records of DNA." *The New York Times*, (2013).
[18] "Federal DNA Collection," *Electronic Frontier Foundation*, 2012.

Even outside the realm of law enforcement, there are numerous occasions in which DNA records could be covertly produced and collected. Many hospitals in the United States are now rapidly amassing blood samples from newborn babies as a result of the many health screenings that babies now all undergo. It is estimated that up to 98% of newborn babies now undergo such genetic screening, and while such residual blood samples were tossed out in previous years, many have since ended up being stored in a national newborn DNA bank.[19] In most cases, parents are unaware of the DNA depositing that occurs after the screening, and efforts to obtain consent are often misleading. For example, many states have an "opt-out" option for screening tests altogether, rather than for the storage of samples in newborn DNA banks.[20] Defendants of the newborn DNA database argue that the samples will be used for the purposes of medical research, although what that precisely entails is largely unknown. Regardless of how the data is used, the generation of the genetic record must be dealt with greater transparency.

Even if genetic material is provided voluntarily, there are currently very little regulations for how such information could be re-used. Over the past decade, a growing public interest in recreational genealogical studies has led to the creation of massive pools of individual's DNA sequences. Recently, FamilyTreeDNA, the fourth largest commercial DNA testing company in the world, admitted to sharing its user's DNA data with the FBI for forensic purposes despite extensive marketing claiming itself as a staunch protector of consumer privacy.[21] Informed consent becomes meaningless if conditions for records reuse change after an agreement is duly

---

[19] Aditi Shah, "Do You Know Where Your DNA Is? Genetic Privacy and Non-Forensic Biobanks." Council for Responsible Genetics (2014): 6.

[20] Ibid.

[21] Matthew Haag, "FamilyTreeDNA Admits to Sharing Genetic Data with FBI," *New York Times*, February 4, 2019, https://www.nytimes.com/2019/02/04/business/family-tree-dna-fbi.html

made, or made apparent to the individuals affected only after a breach of privacy. Rather than terminating their data sharing practices with the FBI, FamilyTreeDNA has instead modified their terms and conditions agreement to provide users the ability to opt-out of having their DNA associated with those of suspected criminals. With information being transmitted with ever-increasing ease, and collected at more and more checkpoints within one's daily life, it seems that the distinction between public and private information has become increasingly liminal.

The issue of unregulated DNA record formation and collection can be viewed as a matter of appraisal and reappraisal. As noted by scholar Anne Gilliland, one aspect in the proper appraisal of a document for archival keeping involves "assessing the kinds and degrees of bureaucratic, legal, research, cultural, community, personal and intrinsic values that are present in extant records in order to arrive at a decision about their eventual disposition."[22] There are numerous stakeholders involved in genetic records: the contributors of the genetic sample, the database that owns it, and the users of that data who wish to employ it for some (often unforeseen) future purpose. In the case of forensic DNA databases, the legal and personal valuations of the DNA record may conflict, meaning the individual's desire to keep their DNA record private should be weighed along with the law enforcement's desire to expand their database of potential criminals. In appraising records obtained from newborn screenings, a similar recognition of the need for obtaining proper consent – as well as providing a way for people to retract their consent if they so desire – should play an integral role in the retention of the document. Furthermore, especially in the case of the forensic database, there should be a

[22] Anne J Gilliland, "Archival Appraisal: Practising on Shifting Sands," in *Archives and Recordkeeping: Theory Into Practice* (Facet Press, 2014), pp.34.

standard of appraisal and reappraisal that could better define the retention periods of the DNA documents. In the absence of explicit and enforced retention periods for DNA records, such records are vulnerable to use in contexts that are outside the initial reasons for collection. The act of including a genetic profile into a database requires a recognition of the circumstances in which the documents were produced or acquired.

Unequal access to DNA data

Related to the issue of largely unregulated DNA collection and the lack of informed consent is the disparities in access to such records. The applications of DNA research, whether it be for medical or forensic purposes, are most relevant when there is a large sample of records to draw from. This has led to the commodification of genetic data and the rise of exorbitantly profitable markets for DNA data brokerage. Archives function on the principles of accessibility, which is dependent on people knowing what types of documents they carry, who is represented in their collection, how they are represented, as well as having processes in place that allow for such representations to be contested. This becomes a particularly difficult issue for DNA databases, particularly those collected from crime sccenes, since individuals would not even be aware that their data has been registered. Even in the case of voluntarily donated samples, the donor is not tend to have a clear idea of what aspects of their data are actually being kept by the repository and the contexts in which it can be used. Medical DNA databases also contain one caveat that restricts their access compared to traditional archives: the record has usually been "anonymized". Ironically, while this is done to safeguard the identity of the donor, numerous studies have emerged wherein researchers were able to re-identify the owner of the sample using

the associated metadata.[23] As mentioned earlier in the paper, there are different types of DNA records. Such records are also often accompanied by other non-genetic metadata regarding an individual's race, economic status, and medical history. Hence, it is possible that the record remains accessible to other third parties while remaining inaccessible to the individual from whom the sample was taken. This makes it difficult for some individuals to request the removal of their data. This imbalance in the technical aspect of accessibility, as well as the oftentimes sheer lack of knowledge of a record's existence, makes DNA databases intrinsically prone to breaches in confidentiality on the part of the database itself as well as to external attacks.

Conclusion

DNA records are a relatively new form of identifying document that should be handled with the same considerations as would be applied to traditional archives. They are material records that take on a variety of digital manifestations, and while there are currently a number of policies that try to govern their use, most of these are geared towards DNA collected for biomedical research purposes. With regards to issues of appraisal, accessibility, and privacy, they present numerous challenges that will invoke the same concerns as do archival ethics.

[23] Aditi Shah, "Do You Know Where Your DNA Is? Genetic Privacy and Non-Forensic Biobanks." Council for Responsible Genetics (2014): 13.

Works Cited

23andMe. 2016. “23andPrivacy: Your Data and Law Enforcement.” 23andMe Blog. March 16, 2016. .https://blog.23andme.com/23andme-and-you/23andprivacy-your-data-law-enforcement/

Briet, Suzanne, Ronald E Day, Laurent Martinet, and Hermina G. B Anghelescu. *What Is Documentation?: English Translation of the Classic French Text*. Lanham, MD: Scarecrow Press, 2006.

Brown, Kristen V., and Kristen V. Brown. "Sen. Chuck Schumer Wants the FTC to 'Take a Serious Look' at DNA Testing Companies." Gizmodo. November 30, 2017. Accessed April 09, 2019. https://gizmodo.com/sen-chuck-schumer-wants-the-ftc-to-take-a-serious-look-1820769045.

“CODIS-NDIS Statistics.” FBI. https://www.fbi.gov/services/laboratory/biometric-ana.lysis/codis/ndis-statistics (accessed November 19, 2017).

Daston, Lorraine. *Science in the Archives: Pasts, Presents, Futures*. Chicago: University of Chicago Press, 2017.

"Giambattista della Porta." Science and Its Times: Understanding the Social Significance of Scientific Discovery. . *Encyclopedia.com.* (April 4, 2019). https://www.encyclopedia.com/science/encyclopedias-almanacs-transcripts-and-maps/giambattista-della-porta

Gilliland, Anne J. “Archival Appraisal: Practising on Shifting Sands,” in Archives and Recordkeeping: Theory Into Practice, Patricia Whatley and Caroline Brown, eds. (Facet Press, 2014), pp.31-61.

Goldstein, Joseph. 2013. “Police Agencies Are Assembling Records of DNA.” *The New York Times*, (2013). http://www.nytimes.com/2013/06/13/us/police-agencies-are-assembling-records-of-dna.html.

