MARISA PURCELL

Portfolio

MLIS Candidate

University of California, Los Angeles
Department of Information Studies

Spring 2020

Advisor: Jean-François Blanchette

Website: mfpurcell.github.io

TABLE OF CONTENTS

ABOUT ME

I am a master's candidate in the Department of Information Studies at the University of California, Los Angeles, specializing in Informatics. I hold a bachelor's degree from the University of California, Berkeley in Rhetoric with a focus in Public Discourse.

My studies include: information-seeking behavior and information use, user-centered approaches to information system design, human-computer interaction, database design and management, and information policy. Topics that are of particular interest to me include: information architecture, user experience, data analytics, and asset management.

Marisa Purcell

(310) 944 - 0475
mfpurcell@ucla.edu
480 Mission Bay Blvd N, 1008
San Francisco, CA 94158

OBJECTIVE

Seeking a position in user experience to develop creative solutions through research and user-driven innovation.

ACADEMIC EXPERIENCE

UCLA Library User Experience Research Project (Mar 2020 – Present).
Employing qualitative and quantitative research methods to test a digital inventory tracking tool that manages the digital and hardware assets of the library and interfaces with the library's IT department.

UCLA Library's Digital Initiatives and Information Technology 3D Model Preservation Plan (Jan– Mar 2020). Produced preservation and asset management strategies for digitized 3D models of 4,000-year-old Mesopotamian tablets housed in UCLA's Library Special Collections**.**

FBI Hate Crime Data Website: https://jcboomatucla.github.io/ (Sep – Dec 2019).
Gathered, structured, and analyzed FBI hate crime data from 1996 – 2018, created data visualizations, and built a website using HTML and CSS to present the findings.

Data and Digital Asset Management for Ancient Egyptian Coffin Research (Jan – Jun 2019).
Created a metadata schema, data dictionary and taxonomy, and presented viable storage options and website interfaces for Dr. Kara Cooney, Chair, UCLA Department of Near Eastern Languages and Cultures.

Software Application to Promote Ethical Surveillance for Smart Assistants (Jan – Jun 2019).
Utilized user experience research and created wireframes, mockups, and user personas to create an application to aid in ethical surveillance practices for users of smart assistants..

Virtual Reality Policy Brief for the UCLA Information Technology Planning Board (Mar 2019).
Researched virtual reality hardware and software and authored a report forecasting the benefits and detriments of integrating virtual reality systems on the UCLA campus.

Photograph Collection and User Manual for Los Angeles Murals using CONTENTdm (Sep—Dec 2018).
Built a digital collection of murals by cataloging photos and adding metadata, including a controlled vocabulary. Co-authored a handbook for users to access and edit these records.

EMPLOYMENT

Metadata Intern, TMZ, Playa Vista, CA (Sep – Dec 2019). Supported digital asset management by managing, researching, and creating metadata and taxonomies for video, audio, and photographs. Used controlled vocabularies, schema governance, and data integrity standards to structure assets.

Information Policy Intern, IRIS.TV, Los Angeles, CA (Jun – Aug 2019). Created internal and client-facing whitepapers to inform data collection practice and compliance regarding data privacy.

NCAA Division I Assistant Women's Water Polo Coach, Marist College, Poughkeepsie, NY (Dec 2017 – Aug 2018). Managed college students in their academic and athletic development through training, game strategy, and administrative duties.

Law Clerk, District Attorney's Office – Organized Crime Division, Los Angeles, CA (Jun – Nov 2015).
Drafted legal documents, organized case files, and gathered evidence to prepare the team for trial.

Researcher, Positive Coaching Alliance, San Jose, CA (Nov 2014 – Feb 2015). Created and managed a database of 9 different youth sports in 11 counties to support outreach.

Research Assistant, VA Greater Los Angeles Healthcare System, Los Angeles, CA (2013 – 2014).
Assisted with data management, data entry and cleaning, summary statistics, and administrative support.

EDUCATION

UCLA
Dept of Information Studies
Master's Library & Information Science, focus Informatics
Cumulative GPA, 3.96
MLIS expected June 2020

Willamette University
College of Law
JD Program
2016 – 2017

UC Berkeley
College of Letters & Science
BA Rhetoric
May 2015

RELEVANT COURSEWORK

UX Research and Design
Description and Access
Management of Digital Records
Data Management and Practice
Systems and Infrastructures
Human Computer Interaction
Economics of Information
Data Curation and Policy
Digital Humanities

COMPUTER SKILLS

Microsoft Office, Confluence, Slack, SQL, Zotero, Dalet Management System, REACH Management System, CONENTdm, OpenRefine, Air Table, MALLET Topic Modeling Tool Cytoscape, RAWGraphs, Tableau, Google Data Studio, Sketch, Google Analytics, CSS, HTML, GitHub, Bootstrap

PROFESSIONAL DEVELOPMENT STATEMENT

My career goal is to work in user experience (UX). I hope to contribute to an organization's strategic decision-making by helping to make products, services, and systems optimally functional for the target user, such as a client visiting an organization's website or an employee accessing a software application or database in the organization's information pipeline. In the short term, I am looking for a position as a UX researcher. The ideal position would involve planning, conducting and evaluating qualitative and quantitative UX research. My longer-term goal is to have a leadership role in an organization working closely with project managers, designers, and engineers to steer product design and business strategies that promote the mission and vision of the organization and build better products for the target user. To achieve this longer-term goal, I plan to pursue studies in business to better understand how I can bring together product management and customer trust to drive revenue through customer engagement, and studies in anthropologic, ethnographic research methods and in human computer interaction to advance my understanding of technology's impact on human beings through design and behavioral studies.

Courses that have been instrumental in the development of my career goals include human computer interaction, digital humanities, data management and practice, and UX design. The human computer interaction course introduced me integral components of UX, including the study of visual design, human behavior, and qualitative research methods such as, wireframing, creating user personas, and literature review. In my digital humanities course, I performed quantitative research with statistical analysis of FBI hate crime data, created data visualizations, and designed a website with a classmate. In a data management and practice course, I collaborated with a classmate and interviewed Dr. Kara Cooney, Chair, UCLA Department of Near Eastern Languages and Cultures and her research team to understand her short and long term needs to preserve her research data. I gathered information surrounding the management of her assets, presented viable storage options, and recommended website interfaces to share her research with others. Lastly, through my course in UX design I applied UX methods to every stage of the UX process, and worked with target users to provide UX-centered recommendations that benefit both target users and the organization. To advance my knowledge and experience in UX, I spent my last quarter at UCLA leading a research project with the UCLA library. A classmate and I used both qualitative and quantitative methods to develop a digital inventory tracking tool. The tool will manage the digital and hardware assets of the library and interface with the library's IT department.

My internship experience at the media company, TMZ and the technology company, IRIS.TV piqued my interest in working to inform business decisions and support target user needs. In both internships, I learned how information flows through an organization touching every employee, and how that information impacts the production and mission of the organization. As an intern at TMZ, I supported the Metadata Department in the digital asset management (DAM) process by managing, researching, and creating metadata and taxonomies for digital video and photograph assets in various content management systems. Understanding how the production, editing, or legal team would access these assets was important in describing, tagging, and cataloging TMZ's photograph and video assets. In my internship at IRIS.TV, I authored internal and client-facing whitepapers documenting the company's compliance of data privacy regulations including California Consumer Privacy Act, General Data Protection Regulation, ePrivacy Regulation, Video Privacy Protection Act, and Children's Online Privacy Protection Rule. These whitepapers described the company's data collection process and were created to inform clients as to how IRIS.TV has designed their services and product around these regulations to protect clients' data.

My experience with the UCLA chapter of the Association for Information Science & Technology (ASIS&T) helped shape my interest in UX. I participated on the board of the student chapter serving first as Vice

President and more recently as President of the board. The group shares ideas and we collaborate with like-minded students interested in UX. My participation in ASIS&T was motivated by the need to bring technology and information tools to students in the UCLA Master of Library and Information Science (MLIS) program. This included coordinating a meet-the-professor session with Dr. Leonard Kleinrock, the UCLA professor considered to be the pioneer of what we know today as the Internet. I also arranged and organized an introduction to coding workshop in which MLIS students learned basic coding skills to build their own website. I plan to continue involvement in ASIS&T and become a member of the UX Professionals Association (UXPA) to connect with and learn from individuals in the user experience field.

My matriculation in the UCLA Information Studies Department has been invaluable to my personal and professional development. The skills and knowledge I have gained during my graduate education have prepared me well for a user experience role.

COURSES

FALL 2018 COURSES

INF STD 211: Artifacts & Cultures

INF STD 240: Management of Digital Records

INF STD 260: Description & Access

WINTER 2019 COURSES

INF STD 262A: Data Management and Practice

INF STD 270: Systems and Infrastructures

INF STD 272: Human/Computer Interaction (HCI)

SPRING 2019 COURSES

INF STD 206: Intro to Economics of Information

INF STD 212: Values and Communities in Information Professions

INF STD 262B: Data Curation and Policy

FALL 2019 COURSES

DH 201: Introduction to Digital Humanities

INF STD 289: Special Issues in Information Studies: Data Informatics

INF STD 498: Internship

WINTER 2020 COURSES

INF STD 241: Digital Preservation

INF STD 400: Portfolio and Professional Development

INF STD: Historical Methodology of Information Studies

SPRING 2020 COURSES

INF STD 279: User Experience Design

INF STD 288: Research Apprenticeship Course

INF STD 289: Special Issues in Information Studies: Digital Asset Management

COURSEWORK

MAJOR PAPER

IS 206: Introduction to Economics of Information - Final Paper

CORE PAPER

IS 212: Values and Communities - Final Paper

ELECTIVE WORK

IS 262B: Data Curation and Policy - Data Management Plan

Marisa Purcell

June 13, 2019

IS 206: Introduction to Economics of Information

Final Brief: The Benefits and Detriments of Recommendation Systems

I. *A Breakdown of the Recommendation System*

A recommendation system is a software tool typically implemented by companies to suggest content or goods to users searching the web.[1] The suggestions recommendation systems generate are also called "items".[2] Recommendation systems utilize Artificial Intelligence (AI) by layering AI onto a company's website, allowing the system to collect and analyze user data and/or item data and then recommend additional content and products to users.[3] The user data that is analyzed to generate recommendations can include a user's "past purchases, demographic info, or their search history."[4] This data is turned into two types of information: char*acteristic information* and *user-item interactions*.[5] Characteristic information is generated by the user, such as searched key words, while user-item interaction information is drawn from user behavior, such as item ratings, number of purchases, and likes.[6] Recommendation systems are typically effective because they give users personalized suggestions through the collection and analysis of their corresponding data.[7]

---

[1] Francesco Ricci, Lior Rokach, and Bracha Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, ed. Francesco Ricci et al. (Boston, MA: Springer US, 2011), 1–35, https://doi.org/10.1007/978-0-387-85820-3_1.
[2] Ibid.
[3] "Recommendation Systems - How Companies Are Making Money," *Sigmoidal* (blog), September 27, 2017, https://sigmoidal.io/recommender-systems-recommendation-engine/.
[4] Ibid.
[5] "Introduction to Recommender Systems in 2019 | Tryolabs Blog," accessed June 8, 2019, https://tryolabs.com/blog/introduction-to-recommender-systems/.
[6] Ibid.
[7] Francesco Ricci, Lior Rokach, and Bracha Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, ed. Francesco Ricci et al. (Boston, MA: Springer US, 2011), 1–35, https://doi.org/10.1007/978-0-387-85820-3_1.

II. Types of Recommendation System

There are several types of recommendation systems that collect characteristic information, user-item interaction data, or combination of both types of data. The two main types of recommendation systems currently in use are *collaborative filtering* and *content-based filtering*. While these systems are the most prevalent, an organization with a custom approach in recommending items to users may consider employing other recommendation systems. [8]

a. Collaborative Filtering Recommender System

Collaborative filtering recommends items based on users' past behaviors by utilizing two types of measurements.[9] The first is a user-based measurement: collecting and assessing the similarities between target users and other users.[10] The second is an item-based measurement: collecting and assessing the similarities between the items that target users interact with and rate.[11] Collaborative filtering assumes that users will both have similar interests and that their past and present preferences will remain in place for the future.[12] The issues with this system are its scalability; a site with more visitors is likely to not provide as accurate or personalized recommendations for a single user because of the need to categorize its many users, and the

[8] Stephanie Blanda, "Online Recommender Systems – How Does a Website Know What I Want?," *AMS Grad Blog* (blog), May 26, 2015, https://blogs.ams.org/mathgradblog/2015/05/25/online-recommender-systems-website-want/.
[9] "Introduction to Recommender System. Part 1 (Collaborative Filtering, Singular Value Decomposition)," accessed June 9, 2019, https://hackernoon.com/introduction-to-recommender-system-part-1-collaborative-filtering-singular-value-decomposition-44c9659c5e75.
[10] Ibid.
[11] Ibid.
[12] "Introduction to Recommender System. Part 1 (Collaborative Filtering, Singular Value Decomposition)," accessed June 9, 2019, https://hackernoon.com/introduction-to-recommender-system-part-1-collaborative-filtering-singular-value-decomposition-44c9659c5e75 and "Types of Recommender Systems | Machine Learning | Bluepi Blogs," *BluePi Blogging* (blog), November 14, 2015, https://www.bluepiit.com/blog/classifying-recommender-systems/.

assumption that human behaviors will not change minimizes the complexity of human behavior.[13]

b. Content-Based Filtering Recommender System

Content-based recommender systems are an extension of collaborative-filtering.[14] These systems analyze and learn a new user's interests based on discrete characteristics of the items the user selects; the item characteristics can include name, location, or description, in order to recommend additional items with similar properties.[15] Content-based recommender systems create a user profile " to provide information about the types of items that the user likes based on keywords used to describe the items."[16] An example of this system is Pandora Radio, which prompts users to enter an artist's name or genre to create a station based on the keyword they input into Pandora's search.[17] Though companies like Pandora have seen success with these systems, the issue with content-based filtering is its lack of dimensionality. These systems are unable to take one type of user preference and recommend a different type of item based on that preference; for example, if a system has profiled the user's music preference, that system cannot recommend books to that user, thus limiting the scope of the system's use.[18]

---

[13] "Introduction to Recommender System. Part 1 (Collaborative Filtering, Singular Value Decomposition)," accessed June 9, 2019, https://hackernoon.com/introduction-to-recommender-system-part-1-collaborative-filtering-singular-value-decomposition-44c9659c5e75

[14] "Types of Recommender Systems | Machine Learning | Bluepi Blogs," *BluePi Blogging* (blog), November 14, 2015, https://www.bluepiit.com/blog/classifying-recommender-systems/.

[15] Stephanie Bl and a, "Online Recommender Systems – How Does a Website Know What I Want?," *AMS Grad Blog* (blog), May 26, 2015, https://blogs.ams.org/mathgradblog/2015/05/25/online-recommender-systems-website-want/.

[16] Ibid.

[17] Ibid.