Haag, Matthew. "FamilyTreeDNA Admits to Sharing Genetic Data With F.B.I." The New York Times. February 04, 2019. Accessed April 06, 2019. https://www.nytimes.com/2019/02/04/business/family-tree-dna-fbi.html.

"Federal DNA Collection." 2012. Electronic Frontier Foundation. December 13, 2012. https://www.eff.org/foia/federal-dna-collection.

Hares, Douglas R. 2015. "Selection and Implementation of Expanded CODIS Core Loci in the United States." Forensic Science International: Genetics 17 (Supplement C): 33–34. https://doi.org/10.1016/j.fsigen.2015.03.006.

Ketelaar, Eric. "The Right to Know, the Right to Forget? Personal Information in Public Archives," *Archives and Manuscripts* 23 (1995): 8-17.

Li, Jun Z., Devin M. Absher, Hua Tang, Audrey M. Southwick, Amanda M. Casto, Sohini Ramachandran, Howard M. Cann, et al. "Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation." *Science* 319 (2008):1100–1104. https://doi.org/10.1126/science.1153717.

Marks, Jonathan M. *What It Means to Be 98% Chimpanzee: Apes, People, and Their Genes*. Oakland: University of California Press, 2005.

Millar, Laura A. Archives Principles and Practices, 2nd. ed (London: Facet, 2017), pp.93-115.

Shah, Aditi. 2014. "Do You Know Where Your DNA Is? Genetic Privacy and Non-Forensic Biobanks." Council for Responsible Genetics.

Yaya, Isabel. "Wonders of America: The Curiosity Cabinet as a Site of Representation and Knowledge." *Journal of the History of Collections* 20, no. 2 (2008): 173-88. doi:10.1093/jhc/fhm038.

Core Paper:

The Infrastructures of Cryptocurrencies

IS 270: Systems & Infrastructures

Winter 2018

Professor Miriam Posner

At the time when I was taking Systems and Infrastructure, the world of cryptocurrency trading had been in the midst of its first major crash since the creation of Bitcoin in the beginning of 2009. The event has since been historicized (on Wikipedia, at least) as the "Bitcoin Crash", or more dramatically, the "Great Crypto Crash". As the value of cryptocurrencies continued to freefall throughout the months of January and February, the topic stirred significant commotion within the pages of the popular press. Having had a lingering interest in the history of money thanks in large part to a Marxist historian I had as a professor in my undergraduate years, I became interested in studying the technical and social infrastructure that underpinned cryptocurrencies. For the final assignment, the class was asked to do a video project and a supplementary write-up. My video focuses on Bitcoin ATMs as material manifestations of the infrastructure involved in this oft-abstracted form of monetary exchange. The paper takes a more historical perspective and studies the ideological, social, and technological changes that have enabled cryptocurrencies to take root.

Even within today's highly polarized news environment, it is hard to find a topic more subject to conflicting viewpoints than that of cryptocurrencies. Proponents of the technology claim that it will usher in a new era of commerce, whereas its critics argue that the technology is inherently flawed and destined to collapse spectacularly. This paper on the infrastructure of this so-called "currency" was born out of a curiosity in the history of money – the oft-termed "root of all evil" – and at the time of writing, heavy publicity regarding the crash of poster cryptocurrency Bitcoin and the new measures being taken by different countries to regulate its activity. While money itself is not inherently a sin, however much people may revile financial institutions, the disproportionate distribution of wealth is one of the most glaring, universal indicators of systemic inequality within the world. While cryptocurrency did not emerge out of a desire to eradicate economic inequality, their emergence was rooted in a desire to topple those institutions that are so instrumental to the persistence of unequal conditions. This paper will explore the political and socioeconomic ideologies that underlie cryptocurrencies, the labor and industrial infrastructure of what is commonly accepted as a nebulous monetary system, and the implications that they hold for future economic systems.

Money as Infrastructure and the Infrastructure of Money

When discussing cryptocurrencies, it may be helpful to look into the historical contexts of money in general and the infrastructural relationships that make it a viable means of exchange. Money – and all the underlying institutions and modes of control that allow it to flow throughout society – provides the transactional power for the perpetuation of economic, social, and political order. While there have been communal societies and movements throughout history that have sustained themselves without utilizing some form of currency for exchange, there are few examples of money-less societies in the present day. Whether it be metal coins, sacks of grains, cattle, or credit, money provides the

standards for exchange between the various entities (i.e. workers, institutions, government bodies, to list a few) that make up society. As argued by Marx, money is the material manifestation of the fetishization of social relationships of material production.[1] This is evidenced through fiat monies, which possess no intrinsic value other than being issued by state authorities as the official way of trade among its subjects.

By defamiliarizing the idea of money, its everyday use can almost be stated as a "madness shared by many", where society at large tacitly agrees that these symbolic assets have value both in the present and in the future. Developing an alternative system of exchange would necessarily pit an individual or community against these very imposing standards and their extensive spheres of influence. Propped up by banks, the US Treasury, and other financial institutions, the dollar (and its digital representation via a credit card or a debit card) is a very prosaic item that embodies many of the defining traits of infrastructure, such as ubiquity, centralization, and territoriality.[2] While essentially forming the basis of a money-based economy, money is also dependent on a variety of legal, social, and financial infrastructures that enable their continued use.

Infrastructural changes with digital forms of payment

The rise of new social networking and information communication technologies such as near-field communication has heralded new methods of mediating monetary transactions. Examples of this trend include the rent-payment-revolutionizing, peer-to-peer payment app Venmo, the digital wallet Apple Pay, and the international money transfer app Xoom, which is backed by Paypal. Suffice to say, many countries are now moving towards a cashless society. While the gradual decline in paper money can be safely described as a global trend, the manner of transition is by no means a homogeneous occurrence. Numerous Scandinavian and eastern European countries like Sweden and

---

1 Karl Marx, *Capital*, vol. 1 (New York: Penguin Books, 1981), p. 168.

2 Nicole Starosielski, "Introduction: Against Flow," in *The Undersea Network* (Durham, NC: Duke University Press, 2015), 1-26.

Estonia, as well as China, have made substantial gains in this regard.[3] Comparatively, the US is behind in the widespread use of such technologies, and the process of updating its material infrastructure to accommodate such changes is often fraught with challenges. One immediate example is the nationwide roll-out of EMV (Europay, Mastercard, and Visa)-compliant terminals and credit cards in 2015 to better protect against credit card fraud. This transition came a full decade after such technologies had been implemented by other countries, and was met with numerous technical setbacks and changes in legal policies concerning retailers' liabilities.[4] The infrastructure of money is intricately interwoven with various others systems, cultures, and modes of production, although, like other infrastructure, this becomes most apparent upon its breakdown.

In all of the cases discussed previously, the basic premise of how money is regulated, distributed, and produced is largely the same – through centralized financial institutions. In undermining such norms, cryptocurrencies have often been touted as serious threats to the nature of current financial systems. On a political level, they threaten the decades-long hegemony of the US dollar. The particularly radical core of cryptocurrencies is their utilization of a decentralized verification system known as blockchain technology. Explanations of the blockchain tend to get rather hairy for the uninitiated and the initiated alike, but they can be generally understood as a transparent, publicly accessible ledger maintained and managed through the consensus of a mass, distributed network of anonymous individuals. In ideal situations, blockchains are irreversible. Every new transaction is added as a "block" onto a "chain" of past transactions, and the exact identities of the parties involved in a transaction are concealed behind an encrypted address. It allows for peer-to-peer exchange without third party oversight. The blockchain cuts out money's usual flows through financial

---

3 Nathaniel Popper, Guilbert Gates, and Sarah Almukhtar, "Will Cash Disappear?" The New York Times, November 14, 2017, accessed April 13, 2019.