[18] Ibid.

c. Demographic-Based Recommender System

The demographic-based system categorizes users into demographic classes.[19] To do so, companies conduct market research and use collaborative-based system relationship data, such as peer recommendations, to understand what items correlate to which particular demographic.[20]

d. Utility-Based Recommender System

Utility-based recommender systems use the multi-attribute utility theory to generate recommendations for users.[21] The utility theory is used to numerically scale preferences of users and "is a systematic approach for quantifying an individual's preferences."[22] In these systems, recommendations are based on the level of utility each item has to the user.[23] Cost is considered a utility attribute but each item can be associated with several utility attributes, such as "location information, available connectivity, performance and reliability requirements, and contractual aspects and costs."[24] With all of these attributes to consider, as well as how each attribute may be valued differently by different users, it is difficult to recommend an item that will be the best option for each user.[25]

---

[19] "Types of Recommender Systems | Machine Learning | Bluepi Blogs," *BluePi Blogging* (blog), November 14, 2015, https://www.bluepiit.com/blog/classifying-recommender-systems/.
[20] Ibid.
[21] Alexander Felfernig et al., "An Overview of Recommender Systems in the Internet of Things," *Journal of Intelligent Information Systems* 52, no. 2 (April 1, 2019): 285–309, https://doi.org/10.1007/s10844-018-0530-7.
[22] "HSOR.Org: What Is OR," accessed June 9, 2019, http://www.hsor.org/what_is_or.cfm?name=mutli-attribute_utility_theory.
[23] "Types of Recommender Systems | Machine Learning | Bluepi Blogs," *BluePi Blogging* (blog), November 14, 2015, https://www.bluepiit.com/blog/classifying-recommender-systems/.
[24] Francesco Ricci, Lior Rokach, and Bracha Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, ed. Francesco Ricci et al. (Boston, MA: Springer US, 2011), 1–35, https://doi.org/10.1007/978-0-387-85820-3_1.
[25] "Types of Recommender Systems | Machine Learning | Bluepi Blogs," *BluePi Blogging* (blog), November 14, 2015, https://www.bluepiit.com/blog/classifying-recommender-systems/.

e. *Knowledge-Based Recommender System*

Knowledge-based systems require greater interaction and agency from users to support decision making.[26] These systems incorporate machine learning to generate more accurate recommendations and improve performance based on experience and interactions.[27] The healthcare industry has utilized knowledge-based systems to "help healthcare providers make decisions and improve patient care… [by examining] a patient's medical history in conjunction with relevant clinical research. Such analysis can then help predict potential events, such as drug interactions, or flag disease symptoms."[28]

f. *Hybrid Recommender System*

Hybrid systems are the combination of two recommendation systems to best suit a particular industry.[29] This is the most sought-after system because it combines the strengths of multiple systems while compensating for the blind spots of a single system.[30] There are three types of hybrid systems:

i. *Weighted Hybrid Recommender*

Weighted hybrid systems give equal consideration of both collaborative-based and content-based filtering systems in the beginning stages of implementation.[31] Gradually, this equally weighted system adjusts to utilize the collaborative or content-based system to conform to the predictions of confirmed and unconfirmed user ratings.[32]

---

[26] "What Is Knowledge-Based Systems (KBS)? - Definition from WhatIs.Com," SearchCIO, accessed June 9, 2019, https://searchcio.techtarget.com/definition/knowledge-based-systems-KBS.
[27] "What Is Expert System? - Definition from WhatIs.Com," SearchEnterpriseAI, accessed June 9, 2019, https://searchenterpriseai.techtarget.com/definition/expert-system.
[28] "What Is Clinical Decision Support System (CDSS)? - Definition from WhatIs.Com," SearchHealthIT, accessed June 9, 2019, https://searchhealthit.techtarget.com/definition/clinical-decision-support-system-CDSS.
[29] "Types of Recommender Systems | Machine Learning | Bluepi Blogs," *BluePi Blogging* (blog), November 14, 2015, https://www.bluepiit.com/blog/classifying-recommender-systems/.
[30] Ibid.
[31] Ibid.
[32] Ibid.

ii. <u>Switching Hybrid Recommender</u>

A switching hybrid system functions by deploying either a content-based system or collaborative-based system and then continually switching between systems. When one system is being utilized, the recommender engages the switch to the system that is not being utilized.[33] *The system will then switch recommenders and deploy the previously used recommender if one system is not effectively and accurately making recommendations.*[34]

iii. <u>Mixed Hybrid Recommender</u>

Mixed hybrid systems are effective when making a large number of recommendations simultaneously.[35] *The mixed system can introduce more than one technique from each system into a single system to offer a wide range of recommendations.*[36]

III. <u>The Economically Beneficial but Psychologically Damaging Nature of Recommendation Systems</u>

American adults, on average, spend "six hours and 43 minutes a day looking at a screen, or 7,956 days of their life."[37] Some of the highest trafficked sites worldwide are Amazon, YouTube, and Facebook, all of which utilize recommendation systems.[38] Recommendation systems offer website visitors a personalized experience, encouraging vistors to return to the site or remain on the site for a longer period.[39] Keeping users on a site increases the likelihood they will make a purchase, absorb content, and view advertisements; reasons companies want airtime on these

---

[33] Ibid.
[34] Ibid.
[35] Ibid.
[36] Ibid.
[37] "Screen Time Now Takes up Half Our Waking Hours," accessed June 10, 2019, https://nypost.com/2018/08/13/americans-spend-half-their-lives-in-front-of-screens/.
[38] Jerri Collins, "The Top 10 Most Popular Sites of 2019," Lifewire, accessed June 10, 2019, https://www.lifewire.com/most-popular-sites-3483140.
[39] "Recommender Systems | Their Impact on Customer Experience," accessed June 10, 2019, https://theappsolutions.com/blog/development/recommender-systems-guide/.

sites. The model of implementing recommendation systems on corporate websites is profitable for several companies. Amazon's revenue, for example, greatly increased when recommendation systems were implemented on their site and via email: "[Amazon] reported a 29% sales increase to $12.83 billion during its second fiscal quarter, up from $9.9 billion during the same time last year. A significant portion of Amazon's growth can be attributed to Amazon's integration of recommendations into nearly every part of the purchasing process."[40] Netflix had similar success with recommendation systems with, "up to 75% of what consumers watch on Netflix… [generated] from the company's recommender system."[41]

Recommendation systems have proven to be economically beneficially for companies, but their impact on consumers has not been as positive. In order to retain a larger proportion of users, companies have resorted to implementing recommendation systems to hook users.[42] The almost addictive nature of these sites has led to a host of behavioral and psychological conditions, as well as a culture of screen dependence wherein over half of millennials feel anxious and irritated when they cannot check their phone.[43] The addictive nature and the lack of agency perpetuated by recommendation systems, should incline users to question a company's motives while they consume content and browse products on the web.

---

[40] "Amazon's Recommendation Secret," Fortune, accessed June 13, 2019, http://fortune.com/2012/07/30/amazons-recommendation-secret/.
[41] "Recommendation Systems - How Companies Are Making Money," *Sigmoidal* (blog), September 27, 2017, https://sigmoidal.io/recommender-systems-recommendation-engine/.
[42]"Why The Brands We All Love Use Online Recommendation Systems," Zeta Global, September 22, 2016, https://zetaglobal.com/blog-posts/online-recommendation-systems-personalization/
[43] "Screen Time Now Takes up Half Our Waking Hours," accessed June 10, 2019, https://nypost.com/2018/08/13/americans-spend-half-their-lives-in-front-of-screens/.

IV. The Effects of the Long Tail Economic Model

Recommendation systems allow companies to take advantage of the long-tail economic model to satisfy and reach more users. The long tail theory states that less popular goods with a lower demand can still increase a company's profitability:

> With no shelf space to pay for and, in the case of purely digital services like iTunes, no manufacturing costs and hardly any distribution fees, a miss sold is just another sale, with the same margins as a hit. A hit and a miss are on equal economic footing, both just entries in a database called up on demand, both equally worthy of being carried. Suddenly, popularity no longer has a monopoly on profitability.[44]

Not only do large companies like Amazon generate higher revenues, but the user benefits from greater exposure to items, products, or services of which they may have been unaware of without recommendation systems.[45] Companies that ascribe to the long tail theory are giving small businesses a platform to which they previously did not have access; Amazon, for instance, has reportedly "helped more than 1.9 million U.S.-based small and medium-sized businesses (SBMs) generate more than $160 billion in 2018."[46] The long tail theory can also prove beneficial for companies that do not sell products, but instead offer free services, like Google. In this information economy, knowledge has become a commodity; as users seek out items outside of the mainstream, this fringe content distinguishes them from other consumers and thereby makes them easier to target. These fringe choices allow companies to create more specific user profiles, ultimately allowing for more effective targeting of users.

---

[44] Chris Anderson, "The Long Tail," *Wired*, October 1, 2004, https://www.wired.com/2004/10/tail/.
[45] https://dataconomy.com/2015/03/an-introduction-to-recommendation-engines/
[46] "Amazon Says Small Business Owners Make $90,000 a Year from Selling in Its Stores," USA TODAY, accessed June 10, 2019, https://www.usatoday.com/story/tech/2019/05/07/small-businesses-selling-amazon-stores-create-1-6-m-jobs-report/1120026001/.

While the long tail model provides benefits to companies, users, and the marketplace, the personal information that users are giving to fuel recommendation systems is arguably not worth the benefits. Google has proven the value of collecting and selling user data at the expense of user privacy: "Upon the acquisition of user data as the raw material for proprietary analyses and algorithm production [Google found] that [it] could sell and target advertising through a unique auction model with ever more precision and success. As Google's revenues rapidly grew, they motivated ever more comprehensive data collection"[47] Google's model has acquired the critical moniker of surveillance capitalism in reference to the monitoring of websites to collect data focused on market, social, physical, and biological behaviors.[48] Though companies like Google have commodified users' privacy to avoid the fees-for-service business model, they have put profit ahead of user privacy by selling data to third-party companies and taking advantage of the unregulated landscape.[49] Facebook similarly released user information, losing the trust of many users: "Mark Zuckerberg, who's expected to act as the trusted custodian of the personal information of more than 2 billion people, has allowed his company's partners — Netflix, Amazon and Spotify, among many others — access to users' most intimate communications."[50] Americans have trusted these companies with their personal information, in part because privacy is included in the Constitution and, as such, considered a hallmark of the country society. Companies like Google and Facebook have taken advantage of this trust by consistently prioritizing profit.

---

[47] Shoshana Zuboff, "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization," *Journal of Information Technology* 30, no. 1 (March 2015): 75–89, https://doi.org/10.1057/jit.2015.5.
[48] Ibid.
[49] "Google Facing Legal Action in EVERY EU Country over 'data Goldmine' Collected about Users | Daily Mail Online," accessed June 10, 2019, https://www.dailymail.co.uk/sciencetech/article-2302870/Google-facing-legal-action-EVERY-EU-country-data-goldmine-collected-users.html.
[50] "How Tracking And Selling Our Data Became A Business Model | On Point," accessed June 10, 2019, https://www.wbur.org/onpoint/2019/01/15/surveillance-capitalism-shoshana-zuboff-facebook-data.

V. The Influence of the Network Effect over Recommendation Systems

Many recommendation systems, such as demographic-based or collaborative filtering systems, use aggregate data to make recommendations and rely heavily on the Network Effect to gather that data. The Network Effect "causes a good or service to have a value to a potential customer which depends on the number of other customers who own the good or are users of the service."[51] Companies like Facebook have perfected the art of utilizing the Network Effect; as more users join Facebook, the more useful it becomes.[52] This culminates in a worldwide network that allows users to exchange and share information, Facebook can collect large amounts of personal information from users, create user profiles based on this information, and recommend new content that aligns with a user's interests. While users may benefit from the scale of this network, both in the quality and nature of the fringe recommendations, Facebook "makes plenty of money — millions daily — by selling access to users' data to advertisers."[53]

Social media platforms and companies with models similar to Facebook that recommending information via the Network Effect can wreak havoc with the spread of misinformation. The lack of information regulation, also referred to as disinformation or fake news, has promulgated conspiracy theories, spurred mass shootings, and advanced election tampering. The untraceable spread of information is problematic because"[d]etecting what's fake in images and video is only getting harder. Misinformation is part of an online economy that weaponizes social media to profit from our clicks and attention."[54] Companies like Facebook are trying to create solutions to

---

[51] Robert M. Metcalfe, "It's All In Your Head," Forbes, accessed June 11, 2019, forbes/2007/0507/052.
[52] A. J. Chavar, "Why You Keep Using Facebook, Even If You Hate It," Vox, April 11, 2018, https://www.vox.com/videos/2018/4/11/17226430/facebook-network-effect-video-explainer.
[53] Ibid.
[54] "I Fell for Facebook Fake News. Here's Why Millions of You Did, Too. - The Washington Post," accessed June 11, 2019, https://www.washingtonpost.com/technology/2018/10/18/i-fell-facebook-fake-news-heres-why-millions-you-did-too/?noredirect=on&utm_term=.285f61272a2b.

battle misinformation, even going as far as considering crowdsourcing users as fact checkers.[55] While a solution may not be clear, it is evident that recommendation systems are increasing the misinformation epidemic.

On December 1, 2016 Edgar Welch walked into a Washington, D.C. pizza shop with an AR-15 semiautomatic rifle, a .38 handgun, and a folding knife because he saw an Info-Wars report stating Hillary Clinton was sexually abusing children in the restaurant's basement.[56] The Info-Wars report was popularized by users and spread to users on Facebook's network through the recommendation systems. The recommendation system is meant to spread popular content that will generate clicks. The proliferation of this story gave it power, and without the ability to filter for misinformation through the recommendation system, Facebook's recommendation system played a role on Welch's actions on December 1st.[57] Though this may be considered an uncommon example of the effect of misinformation, company's like Facebook have received a great deal of backlash for ads and posts that have shown up on its site, including the implication that Facebook did not deter foreign meddling in the U.S. presidential election. Facebook is not alone; Amazon recommends products to users every time they enter the site; each product contains reviews allowing users to ask questions and garner a better understanding of the item. Amazon has even allowed reviews to be incentivized, which leads to bias through fake reviews, until this process was banned on the site in 2016.[58] To get around this ban, "[f]ake reviews… often appear as "verified purchases", just like real reviews, with no indication of a connection

---

[55] Sam Levin and Julia Carrie Wong, "'He's Learned Nothing': Zuckerberg Floats Crowdsourcing Facebook Fact-Checks," *The Guardian*, February 20, 2019, sec. Technology, https://www.theguardian.com/technology/2019/feb/20/facebook-fact-checking-crowdsourced-mark-zuckerberg.
[56] "Pizzagate: Anatomy of a Fake News Scandal – Rolling Stone," accessed June 11, 2019, https://www.rollingstone.com/politics/politics-news/anatomy-of-a-fake-news-scandal-125877/.
[57] Ibid.
[58] Chris McCabe, "Amazon's Fake Review Problem Is Worse Than Ever. Here's Why.," *Web Retailer* (blog), June 15, 2018, https://www.webretailer.com/lean-commerce/amazon-fake-reviews/.

between the buyer and the seller… In reality, the purchase was funded by the seller using PayPal, an Amazon gift voucher, or other means."[59] The spread of misinformation through the use of recommendation systems in a network system is an incentive for users to distrust the information on those platforms. Until platforms that utilize the Network Effect through the use of recommendation systems can find a way to verify information sources, companies should be aware of the societal implications recommendation systems can have through the spread of misinformation.

VI. Conclusion

The benefits of recommendation systems must be weighed against the concerns for both individual privacy and widespread misinformation in an era of unlimited access to information. To strike an equitable balance, companies and users must broker an open and honest dialogue. Currently, companies such as Facebook are taking advantage of users' trust because, "Facebook knows that as long as your 2 billion friends are online, you're probably not going anywhere."[60] Recommendation systems are the driving force behind this dynamic, as they bridge the gap between users and companies through the collection of their users' personal data. Change can only come through regulation. Companies have little motivation to change as they continue to maximize profits using recommendation systems. Presently, antitrust laws are not established to regulate companies that are monopolizing the information economy, to establish oversight, legislators must prioritize regulation for companies using the recommendation system economic

[59] Ibid.
[60] A. J. Chavar, "Why You Keep Using Facebook, Even If You Hate It," Vox, April 11, 2018, https://www.vox.com/videos/2018/4/11/17226430/facebook-network-effect-video-explainer.

model.[61] Policy changes must be executed to protect users and prevent regulatory bodies from falling even further behind in the information economy.

---

[61] "Yale Law Journal - Amazon's Antitrust Paradox," accessed June 11, 2019, https://www.yalelawjournal.org/note/amazons-antitrust-paradox.

Work Cited

"Amazon Says Small Business Owners Make $90,000 a Year from Selling in Its Stores." USA TODAY. Accessed June 10, 2019. https://www.usatoday.com/story/tech/2019/05/07/small-businesses-selling-amazon-stores-create-1-6-m-jobs-report/1120026001/.

"Amazon's Recommendation Secret." Fortune. Accessed June 10, 2019. http://fortune.com/2012/07/30/amazons-recommendation-secret/.

Anderson, Chris. "The Long Tail." *Wired*, October 1, 2004. https://www.wired.com/2004/10/tail/.

Bianda, Stephanie. "Online Recommender Systems – How Does a Website Know What I Want?" *AMS Grad Blog* (blog), May 26, 2015. https://blogs.ams.org/mathgradblog/2015/05/25/online-recommender-systems-website-want/.

Chavar, A. J. "Why You Keep Using Facebook, Even If You Hate It." Vox, April 11, 2018. https://www.vox.com/videos/2018/4/11/17226430/facebook-network-effect-video-explainer.

Collins, Jerri. "The Top 10 Most Popular Sites of 2019." Lifewire. Accessed June 10, 2019. https://www.lifewire.com/most-popular-sites-3483140.

Felfernig, Alexander, Seda Polat-Erdeniz, Christoph Uran, Stefan Reiterer, Muesluem Atas, Thi Ngoc Trang Tran, Paolo Azzoni, Csaba Kiraly, and Koustabh Dolui. "An Overview of Recommender Systems in the Internet of Things." *Journal of Intelligent Information Systems* 52, no. 2 (April 1, 2019): 285–309. https://doi.org/10.1007/s10844-018-0530-7.

"Google Facing Legal Action in EVERY EU Country over 'data Goldmine' Collected about Users | Daily Mail Online." Accessed June 10, 2019. https://www.dailymail.co.uk/sciencetech/article-2302870/Google-facing-legal-action-EVERY-EU-country-data-goldmine-collected-users.html.

"How Tracking And Selling Our Data Became A Business Model | On Point." Accessed June 10, 2019. https://www.wbur.org/onpoint/2019/01/15/surveillance-capitalism-shoshana-zuboff-facebook-data.

"HSOR.Org: What Is OR." Accessed June 9, 2019. http://www.hsor.org/what_is_or.cfm?name=mutli-attribute_utility_theory.

"I Fell for Facebook Fake News. Here's Why Millions of You Did, Too. - The Washington Post." Accessed June 11, 2019. https://www.washingtonpost.com/technology/2018/10/18/i-fell-facebook-fake-news-heres-why-millions-you-did-too/?noredirect=on&utm_term=.285f61272a2b.

"Introduction to Recommender System. Part 1 (Collaborative Filtering, Singular Value Decomposition)." Accessed June 9, 2019. https://hackernoon.com/introduction-to-recommender-system-part-1-collaborative-filtering-singular-value-decomposition-44c9659c5e75.