4 Ian Kar, "The Chip Card Transition in the US Has Been a Disaster," Quartz, July 29, 2016, , accessed April 13, 2019,

institutions, speeding up transactions and cutting out processing fees. Bitcoin was the first proof-of-concept for blockchain technology.

Trials in Realizing Crypto-Utopias

The ideological underpinnings of cryptocurrencies can be traced back to 1980s social movements that advocated the use of cryptographic and privacy-enhancing means to secure individual freedoms from government and corporate control. In the wake of the 2008 financial crisis, an individual operating under the pseudonym Satoshi Nakomoto released a seminal white paper detailing the implementation of blockchain technology for a new digital currency, later termed the "bitcoin". In his article, Nakamoto cites a work by computer engineer Wei Dai, who references a late 1980s FAQ document by activist Timothy May titled "The Cyphernomicon". May's document contained a stirring section subtitled "The Crypto Anarchist Manifesto", written for the "cypherpunks of the world".[5,6] May's manifesto formed the ideological backbone of the cypherpunk movement, heralding the arrival of a new age of digital egalitarianism rooted in anonymity and privacy. Cypherpunks sought to leverage the development and greater availability of cryptographic technologies to upend the entire apparatus of a centralized state, preventing oppressive institutions from stifling the opinions and wills of the masses. May boldly proclaimed a future where governments "will have a hard-time collecting taxes, regulating the behavior of individuals and corporations (small ones at least), and generally coercing folks when it can't even tell [sic] what _continent_ folks are on"[7] For many in the cypherpunk movement, cryptography would help make possible the creation of an autonomous network.

On the other hand, it should be noted that this surge of crypto-infused optimism stemmed from technology developed largely by the National Security Agency.[8] SHA-256 is one of most important

---

5 Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," Bitcoin.org, 2008

6 Timothy C. May, "The Cyphernomicon: Cypherpunks FAQ and More, Version 0.666," Cypherpunks.to.

7 Ibid.

8 Sam Biddle, "The NSA Worked to 'Track Down' Bitcoin Users, Snowden Documents Reveal," *The Intercept*, 2018.

secure hash algorithms underlying the Bitcoin network – including numerous other cryptocurrencies – and had been produced by the NSA in the early 2000s. In fact, numerous conspiracy theories abound around the exact origins of Bitcoin, with some believing it to be an elaborate psychological operation on the part of the NSA to roll out a "one-world currency". A recent report by *The Intercept* has revealed that the NSA had been actively tracking the activities of Bitcoin owners, seeking to uncover users' internet addresses, network ports, and timestamps to identify their targets.[9] Absolute anonymity also engenders its own slew of problems, providing the means for illegal activity. One of Bitcoin's earliest meetings with fame came after the arrest of notorious Silk Road drug trafficker Ross William Ulbricht (aka Dread Pirate Roberts), who was discovered to have possessed about $28.5 millions worth in Bitcoins at the time of his arrest.[10] Even today, the general perception of cryptocurrencies is that they facilitate illegal activity, rather than provide any challenge to legitimately oppressive financial structures. Many cryptocurrency advocates associate themselves with a controversial term known as "anarcho-capitalism", a derivative of capitalism free from government regulation. Combined with the libertarian values of absolute privacy and individualism, values shared by a significant proportion of the cryptocurrency community, this definition of anarcho-capitalism ends up echoing the core tenets of neoliberalism, which stresses the superiority of the free market in resolving all financial and social ailment.[11]

Coin Communities

Given that cryptocurrencies do not have any state backing, their transactional power and viability is dependent on the formation of a community willing to utilize them as units of exchange. In other words, cryptocurrencies require dedicated investors, making them not too different from Silicon

---

9 Ibid.

10 Sam Biddle, "The NSA Worked to 'Track Down' Bitcoin Users, Snowden Documents Reveal," *The Intercept*, 2018.

11 David Harvey, "Freedom's Just Another Word…," in *A Brief History of Neoliberalism* (Oxford; New York: University Press, 2005), 14.

Valley tech start-ups. Much of the publicity regarding cryptocurrencies centers around the first crypto Bitcoin, but in the decade since its inception there have been hundreds of altcoins (Bitcoins spin-offs) that have boomed and busted, each appealed (and, oftentimes, scammed) a certain niche group of users. Coins such as Monero (which has an overwhelmingly negative public image stemming from its predominant usage in dark web and black market purchases) were made with privacy as its core principle of design. Others have very specific scenarios of use, such as Potcoin, a cryptocurrency used exclusively for transactions concerning legalized marijuana.[12] Dogecoin, which started out purely as a "joke" coin, has been surprisingly resilient to obsolescence despite reports of it becoming a fad as members of the community move onto the next "meme-coin".[13] Most cryptocurrencies are not started with extensibility in mind, lending them an ephemeral quality. With regards to the many legitimate cryptocurrencies that have fallen to the wayside, the issue seems not to be that of a subconscious prizing for "innovation" over "maintenance" as Russell and Vinsel identified in other infrastructures, but to shifts in community interests and the coins' limited use scenarios.[14] For a system that was designed to enable "electronic transactions without relying on trust" – as sought by Satoshi Nakamoto white paper – there is a considerable amount of trust and community-developing involved.

The labor and materiality of cryptocurrencies

The labor that sustains a cryptocurrency's operations varies between coin to coin. At the basic level, all cryptocurrencies have the same hardware needs, namely high-performance graphic cards and ASICs (application specific integrated chips), cooling systems, a constant source of cheap energy, and secure networks over which transactions are made. The role of cryptocurrency "miners", then, is to utilize high-capacity CPUs, GPUs, and special mining chips to solve the complex math problems

---

12 Maria Bustillos and Maria Bustillos, "The Bitcoin Boom," The New Yorker, June 19, 2017, , accessed February 10, 2018,

13 Sarah Jeong, Lana Swartz, and Bill Maurer, "Dogecoin" in *Paid: Tales of Dongles, Checks, and Other Money Stuff* (Cambridge: MIT Press, 2017), 71.

14 Andrew Russell and Vinsel Lee, "Hail the Maintainers," *Aeon*, 2016.

needed to verify these transactions. A cursory tracking of the supply chain that belies the production of these devices and resources, draws up industries that have become increasingly strained by the exponential growth in cryptocurrency mining activity. While mining was once a process that could be done by an individual working on their laptop, rising energy costs and increasingly difficult cryptographic problems have led to the development of large scale miner farms. In such farms, individuals can purchase servers on remote mining farms, cutting down costs in electricity and maintenance to maximize profit. The mining process brings to light the sheer material foundation needed to keep the cryptocurrency system running. Every step of the infrastructure depends on having a strong, reliable Internet connection as well as sophisticated computer hardware/software capable of performing high-throughput processing of cryptographic algorithms.

Cryptocurrency mining incurs heavy environmental costs. According to one 2017 report, the collective energy consumption of mining ventures worldwide exceeds that of over 159 countries, a phenomenon that has received much attention in the press. A 2018 *Engadget* article reported that China is now seeking to curtail the activities of "cryptocurrency mining enterprises" by increasing electricity rates and introducing stricter environmental regulations. China remains heavily dependent on coal for meeting most of its energy needs, and as the world's leader in large scale cryptocurrency mining, this gives the terminology of "mining" perhaps an unintended extra layer of meaning. Within the US, cryptocurrency mines have been cropping up in rural areas such as Wenatchee, Washington, often taking root in abandoned warehouses located close to cheap power sources. As expected, the diversion of resources to these ventures has led to significant tension between the mining industry and local communities.

Conclusion

Cryptocurrencies rely on complex technological and social infrastructures that have very strong material and environmental impacts. In seeking an alternative to the centralized control of money, they run contrary to various financial and political institutions. As with most technology ill-understood, cryptocurrencies simultaneously carry the power to dismantle and prop up the status quo. The technology has received support from significant figures in the conversation regarding privacy and state control, such as Wikileaks founder Julian Assange, whose organization subsisted substantially off of Bitcoin donations during its inception during the 2011 Wikileaks Banking Blockade. Assange even dubs it as "the real Occupy Wall Street".[15] At the same time, the coveted blockchain technology that is so central to cryptocurrencies has also drawn the interest of former chairman of the US Federal Reserve during the 2008 fiscal crisis Ben Bernanke, who gave a keynote presentation for a cryptocurrency conference in late 2017. Such a figure would have been viewed as "Public Enemy #1" for cryptocurrency traditionalists, and it is unlikely he would have received such a platform during the concept's earlier years. Capitalism is said to have the capacity to subsume and appropriate other modes of exchange and value. Currently, cryptocurrencies exist in an uncomfortable liminal space between the ideals of 1980s cryptography enthusiasts and the existing structures and institutions of state power.

---

[15] Paul Roberts et al., "This Is What Happens When Bitcoin Miners Take Over Your Town," POLITICO Magazine, March/April, 2018, , accessed March 10, 2018,

Works Cited

Appelbaum, Binyamin. 2018. “Is Bitcoin a Waste of Electricity, or Something Worse?” *The New York Times*, February 28, 2018, sec. Economy. https://www.nytimes.com/2018/02/28/business/economy/bitcoin-electricityproductivity.html.

Biddle, Sam. “The NSA Worked to ‘Track Down’ Bitcoin Users, Snowden Documents Reveal.” n.d. Accessed March 24, 2018. https://theintercept.com/2018/03/20/the-nsa-worked-to-track-down-bitcoin-users-snowden-documents-reveal/.

Bustillos, Maria. 2013. “The Bitcoin Boom.” *The New Yorker*, April 1, 2013. https://www.newyorker.com/tech/elements/the-bitcoin-boom.

“China Reportedly Wants to Curtail Wasteful Bitcoin Mining.” n.d. Accessed February 3, 2018. https://www.engadget.com/2018/01/08/china-to-curtail-bitconi-mining/.

Harvey, David. “Freedom’s Just Another Word…,” in *A Brief History of Neoliberalism* (Oxford; New York: Oxford University Press, 2005), 5–38.

Jeong, Sarah, Lana Swartz, and Bill Maurer. n.d. “Dogecoin.” In *Paid: Tales of Dongles, Checks, and Other Money Stuff*, 53–70. MIT Press.

Kar, Ian. "The Chip Card Transition in the US Has Been a Disaster." Quartz. July 29, 2016. Accessed April 13, 2019. https://qz.com/717876/the-chip-card-transition-in-the-us-has-been-a-disaster/.

May, Timothy C. "The Cyphernomicon: Cypherpunks FAQ and More, Version 0.666". Cypherpunks.to. *February 28, 2011*, http://www.cypherpunks.to/faq/cyphernomicron/chapter2.html#2

Marx, Karl. *Capital*. Vol. 1. New York: Penguin Books, 1981.

"Mining Bitcoin Costs More Energy than What 159 Countries Consume in a Year." 2017. *Futurism* (blog). November 27, 2017. https://futurism.com/mining-bitcoin-costs-more-energy-159-countries-consume-year/.

Nakamoto, Satoshi. "Bitcoin: A Peer-to-Peer Electronic Cash System." Bitcoin. October 2008. https://bitcoin.org/bitcoin.pdf.

Nowviskie, Bethany, and Dot Porter. "Graceful Degradation Survey Findings: Managing Digital Humanities Projects Through Times of Transition and Decline?" Digital Humanities 2010 Conference Abstract, June 2010. http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/ html/ab-722.html.

Starosielski, Nicole. *The Undersea Network*. Durham: Duke University Press, 2015.

Popper, Nathaniel, Guilbert Gates, and Sarah Almukhtar. "Will Cash Disappear?" The New York Times. November 14, 2017. Accessed April 03, 2019. https://www.nytimes.com/interactive/2017/11/14/business/dealbook/cashless-economy.html.

Roberts, Paul. "This Is What Happens When Bitcoin Miners Take Over Your Town." POLITICO Magazine. March/April, 2018. Accessed March 10, 2018. https://www.politico.com/magazine/story/2018/03/09/bitcoin-mining-energy-prices-smalltown-feature-217230.

Russell, Andrew and Vinsel Lee, "Hail the Maintainers," *Aeon*, April 7, 2016, https://aeon.co/essays/innovation-is-overvalued-maintenance-often-matters-more.

Elective Paper:

Digital Asset Management Plan for the William Andrews Clark Library

IS 289: Digital Assets Management

Spring 2018

Professor Linda Tadic

I decided to take Digital Assets Management in Spring 2018 with Professor LInda Tadic out of a curiosity in the UCLA Clark Library's various digital projects. Our assignment was to work with an institution that was in need of or interested in adopting a new digital asset management system and to identify the functional requirements needed of a DAM to best serve the library's purposes. Since I worked at the Clark Library and was already somewhat familiar to the types of assets they had and their need for a centralized DAM, I decided to use the Clark as my site of study.

Functional Requirements for a DAM: The William Andrews Clark Memorial Library

Description of Site

The William Andrews Clark Memorial Library is a UCLA off-campus, non-circulating library located on a historic five-acre property in West Adams that had been bequeathed to the university by its former owner William Andrews Clark in 1934. The library specializes in English rare books and manuscripts dating from the Tudor period to the long eighteenth century, as well as materials from the rest of early modern continental Europe. Though originally part of the UCLA library system, the Clark is now administered by the Center for 17th & 18th Century Studies. From a technological standpoint, the library is distinctive in its continued use of the card catalog (as a supplement to the regular UCLA-wide online library catalog) and its lack of a workflow management system for requesting materials. As a Los Angeles Historic-Cultural Monument, the ground's features (furniture, facilities, and artwork) are not to be altered, and must be preserved accordingly to the appropriate HPOZ (Historic Preservation Overlay Zone) guidelines. The Library currently receives grants from institutions such as the NEH and CLIR to perform large-scale digitization projects of their early modern manuscripts and annotated books, many of which are currently hosted on Calisphere through a partnership with the UCLA Digital Library.

Staff Members Interviewed:

Anna Chen
Head Librarian
achen@humnet.ucla.edu

Philip S. Palmer
Research Services
crobinson@humnet.ucla.edu

Scott Jacobs
Reading Room Supervisor
sjacobs@humnet.ucla.edu

Alejandro Sanchez Nuñez
Cataloging Assistant
sanchez@humnet.ucla.edu
s

Carole Robinson
Site Manager
crobinson@humnet.ucla.edu

1. USERS

The hypothetical DAM system will primarily be utilized by the Research Services Department of the library (composed of 1-2 people), which will be creating most of the digital assets to be stored in the system. The digital assets emerging from this department, generally high quality scans of entire rare books and manuscripts, are generally amongst the more valuable

digital assets of the library, resulting from projects funded by institutions such as the NEH and CLIR (Council on Library and Information Resources). Currently, the scanning is performed by the third-party service Luna Imaging, with the high-resolution scans of the documents being sent back to the Clark Library on a single hard disk drive (no redundancies, though the images are backed up on tape at the imaging center) that is then taken over to the Digital Library, where the files are ingested into Nuxeo and uploaded onto Calisphere. The Head of Research Services has expressed that the Library's use of Calisphere was a decision born out of convenience rather than it being an ideal fit, and that they would be open to an alternative system that would grant them more flexibility with what could be done with the assets. Research Services also handles quality control of the assets while the items are being scanned and after they have ingested into the system. checking to see if the metadata matches up with the physical items. Some metadata management may also be performed by those in the Cataloging department.