"Introduction to Recommender Systems in 2019 | Tryolabs Blog." Accessed June 8, 2019. https://tryolabs.com/blog/introduction-to-recommender-systems/.

Levin, Sam, and Julia Carrie Wong. "'He's Learned Nothing': Zuckerberg Floats Crowdsourcing Facebook Fact-Checks." *The Guardian*, February 20, 2019, sec. Technology. https://www.theguardian.com/technology/2019/feb/20/facebook-fact-checking-crowdsourced-mark-zuckerberg.

McCabe, Chris. "Amazon's Fake Review Problem Is Worse Than Ever. Here's Why." *Web Retailer* (blog), June 15, 2018. https://www.webretailer.com/lean-commerce/amazon-fake-reviews/.

Metcalfe, Robert M. "It's All In Your Head." Forbes. Accessed June 11, 2019. forbes/2007/0507/052.

"Why The Brands We All Love Use Online Recommendation Systems." Zeta Global, September 22, 2016. https://zetaglobal.com/blog-posts/online-recommendation-systems-personalization/.

"Pizzagate: Anatomy of a Fake News Scandal – Rolling Stone." Accessed June 11, 2019. https://www.rollingstone.com/politics/politics-news/anatomy-of-a-fake-news-scandal-125877/.

"Recommendation Systems - How Companies Are Making Money." *Sigmoidal* (blog), September 27, 2017. https://sigmoidal.io/recommender-systems-recommendation-engine/.

"Recommender Systems | Their Impact on Customer Experience." Accessed June 10, 2019. https://theappsolutions.com/blog/development/recommender-systems-guide/.

Ricci, Francesco, Lior Rokach, and Bracha Shapira. "Introduction to Recommender Systems Handbook." In *Recommender Systems Handbook*, edited by Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, 1–35. Boston, MA: Springer US, 2011. https://doi.org/10.1007/978-0-387-85820-3_1.

Schmall, Tyler, and SWNS. "Americans Spend Half Their Lives in Front of Screens." *New York Post* (blog), August 13, 2018. https://nypost.com/2018/08/13/americans-spend-half-their-lives-in-front-of-screens/.

"Screen Time Now Takes up Half Our Waking Hours." Accessed June 10, 2019. https://nypost.com/2018/08/13/americans-spend-half-their-lives-in-front-of-screens/.

"Types of Recommender Systems | Machine Learning | Bluepi Blogs." *BluePi Blogging* (blog), November 14, 2015. https://www.bluepiit.com/blog/classifying-recommender-systems/.

“What Is Clinical Decision Support System (CDSS)? - Definition from WhatIs.Com.” SearchHealthIT. Accessed June 9, 2019. https://searchhealthit.techtarget.com/definition/clinical-decision-support-system-CDSS.

“What Is Expert System? - Definition from WhatIs.Com.” SearchEnterpriseAI. Accessed June 9, 2019. https://searchenterpriseai.techtarget.com/definition/expert-system.

“What Is Knowledge-Based Systems (KBS)? - Definition from WhatIs.Com.” SearchCIO. Accessed June 9, 2019. https://searchcio.techtarget.com/definition/knowledge-based-systems-KBS.

“Yale Law Journal - Amazon’s Antitrust Paradox.” Accessed June 11, 2019. https://www.yalelawjournal.org/note/amazons-antitrust-paradox.

Zuboff, Shoshana. “Big Other: Surveillance Capitalism and the Prospects of an Information Civilization.” *Journal of Information Technology* 30, no. 1 (March 2015): 75–89. https://doi.org/10.1057/jit.2015.5.

Marisa Purcell

June 12, 2019

IS 212: Values and Communities

Making America Unsafe Again: The Use of Discourse to Perpetuate Systemic Racism through Facial Recognition

I. *Introduction*

The current political climate of the United States is fraught with fearmongering and hostile discourse, leading to the further marginalization of minorities in communities across the country. This discourse is reflective of more than words, seeping into the constructs of our society and expressing itself as racial biases. Technology company's utilization of discourse is a primary example of discourse as a tool to perpetuate racism and socio-economic inequities. Facial recognition technology is not exempt from these biases, this is of particular concern because of the technology's rising popularity. Culturally impactful institutions and government agencies alike have supported the use of facial recognition systems. These organizations are using a discourse of safety to justify the use of facial recognition while the technology perpetuates racial hierarchies and biases, endangering minorities across the United States that are underrepresented in the technology industry.

II. *The Innerworkings of Facial Recognition and its Creators*

a. *How These Systems Work*

Facial recognition systems are used to identify individuals using computer vision and algorithms, making them a threat to privacy for many Americans.[62] These systems use software to "scan an image or live video for a person's face and then match it with a similar, previously

[62]"What Is Facial Recognition Technology, and Why Is It so Controversial?," accessed June 10, 2019, https://finance.yahoo.com/news/facial-recognition-amazon-114720161.html.

taken image or video of that same person."[63] Machine learning is used to teach facial recognition systems how to identify and scan an individual's face. Company's rely on large photograph databases to train artificial intelligence during machine learning, "[t]o find a single person… an operator uploads a photo of whoever they are trying to identify, the computer then looks at the person's facial landmarks, such as the distance between their eyes… and compares that against the other images in its stockpile."[64] Facial recognition is rapidly improving and spreading, and its success is attributed to new databases with vast amounts of records containing images of faces from Americans across the country.[65]

b. The History of Facial Recognition

The U.S. government developed facial recognition in the 1960s, steadily working on its accuracy until introducing the technology to the commercial market in 1993.[66] After emerging from behind the government's closed doors, facial recognition has been used by law enforcement agencies at the Super Bowl, to unlock iPhones at Apple, to tag photos on Facebook, and for military missions like identify Osama Bin Laden.[67] The government has played a large role in the development of facial recognition, but American companies have begun creating their own facial recognition software and databases, such as Amazon's Rekognition API, which allows individuals to purchase facial recognition technology for their business.[68] With the tech industry

---

[63] Ibid.
[64] Ibid.
[65] "Analysis | Facial Recognition," Washington Post, accessed June 10, 2019, https://www.bloomberg.com/view/quicktake/facial-recognition.
[66] "History of Face Recognition & Facial Recognition Software," *FaceFirst Face Recognition Software* (blog), August 1, 2017, https://www.facefirst.com/blog/brief-history-of-face-recognition-software/.
[67] Ibid.
[68] "Amazon Rekognition – Video and Image - AWS," Amazon Web Services, Inc., accessed June 10, 2019, https://aws.amazon.com/rekognition/.

finetuning facial recognition systems with little to no regulation, biases and inequities are blending into the technology and changing the trajectory of facial recognition's future.

c. The Faces Behind Facial Recognition Technology

The lack of diversity in the technology sector has perpetuated spawned racial biases in technology systems and products. A 2014 report from the United States Equal Employment Opportunity Commission found that only 7.4% of the technology sector was made up of African Americans, while African Americans made up 13.2% of the entire country's population.[69] In contrast, white Americans made up 62.1% of the population and 68.5% of the technology sector; furthermore, whites comprised 83.3% of tech industry executives, and 80% of those executives were male.[70] Without minority representation in the creation of these systems, as Safiya Nobel writes, racial biases fall through the cracks and embed themselves into technologies: "[T]he public still struggles to hold tech companies accountable for the products and errors of their ways. These errors increasingly lead to racial and gender profiling, misrepresentation, and even economic redlining."[71] She goes onto explain that when minorities are left out of the creation process, the final product does not equally represent the population:

> There is a missing social and human context in some types of algorithmically driven decision making, and this matters for everyone engaging with these types of technologies in everyday life. It is of particular concern for marginalized groups, those who are problematically represented in erroneous, stereotypical, or even pornographic ways in search engines and

[69] "Tech Is Dominated by Even More White Dudes than the Rest of the Private Sector," accessed June 10, 2019, https://mashable.com/2016/05/19/diversity-report-silicon-valley-white-men/.
[70] Ibid.
[71] Safiya U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018).

who have also struggled for non-stereotypical or nonracist and nonsexist depictions in the media and in libraries.[72]

This lack of representation leads to technologies and algorithms that can overlook minorities entirely or acknowledge minorities through a stereotypical lens.

As stated, facial recognition systems rely heavily on photograph databases to train artificial intelligence. Although, when an industry lacks diversity, foundational components used to develop a technology can be completely overlooked. Facial recognition systems ran into a problem in which the systems were misidentifying people of color because of the absence of diversity in photographs used to train the system, "[t]he massive sets of facial images they train on skew heavily toward white men."[73] The technology sector did not ensure that these databases equally mirrored communities around the United States, and releasing flawed technology with no oversight or penalties. This is in part attributed to the concept of whiteness. Whiteness covertly holds its power "by seeming not to be anything in particular."[74] Whiteness is the unquestioned norm and is always "the unmarked category against which difference is constructed, whiteness never has to speak its name, never has to acknowledge its role as an organizing principle in social and cultural relations."[75] Communities like the technology sector inadvertently perpetuate whiteness due to a lack of diversity, and whiteness's silence helps it covertly blend into the

---

[72] Ibid.

[73] "Will Updates to Facial Recognition Software Be Used against Immigrants? - The Washington Post," accessed June 10, 2019, https://www.washingtonpost.com/technology/2018/06/28/facial-recognition-technology-is-finally-more-accurate-identifying-people-color-could-that-be-used-against-immigrants/?utm_term=.f9b6da9865d2.

[74] Nicole A. Cooke, Miriam E. Sweeney, and Safiya Umoja Noble, "Social Justice as Topic and Tool: An Attempt to Transform an LIS Curriculum and Culture," *The Library Quarterly* 86, no. 1 (January 2016): 107–24, https://doi.org/10.1086/684147.

[75] "The Possessive Investment in Whiteness: Racialized Social Democracy and the 'White' Problem in American Studies on JSTOR," accessed June 10, 2019, https://www.jstor.org/stable/2713291?seq=1#metadata_info_tab_contents.

community. The lack of minority representation leads to the furtherance of inequities and perpetuates whiteness in the American culture to retain racial hierarchies.

III. Discourse Perpetuated by Big Tech and Government

In the current political climate, industry and government leaders are capitalizing on the discourse of fear. From the inception of his campaign, President Donald J. Trump has used the "word animals to describe people crossing the border," recognized members of the alt-right and Nazi sympathizers as "fine people on both sides," and was a loud supporter of the birther movement.[76] Such discourse, whether believed or not, has impacted the American culture, including everyday technologies. Retail of facial recognition systems has increased at an almost exponential rate, while companies and government agencies have touted the software as a way to bring safety to communities across the country.[77] This is a concern for people of color, as facial recognition can be used as yet another tool to target and disenfranchise minorities.

IV. Facial Recognition Case Studies

a. Facial Recognition in Corporate America

i. Amazon

Amazon created its own facial recognition system that fell short when used on women and people of color.[78] When Amazon's facial recognition devices are used on people with dark skin or women, the system is "more likely to find an error, it's more likely to find a mismatch, it's more likely to fail to identify you, it's more likely to identify you as someone you're not."[79] Joy

---

[76] "10 Times President Trump's Comments Have Been Criticized as Racist," accessed June 10, 2019, https://www.usatoday.com/story/news/politics/onpolitics/2018/08/14/times-president-trump-comments-called-racist/985438002/.
[77] "History of Face Recognition & Facial Recognition Software," *FaceFirst Face Recognition Software* (blog), August 1, 2017, https://www.facefirst.com/blog/brief-history-of-face-recognition-software/.
[78] "Amazon Refuses To Quit Selling 'Flawed' And 'Racially Biased' Facial Recognition," accessed June 10, 2019, https://www.forbes.com/sites/zakdoffman/2019/01/28/amazon-hits-out-at-attackers-and-claims-were-not-racist/#abd7e8a46e78.
[79] Ibid.

Buolamwini noticed this trend in her research as a graduate student at the Massachusetts Institute of Technology (MIT). Buolamwini tested facial recognition software and found that it was unable to detect darker-skinned faces:

> [Software] created by brand-name tech firms such as Amazon uncovered much higher error rates in classifying the gender of darker-skinned women than for lighter-skinned men. Along the way, Buolamwini has spurred Microsoft and IBM to improve their systems and irked Amazon, which publicly attacked her research methods.[80]

Even with research from Buolamwini's and leading AI scholars, companies, police departments, and government agencies continue to create, use, and sell flawed facial recognition technology. As a result, companies and institutions are deepening the socio-economic inequities in communities around the country, creating databases and targeting minorities because of structural biases.

Amazon recently acquired Ring doorbell, a popular doorbell that connects to a customer's phone and automatically records individuals who interact or walk up to the doorbell.[81] Ring Doorbell's mission is to "reduce crime in neighborhoods [by giving customers the ability to] safely answer the door from anywhere".[82] Ring's savvy discourse surrounding safety caught Amazon's eye, enough so that Amazon filed a patent for a "suspicious persons database" after soon after acquiring Ring.[83] Amazon's patent application suggests that home safety is extremely important for homeowners:

---

[80] "MIT Researcher Exposing Bias in Facial Recognition Tech Triggers Amazon's Wrath," Insurance Journal, April 8, 2019, https://www.insurancejournal.com/news/national/2019/04/08/523153.htm.
[81] "About | Ring," accessed June 10, 2019, https://shop.ring.com/pages/about.
[82] Ibid.
[83] "This Patent Shows Amazon May Seek to Create a 'Database of Suspicious Persons' Using Facial-Recognition Technology - The Washington Post," accessed June 10, 2019, https://www.washingtonpost.com/technology/2018/12/13/this-patent-shows-amazon-may-seek-create-database-suspicious-persons-using-facial-recognition-technology/?utm_term=.1c30489bc222.

> The presence of doorbell recording devices can be a powerful deterrent against would-be burglars… The application also posits other potential uses for cameras equipped with facial-recognition technology, such as comparing such facial images to a database of suspicious persons. If a suspicious person showed up on a homeowner's doorstep, for example, the technology would then retrieve information about that person from the database.[84]

Considering the structural racism that permeates American culture, the technology industry, and racial biases found in Amazon's facial recognition system, the problem with categorizing people into this database are apparent. As previously noted, structural biases contribute to a higher rate of incarceration for African Americans, but those released may face additional discrimination upon re-entering the workforce: "if a person who has a criminal record is delivering a package, but the system has been set to automatically recognize anyone who has a prior criminal history as a 'suspicious person' and then the cops show up" the cyclical nature of inequity and discriminatory targeting will never cease for a person trying to escape the cycle.[85] Amazon is capitalizing off of fear birthed from discourse perpetuated by American leaders and media outlets.

ii. *Uber*

Uber uses facial recognition as a safety precaution to verify its drivers. The ride hailing application "randomly asks drivers to verify their identity to ensure that the same person that they have background checked and cleared to drive is the person using the app."[86] Uber introduced this safety feature in 2016 and it works "by asking the driver to take a selfie on their

---

[84] Ibid.
[85] Ibid.
[86] "Uber Sued Over 'Racist' Facial Recognition Software," accessed June 10, 2019, https://digit.fyi/uber-sued-over-racist-facial-recognition-software/.

phone, which is then checked using Microsoft's Face API software."[87] Because of the racial biases embedded within facial recognition systems, Uber employees trying to access the application have had trouble logging in. In the case of one employee, Uber contacted him "and said his images were fraudulent and terminated his account."[88] Though Uber is disputing these claims of flawed software, the research regarding the defective nature of facial recognition systems in recognizing minorities, as outlined above, indicates otherwise.

b. *Government Agency's Use of Facial Recognition within the U.S.*

i. *Police Targeting*

Police departments across the country have been cited for racial biases, and the addition of facial recognition could add another layer of tension between to police and minorities. Studies have found that implicit bias exist, as "[p]olice officers of all races—not just white ones—disproportionately kill African American suspects."[89] Predictive policing, and the idea that technology and data cannot contain biases, has become attractive to police departments. The benefits of facial recognition in police departments is evident, "[i]t has been used to catch violent criminals and fugitives… [with] law enforcement face recognition networks include over 117 million American adults."[90] Currently, the standard protocol surrounding facial recognition in law enforcement is as follows:

> [O]fficers can use mobile devices to capture face recognition-ready photographs of people they stop on the street; surveillance cameras boast real-time face scanning and identification

---

[87] Ibid.
[88] Ibid.
[89] "Killing of Black Suspects Is More than a 'white Police Problem' - Futurity," accessed June 10, 2019, https://www.futurity.org/police-killings-african-americans-1836722/.
[90] "The Perpetual Line-Up," Perpetual Line Up, accessed June 10, 2019, https://www.perpetuallineup.org/.