The second group of internal users will be those in Reader Services (an operation run predominantly by one person), who seek a DAM system to manage reprographic requests from a variety of sources, ranging from museum institutions to individual researchers. Reprographic requests are fulfilled using simpler means – materials are usually photographed using a camera, rather than professionally scanned. Currently, all the assets for this department are stored on the computer of a single staff member (who serves as the asset's main point of access). The DAMS system would allow this individual to better keep track of which assets have already been created, as well as allow them to retrieve requests more quickly. With the acquisition of a DAMS, Reader Services has posited the possibility of making a greater variety of assets (i.e. video and audio content).

The third internal user group is the Site Manager, who believes it would be beneficial for the library to have a system that could be used to track the conditions of the library's furniture and other decorative features (i.e. artwork). The DAM system would help the library keep track of such assets according to the Clark's needs as a historical preservation site. For example, images could be taken of the furniture, placed into the DAMS, and be used to track upholstery changes overtime.

External users include the staff at the Center for 17th & 18th Century Studies, who will need to find the necessary materials for predominantly promotional needs, but they will primarily by interacting with the DAM through the content managements system that it is integrated with, although. Public users – the library patrons and other scholars – will likely interact with the DAM in a similar fashion.

The putative DAM system will have to support a minimum of 3-4 internal users (with differential levels of access) and be able to integrate with a CMS for public access.

2. FORMATS

The Clark's assets are predominantly photographic, although, as mentioned there are plans to include a greater variety of assets should there be the existence of a system to manage them all. Currently, the files for the digital projects are distributed across several hard disk drives, with copies existing on the servers for the Digital Library, so size estimates are hard to make. Based on rough estimates, as well as a survey of the library's shared drives:

Current file formats	File storage (in GB)	Estimate 5-year growth
TIFF	3500	variable, depending on project
JPEG2000	500	100
PNG	400	80
CR2	< 1	< 1
PDF	300	50
.psd (Adobe Photoshop Image)	10	< 1
.doc	< 1	< 1
.xlsx	< 1	< 1
MP4 (H.264 codec)	< 1	50

Total storage size: ~4500 GB, ~4.5 TB
Estimated growth in 5 years: 0.5-1 TB

3. CONTENT

The DAM system will primarily be used to organize the following digital asset contents:

- Professional grade high-resolution pages of early modern books and manuscripts
    - (~85%)
- Digitized materials (manuscripts, parts of books, illustrations, and images from the Clark Library collections aggregated from a history of reprographic requests)
    - (~10%)
- Facilities management (photograph of the grounds, repairs, furniture, artwork)
    - (~1%)

Potential secondary assets for inclusion in the DAM:

- Programming (i.e. invitations for speakers, conference materials, itineraries, hotel checklists, anniversary events)
    - (< 1%)
- Library workflow documents (i.e. acquisition information and collection files, purchase receipts, correspondence, deed of gifts, loans)
    - (< 1%)
- Photographs (i.e. from exhibitions, events)
    - (2%)
- Promotional materials (i.e. event flyers, newsletters)
    - (< 1%)

Looking forward, the library is considering the creation of short documentary-type videos regarding books and authors that are prominently featured in their collections.

4. CONVENTIONS

Currently, file-naming conventions are employed inconsistently across the different asset types and are mostly exclusive to the creator of the file. Depending on the asset, there is usually some indicator for the date of the content matter (i.e. the year a copper-plate engraving was made) in the asset as well as a descriptive statement.

- Sample naming conventions:
    library_gatehouse_history_1987
    MS_2015_014_bookseller_description_item39
    Bruman_Payment_iPalpiti Soloists_2 Performance Payments
    UCLA_Clark_Library_10.24.13
    Collection#_DescriptionOfItem
    uclaclark_call#_page#

Assets for Digitization Project

The file naming conventions for assets produced as a result of the digital projects have a decent file naming convention: uclaclark_call#_page#. One element to add may be the project ID under which the asset was made, if this information is not already made available in the metadata.

Example: uclaclark_projectID_call#_page#

Since the call number that comprises the bulk of these filenames is oftentimes long and convoluted (for example “QL673.A31779*”), it might be beneficial if the DAM system could

have some system in place to check that the correct call # was entered in its entirety into the filename. Within the catalog, the call number of a lot of the rare books have been appended with an asterisk to indicate that it is a rare book. In such cases, the asterisk should be removed since asterisks have their own meaning as command functions in many DAM systems, and be replaced with a different code like “RB”.

Assets for Reprographic Services:

Currently, the file names follow the schema “Collection#_DescriptionOfItem”. It is recommended that some extra information be added in regards to its physical location (i.e. box and folder number), and that the description of the item be kept in a separate spreadsheet. For reprographics of archival materials, it might be beneficial to include information such as the box and folder numbers. For books, the call number and a page number should be included. If edits have been made to a particular file, they should be indicated with information regarding the version number; the master copy should be clearly labeled and kept as well. Same as with the files for the digitization project, a way to easily cross-reference the materials in the DAM to the library’s existing catalog records would be helpful as well.

Assets for Furniture/Artwork Management

Furniture inventorying currently utilizes unique IDs (termed property IDs), which are physically noted on the furniture. Photographs taken of the furniture have their own separate numbering system: a simple “001, 002, 003…” numbering scheme that seems to reflect the order in which the photo was taken. Since a lot of the furniture has yet to be systematically catalogued/identified, the use of these two IDs (property IDs and photograph IDs) persists, and the two different numbering systems are corresponded with the item description using an Excel spreadsheet. Upholstery changes are also not tracked in an organized manner. The artwork at the Clark is currently being catalogued and will be assigned unique bibliographic numbers and call numbers based on their location. Within the DAMS, the file names for the photographed/scanned artworks should include information on who created the artwork, its title (if available), and information regarding the digital version made of the work (publication-grade, low resolution)

The DAMs system could help to streamline this task, keeping track of multiple numbering schemas as well as manage “version control” resulting from different manifestations of a piece of furniture throughout the course of upholstery changes. It is advised that the photographic files of the furniture – those with the numbering system “001, 002, 003” – be changed to a more descriptive scheme that includes information such as the type of upholstery work that has been done, a date of when the photograph was taken, and a brief description of its location. A controlled vocabulary should be created for the different types of upholstery work.

Example: photograph#_upholsteryWork_mmyyyy_location

If the item has been properly identified and assigned a property ID, the photographic manifestation of the piece of furniture should be given a filename that reflects this, instead of the "001, 002, 003..," numbering system.