> capabilities; and federal, state, and local law enforcement agencies have access to hundreds of millions of images of faces of law-abiding Americans.[91]

The future of law enforcement is even more entangled with facial recognition, as they "would like to use face recognition with body-worn cameras, to identify people in the dark, to match a person to a police sketch, or even to construct an image of a person's face from a small sample of their DNA."[92] The databases that these facial recognition systems use to cross-reference mugshots is concerning "due to disproportionately high arrest rates, systems that rely on mug shot databases likely include a disproportionate number of African Americans."[93] The goal of any police department is to keep a community safe and, while facial recognition has the potential to do so, the technology needs to be reevaluated. Safety cannot be achieved if implicit and systematic biases exist within officers, technologies, and databases, particularly in such a discursively charged environment.

ii. *United States Customs Targeting*

The United States Customs and Boarder Protection (CBP) began using facial recognition to protect against criminals and undocumented immigrants from entering the country. The importance of safety is highlighted in the CBP mission statement to "serve as the premier law enforcement agency enhancing the Nation's safety, security, and prosperity through collaboration, innovation, and integration."[94] Facial recognition technologies are being used at airports by CBP to catch imposters at the U.S. and Mexico border to by monitoring people crossing into the United States. Microsoft has started working with CBP and claims its facial

[91] Ibid.
[92] Ibid.
[93] Ibid.
[94] "About CBP | U.S. Customs and Border Protection," accessed June 10, 2019, https://www.cbp.gov/about.

recognition has improved the accuracy of the agency's identification of undocumented immigrants.[95]

Implementing faulty technology at the border could lead to racial targeting and, ultimately, endanger the lives of those returning home or even seeking sanctuary. The CBP has a history of targeting minorities, underscored by a 2018 report documenting the excessive targeting of Muslims.[96] The CBP concluded that individuals with similar characteristics to these radical Suni Islamists should be continually evaluated because the agency believes they are "individuals who might have a higher risk of becoming radicalized."[97] Following this methodology seems to be the safe choice, and groups like CBP use a discourse of fear to justify their actions. Although this method can increase inequities for Muslim people, unfairly targeting individuals who live their lives lawfully. Adding racially bias facial recognition into the mix can lead to hyper targeting of American-Muslim people, forcing them to justify their innocence and live their lives as the other.

Furthermore, all databases are vulnerable to hacking, but hacking is of particular concern when a facial recognition photo database has millions of American biometrics. In the last week, CBP officials disclosed that "photos of travelers had been compromised as part of a malicious cyber-attack."[98] Cyber-attacks like this should provoke questions causing agencies like CBP weighing the benefits of safety versus the dangers of these systems. The message from government agencies that facial recognition has become more accurate, and therefore safer, is

---

[95] "Will Updates to Facial Recognition Software Be Used against Immigrants? - The Washington Post," accessed June 10, 2019, https://www.washingtonpost.com/technology/2018/06/28/facial-recognition-technology-is-finally-more-accurate-identifying-people-color-could-that-be-used-against-immigrants/?utm_term=.f9b6da9865d2.

[96] "Leaked DHS Report Uses Junk Science to Argue for Surveillance of Muslims," American Civil Liberties Union, accessed June 10, 2019, https://www.aclu.org/blog/national-security/discriminatory-profiling/leaked-dhs-report-uses-junk-science-argue.

[97] Ibid.

[98] "U.S. Customs and Border Protection Says Photos of Travelers Were Taken in a Data Breach - The Washington Post," accessed June 10, 2019, https://www.washingtonpost.com/technology/2019/06/10/us-customs-border-protection-says-photos-travelers-into-out-country-were-recently-taken-data-breach/?utm_term=.2da5c332ca8a.

misleading. Making these statements does not account for the software's implicit racial biases or its susceptibility to cyber-attacks.[99]

V. Conclusion: Presence of Mind Can Have Great Impact

Though structural racism may never be eliminated from the American social structure, any hope of lessening its grip must begin with social awareness. This awareness can grow through the analysis and identification of racial biases and the deconstruction of structural racism through the use of Walter Benjamin's "presence of mind." [100] Benjamin's views presence of mind as "an abstract of the future, and precise awareness of the present moment more decisive than foreknowledge of the most distant events."'[101] Understanding the ramifications of technology for communities across the country could give a "voice to communities who have been forced into silence."[102] Cities like San Francisco have already looked past the fiery discourse and prioritized social justice by banning facial recognition throughout the city.[103]

Facial recognition has the opportunity to bring safety to communities but cannot do so in its current state. While organizations and government agencies are propagating the use of facial recognition, this technology is furthering inequities continually faced by minorities. Users of these systems and individuals who interact with them unwillingly are powerless when trying to insight changes in facial recognition systems because of socio-economic constructs that silence

---

[99] "Facial Recognition Technology Is Finally More Accurate in Identifying People of Color. Could That Be Used against Immigrants?," Washington Post, accessed June 10, 2019, https://www.washingtonpost.com/technology/2018/06/28/facial-recognition-technology-is-finally-more-accurate-identifying-people-color-could-that-be-used-against-immigrants/.

[100] "The Possessive Investment in Whiteness: Racialized Social Democracy and the 'White' Problem in American Studies on JSTOR," accessed June 10, 2019, https://www.jstor.org/stable/2713291?seq=1#metadata_info_tab_contents.

[101] Ibid.

[102] Ibid.

[103] Kate Conger, Richard Fausset, and Serge F. Kovaleski, "San Francisco Bans Facial Recognition Technology," *The New York Times*, May 16, 2019, sec. U.S., https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html.

their voice. Making it the responsibility of regulators and companies to create an equitable facial recognition system. Until that time, users should be wary of companies and governments alike that implement facial recognition technology without confronting the risks.

Work Cited

"10 Times President Trump's Comments Have Been Criticized as Racist." Accessed June 10, 2019. https://www.usatoday.com/story/news/politics/onpolitics/2018/08/14/times-president-trump-comments-called-racist/985438002/.

"About | Ring." Accessed June 10, 2019. https://shop.ring.com/pages/about.

"About CBP | U.S. Customs and Border Protection." Accessed June 10, 2019. https://www.cbp.gov/about.

"Amazon Refuses To Quit Selling 'Flawed' And 'Racially Biased' Facial Recognition." Accessed June 10, 2019. https://www.forbes.com/sites/zakdoffman/2019/01/28/amazon-hits-out-at-attackers-and-claims-were-not-racist/#abd7e8a46e78.

"Amazon Rekognition – Video and Image - AWS." Amazon Web Services, Inc. Accessed June 10, 2019. https://aws.amazon.com/rekognition/.

"Analysis | Facial Recognition." Washington Post. Accessed June 10, 2019. https://www.bloomberg.com/view/quicktake/facial-recognition.

Conger, Kate, Richard Fausset, and Serge F. Kovaleski. "San Francisco Bans Facial Recognition Technology." *The New York Times*, May 16, 2019, sec. U.S. https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html.

Cooke, Nicole A., Miriam E. Sweeney, and Safiya Umoja Noble. "Social Justice as Topic and Tool: An Attempt to Transform an LIS Curriculum and Culture." *The Library Quarterly* 86, no. 1 (January 2016): 107–24. https://doi.org/10.1086/684147.

"Does Inequality Cause Crime? - The Atlantic." Accessed June 10, 2019. https://www.theatlantic.com/business/archive/2014/10/does-inequality-cause-crime/381748/.

"Gap between Number of Blacks, Whites in Prison Narrows | Pew Research Center." Accessed June 10, 2019. https://www.pewresearch.org/fact-tank/2019/04/30/shrinking-gap-between-number-of-blacks-and-whites-in-prison/.

"History of Face Recognition & Facial Recognition Software." *FaceFirst Face Recognition Software* (blog), August 1, 2017. https://www.facefirst.com/blog/brief-history-of-face-recognition-software/.

"Facial Recognition Technology Is Finally More Accurate in Identifying People of Color. Could That Be Used against Immigrants?" Washington Post. Accessed June 10, 2019. https://www.washingtonpost.com/technology/2018/06/28/facial-recognition-technology-is-finally-more-accurate-identifying-people-color-could-that-be-used-against-immigrants/.

"Killing of Black Suspects Is More than a 'white Police Problem' - Futurity." Accessed June 10, 2019. https://www.futurity.org/police-killings-african-americans-1836722/.

"Leaked DHS Report Uses Junk Science to Argue for Surveillance of Muslims." American Civil Liberties Union. Accessed June 10, 2019. https://www.aclu.org/blog/national-security/discriminatory-profiling/leaked-dhs-report-uses-junk-science-argue.

Michael Brown. "Whitewashing Race: The Myth of a Color-Blind Society." University of California Press, 2003.

"MIT Researcher Exposing Bias in Facial Recognition Tech Triggers Amazon's Wrath." Insurance Journal, April 8, 2019. https://www.insurancejournal.com/news/national/2019/04/08/523153.htm.

Safiya U. Noble. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press, 2018.

"Tech Is Dominated by Even More White Dudes than the Rest of the Private Sector." Accessed June 10, 2019. https://mashable.com/2016/05/19/diversity-report-silicon-valley-white-men/.

"The Perpetual Line-Up." Perpetual Line Up. Accessed June 10, 2019. https://www.perpetuallineup.org/.

"The Possessive Investment in Whiteness: Racialized Social Democracy and the 'White' Problem in American Studies on JSTOR." Accessed June 10, 2019. https://www.jstor.org/stable/2713291?seq=1#metadata_info_tab_contents.

"The Source of Black Poverty Isn't Black Culture, It's American Culture - The Atlantic." Accessed June 10, 2019. https://www.theatlantic.com/politics/archive/2014/04/the-source-of-black-poverty-isnt-black-culture-its-american-culture/359937/.

"This Patent Shows Amazon May Seek to Create a 'Database of Suspicious Persons' Using Facial-Recognition Technology - The Washington Post." Accessed June 10, 2019. https://www.washingtonpost.com/technology/2018/12/13/this-patent-shows-amazon-may-seek-create-database-suspicious-persons-using-facial-recognition-technology/?utm_term=.1c30489bc222.

"Uber Sued Over 'Racist' Facial Recognition Software." Accessed June 10, 2019. https://digit.fyi/uber-sued-over-racist-facial-recognition-software/.

"U.S. Customs and Border Protection Says Photos of Travelers Were Taken in a Data Breach - The Washington Post." Accessed June 10, 2019. https://www.washingtonpost.com/technology/2019/06/10/us-customs-border-protection-says-photos-travelers-into-out-country-were-recently-taken-data-breach/?utm_term=.2da5c332ca8a.

"What Is Facial Recognition Technology, and Why Is It so Controversial?" Accessed June 10, 2019. https://finance.yahoo.com/news/facial-recognition-amazon-114720161.html.

“Will Updates to Facial Recognition Software Be Used against Immigrants? - The Washington Post.” Accessed June 10, 2019. https://www.washingtonpost.com/technology/2018/06/28/facial-recognition-technology-is-finally-more-accurate-identifying-people-color-could-that-be-used-against-immigrants/?utm_term=.f9b6da9865d2.

Marisa Purcell, Savannah Lake, and Ashton Prigge

IS 262B | Dr. Jillian Wallis

June 12, 2019

Data Management Plan: Implementation Report

Dr. Kara Cooney | Coffin Reuse Research

I. Overview

Dr. Kara Cooney is a professor of Egyptian art and architecture at UCLA, as well as the chair of the Department of Near Eastern Languages & Cultures. Specializing in craft production, coffin studies, and economies in Ancient Egypt, Dr. Cooney has extensively researched Ancient Egyptian funerary practices, contextual architecture, funerary arts, and material culture.

Her current research explores coffin reuse in Ancient Egypt. For Ancient Egyptians, coffins were an integral part of the afterlife, facilitating the transformation of the dead. However, during the 21st Dynasty (circa 1150 BCE), Egypt and its neighboring civilizations in the Mediterranean and the Near East experienced socioeconomic and political instability, afflicted with drought, famine, and foreign invasion.[104] As a result, trade networks collapsed, and Egypt no longer had access to the materials necessary to create coffins. As coffins were not believed to protect the dead in the long term (instead seen as enabling rebirth immediately after death), Ancient Egyptians resorted to reusing older coffins in order to ensure safe passage to the afterlife for the recently deceased.[105] Dr. Cooney's research revolves around evaluating these coffins for signs of reuse, looking for modifications in decoration, names, and coffin parts.

[104] "Update from ARCE: Current Research, Excavation and Conservation Projects in Egypt," *NILE* Magazine, October-November 2018, 59.
[105] "Update from ARCE," *NILE* Magazine, 59.

Dr. Cooney's research on coffin reuse is an expansive endeavor, involving eight years of data collection. All in all, Dr. Cooney has examined over 300 coffins in over 20 museums and private collections around the world, generating over 100,000 photographs as well as qualitative data on each coffin. This qualitative data was collected and saved as PDF field note files, which were then manually entered into an Excel database. The photos and field notes total one terabyte of data. Dr. Cooney is no longer collecting data. At this stage of her research she is interested in cleaning, curating, and sharing her data via publications and her research and educational website called WikiArtifact.

At the start of this quarter, we met with Dr. Cooney and her research team to review our data management plan from the winter quarter to determine next steps for data management and WikiArtifact. Dr. Cooney expressed concern about the financial sustainability of WikiArtifact, and wanted to ensure WikiArtifact can be financed and maintained for the long term. Last quarter, Dr. Cooney envisioned WikiArtifact as a visual, interactive online database of her research data, complete with visual tags that users could click on to view and search multiple data elements of the coffins, such as location, mythical imagery, and coffin materials. Dr. Cooney wanted WikiArtifact to be accessible, usable, and dynamic, with 3D imaging of her coffins and easy sharing via social media. The database would be collaborative, with vetted researchers adding their coffin-related data.

Accordingly, we recommended Omeka as a platform that could accommodate all of these functionalities. The platform is available via two routes: Omeka.org and Omeka.net. Omeka.org is a free, open-source platform, but requires the user to have their own server space.[106]

[106] "Omeka Classic," https://omeka.org/classic/. (Accessed June 6, 2019).

Maintaining a server can be complicated and costly, so we investigated Omeka.net. With Omeka.net, users pay for server space hosted by Omeka based on their storage needs.[107] Omeka's support team gave us a quote of $5000 a year based on Dr. Cooney's need to store one terabyte of data. Even if Dr. Cooney were able to eliminate duplicate photos and reduce her storage needs to half as much (500 gigabytes), Omeka would cost $3000 a year.[108] Though Omeka checked several boxes off of Dr. Cooney's must-have list, it proved to be too costly and the desired qualities of WikiArtifact needed to be reevaluated.

Accordingly, instead of envisioning WikiArtifact as a visual database, Dr. Cooney expressed interest in sharing her data through photo essays.[109] Dr. Cooney still envisions the project as being collaborative, and would like it to be shareable through social media. We took these concerns into consideration as part of this project, and looked into how to facilitate more cost-effective WordPress and Drupal sites for WikiArtifact (see the section entitled "Objective 2: Identify Server and Platform Options for WikiArtifact," starting on page 15).

II. Project Objectives

Through our conversations with Dr. Cooney and her research team, we identified two objectives for this project that will best prepare Dr. Cooney's data for the creation of WikiArtifact. First, we will clean the data in Dr. Cooney's Excel database. Currently, data is difficult to extract from the spreadsheet, in large part due to an inconsistent and unwieldy metadata schema as well as a lack of controlled vocabulary. A more strategic, consistent

[107] "Pricing," Omeka.Net, https://www.omeka.net/signup. (Accessed June 6, 2019).
[108] "Omeka.Net Price List," 2018, Corporation for Digital Scholarship.
[109] Kara Cooney, Interview with Dr. Kara Cooney and her research team, In-person, April 18, 2019.

approach to metadata schemata and controlled vocabularies will correct these issues, and get the data in a more contextualized, standardized state. This will both aid in current use of the spreadsheet and it will prepare the data for safe upload into WikiArtifact.

Our second objective is to identify server options and platforms for WikiArtifact. Dr. Cooney and her team were very interested in understanding the technical options for WikiArtifact. Our research and recommendations for server and platform options will prioritize the long-term stability of the data, funding limitations, as well as Dr. Cooney's vision for the project.

III. Objective 1: Cleaning the Excel database

The current Excel database houses all of the qualitative data Dr. Cooney has collected on 300 coffins throughout the world. While robust and largely functional, Dr. Cooney's research team has encountered problems when trying to extract data from the spreadsheet. We have identified two elements of the spreadsheet—the metadata schema and the controlled vocabulary—that could be expanded upon and standardized to streamline the spreadsheet, making it more navigable and usable.

A. Metadata Schema

The Excel database currently uses a homegrown metadata schema, tailored to the specific needs of Dr. Cooney's research. The schema is comprised of 15 fields, including information on the holding institution of the coffin (city, museum, accession number), descriptive metadata about the coffin (coffin type, coffin part, dating, provenance, Niwinski number, name(s) of the deceased, title), reuse information about the coffin (date examined, reuse score, type(s) of reuse), and miscellaneous notes (notes and other notes). These fields set a strong foundation for us to work off of. Ultimately, we created a revised schema that more accurately captures the complexity of Dr. Cooney's data, as well as created a crosswalk to Dublin Core to enable potential data sharing on a larger scale.

a. Revised metadata schema

We identified three issues in Dr. Cooney's homegrown schema that we sought to address with a revised schema: inconsistency, a lack of specificity, and insufficient description. When reviewing the spreadsheet, we saw that metadata fields were used inconsistently in some cases. The "provenance" field, for example, sometimes included the museum the coffin was housed at, the name of the excavation team, and information about the buyer and seller of the coffin.[110] This inconsistent entry is in part because the field of "provenance" is too broad. Many elements contribute to an item's provenance, include acquisition details, creation information, and the item's holding information. Broad terms ellide this specificity. This complicates data retrieval, as it is not clear what information will be found in each field. Further, during the initial stages of data collection, having more specific, granular metadata fields encourages more thorough and consistent data collection. In addition to the overly broad metadata fields that are inconsistently used, we found that Dr. Cooney's spreadsheet lacked some critical metadata fields, such as administrative data on access rights or provenance metadata for the data itself (not the coffin).

[110] Kara Cooney and Amber Wells, 2018, "Coffin List (FULL) 3.0."

Ultimately, we wanted to create a metadata schema that was useful to Dr. Cooney, and streamlined features of her existing database. Our fully revised scheme can be found in the appendix on page 30. Our implementation of the schema on a set of data can be found in the appendix on page 37. As noted on both of these pages, the new scheme provides for provenance, descriptive, and administrative metadata. We have also defined each element in the revised metadata schema, so Dr. Cooney and her team can fully understand the new schema (appendix, page 30). Starred elements on this appendix item represent Dr. Cooney's initial elements that we carried forward to the revised schema. The other elements that are not starred, represent revisions and additions we made.

First, our revised scheme breaks apart several of Dr. Cooney's original fields in order to capture the nuance and complexity of the data. This granularity improves retrievability and entry consistency, and follows the principle of "atomizing" information—essentially, breaking up information so that each metadata field only contains one type of data.[111] An example of this can be found with the revisions we made to the "dating" field. Previously, Dr. Cooney's schema only had one field for describing the date of the coffins. This meant that the field was often filled with long, continuous blocks of texts, such as "19th dynasty to mid 20th dynasty ???"[112] In the revised schema, the "dating" field has been broken into five distinct fields: coffin start time period and coffin end time period, both accompanied by fields to add period descriptors (such as early, mid, late), as well as field to denote ambiguity about the period. From the above example of "19th dynasty to mid 20th dynasty ???", the information would be "atomized" into distinct chunks:

---

[111] Carly Strasser, 2015, "Research Data Management," NISO Primer Series, Baltimore, MD: National Information Standards Organization, 5.
[112] Cooney and Wells, "Coffin List (FULL) 3.0."

19th dynasty, mid, 20th dynasty, and ambiguous. Breaking “dating” into these five categories makes the Excel noticeably more navigable, as you can sort and filter for these facets by column.

Similarly, we broke out the “provenance” field to be more specific, adding fields such as “buyer,” “seller,” and “acquisition date.” We also added a few categories for data provenance, including “data collector” and “reuse observations and explanation.” These fields were missing from the previous spreadsheet, but are important for tracking the provenance of the data, thereby engendering trust in the data for external researchers who access WikiArtifact and wish to reuse the data.

The final two fields from Dr. Cooney’s initial schema that we reworked were “notes” and “other notes.” These sections functioned as catch-all fields within the database for miscellaneous information related to the coffins. We reviewed the content of these fields, and identified three commonalities: notes on file location, information on related coffins, and reuse observations for the coffin. We thus eliminated these two categories, which were vague and inconsistently utilized, and created these three new metadata categories. It is advised, however, that Dr. Cooney and her research team do their own review of the “notes” and “other notes” sections, to determine if there was any other categories information that should be added to the revised metadata schema. We are neither Egyptologists nor were we deeply entrenched in the data collection process. Dr. Cooney and her team may be able to identify additional patterns and commonalities within these two sections.

b. Dublin Core Crosswalk

In addition to revising Dr. Cooney’s metadata schema to make the data easily retrievable and navigable, we also wanted to provide Dr. Cooney with the option to make her data more interoperable via a standardized schema. Last quarter, we suggested reviewing two potential standardized schemata for Dr. Cooney's data: MIDAS Heritage and Dublin Core. Developed by the Forum on Information Standards in Heritage and recommended by the Digital Curation Centre, MIDAS Heritage is a robust metadata standard for describing archaeological buildings, sites, and artifacts.[113] When reviewing the MIDAS Heritage standard, however, we found their categories and subcategories too detailed, and perhaps overwhelming to a research team with limited resources for information management.

Conversely, the Dublin Core metadata standard is a simple, low-cost metadata standard for digital objects. The schema was "designed to be extremely simple, flexible, and extensible" to encourage as wide adoption as possible. Dublin Core is comprised of just fifteen core elements, which are all optional and repeatable.[114] After evaluating its core elements, we determined that Dr. Cooney and her team would feel comfortable with the schema, especially as compared to other standardized schemata. Further, because the basic elements are simple and flexible, a wide variety of communities are more likely to use it—making the schema a good fit for Dr. Cooney's data, which spans the disciplines of art history, Egyptology, and archaeology.

Our Dublin Core crosswalk can be found in the appendix, on page 33. Should Dr. Cooney ever wish to make her data more shareable or interoperable, she now has a clear roadmap for doing so. Further, Dr. Cooney could encode these Dublin Core elements into the back-end of the WikiArtifact site—not the front-end—so that this metadata is searchable and retrievable without ruining the aesthetic or preferred term usage for metadata fields within the photo essays on

---

113 "MIDAS Heritage: The UK Historic Environment Data Standard," 2012, Forum on Information Standards in Heritage, https://historicengland.org.uk/images-books/publications/midas-heritage/midas-heritage-2012-v1_1/, 22 and 27.
114 Stephen J. Miller, 2011, *Metadata for Digital Collections: A How-to-Do-It Manual*, London, UK: Facet Publishing, 51.

WikiArtifact (for example, if Dr. Cooney would like to maintain the metadata field "types of reuse" instead of the more generic Dublin Core field "subject").

We encountered a few roadblocks when mapping Dr. Cooney's revised schema to Dublin Core. Overall, there were instances of unclear mapping. More specifically, we often found that we were unclear about whether we were describing the coffin, the photo of the coffin, or the dataset about the coffin. This problem arose in mapping to Dublin Core fields such as "contributor", "date," "type", and "language." Generally speaking, we prioritized describing the coffin, not the image of the coffin or the dataset. However, for "contributor," for example, we felt it was important to list Dr. Cooney as a "contributor" for her role in data collection, even though we were technically describing the coffin, not the data set about the coffin. While somewhat inconsistent, omitting this metadata would erase a lot of the context for the coffin within the setting of WikiArtifact. Similarly, the "type" of object per Dublin Core could be a physical object (the coffin itself) or a data set. The "language," too, could refer to the language on the coffins, or the language of the dataset. This ambiguity is a natural consequence of using Dublin Core, whose simplicity does not allow for nuanced description that could delineate these relationships. Nevertheless, the strengths of Dublin Core with regard to wide-scale adoption, simplicity, and interoperability make it the best fit for Dr. Cooney's data.

Additionally, some fields from Dr. Cooney's schema mapped onto several Dublin Core elements. For example, "name of the deceased" maps onto both "subject" and "description." Per Dublin Core's usage guideline, "subject" refers to "the topic of the content of the resource," often described in keywords or key phrases, while the "description" field is "an account of the content of the resource," serving as "a potentially rich source of indexable terms" that can use full sentences.[115] The “name of the deceased” sits between these two categories, without a clear-cut home.

It was also difficult to create the crosswalk without knowing the new context in which the data would be used. When crosswalking, it is not always necessary to map every element from the old schema to the new schema. It is only necessary to map the elements that are relevant to the new context. For example, the category “reuse score” would not always need to be mapped onto “description,” if the new context in which the data is being used is not concerned with this particular metric. In mapping the schema, we tried our best to be as inclusive as possible, to account for whatever new contexts the data may be used in.

There were also a few Dublin Core elements that we were not able to map to, as Dr. Cooney had not collected data on that front. “Format,” for example, was unmappable, as Dr. Cooney did not collect data on dimensions of coffins or materials. Since Dublin Core does not require usage of each element, this is not a serious hindrance. But it does show how schemata are not always easily matched or aligned.

Finally, in completing the crosswalk we noticed that some categories from Dr. Cooney’s schema do not exist in Dublin Core. This meant that some data collected is lost in the crosswalk to Dublin Core. For example, the location of the holding institution for the coffin, which in Dublin Core would be the location of the “publisher,” did not make it during the mapping process. Fortunately, none of the affected elements were especially significant to understanding Dr. Cooney’s research.

[115] “DCMI: Using Dublin Core,” http://www.dublincore.org/specifications/dublin-core/usageguide/elements/. (Accessed April 27, 2019).

All of these issues are endemic to mapping and Dublin Core in general.[116] Despite this, the benefits of potential widespread sharing and interoperability make mapping a valuable exercise and potential option for Dr. Cooney's data.

B. Controlled Vocabulary

Dr. Cooney's Excel database suffers from inconsistent naming practices and data entry, in part due to turnover in research assistants and in part due to the nature of her data, which is subjective and thus can generate ambiguous descriptors. This has made it difficult to systematically and efficiently extract information from the database, since each column had so many variant terms that filtering columns did not always retrieve accurate or complete results. Often when research assistants were attempting to create charts from the data, they had to manually review cells.

We took several steps to standardize vocabulary use within the spreadsheet. First, we collocated the terms. We ran the entire Excel spreadsheet through OpenRefine, a data cleaning application. This allowed us to collocate terms, and determine authority terms amongst variant terms. This provided a holistic view of what sort of terms were used and where the variation took place. We completed this for categories that required little Egyptology expertise, and have shared a list with Dr. Cooney and her research team to review and approve. However, there were quite a few categories that would require Egyptology expertise to understand the variant terms and their relationships. Accordingly, Dr. Cooney's research team will need to review these categories, and determine authority terms directly. Dr. Cooney has communicated that this will likely happen in

[116] Mary S. Woodley, 2016, "Setting the Stage," In *Introduction to Metadata*, edited by Murtha Baca, http://www.getty.edu/publications/intrometadata/metadata-matters/.

Fall 2019, when they have more graduate student researcher support. This is a critical component of the data clean-up, and should be prioritized.

Our next step toward standardizing Dr. Cooney's vocabulary was to integrate the Getty Vocabularies, where applicable. Although not a perfect fit for her very specific data regarding coffin reuse, there are a number of fields within her spreadsheet that have been standardized via Getty Vocabularies. The Thesaurus of Geographic Names could be used for locations, while the Union List of Artist Names could be used for museum names. A full list of these recommendations, as well as other style and naming conventions, can be found in the appendix on page 35. Integrating the Getty Vocabularies into the spreadsheet will make Dr. Cooney's data more interoperable; however, this is not as high of a priority as cleaning up the variant terms via OpenRefine or migrating the spreadsheet to the revised metadata schema, as the Getty Vocabulary standardization will only help the external researchers who are using these vocabularies. The other two data standardization tasks will help all users, including Dr. Cooney and her team. Accordingly, Dr. Cooney should only implement this recommendation if she has the time and resources to do so.

Finally, we found that our revised metadata schema solved some of our vocabulary issues, particularly with regard to ambiguity. The coffin dates, for example, are now broken into five columns that are granular enough to avoid variance. Where before, it was common to have variant dates like "early to mid 21st Dynasty," "early-mid 21st Dynasty," and "early/middle 21st Dynasty," now, the five metadata fields in the revised metadata schema encourage more consistent and standard inputs.[117]

[117] Cooney and Wells, "Coffin List (FULL) 3.0."

C. Workflow and Implementation Recommendations

In order to jumpstart the database clean-up, we have revised a subset of Dr. Cooney's data with both the new metadata schema as well as our controlled vocabulary recommendations. These entries will also serve as an example of clean data, which Dr. Cooney and her team can review and refer to during their data cleaning process.

Dr. Cooney's team selected the coffins from Museo Egizio in Turin, Italy as the ideal starting point for WikiArtifact as they believe that Museo Egizio will be the most flexible in terms of image and data sharing.[118] Accordingly, we transformed the coffin entries from this museum, devising the following workflow for data-cleanup:

Migrate data from the old metadata schema to the revised schema.

a. Review and understand the new metadata schema. Read the definitions of the new metadata fields (page 30).
b. Understand the differences between the old schema and the revised schema.
c. Create a new Excel spreadsheet, with the new metadata fields as the header.
d. Copy and paste data from the old metadata spreadsheet to the new spreadsheet with the revised schema, moving data to the new metadata fields.

Standardize data through controlled vocabularies.

a. Import the spreadsheet into OpenRefine.

[118] Cooney, Interview with Dr. Kara Cooney and her research team.

b. Correct for variant terms: select a column and then filter by text facet. On the left side panel, all variations within the column will appear. Cluster the terms, and then enter your preferred authority term. Merge and recluster the items.
c. Integrate the Getty vocabularies: review which columns should use the Getty vocabularies (page 35). Add these terms to the spreadsheet. For Getty terms that can be applied to multiple cells, you can apply them at scale in OpenRefine through clustering and assigning a preferred authority term again.

Clean the data. In addition to clustering variant terms, OpenRefine is a powerful tool for editing and transforming data.

a. Eliminate white space. White space is extra spacing within a cell that is invisible to the eye, but can cause problems in data curation. Best practice for cleaning data is to eliminate white space.
b. Clean up the "types of reuse" field. In OpenRefine, you can break apart cells into multiple rows. This will allow for better manipulation, sorting, and filtering of this column, improving data retrieval. These edits do not export well into Excel, so this search functionality may be best done exclusively within OpenRefine.

We encountered a few difficulties when carrying out the data transformation for these 30 entries, which should be noted so that Dr. Cooney and her team can do their best to avoid them. During the initial steps of moving data from the old metadata schema to their new fields, we found the Excel spreadsheet to be somewhat cumbersome and not user friendly due to the sheer number of columns. We would recommend freezing the header so that team members do not need to scroll up to remember each column name. Column order could also be adjusted to best suit the team members' workflows. With regard to the controlled vocabularies, the Getty vocabularies rely on hierarchies to mark relationships. This may not be the most intuitive structure for new users. Fortunately, the two Getty vocabularies we are recommended—Union List of Artist Names (ULAN) and Thesaurus of Geographic Names (TGN)—are more straightforward on this front than a vocabulary like the Getty Art & Architecture Thesaurus, whose subject and topics are highly interconnected and more hierarchical than names and locations.

As for implementing the above workflow, we would highly recommend that Dr. Cooney and her team attend a training on OpenRefine. Utilizing OpenRefine as part of their data clean-up will save Dr. Cooney and her research team an immense amount of time as the program is intuitive, powerful, and can transform data on a large scale. The Data Science Center at UCLA Library regularly hosts workshops on OpenRefine.[119] We would recommend attending one of these workshops or contacting the Data Science Center directly for training.

As described in the previous data management plan, data stewardship is an active, ongoing responsibility. The initial clean-up and transformation of data does not signal the end of data management practices. There are a few fields that are likely to change throughout the lifetime of the data, including "file location notes" and "access rights." Any changes to the data on these fronts should also happen to the Excel database. In fact, it is recommended that researchers revisit all data management documentation, including the previous plan and this report, on a weekly basis, to ensure follow-through and consistency as well as to record any updates.[120]

[119] "UCLA Library Events," UCLA Library, https://www.library.ucla.edu/events/data-cleaning-openrefine. (Accessed June 8, 2019).
[120] Strasser, "Research Data Management," 5.

IV. Objective 2: Identify server options and platforms for WikiArtifact

A. UCLA IT Options

As addressed above, new server and platform options were needed in order to cut down on cost and fit the WikiArtifact's new format as photo essays. UCLA's IT department offered three options to build and host a website.

1. UCLA IT Option #1:

The most cost-effective option would be for Dr. Cooney to create her own WordPress site that could be coupled or uncoupled with a CPanel hosted by UCLA's IT department. A CPanel is a "web based hosting control panel provided by many hosting providers to website owners allowing them to manage their websites from a web based interface. This program gives users a graphical interface from which they can control their portion of the... server."[121] Hosting their own website on the CPanel means that Dr. Cooney's team would be responsible for maintaining the server; as such, they would be responsible for any security threats that UCLA deems concerning. This also means that Dr. Cooney and her team would need to use a platform like Drupal or WordPress for the site, and potentially pay a developer to add and maintain any plugins they want the site to have. Though this server option is the most cost effective at $28 a month, the responsibility of maintaining their CPanel could take up valuable resources like the time of graduate student researchers and funding for a developer to perform security maintenance.[122]

[121] "What Is CPanel? How to Use CPanel for WordPress Hosting," WPBeginner, https://www.wpbeginner.com/glossary/cpanel/. (Accessed June 11, 2019).
[122] Damon Wolf, Interview with Technical Account Manager | Information Technology Services at the University of California, Los Angeles, Phone, April 17, 2019.

2. UCLA IT Option #2:

The second option UCLA IT offers is a service called Site Factory.[123] This option is $300 a month but comes with much more than server space.[124] Site Factory offers templates to create an interactive website, and UCLA's IT department used Site Factory to create several websites.[125] The style guide and templates use Drupal.[126] This option provides for a fast launch: through Site Factory, websites can be launched within 8 hours, versus 180 hours when creating your own custom site on cPanel. The downside of the Site Factory option is that it is not as customizable as Dr. Cooney may hope for. Additionally, this option is more costly per month and would require hiring a Drupal developer for roughly $110 an hour to develop the site and potentially maintain the site depending on the level of comfort Dr. Cooney and her team.[127]

3. UCLA IT Option #3:

The third option would be for Dr. Cooney and her team to manage their own server. To run a platform like Omeka with several plugins on a dedicated server, the cost would be nearly $200 per month for the server, and $0.09 for every GB of image storage (1TB would therefore be around $90/month).[128] This does not include the need for a web developer to create the site. This option is not recommended because it can be time-intensive, costly, and require some expertise.