Example: propertyID_upholsteryWork_mmyyyy_location

General Suggestions:

- Regardless of the format/content of file, avoid the use of special characters in file names, in particular: " / : * ? ". Use underscores instead of spaces to separate elements when naming files
- Indicate version types in file names and keep original master copy (should be the one with the best quality)
- For any video related assets that emerge in the future, here are some potential file naming conventions:
    - If the video is regarding a specific book/manuscript in the Clark collection
        Example: VID_DescriptionOfVideo_Call#OfBook
    - If the video is regarding an individual, use their standard names as listed by ULAN (Union List of Artists Names)
        Example: VID_DescriptionOfVideo_ULANentry

5. METADATA

The Library's digital assets are currently described at varying levels depending on the group of assets that they belong to – digital projects, reprographics, or site management. The reprographic materials and photos of site facilities/furniture do not need to be described to the extent as the materials produced for formal digital projects, although the artwork is formally cataloged. Since there will be different types and extents of metadata, the DAM must support customizable metadata fields.

Assets for Digitization Projects

Currently, the metadata for the rare books and manuscripts can be understood as falling under "item level" and "page level" metadata. Page level metadata is currently being entered into an Excel spreadsheet, and quality control is performed to make sure that the data in the spreadsheets correspond with what is uploaded into Nuxeo at the Digital Library. Item level metadata (referring to records for the entire book or manuscript) can be found in the Voyager cataloging system and is publicly accessible through the UCLA Library catalog. One important feature for the Clark's own DAM would be to have a straightforward way for item-level metadata in the cataloging record to be transferred over. The MARC records would have to exported in MODS/XML before they can be ingested into the DAMs. The metadata for the

materials can sometimes be in languages that utilize diacritics, and there are oftentimes special symbols (superscripts, subscripts, double daggers ‡, sections §, etc.) that are used to describe certain aspects of the book such as their signatures and pagination. As such, the DAM system should also support Unicode. The DAM should also ensure that metadata for the assets in this category can be mapped onto DublinCore, with such information made exportable at the front end (CMS).

Assets for Reprographic Services:

The individual managing reprographics utilizes a local metadata schema consisting of a relatively limited number of fields:

- collection number
- item's year of publication
- page number
- document type
- a brief title describing the content (i.e. "Oscar Wilde American Tour")

While these assets do not warrant extensive technical metadata extraction, it would be helpful to include information about the file resolution, as well as some metadata regarding the process of asset creation (i.e. who scanned the photo, who made edits to it). Since many of the assets are requested for publication elsewhere, some administrative metadata (addressing issues of copyright and licensing) would also be necessary. Having a formal system through which public users can make requests for items could also help facilitate the process of monetizing reprographic services.

For a time being, there was an attempt to use Adobe Bridge to attach metadata onto an asset, but the entire process turned out to be rather cumbersome as there were some difficulties performing bulk ingest. Reinstating the use of Bridge to manage new assets created could serve as a short-term solution and could probably ensure that the assets are better prepared for transfer into a formal DAM system.

Assets for Furniture/Artwork Management

The furniture is described based on a local metadata schema including information regarding its:

- location
- dimensions
- creator
- monetary value
- physical description
- provenance

There is no need for technical metadata extraction with these assets, as they will only be for internal use and to track upholstery changes and locations of the furniture within the library.

Regarding the artwork, the DAM system should support existing thesauri such as the AAT (Arts and Architecture Thesaurus) as well as a local thesaurus creating tool. Similar to the digitized rare books/manuscript assets, the artwork will have a cataloging record. Information in the catalog record should also be present in the DAMs, with the addition of fields for describing specifics for the digital manifestation of the art:

- Resolution of file
- Date when photographic image was taken
- Usage in exhibitions

Overall, given that many of the digital assets will relate in some way to a singular physical item (a book or a collection), the ideal DAM system should support parent/child relationships where all the different manifestations (i.e. different scanned versions of a photograph or the pages of a book) can be linked back to a parent work (the book or archival collection). Most of the metadata for the assets is encoded in MARC, so the putative DAM system should be one that can take in XML-rendered information from the original MARC records. Since much of the metadata is also currently being entered into an Excel file, the Clark would need a system that can take in CSV files.

6. SEARCH

The putative DAM system will need to support faceted and keyword searching. Oftentimes, patrons make very specific image requests and faceted searching based on a specific metadata field would make it easier to find items. Other times, external/public users may request images based on subject tagging (i.e. “alchemical imagery”), so general keyword searching is also desirable. A sizeable portion of the assets, particularly those resulting from early 2000s reprographic requests, has little to no metadata attached, so the DAMS should support some manner of visual recognition, OCR (optical character recognition), or HTR (handwritten text recognition) to make them findable if retroactively attaching metadata is no longer feasible.

7. DISPLAY AND ACCESS

Currently a chunk of the library's digital assets is all on a shared drive that is close to capacity, making the folders and files within the drive hard to open. The DAMS should support the ability to preview assets without having to open the entire file, as many of the staff members have remarked that opening the files takes up a sizable portion of the time spent on searching. All assets should be downloadable (and with future inclusion of video assets, streamable). With regards to the digitized book and manuscripts assets, materials should be downloadable at both the parent (entire book) and child level (individual pages).

Public access to the Library's digital assets, asides for the digital projects hosted on Calisphere, is all currently mediated through email or phone requests with one library staff member. Through the CMS, public users should be able to make requests for certain assets, which can be approved by the appropriate staff member.

8. WORKFLOW

Assets for Digitization Projects

After being scanned by a third-party imaging vendor, the files will be bulk-ingested into the DAM system, where they can be identified as being in a "pending" status. The system will also take in the appropriate metadata from exports of library MARC records, either from CSV or XML source formats, and allow for bulk-metadata editing. After the appropriate quality control measures are taken (comparison with the physical item), the status of the assets can be changed to "approved". Digitization project assets are intended to be free for public use, and such users should be able to download entire copies of books/manuscripts, as well as metadata information by requesting from the CMS.

Assets for Reprographic Services

After the images are photographed on-site, they will be ingested into the DAM system. The DAM will facilitate the bulk conversion of the assets into 3 different image file types – TIFF, PDF, and JPEG2000. When assets are requested, the user should be able to easily search for relevant items based on a specific metadata field or general keyword search. Since the items would be in a database accessible to a number of users, the task of answering reprographic requests could be distributed. Through the DAMs, the staff member should be able to issue the client a link to the file (through which they will be able to download the asset). The DAM system should cut down on the creation of redundant files and make it easier to manage copyright and intellectual property issues if relevant.

There is currently a sizeable backlog of digitized collection materials from staff members who have long since left the organization. Since many of these assets come from the same collection or deal with similar themes, a way to bulk ingest files and to perform bulk editing would significantly expedite the processing of the backlog. As mentioned previously, visual recognition, OCR (optical character recognition), or HTR (handwritten text recognition) could perform auto-tagging of such assets and make them searchable even with limited metadata.

Assets for Furniture/Artwork Management

Using the DAM system, the site manager will be able to create assets that help track preservation changes to furniture/artwork and link photographs to the physical object and their locations. The system could also help to manage the status of certain art pieces, which may sometimes get loaned to other institutions.

9. PUBLISHING

The Library is interested in a content management system that would help them push their assets onto a public platform – for example, a CMS such as Omeka. With future video assets, it might be necessary to publish them onto a platform such as YouTube or Vimeo.

10. DIGITAL PRESERVATION

The Clark's digital assets are not currently preserved in a systematic way, although there could be a policy in place for the materials that were sent to the Digital Library. Particularly regarding reprographic requests, a lot of the files made preceding the mid-2000s have become largely obsolete or have been corrupted with time.

The organization will need to have a DAMS that automatically generates MD5 checksums, although such checksums do not seem to have immediate use to the organization's day-by-day operations (at least, not given in current procedures).