B. Server and Platform Options Based on Dr. Cooney's New WikiArtifact Vision

[123] Ibid.
[124] Ibid.
[125] Ibid.
[126] Ibid.
[127] Ibid.
[128] Damon Wolf, Interview with Technical Account Manager | Information Technology Services at the University of California. .

In speaking with UCLA's IT department and Dr. Cooney, as well as through researching platforms, it became clear that the vision of WikiArtificat had to change based on the resources available to Dr. Cooney. All server and platform options had trade-offs such as functionality, budget, or sustainability. Dr. Cooney and her team made it clear that sustainability and usability are the top priorities.[129] After understanding the budget needed to support running a large amount of data on a server as well as customized APIs, Dr. Cooney and her team decided WikiArtifact should be a more streamlined site formatted as photo essays.[130] They also determined that WordPress is a good option to host WikiArtificat because Dr. Cooney's team has used and is comfortable with the platform.[131] Our team balanced the priorities of budget, accessibility, and sustainability, ultimately deciding that WikiArtifact should be built using WordPress, and Dr. Cooney should use WordPress server space to host the site.

WordPress offers a business option for $25 a month with unlimited storage space, which is necessary for Dr. Cooney's terabyte of photographs.[132] This pricing structure means a WordPress platform and hosting option would cost $300 year, a far cry from Omeka's $5000 a year. The pricing of WordPress combined with the team's general familiarity makes WordPress the most sustainable option for WikiArtifact. Furthermore, WordPress features that come with the business plan will be helpful to Dr. Cooney and her team, including social media compatibility, Google Analytics, live chat setup support, unlimited premium themes, and the option to install customized themes and plugins.[133] If Dr. Cooney does decide she wants to customize her WordPress site beyond the templates that WordPress offers, she may need to

---

[129] Cooney, Interview with Dr. Kara Cooney and her research team.
[130] Ibid.
[131] Ibid.
[132] "WordPress.Com Plans and Pricing – Get Started for Free Today!," *WordPress.Com* (blog), February 23, 2016, https://wordpress.com/pricing/.
[133] Ibid.

consider putting funding aside for a developer. Lastly, Dr. Cooney could start with a smaller and less costly WordPress site, and upgrade the site in the future when she uploads the full terabyte of data.[134]

V. Roadblocks and Future Recommendations

We have identified potential roadblocks on the data management side as well as with the technical implementation of WikiArtifact. Proactive and thoughtful stewardship will go a long way in addressing these issues.

A. Data Management

a. Photo permissions

During our meeting with Dr. Cooney and her research team at the start of the quarter, Dr. Cooney communicated that museums can be protective around the dissemination of images and information about their holdings.[135] While Dr. Cooney took photos of the coffins personally and thus technically has copyright over that creative work, Dr. Cooney is very sensitive to museum needs and wants to maintain good working relationships with them. Accordingly, Dr. Cooney and her team will need to secure permissions from each museum in order to distribute coffin photos on a public website like WikiArtifact. Dr. Cooney could also watermark the photographs to prevent unauthorized dissemination, but securing photo permissions is the highest priority.

Dr. Cooney indicated that some museums may be more open to the idea of sharing the photos on WikiArtifact, such as Museo Egizio. Contacting such institutions should be prioritized

---

[134] Ibid.
[135] Cooney, Interview with Dr. Kara Cooney and her research team.

in order to get WikiArtifact started. Once there is a critical mass of participants, other institutions may be more motivated to join and permit photo distribution.

In addition to recording these permission rights in the Excel database, it is best practice to also make it clear to external users how these photos can be used. Creative Commons and Rightsstatements.org both offer simple, standardized language regarding reuse rights that will help external users of WikiArtifact understand how they can use these coffin images.[136][137] Dr. Cooney could review these licenses and statements, determine which best fits for each rights situation, and then describe or directly link to the appropriate license or rights statement.

b. Codex

Dr. Cooney's data is qualitative, as it consists of Dr. Cooney's observations and determinations regarding signs of coffin reuse. Further, her research covers a very niche subject, with terms like "decorative reuse" and various reuse scores not necessarily ubiquitously used or known. Dr. Cooney's research would benefit greatly from a codex, which defines every data entry within the sheet. With regard to dates, for example, a codex would define what "19th Dynasty" means, or "mid." With regard to types of reuse, a codex would define "name reuse" and "decorative reuse," for example. The codex should be made available on the WikiArtifact website, to help users of the data as well as potential contributors to WikiArtifact, better understand the data and ensure consistent usage of terms.

B. Technical Rollout of WikiArtifact

a. Staffing limitations

[136] "About The Licenses," Creative Commons, https://creativecommons.org/licenses/. (Accessed June 6, 2019).
[137] "Rights Statements," https://rightsstatements.org/page/1.0/?language=en. (Accessed June 6, 2019).

While Dr. Cooney's research team does have some background with WordPress, it is possible that Dr. Cooney may need to hire a WordPress developer to create WikiArtifact. WordPress has a strong user community and offers robust technical documentation for setting up and running WordPress sites; however, if Dr. Cooney wants to make best use of advanced plug-ins, beyond what WordPress templates offer, she may need to hire a software engineer.

b. Collaborative workflow controls

Dr. Cooney wants WikiArtifact to be a collaborative site where researchers can contribute their findings and data. Dr. Cooney and her team will need to vet researchers to confirm both the veracity and style of the data contributed. For instance, Dr. Cooney may wish to ensure that external data does not exhibit problems with term variance, and enforce authority terms for certain topics. Depending on the volume of data received and the resources at Dr. Cooney's disposal, she may not be able to do this on the largest scale, but she could narrow vocabulary control to a preferred term for specific topics, such as dating or names. Dr. Cooney and her team should consider implementing a vetting form, such as Google Forms or the WordPress API Gravity Forms.[138] These forms allow site administrators to control content that goes onto WikiArtifact before it is published on the site.

c. Maintaining the WordPress site and its plug-ins

WordPress is an open source platform, which allows programmers to create plug-ins to integrate into WordPress sites.[139] The WordPress platform has regular updates to "fix bugs and ensure speed, security, and compatibility."[140] Similarly, many programmers who have created

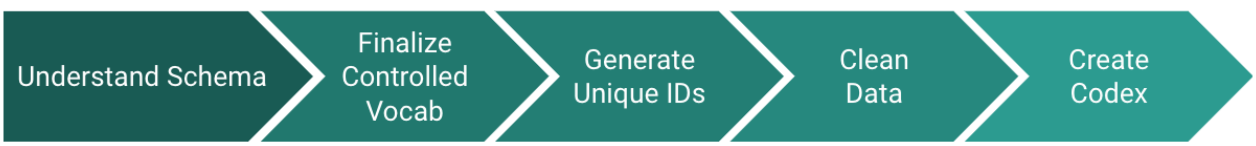
[138] "Using the API Lead Form with Gravity Forms in Wordpress," Tripleseat Support, accessed June 7, 2019.
[139] "Why Is WordPress Free? What Are the Costs? What Is the Catch?," WPBeginner, January 22, 2019, https://www.wpbeginner.com/beginners-guide/why-is-wordpress-free-what-are-the-costs-what-is-the-catch/.
[140] "Why Is WordPress Free? What Are the Costs? What Is the Catch?," WPBeginner.

WordPress plug-ins will update those APIs when WordPress runs a site-wide update.[141] However, if Dr. Cooney utilizes plug-ins that are not frequently updated by the developers who created them, she may need to consider hiring a programmer to fix the code. This would be an added cost, and it should be a consideration when integrating plug-ins not offered by preset templates.

V. Project Timeline

Dr. Cooney's timeline is greatly dependent on graduate student researcher availability and funding. We have created two project timelines that Dr. Cooney can follow at whatever pace her staffing and resources allow. We suggest Dr. Cooney first implement the data management recommendations (in turquoise), and then begin creating WikiArtifact (in pink).

Phase 1: Data Management



A. Understand the Schema: In order to implement the new schema that we have created, Dr. Cooney and her team should review the schema and its definitions to better understand where information falls in the database.
B. Finalize the Controlled Vocabulary: There were many variant terms that we were not able to create authority terms for because we did not have the Egyptology expertise to do so. We have shared these lists of variant terms with Dr. Cooney and her team. They will need to review these and come up with a controlled vocabulary they feel comfortable with.

[141] Ibid.

C. Generate Unique Identifiers: Creating unique identifiers for each coffins will greatly facilitate data retrieval and long-term stewardship. Currently, data regarding each coffin is located in the Excel, photo files, and WikiArtifact. Generating unique identifiers for each coffin—and integrating these unique IDs into all three of these spaces—will allow Dr. Cooney and her team to uniquely, efficiently, and unambiguously identify each coffin, no matter where they are digitally stored.[142] Linking photos, Excel entries, and WikiArtifact in this way will also be extremely helpful for new graduate student researchers who are not familiar with the coffin reuse database. Additionally, these unique IDs can be incorporated into a preferred citation for coffins within WikiArtifact, so that external users of the data can easily locate and access the correct coffin.[143]

Unique identifiers can be created through the Online UUID Generator, a site that generates a unique identifier using a timestamp and the MAC address of the computer to make it truly unique.[144] Dr. Cooney and her team may want to create unique identifiers that have visible characteristics tied to the coffin; for example, each unique ID could begin with the initials of the city where the museum is located.

D. Cleaning the Data: The data within the database must be placed in the new schema using the controlled vocabulary once the previous steps are completed. When we tested out implementing the schema, we found that transforming ten coffin entries takes around 30 minutes. Therefore, 300 coffin entries will take approximately 15 hours to input.

---

[142] "On the Utility of Identification Schemes for Digital Earth Science Data: An Assessment and Recommendations | SpringerLink," accessed June 8, 2019, https://link.springer.com/article/10.1007%2Fs12145-011-0083-6.
[143] Ibid.
[144] "Online UUID Generator Tool," https://www.uuidgenerator.net/. (Accessed June 7, 2019).

E. Create a Codex: This step will also be very time consuming. It is not the most time-sensitive item, as it will be most helpful for external users of WikiArtifact, in understanding what the data means. It is, however, an important element for creating the necessary context needed for resharing data.

Phase 2: WikiArtifact Implementation

A. Contact Museums for Permissions: Without permissions from the museums, the vision of WikiArtifact as a public site is not viable. Accordingly, this is a very important first step in setting up WikiArtifact.

B. Update the Database with Museum Permissions: Once the museums are contacted, the Excel should be updated with the museum's response or terms for sharing so that this information is in a centralized location. Keeping the museums' responses organized will also allow Dr. Cooney to understand if her current vision of sharing coffin photos publicly is achievable in the format she wants.

C. Create a WordPress Site: During this phase, Dr. Cooney can determine what plug-ins are needed and if she would like to hire a programmer to facilitate this process. A good starting point would be to eliminate duplicate photos of coffins, to streamline storage and identify which photos should be featured on the website. Duplicate photo sorting software exists for both Windows and Mac computers.[145]

D. Batch Upload Excel and Photos: Excel files can be batch uploaded to WordPress using a plug-in.[146] Batch uploading will prevent Dr. Cooney's team from manually inputting data. Similarly, WordPress offers a plug-in that batch uploads media files, including photos.[147]

E. Implement Vetting Forms: A form needs to be created that vets researchers who will contribute to WikiArtifact. Dr. Cooney can determine which metadata fields are necessary for each contributor to include, and embed these into the form. She can also determine which fields should comply with a controlled vocabulary, and either provide for that via a dropdown menu where applicable or send documentation to contributors with the authority terms. Such a controlled form will give Dr. Cooney and her team time to confirm the accuracy of contributed data before it goes live on the site, as well as introduce some standardization via the schema elements and controlled vocabulary.

VII. Conclusion

Dr. Cooney's research on coffin reuse in Ancient Egypt provides a fascinating look into the economies, social conditions, artistry, and spiritual beliefs of Ancient Egypt. Dr. Cooney's research has revealed some stunning discoveries—for instance, coffin reuse rates during the 21st

[145] "Photos Duplicate Cleaner on the Mac App Store," accessed June 8, 2019, https://itunes.apple.com/us/app/photos-duplicate-cleaner/id592704001?mt=12;%20https://www.ashisoft.com/blog/top-5-best-duplicate-photo-finder-to-delete-duplicate-photos/.

[146] "Import Spreadsheets from Microsoft Excel – WordPress Plugin | WordPress.Org," accessed June 9, 2019, https://wordpress.org/plugins/import-spreadsheets-from-microsoft-excel/.

[147] "How to Bulk Upload WordPress Media Files Using FTP," WPBeginner, January 10, 2018, https://www.wpbeginner.com/plugins/how-to-bulk-upload-wordpress-media-files-using-ftp/.

averaged 60%, suggesting that the practice was socially acceptable and legal. Some coffins even reveal multiple reuses.[148] After nearly a decade of traveling the globe, Dr. Cooney's findings offer a rare and invaluable look at Ancient Egyptian society, culture, and beliefs.

Active data management will help ensure the usefulness and preservation of all of the valuable data Dr. Cooney has collected. In addition to facilitating preservation, the recommendations within this report will help foster structured data sharing and peer-to-peer collaboration. The new schema and controlled vocabulary are crucial steps to increase uniformity and consistency in the data management process.

The collaborative, visual nature of WikiArtifact has "the potential to revolutionize how we approach object studies in archaeology, art history, and Egyptology."[149] The use of WordPress will make WikiArtifact a space for collaboration, centralized data sharing, and equitable access to cultural heritage materials, while allowing Dr. Cooney to stay within her budget. Furthermore, Dr. Cooney's team's familiarity with WordPress will play an important role in the launch and longevity of the project. Ultimately, conscientious and proactive data management is critical to facilitating WikiArtifact's goals and long-term success.

VIII. References

"About The Licenses." n.d. Creative Commons. Accessed June 6, 2019. https://creativecommons.org/licenses/.

[148] "Update from ARCE: Current Research, Excavation and Conservation Projects in Egypt," *NILE* Magazine, 60.
[149] Cooney, Wells, and Campbell, 2018, "National Geographic: Storytelling and Technology," 2.

Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge: MIT Press, 2015.

Cooney, Kara. Interview with Dr. Kara Cooney and her research team. In-person, April 18, 2019.

Cooney, Kara, and Amber Wells. "Coffin List (FULL) 3.0," 2018.

Cooney, Kara, Amber Wells, and Rose Campbell. "National Geographic: Storytelling and Technology," 2018.

"DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description." Accessed April 27, 2019. http://www.dublincore.org/specifications/dublin-core/dces/.

"DCMI Type Vocabulary." n.d. Dublin Core Metadata Initiative. Accessed June 7, 2019. http://www.dublincore.org/specifications/dublin-core/dcmi-type-vocabulary/.

"DCMI: Using Dublin Core." n.d. Accessed April 27, 2019. http://www.dublincore.org/specifications/dublin-core/usageguide/elements/.

"Getty Thesaurus of Geographic Names." n.d. Getty Research Institute. Accessed June 8, 2019. https://www.getty.edu/research/tools/vocabularies/tgn/.

"Getty Union List of Artist Names." n.d. Getty Research Institute. Accessed June 8, 2019. https://www.getty.edu/research/tools/vocabularies/ulan/.

"How to Bulk Upload WordPress Media Files Using FTP." WPBeginner, January 11, 2018. https://www.wpbeginner.com/plugins/how-to-bulk-upload-wordpress-media-files-using-ftp/.

"Import Spreadsheets from Microsoft Excel – WordPress Plugin | WordPress.Org." Accessed June 11, 2019. https://wordpress.org/plugins/import-spreadsheets-from-microsoft-excel/.

"MIDAS Heritage: The UK Historic Environment Data Standard." Forum on Information Standards in Heritage, October 2012. https://historicengland.org.uk/images-books/publications/midas-heritage/midas-heritage-2012-v1_1/.

Miller, Stephen J. 2011. *Metadata for Digital Collections: A How-to-Do-It Manual*. London, UK: Facet Publishing.

Niwinski, Andrzej. 1988. *21st Dynasty coffins from Thebes: chronological and typological studies*. Mainz am Rhein: P. von Zabern.

"Omeka Classic." n.d. Accessed June 6, 2019. https://omeka.org/classic/.

"Omeka.Net Price List." 2018. Corporation for Digital Scholarship.

"On the Utility of Identification Schemes for Digital Earth Science Data: An Assessment and Recommendations | SpringerLink." Accessed June 9, 2019. https://link.springer.com/article/10.1007%2Fs12145-011-0083-6.

"Online UUID Generator Tool." n.d. Accessed June 7, 2019. https://www.uuidgenerator.net/.