Redundant copies of the hard drives containing files for the digitization project should also be made, particularly if the Clark is considering moving those assets off of the Digital Library's servers in the future. Since the Clark's current servers are shared with the Center, it may be beneficial to have s SaaS system that handles storage through cloud-based servers and conducts preservation actions.

11. USER RIGHTS AND SECURITY

All files are currently on a shared drive between the Center for 17th & 18th Century Studies and the Clark, with everyone having basically the same access and editing rights.

Historically, this has lead to files getting improperly edited or deleted, as well as numerous cases of creating redundant copies of an asset. The DAM will need to support different levels of access, with some people having the rights to create, edit and view the files, and others to only view them.

12. INTEROPERABILTIY

The proposed DAM will need to support International Image Interoperability Framework (IIIF) functionality, particularly the Image API and Presentation API. The DAM system should also interoperate with a CMS.

13. TECHNICAL

The Clark library utilizes both the Windows and Apple platform. Since the Library is unlikely to hire a programmer to create a tailored instance of an open-source DAMS, it may be necessary to seek out a proprietary SaaS DAM services. This would also help with storage concerns.

Advising History

I have many people to thank for my professional and academic development throughout my time in this program. In my first year, I had quarterly meetings with my advisor Professor Christine Borgman, who helped me map out my coursework and encouraged me to explore the different subfields within the information science discipline. She provided me with helpful suggestions for ways to integrate my background in microbiology, immunology, and molecular genetics with the program curriculum.

Throughout my first year in the program, I was also incredibly fortunate to have had the chance to work with Professor Johanna Drucker on an inventorying project at the community poetry center Beyond Baroque. Along with teaching me the best practices for handling, appraising, and describing materials, she also taught me by example to approach such tasks with a great deal of humor and persistence. It was always a great confidence boost to see the way valued my suggestions and thoughts on the work at hand.

When Professor Borgman retired following my first year, Professor Miriam Posner generously accepted my request to become one of her advisees, I have had numerous meetings with her over the past few quarters regarding career choices, building my portfolio, balancing academics with work, and how to deal with my own sense of imposter syndrome within the LIS profession after coming from an undergraduate background in molecular biology. She has responded to my indecisiveness and anxieties over the past year with infinite patience and understanding, and has always provided me concrete steps to move forward in whether it be for class assignments or for professional development.

I must also thank everyone at the William Andrews Clark Library. Being geographically distant from the main campus, as well as being overseen by the Center for 17th and 18th Century Studies as opposed to the UCLA Library, there is a unique sense of community there. I have learned various technical and professional skills - from cataloging in Voyager and automating tasks with macros to navigating office politics - from the librarians and staff members there including Rebecca Fenning Marschall, Alejandro Sanchez Nuñez, Nina Schneider, Philip Palmer, David Eng, Scott Jacobs, and Anna Chen. I have drawn upon their expertise numerous times for my academic work and am thankful for their indulging me in the process. Through their digitization projects, early modern cookery workshops, and budding seed exchange program, the Clark has taught me what it means to run a community-oriented special collections library.

I thank Dawn Childress and Geno Sanchez at the UCLA Digital Library, to whom I owe much of my knowledge of the hard skills for data management and digital humanities projects such as data cleaning, metadata standards, and using markup languages. Dawn has also helped me workshop resumes and job applications, and always approved of her student workers spending some of their work time attending job talks and workshops. She has also introduced me to many of her colleagues within the field and helped me expand my network. Geno has imparted upon me some of his considerable user experience expertise to help me with some of my own user experience side projects.

I have also had the great experience of doing volunteer work for Yolanda Bustos at the Natural History Museum. From her, I have learned some of the best practices for digital asset management and received advice for job hunting and professional development.

Finally, I would like to take some time to thank some of the individuals who have been responsible for my being in the program in the first place and who I have continued to seek out advice from. This includes Jillian Cuellar (Tulane U) and Courtney Dean (UCLA), who welcomed me into YRL Special Collections as an undergraduate student assistant at the Center for Primary Research and Training. Throughout the program, they have encouraged me to apply to scholarships and provided numerous letters of recommendation. I thank Professor Katsuya Hirano (History, UCLA) for showing me the potential of historiographical inquiry for producing social change, and the central roles that archives play in the endeavor.

Course History

Fall 2017
IS 211 | Artifacts and Cultures
IS 260 | Description and Access
IS 431 | Archives, Records, and Memory

Winter 2018
IS 270 | Systems and Infrastructures
IS 272 | Human-Computer Interaction
IS 289 | Public Health Informatics

Spring 2018
IS 212 | Values and Communities
IS 279 | User Experience Design
IS 289 | Digital Asset Management

Fall 2018
IS 240 | Management of Digital Records
IS 289 | Digital Methods for Research and Scholarship
IS 400 | Professional Development and Portfolio Design

Winter 2019
DGT HUM 201 | Introduction to Digital Humanities
HIST 201Q | Theory of History: Marxism and Culture
IS 262A | Data Management and Practice

Spring 2019
IS 262B | Data Curation and Policy
IS 289 | Web Development
IS 462 | Subject Cataloging and Classification
IS 464 | Metadata

Professional Development Statement

My entry into this field was an accidental one, inspired in part by an opportunity to work with special collections materials during an undergraduate history capstone seminar and a deep interest in historiography. After spending my senior year working as an undergraduate processing assistant in the Center for Primary Research and Training at YRL, I applied to this program declaring a specialization in archives. Going into the program, I had the good fortune to secure a technical services student work position at the William Andrews Clark Memorial Library, where I was able to work closely with the archivist and catalogers to organize, describe, and appraise materials according to best practices. Having had some experience working in the hodgepodge space that characterizes special collections, and having the guidance of individuals who were more than willing to teach me skills that I had no prior coursework experience in, such as cataloging, I decided that I could afford to branch out in terms of my academic coursework and take courses outside the field of archives to develop a more interdisciplinary perspective.

Over the course of this program, I have selected classes primarily in informatics, an act which I will not deny was due in some significant part to the four years I spent prior to this program as an undergraduate majoring in microbiology and molecular genetics. Taking Public Health Informatics was one such decision, as was deciding to join a team involved with the information architecture redesign of an environmental DNA sample collection research team in my user experience class. I am always trying to incorporate elements of my previous educational upbringing into my work, but the shift in career track has not always been a comfortable one. As anyone who looks at my curriculum vitae would be able to surmise, I have often had trouble pinning down a specific area of interest. Even now, after having recently taken a course on the

Marxist theory of history, I feel the stirrings of a desire to pursue further studies in history and critical theory. However, my time in this program has also taught me the sheer impossibility of describing anything in its totality. I am fascinated by the ways in which the sciences, social sciences, and the humanities overlap, and I feel that the LIS profession remains the best place to explore and to get others to engage in these liminal spaces.

Over the past year, I have steadily developed an interest in the field of digital humanities and digital libraries. I had taken a few computer science courses as an undergraduate and the idea of working with data and programming has always appealed to me. In my time working at the UCLA Digital Library, I have learned how to work with different markdown languages, automating tasks using bash scripts, working with Github, and digital asset management, all the while taking numerous workshops on a variety of technologies such as geospatial mapping and Jupyter Notebook. My methodologies class on Digital Research and Scholarship exposed me to even more tools, including natural language processing (NLP), topic modeling, and website development and hosting. When I took Professor Posner's class in the following quarter, I was able to further solidify my skill sets in data visualization and web development. In the immediate future, I plan to continue to develop greater proficiency in languages such as python and javascript. I feel that the introduction of these tools to humanities and social science research poses a transformative development in the field, and that over time these tools will become more accessible to a larger audience. As such, I also hope to be able to acquire some pedagogical skills in this regard, perhaps by getting involved with Library Carpentry.