"Photos Duplicate Cleaner on the Mac App Store." Accessed June 8, 2019. https://itunes.apple.com/us/app/photos-duplicate-cleaner/id592704001?mt=12;%20https://www.ashisoft.com/blog/top-5-best-duplicate-photo-finder-to-delete-duplicate-photos/.

"Pricing." n.d. Omeka.Net. Accessed June 6, 2019. https://www.omeka.net/signup.

"Rights Statements." n.d. Accessed June 6, 2019. https://rightsstatements.org/page/1.0/?language=en.

Strasser, Carly. 2015. "Research Data Management." NISO Primer Series. Baltimore, MD: National Information Standards Organization.

"UCLA Library Events." n.d. UCLA Library. Accessed June 8, 2019. https://www.library.ucla.edu/events/data-cleaning-openrefine.

"Update from ARCE: Current Research, Excavation and Conservation Projects in Egypt." *NILE Magazine*, November 2018.

"Using the API Lead Form with Gravity Forms in Wordpress." Tripleseat Support. Accessed June 9, 2019. http://tripleseat.zendesk.com/hc/en-us/articles/219006788-Using-the-API-Lead-Form-with-Gravity-Forms-in-Wordpress.

"What Is CPanel? How to Use CPanel for WordPress Hosting." WPBeginner. Accessed June 7, 2019. https://www.wpbeginner.com/glossary/cpanel/.

"Why Is WordPress Free? What Are the Costs? What Is the Catch?" WPBeginner, January 22, 2019. https://www.wpbeginner.com/beginners-guide/why-is-wordpress-free-what-are-the-costs-what-is-the-catch/.

Wolf, Damon. Interview with Technical Account Manager | Information Technology Services at the University of California, Los Angeles. Phone, April 17, 2019.

Woodley, Mary S. 2016. "Setting the Stage." In *Introduction to Metadata*, edited by Murtha Baca. http://www.getty.edu/publications/intrometadata/metadata-matters/.

"WordPress.Com Plans and Pricing – Get Started for Free Today!" *WordPress.Com* (blog), February 23, 2016. https://wordpress.com/pricing/.

IX. Appendix

A. Revised Metadata Schema Definitions

The following are definitions of each element of the revised metadata schema. Understanding these will facilitate consistent data entry. We have starred (*) elements that were carried forward from Dr. Cooney's initial metadata schema. Other elements that are not starred represent revisions and additions we made.

Provenance metadata

- **City of holding institution***: City in which the institution that holds the coffin is located. Best practice is to use the Getty Thesaurus of Geographic Names (TGN) vocabulary.[150]
- **Holding institution***: Institution that holds the coffin. Best practice is to use the Getty Union List of Artist Names (ULAN) vocabulary.[151]
- **Accession number***: The holding institution's unique identifier for the coffin.
- **Niwinski number***: Correlating coffin number in the Niwinski study.[152]
- **Acquisition date**: Date in which the coffin was acquired by the museum. Best practice is to use the format MM-DD-YYYY.
- **Purchase location**: Location in which the coffin was acquired by the museum. Best practice is to use the TGN vocabulary.
- **Seller**: Agent who sold the coffin to the holding institution. Best practice is to use the ULAN vocabulary.
- **Buyer**: Agent who acquired the coffin for the holding institution.. Best practice is to use the ULAN vocabulary.
- **Current collection**: Collection in which the coffin is currently housed in within the holding institution.
- **Excavation location**: Location in which the coffin was excavated. Best practice is to use the TGN vocabulary.
- **Excavation date**: Date(s) in which the coffin was excavated. Dr. Cooney can make a decision as to whether the date range or final date of excavation is preferred. Best practice is to use the format MM-DD-YYYY–MM-DD-YYYY.
- **Excavation team/agent**: Team or agent that excavated the coffin. Best practice is to use the ULAN vocabulary.
- **Date examined***: Date in which the coffin was examined for the purposes of this coffin reuse study. Best practice is to use the format MM-DD-YYYY.

---

[150] "Getty Thesaurus of Geographic Names," Getty Research Institute, https://www.getty.edu/research/tools/vocabularies/tgn/. (Accessed June 8, 2019).
[151] "Getty Union List of Artist Names," Getty Research Institute, https://www.getty.edu/research/tools/vocabularies/ulan/. (Accessed June 8, 2019).
[152] Andrzej Niwinski, 1988, *21st Dynasty coffins from Thebes: chronological and typological studies*, Mainz am Rhein: P. von Zabern.

- **Data collector**: Agent who collected qualitative data on the coffin as part of this coffin reuse study.
- **Reuse observations and explanation**: Observations about the coffin's reuse, and justification for why the coffin received its reuse score.

Descriptive metadata

- **Coffin type***: When we queried Dr. Cooney's team for this definition, we received the following: "if it's inner or outer of board or mask." This is an example of "definition by example," and is not best practice because defining something by listing examples of it does "not establish clear boundaries between what is and is not included in a concept."[153] We recommend that Dr. Cooney offer a more robust definition to facilitate better understanding of this metadata field.
- **Ambiguity and/or notes on Coffin Type**: Any additional notes on the coffin type that go beyond defining the actual coffin type.
- **Coffin part***: When we queried Dr. Cooney's team for this definition, we received the following: "is if it's only lid or case or a fragment." Again, this is "definition by example," and is not best practice. We recommend that Dr. Cooney offer a more robust definition to facilitate better understanding of this metadata field.
- **Ambiguity and/or notes on Coffin Part**: Any additional notes on the coffin type that go beyond defining the actual coffin type.
- **Coffin time period descriptor (early, mid, late)**: A general descriptor for the time period of the coffin as being either "early," "mid," or "late." Dr. Cooney and her team may wish to define these more specifically, either here or in the project's codex (by number of decades, for example, represented in each stage).
- **Coffin start time period**: The starting Dynasty in which the coffin could be dated.
- **Coffin end time period**: The ending Dynasty in which the coffin could be dated.
- **Coffin date ambiguity**: A field to denote any ambiguity or uncertainty about the dating. Use the term "Yes" to denote that the period is uncertain and "No" to denote that it is not.
- **Additional dates**: Any other dates associated with the coffin and its items. For example, if mummy linens placed on the body found inside the coffin are inscribed with dates that differ from the coffin date (ie. the date of the interment of the body differs from the date of the coffin).
- **Name(s) of the deceased***: Names of deceased person(s) who have used the coffin.
- **Title***: Where applicable, the coffins reflect the museum's title for the object. Sometimes the title is chosen by Dr. Cooney and her research team. An explanation here of when and why Dr. Cooney sometimes chooses a new title would be helpful.
- **Reuse score***: A rating of Dr. Cooney's confidence in her ability to see coffin reuse on a scale from 0 to 3, with 3 being obvious and clearly visible evidence of reuse, 1 being only circumstantial, and 0 being no visible evidence of reuse. To clarify, a 0 score does not mean that a given coffin was not reused; it just means that Dr. Cooney cannot see evidence of that (evidence of reuse could be carefully covered by a carpenter, for example).

[153] Christine L. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World* (Cambridge: MIT Press, 2015), 19.

- **Type(s) of reuse***: Dr. Cooney's determination of the types of reuse in the coffin. Terminology should come from a controlled list of terms determined by Dr. Cooney.
- **Relation**: Any notes regarding the coffin's relation to other coffins on the list. Include the unique identifier of the other coffin in this field as well.

Administrative metadata

- **Unique identifier**: Unique identifier for each coffin generated by Dr. Cooney's team.
- **Rightsholder**: The institution or agent with which Dr. Cooney and her team correspond concerning access rights for the coffin photos and data.
- **Access rights**: Information from museum permissions concerning who is allowed to see coffin photos and data.
- **File location notes**: Notes about locations of photos and field note files, including when these files are missing.

B. Dublin Core Crosswalk

The following is a crosswalk that maps Dr. Cooney's revised metadata schema onto Dublin Core.

- DC: Title
    - KC: Title
- DC: Creator
    - Not listed on Dr. Cooney's current spreadsheet. It is recommended that this field be populated the "Unknown," as the creators of these coffins are not known.
- DC: Subject
    - KC: Types of reuse
    - KC: Name of the deceased
- DC: Description
    - KC: Coffin type
    - KC: Coffin part
    - KC: Name of the deceased
    - KC: Reuse score
    - KC: Acquisition date
    - KC: Purchase location
    - KC: Seller
    - KC: Buyer
    - KC: Current collection
    - KC: Excavation location
    - KC: Excavation date
    - KC: Excavation team/agent
- DC: Publisher
    - KC: Holding institution
- DC: Contributor
    - KC: Data collector
- DC: Date
    - KC: Coffin time period descriptor (early, mid, late)
    - KC: Coffin start time Period
    - KC: Coffin end time period
    - KC: Coffin date ambiguity
    - KC: Acquisition date
    - KC: Excavation date
    - KC: Date examined
- DC: Type
    - Not listed on Dr. Cooney's current spreadsheet. It is recommended that this field be populated with both PhysicalObject and Dataset. Both of these terms are taken from the recommended DCMI Type vocabulary.[154]
- DC: Identifier

[154] "DCMI Type Vocabulary," Dublin Core Metadata Initiative, http://www.dublincore.org/specifications/dublin-core/dcmi-type-vocabulary/. (Accessed June 7, 2019).

    - KC: Unique identifier
- DC: Language
    - Not listed on Dr. Cooney's current spreadsheet. Where there is writing on the coffin, it is recommended that Egyptian is cited using the recommended standards for Dublin Core: RFC 3066 and ISO 39, which define primary language tags and subtags.[155]
- DC: Relation
    - KC: Relation
- DC: Coverage
    - KC: Coffin time period descriptor (early, mid, late)
    - KC: Coffin start time Period
    - KC: Coffin end time period
    - KC: Coffin date ambiguity
    - KC: Excavation location
- DC: Rights
    - KC: Access rights
    - KC: Rightsholder

---

155 "DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description," http://www.dublincore.org/specifications/dublin-core/dces/. (Accessed April 27, 2019).

C. Controlled Vocabulary Guidelines & Examples

In addition to eliminating variant terms via OpenRefine, we implemented the following general guidelines for standardizing and controlling the vocabulary throughout the spreadsheet.

Style Conventions:

- Use sentence case throughout
- Use commas to denote multiple items (not plus signs or ampersands)
- For specific dates or date ranges in which the month, day, and/or year are known, use MM-DD-YYYY or MM-DD-YYYY–MM-DD-YYYY.
- For the “Coffin Date Ambiguity” field, enter either Yes or No
- For the “Coffin time period descriptor” field, enter either “early,” “mid,” “late,” or “N/A”
- Do not leave any column blank. For data that was not recorded, use “N/R”

External Thesauri and Controlled Vocabularies:

The following thesauri should be used in the denoted fields where applicable. It is marked when use of these controlled vocabularies is required; otherwise, it is only considered best practice.

The Getty Thesaurus of Geographic Names:

- City of Holding Institution
- Purchase Location
- Excavation Location

The Getty Union List of Artist Names:

- Holding Institution
- Seller
- Buyer
- Excavation Team/Agent

When mapping elements onto Dublin Core, the following vocabularies should be used:

- DCMI Type Vocabulary:[156] Type (required)
- RFC 3066 and ISO 39:[157] Language

[156]“DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description,” http://www.dublincore.org/specifications/dublin-core/dces/. (Accessed April 27, 2019).

[157] “DCMI: Using Dublin Core,” http://www.dublincore.org/specifications/dublin-core/usageguide/elements/.

The following is an example of controlled vocabulary work through OpenRefine, taking these stylistic guidelines into account.

The unbulleted terms are the authority terms that should be used in place of all of the variant terms, which are bulleted underneath that term. Items marked with an asterisk (*) denote data entries with extra details that would be placed in the newly created "Ambiguity and/or Notes on Coffin Type" column.

Dr. Cooney's team will need to review these recommendations, given their Egyptology expertise.

COFFIN TYPE

N/R

- Blank

Inner coffin

- Inner Coffin
- inner coffin
- Inner coffin (seems to be Stola)*

Inner coffin, mummy board

- Inner coffin + mummy board
- Inner coffin/Mummy board

Outer coffin

- Outer Coffin
- Outer coffin (?)*
- Outer(?) coffin*

Outer coffin, inner coffin

- Outer coffin + inner coffin

Outer coffin, inner coffin, mummy board

- Outer coffin + inner coffin + mummy board

Stola coffin, inner coffin

- Stola, Inner coffin

ADVISORS

Professor Jean-François Blanchette

Chair, Department of Information Studies, UCLA

Professor Blanchette served as my advisor throughout my MLIS candidacy. We met regularly to discuss my course work, professional goals, and development of my portfolio.

Field Garthwaite

Co-Founder and CEO, IRIS.TV

Field Garthwaite supervised me during my internship at IRIS.TV in the Summer of 2019.

Michelle Loran

Metadata Manager, TMZ

Michele "Mimi" Loran was my supervisor at TMZ during my Fall 2019 internship in the company’s Metadata Department.

ISSUE PAPER

The presence of disinformation on crowdfunding platforms has harmful consequences for both users and internet companies. Disinformation can destabilize a user’s trust while jeopardizing a company's growth. As the internet enabled stronger lines of communication and eased access to larger audiences, crowdfunding companies capitalized on this phenomenon to widen their scope. The greater accessibility to donations, however, ultimately facilitated the perpetuation of fraudulent schemes. Fraudulent postings pose a financial risk to users, while undermining the mission and reputation of a crowdfunding company, draining its finances, and leaving it legally liable. This paper will examine disinformation on crowdfunding platforms using GoFundMe.com as a case study. To deter misinformation on crowdfunding sites, I will argue that content moderation and analysis, transparency, and regulation are essential in maintaining user trust, and information professionals are best suited to implement these recommendations of fraud detection and mitigation.

Combating Disinformation and Building Trust on Crowdfunding Platforms: A Framework to Mitigate and Detect Fraud

Marisa F. Purcell

UCLA MLIS Candidate

UCLA iSchool, Department of Information Studies

May 8, 2020

Introduction

Disinformation has harmful consequences for both Internet-based crowdfunding companies and their users. The lucrative and connective nature of crowdfunding has led to the success of these companies, but made crowdfunding sites a target for fraudsters using tactics of disinformation. Disinformation can jeopardize the reputations of crowdfunding companies and compromise the trust of users. This paper will provide background on crowdfunding, using GoFundMe.com as a case study to illustrate the problem of disinformation. The proposed framework to mitigate and detect crowdfunding fraud will provide recommendations for implementation, and demonstrate the critical role of information professional in applying this approach.

Background

Crowdfunding[158] platforms assist users in donating and receiving money from anywhere and anyone around the world. In 2016 alone, United States donors trafficked $738.9 billion across crowdfunding sites.[159] Companies that use donation-based crowdfunding are at risk for fraud, and donors have no means to validate how funds are spent.[160] This lack of accountability is largely due to crowdfunding's reliance on the Internet, which diminishes interpersonal connections that can otherwise thwart fraudulent actors. Face-to-face interactions expose emotions of guilt, shame, and fear when people act deceptively, but researchers have found that digital interactions can alter individuals' judgement, morals, and ethical sensibility.[161] The

[158] The Federal Trade Commission describes crowdfunding as "a way to raise funds online by convincing a large number of people to each give money for a specific project or cause." See "FTC examines crowdfunding."
[159] "Key Crowdfunding Statistics."
[160] Ibid.
[161] Caspi & Gorsky, "Online Deception: Prevalence, Motivation, and Emotion," 58.

depersonalized and lucrative nature of crowdfunding makes these platforms profitable for creative fraudsters.

GoFundMe.com

GoFundMe is of particular importance in the crowdfunding industry, largely due to its immense popularity and wide reach. This platform is known for donation-based crowdfunding, touted as the preeminent site in free fundraising. GoFundMe supports a variety of crowdfunding needs, with 10,000 people starting a GoFundMe each day in categories including medical, memorial, emergency, education, and animals.[162] In 2018, the company's website documented more than 1.9 million views on a single fundraiser page in 24 hours and more than 7.2 million comments on all fundraiser pages.[163] In 2019, there were more than 120 million donations, raising a total of $9 billion.[164] GoFundMe relies on repeat donors, with 40% of donors making subsequent donations on the site.[165]

GoFundMe has installed some security checks to vet recipients prior to collecting donations. Recipients are required to give a small amount of personal information before publishing their fundraisers. However, recipients can begin fundraising within a few minutes of signing up on the website. After publishing a campaign, recipients are required to login to their personal email account to verify their email address.[166] All incoming donations are paused if a recipient's email address is not verified within fourteen days of the first donation, and all donations are refunded to donors if email verification is not completed within thirty days.[167] After a recipient verifies his

[162] "GoFundMe: #1 Free Fundraiser Platform - Crowdfund Online."
[163] "GoFundMe 2018."
[164] "GoFundMe 2019."
[165] Ibid.
[166] "GoFundMe Terms of Service."
[167] "Important Verification and Withdrawal Deadlines."

email, he must withdraw donations within the first thirty days of setting up the campaign or his campaign will be paused.[168] Recipients must provide bank account information to receive donor funds within ninety days of publishing a campaign; GoFundMe controls the funds until a recipient provides bank account information.[169] Prior to disbursement of the funds, GoFundMe retains a 2.9% transaction fee and transfers $0.30 of every donation to a third-party payment processor.[170] If no withdrawal has been made after ninety days, GoFundMe refunds donations to donors.[171]

GoFundMe protects donors in three ways. The platform retains a third-party payment processor to "keep GoFundMe a safe place to donate"—presumably by verifying recipients.[172] Anyone visiting GoFundMe.com can report fraud by filling out a short online form.[173] Donors who are victims of fraud are protected by GoFundMe's Donor Protection Guarantee, and only need to fill out a form and upload any evidence of fraud in order to receive a refund.[174]

Disinformation and Trust on Crowdfunding Platforms

A subcategory of misinformation, disinformation is an act of intentional deception.[175] There are two types of disinformation found on crowdfunding sites: fraudulent stories and fraudulent identities. Donation recipients will use either tactic, or a combination of both, to elicit funds.