My immediate goals upon graduation are to apply for positions in digital libraries. As someone who would never had entered the profession if I had not already had an interest in history, I am interested in utilizing digital libraries as a way to draw in students from the math and sciences field, increasing the visibility of the work being done in the information science field and promoting more interdisciplinary research. I would also like to build upon my work with user experience and to be involved in projects that seek to re-envision the way digital libraries and archives engage with their users and include them in the process of knowledge making.

In my time as a graduate student, I have made the oversight of not being proactive with joining professional organizations in my field. However, I am now actively working on developing projects/case studies that I could present at the Association for Computers and Humanities (ACH) Conference and at the Digital Library Federation Forum. My advisor Miriam Posner has also suggested that I apply for funding to attend the Digital Humanities Summer Institute (DHSI). The DHSI, which takes place in June of every year, provides numerous week and week-end long courses that teach valuable practical skills in data curation, linked data, and managing digital content as well as courses in critical pedagogy and research ethics, and would certainly provide a nice way to ease out of the academic rigor graduate school before easing into the professional arena.

Joyce Wang
asujok@g.ucla.edu

Education:

2017 – 2019: UCLA MLIS program

2013 – 2017: UCLA undergraduate student,
Major: Microbiology, Immunology, and Molecular Genetics Minor: History

Work Experience:

- *December 2018 – present: User Experience Team coordinator for Nagoya Protocol Learning Portal*

  Invited back by client from User Experience class to be part of the design team for a separate project involving the creation of an educational website for the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization. Conducted stakeholder interviews, user interviews, and competitive reviews, as well as created wireframes and presentation decks with a team of 3 other people.
- *June 2018 – present: UCLA Digital Library student worker*

  Managed the Clark Library's 250+ book large Early Modern Annotated Books digitization project, which involved file management for digital assets in Nuxeo, managing communications between the Digital Library and CDL, and performing quality control of digital assets. Performed metadata cleanup as well as produced metadata for other on-going projects at the Digital Library such as their collections of 20th century Cuban periodicals, Soviet Armenian posters, and Luxor Temple images. Familiarized myself with the use of terminals to run bash scripts to automate tasks like extracting technical metadata. Also worked on creating tutorials for Zooniverse crowdsourcing projects. Routine tasks include file-renaming, data transformations, and image processing in Photoshop. Created documentation for various workflows.
- *October 2017 – present: Rare Book Cataloging and Archival Processing Assistant at the William Andrews Clark Memorial Library*

  Archival processing of manuscripts and small archival collections related to a variety of subjects, including fine press materials, the academic research of early-modern scholars, and theatrical performers. Created online finding aids using ArchivesSpace. Updated and created new catalog records for fine press and reference materials in Voyager and OCLC. Performed rehousing of rare materials, book-plating, and other physical processing tasks, as well as taking the occasional reference desk shift.
- *July 2017 – June 2018: Assistant at Beyond Baroque Archives*

  Worked on creating inventories for the Beyond Baroque literary center's collection of zines, chapbooks, ephemera, serials, and institutional records. Helped with the front-end design, information architecture, and content creation of a digital archive prototype to promote materials within the collection.
- *June 2016 – June 2017: Undergraduate assistant at the Center for Primary Research and Training at UCLA's YRL Special Collections*

  Created finding aids for different collections on the Online Archive of California using Archivist's Toolkit, performed data cleanup on archival databases, scanned materials for digitization projects, and helped to plan exhibits and promotional events. Helped to process the V. Vale collection of punk materials and the faculty papers of biochemist Paul D. Boyer.

Volunteer Experience:

- *July 2018 – present: Volunteer in the Museum Archives of the Natural History Museum of Los Angeles*

  Digitized and performed quality control for a nitrate negatives digitization project. Created metadata to scans of the museum's historical publications to make them searchable within the museum's digital asset management system. Worked on preserving photos from the NHM's old Photo Department, coming up with file naming conventions and converting items to sustainable file types.
- *July 2015 – June 2017: Research Assistant for UCLA CART's Autism Genetics and Human Diversity Project*

  Assisted with data entry into the Internet System for Assessing Autistic Children (ISAAC) and Autism Genetic Research Exchange (AGRE) databases, shadowed psychologists and social workers, and recruited participants for the study through Facebook and Instagram ads. The project sought to better resolve both the genetic basis of autism as well as the socioeconomic barriers faced by African American families with autistic children in the Los Angeles area.
- *June 2015 – March 2016: Volunteer student researcher in Dr. Zhefeng Guo's prion lab*

  Involved with performing protein purification and spin-labeling of the yeast prion protein Ure2 in order to resolve the structure of the amyloid fibrils formed upon contact with prions and translate findings onto human diseases such as Alzheimer's and Creutzfeldt- Jakob.

Awards & Honors:

- 2018 – Theresa and Roy Aaron Fellowship
- 2016 – Ahmanson Undergraduate Research Scholarship

Technical Skills:

- Basic knowledge of Python and object-oriented programming in C++
- Archival Data Management and Cataloging
  - Archivists' Toolkit
  - ArchivesSpace
  - Voyager 10.0
  - OCLC Connexion
- Data Cleaning
  - OpenRefine
  - Microsoft Excel
- Digital Asset Management & Image Processing
  - Adobe Bridge, Photoshop
  - Nuxeo
- Data analysis and visualization
  - MALLET topic modeling
  - ArcGIS mapping
  - Tableau
- Jupyter's Notebook
- Basic web development (HTML, CSS)

Language Skills:

- Mandarin (Speaking: Professional working proficiency, Reading: Limited working proficiency)
- Spanish (Reading: Limited working proficiency)
- Taiwanese (Speaking: Limited working proficiency)

Accessibility Statement

Accessible design is essential for ensuring that all users, regardless of their disabilities or inexperience with utilizing an information technology platform, are able to find and understand the information that they need.

In my website, I have implemented the principles of accessible design in the following ways.

- I have a clearly labeled and persistent navigation bar (extended by the hamburger menu dropdown), so no matter which page the user is on, they will be able to have an overview of all the content of the site.
- I have broken down longer pages such as the "Major Paper" and "Course Work" into discrete sections that can be accessed in a non-hierarchical manner using the sticky "table of contents" sidebar on the right side of the page, reducing the need for scrolling.
- I have done my best to make my site responsive and adaptable to different platforms such as mobile phones, tablets, and resized browser windows. This goes for all the iframes that I have used to embed the different forms of content that I have on my site (i.e. the PDF viewer and the video hosted on Box).
- Sometimes iframes may not work as well on mobile platforms. In such cases, I have provided a more suitable alternate link to the material.

In building my website, I have considered how the site might be experienced by individuals with different disabilities. I have tried to alleviate their experiences by doing the following:

- Including alternate text to all images used in the site
- Making sure that all clickable elements in my pages support keyboard navigation and can be read by a screen reader
- Not employing color as a way to distinguish elements on a page so as not to put individuals with colorblind conditions at a disadvantage

In my digital portfolio, I was able to include substantially more information than in my PDF version. Some of my work, such as samples of the projects I have worked on at the Digital Library, my final project for my Digital Humanities course last quarter, and the video component of my core paper are not suited for presenting in a static format. My digital portfolio also grants users greater control over what they are paging through.

Along with wanting to challenge myself a bit in web development, I also decided to build my site using Github Pages to ensure that the final product would be as minimalistic and free of distractions as possible by having no ads or distracting site banners. Being hosted on Github means that the backend to my website is fully accessible and can be forked by anyone who is so inclined.