Fraudulent Stories

Prior to publishing a campaign, recipients are asked to write about themselves and why they are seeking donations.[176] This gives recipients the freedom to write a compelling story within the

[168] Ibid.
[169] Ibid.
[170] "Everything You Need to Know About GoFundMe's Fees."
[171] Ibid.
[172] Ibid.
[173] "Reporting a Campaign."
[174] "The GoFundMe Guarantee."
[175] "Disinformation: Oxford English Dictionary."
[176] "Creating a GoFundMe From Start to Finish."

categorical bounds of the platform. A highly publicized example of a fraudulent story was published in late 2017. Katelyn McClure created a campaign on GoFundMe.com for a homeless man, Johnny Bobbitt, claiming that he helped her and her boyfriend by giving them gas money.[177] After raising over $400,000, McClure gave Bobbitt $75,000 and spent the remaining donations.[178] The narrative was fabricated and, in March 2019, Bobbitt pled guilty to one count of conspiracy to commit money laundering and McClure pled guilty to one count of conspiracy to commit wire fraud.[179]

Fraudulent Identities

Recipients can use two tactics to assume a fraudulent identity. The first is to act as a proxy to raise money for those in need.[180] Police departments have found scammers gravitate towards this tactic after horrific events occur, such as a mass shooting.[181] For example, a man found a memorial fund had been established in his honor, claiming that he died in the Parkland High School mass shooting.[182] The second approach is to act as a victim of a real event, but use the donations for an entirely separate purpose. This type of identity fraud occurred when a baby passed away and a family member created a campaign to pay for funeral expenses.[183] Though the child did pass away, the person who created the account was not related to the family. After an investigation, it became clear that the fraudster found photos of the child on the parents' social media, posed as a relative, and used the photos to create a GoFundMe campaign to request donations.[184]

---

[177] Garcia, "Couple and Homeless Man Behind Viral GoFundMe Campaign Are Charged With Conspiracy.
[178] Victor, "Woman and Homeless Man Plead Guilty in $400,000 GoFundMe Scam."
[179] Ibid.
[180] "Creating a GoFundMe From Start to Finish."
[181] "AP Uncovers 'Scams and Waste' Rampant in Pulse Shooting GoFundMe Campaigns."
[182] "After Florida Shooting, GoFundMe Scams Sprout Up."
[183] "Expenses for Baby Henley."
[184] "Kentucky Parents 'Angry' Over Fake GoFundMe Campaign For Their Deceased Baby."

Trust on Crowdfunding Sites

Trust is key to battling online disinformation. The concept of trust "concerns a positive expectation regarding the behavior of somebody or something in a situation that entails risk to the trusting party."[185] Trust is a high value currency in the current tech landscape.[186] Establishing a trustworthy relationship with donors can be financially beneficial for Internet companies, leading to "improved Web sites, sales revenues, profitability, and ultimately shareholder value."[187] Without trust, donors may not revisit a site—making this value fundamental to a crowdfunding site's operations. Donors expect fundraising platforms to protect against and deter fraudulent users. A survey of 200 GoFundMe users, however, found that only 27.9% of respondents believe fundraiser recipients are "usually honest people," and only 20.1% think visitors to the site effectively prevent fraud.[188] These opinions underscore the need for crowdfunding companies to effectively mitigate and detect disinformation in order to grow and sustain trust.

A Framework to Mitigate and Detect Fraud

The following framework proposes strategies to mitigate and detect fraud on crowdfunding sites. This approach consists of three interrelated components: content analysis and moderation, transparency, and regulation.

Content Analysis and Moderation

Content analysis and moderation is comprised of machine learning, commercial moderators, and user moderators.

---

[185] Marsh & Dibben, "The Role of Trust in Information Science and Technology."
[186] "An Economy of Trust: How Transparency Is Changing the Tech Industry."
[187] Shankar et al., "Online trust: a stakeholder perspective, concepts, implications, and future directions."
[188] "Worries About Fraud Top List Of Crowdfunding Concerns."

Machine Learning

Classification machine learning techniques can analyze text and photograph content to differentiate between fraudulent and genuine content. Researchers have proven the capabilities of machine learning classifiers in the detection of malicious and untrustworthy Internet content,[189] such as through spam detection filters used by email platforms.[190] These tools may include those that establish language patterns, such as concordances, which could be useful in differentiating between fraudulent and authentic fundraisers.[191]

Commercial Content Moderators

Commercial content moderators are hired to detect fraudulent content on Internet platforms using linguistic and cultural cues—an instinct lacked by a machine.[192] Using trained individuals to scrutinize campaigns may be the best way to differentiate fact from fiction. Fraudulent campaigns are often identical to authentic campaigns, requiring additional research or human insights to distinguish them. Unlike unassuming donors, content moderators can look at content analytically, distancing themselves from the emotional appeal that content scammers use to prey on donors.[193]

Users as Content Moderators

Human intuition is a valuable resource in uncovering fraud, and site visitors can act as a first line of defense for crowdfunding sites.[194] GoFundMe encourages its users to practice vigilance,

---

[189] Asiri, “Machine Learning Classifiers.”
[190] Hou et al., “Malicious web content detection by machine learning.”
[191] Koteyko, “Mining the internet for linguistic and social data: An analysis of ‘carbon compounds’ in Web feeds.”
[192] Dr. Sarah Roberts explains content moderation to be “ the organized practice of screening user-generated content (UGC) posted to Internet sites, social media and other online outlets to determine the appropriateness of the content for a given site, locality, or jurisdiction.” See Roberts, “Content Moderation,” 1.
[193] Luu, “The Life Changing Linguistics of. Nigerian Scam Emails,” 129-130.
[194] Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*.

but warns visitors against abusing reporting power.[195] Like commercial content moderators, users can detect fraud in ways that computers fall short. As GoFundMe suggests, potential donors can take several steps to investigate a campaign before donating: research the organizer on the Internet, [196] reverse image search photos from the campaign,[197] read comments on the page,[198] look for misleading or suspicious statements in the campaign description,[199] search for a clear answer as to how donations will be spent,[200] and determine if the recipient of funds is in control of withdrawals.[201]

<u>Transparency</u>

A transparent relationship between users and online platforms is essential to establish trust. A 2018 European study found that transparency was important to decision making, useful to establish trust and confidence in online environments, and increased the probability that individuals would use or select a product.[202] In the long-term, lack of transparency can impact a company's reputation because damaging trust can lead to backlash and stagnant growth.[203] Beyond the benefits to users, a prioritization of transparency improves internal collaboration by increasing ethical standards. [204] GoFundMe's site, however, has little information regarding fraud. This opacity may stem from the fear that addressing fraud will result in more payouts to donors who have been defrauded on the site.

[195] "Reporting a Campaign."
[196] "How to Determine if it is Safe to Donate to a Campaign."
[197] "Crowdfunding fraud: How to spot fake online fundraising campaign."
[198] "How to Determine if it is Safe to Donate to a Campaign."
[199] "Reporting a Campaign."
[200] "Here's how to spot a fake crowdfunding page."
[201] "How to Determine if it is Safe to Donate to a Campaign."
[202] Lupiáñez-Villanueva, et al., "Behavioural study on the transparency of online platforms."
[203] Weisbaum, "Trust in Facebook has dropped by 66 percent since the Cambridge Analytica scandal."
[204] "An Economy of Trust: How Transparency Is Changing the Tech Industry."

Regulation and Enforcement

Not all fraud can be prevented, but strict regulation and enforcement can deter fraudulent actors and make crowdfunding sites safer. The United States Department of Justice, U.S. Attorney General's Office, and Federal Trade Commission have no power to mitigate fraud through regulation or oversight of Internet sites and can only warn users of the possibility of scams. GoFundMe, like most websites, is a self-regulated entity and, therefore, has sole control over the platform's data. A lack of regulation has led third parties to launch efforts to curb fraud on crowdfunding sites. GoFraudMe.com, for example, was developed to catalog cases of fraud that have occurred on GoFundMe. The owner of GoFraudMe researches, collects, and archives fraudulent GoFundMe campaigns published by reputable news outlets and law enforcement agencies, giving donors the information they need to make informed decisions.

Recommendations

Crowdfunding companies like GoFundMe could effectively apply the above framework to mitigate and detect fraud with the help of information professionals. These trained custodians are experts in managing, filtering, and appraising content throughout the information pipeline. Information professionals specializing in informatics are authorities in "information-seeking behavior and information use; user-centered approaches to information system design; human-computer interaction; database design and management; and information policy, including intellectual property, informational privacy, and internet governance."[205] Each facet of an information professional's education is applicable to the recommendations below, and confirms

[205] "Areas of Specialization – UCLA GSEIS Information Studies."

the impact that information professionals can bring to organizations beyond the library and archive.

Recommendations to Effectively Implement Content Analysis and Moderation

1. Information professionals should be hired to advise crowdfunding companies regarding storage and organization processes for data and assets that are critical to content moderation and analysis.
2. Algorithmic bias is prevalent due to a lack of diversity in the technology industry as a whole.[206] Crowdfunding companies should consider diverse perspectives to reduce lopsided trends when creating algorithms, implementing machine learning, and hiring content moderators.[207]
3. Crowdfunding companies could create a variety of flags dedicated to different types or levels of fraud.[208] This would help companies systematically sift through flagged campaigns, utilize site visitors as a resource, and encourage reporting.

Recommendations to Increase Transparency

1. Changes in the user interface can effectively generate new avenues of trust.[209] This may include: verified user badges; prioritized placement of single-click access to flag fraudulent or suspicious campaigns; documentation that alerts donors as to whether the campaign's creator is or is not the recipient of donations; single-click access on all campaigns that redirects to the crowdfunding company's fraud and reporting policy; and a required photograph of the organizer.

---

[206] "Diversity in High Tech."
[207] Noble, *Algorithms of Oppression*, 3.
[208] Gillespie, *Custodians of the Internet*.
[209] Marsh & Dibben, "The Role of Trust in Information Science and Technology."

2. Information architecture professionals, such as user experience researchers and designers, can be critical in enhancing a website's transparent interface. User testing can help crowdfunding companies understand user needs and, ultimately, create a safer site.
3. Beneficiaries[210] should be asked for personal information that would allow for effective vetting of all individuals associated with a campaign. Such information may include the successful passage of a background check and the publishing of more information on the campaign page.
4. Crowdfunding platforms should explicitly state what proportion of each donation is used to keep the site safe. Explaining how funds are used to mitigate and detect fraud will increase transparency and boost the reputation of the company.

Recommendations to Establish External Regulation and Oversee Enforcement

1. Creating an open channel of communication between crowdfunding companies and government agencies could help the latter analyze campaigns while increasing data accuracy and reporting.
2. Seeking oversight from regulatory groups such as the Financial Action Task Force (FATF) could protect international users. The FATF is "an inter-governmental body that sets standards and promotes effective implementation of legal, regulatory and operational measures for combating money laundering and terrorist financing."[211]
3. Crowdfunding companies should hire information professionals to inform user protection policies and support platforms' compliance with regulation.

---

[210] An individual receiving donations, but not necessarily the administrator of the campaign.
[211] Miralis, "Cyber laundering."

Summary

Crowdfunding platforms provide a valuable resource, but the risk of disinformation can outweigh the positive impact these sites have on society. The sustained growth and popularity of crowdfunding is dependent on establishing and maintaining trust with users. This paper aims to highlight the infiltration of disinformation on crowdfunding sites and provide recommendations to mitigate and detect fraud through a three-pronged approach: content analysis and moderation, transparency, and regulation. Motivation on the part of crowdfunding companies is required to implement this framework, as eroding trust between platform and user can risk the moral and financial existence of a crowdfunding company. Information professionals would play a crucial role in implementing this approach, as threats of fraudulent content demonstrate the need for high-level information management and policy reform. Ultimately, information professionals' expertise could be key to diminishing disinformation's hold on crowdfunding sites.

Bibliography

"After Florida Shooting, GoFundMe Scams Sprout Up." (n.d.). https://www.inverse.com/article/41343-florida-gofundme-scams

"An Economy of Trust: How Transparency Is Changing the Tech Industry." (n.d.). Wired. https://www.wired.com/insights/2015/03/economy-trust-transparency-changing-tech-industry/

"AP Uncovers 'Scams and Waste' Rampant in Pulse Shooting GoFundMe Campaigns." (2016, August 29). GoFraudMe. http://gofraudme.com/ap-uncovers-scams-waste-rampant-pulse-shooting-gofundme-campaigns/

"Areas of Specialization – UCLA GSEIS Information Studies." (n.d.). https://is.gseis.ucla.edu/programs/mlis-degree/mlis-specializations/

Asiri, S. (2018, June 11). "Machine Learning Classifiers." Medium. https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623

Caspi, A., & Gorsky, P. (2006). "Online Deception: Prevalence, Motivation, and Emotion." *CyberPsychology & Behavior*, *9*(1), 54–59. https://doi.org/10.1089/cpb.2006.9.54

"Creating a GoFundMe From Start to Finish." (n.d.). GoFundMe Help Center. http://support.gofundme.com/hc/en-us/articles/360001992627

"Crowdfunding fraud: How to spot fake online fundraising campaigns." (2018). WKYC. https://www.wkyc.com/article/news/crowdfunding-fraud-how-to-spot-fake-online-fundraising-campaigns/95-522605592

"Disinformation: Oxford English Dictionary." (n.d.). https://www.oed.com/view/Entry/54579?redirectedFrom=disinformation#eid

"Diversity in High Tech." (n.d.). U.S. Equal Employment Opportunity Commission. https://www.eeoc.gov/special-report/diversity-high-tech-figure-2

"Everything You Need to Know About GoFundMe's Fees." (2018, December 11). GoFundMe. https://www.gofundme.com/c/blog/gofundme-fees

"Expenses for Baby Henley." (n.d.). GoFundMe Charity. https://charity.gofundme.com/expenses-for-baby-henley/fundraiser/charlesleber

"FTC examines crowdfunding." (2016, October 26). Consumer Information. https://www.consumer.ftc.gov/blog/2016/10/ftc-examines-crowdfunding

Garcia, S. E. (2018, November 15). "Couple and Homeless Man Behind Viral GoFundMe Campaign Are Charged With Conspiracy." *The New York Times*. https://www.nytimes.com/2018/11/15/nyregion/homeless-man-couple-gofundme-scam.html

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press. https://books.google.com/books?hl=en&lr=&id=cOJgDwAAQBAJ&oi=fnd&pg=PA1&dq=user+content+moderation&ots=PgTJBVQQo7&sig=Ky-SHuaIKyMsrbMffF8A56al3yE#v=onepage&q=user&f=false

"GoFundMe: #1 Free Fundraiser Platform—Crowdfund Online." (n.d.). GoFundMe. https://www.gofundme.com/

"GoFundMe 2018: A Year in Giving." (n.d.). GoFundMe. https://www.gofundme.com/2018

"GoFundMe 2019: A Year in Giving." (n.d.). GoFundMe. https://www.gofundme.com/2019

"GoFundMe Terms of Service." (n.d.). GoFundMe. https://www.gofundme.com/terms

"Here's how to spot a fake crowdfunding page." (2017, April 3). *BBC Newsbeat*. http://www.bbc.co.uk/newsbeat/article/39478846/heres-how-to-spot-a-fake-crowdfunding-page

Hou, Y.-T., Chang, Y., Chen, T., Laih, C.-S., & Chen, C.-M. (2010). "Malicious web content detection by machine learning." *Expert Systems with Applications*, *37*(1), 55–60. https://doi.org/10.1016/j.eswa.2009.05.023

"How to Determine if it is Safe to Donate to a Campaign." (n.d.). GoFundMe Help Center. http://support.gofundme.com/hc/en-us/articles/115015913668

"Important Verification and Withdrawal Deadlines." (n.d.). GoFundMe Help Center. http://support.gofundme.com/hc/en-us/articles/204119238

"Kentucky Parents 'Angry' Over Fake GoFundMe Campaign For Their Deceased Baby." (2016, August 6). GoFraudMe. http://gofraudme.com/kentucky-parents-angry-fake-gofundme-campaign-deceased-baby/

"Key Crowdfunding Statistics." (n.d.). https://www.startups.com/library/expert-advice/key-crowdfunding-statistics

Koteyko, N. (2010). "Mining the internet for linguistic and social data: An analysis of 'carbon compounds' in Web feeds." *Discourse & Society*, *21*(6), 655–674. https://doi.org/10.1177/0957926510381220

Lupiáñez-Villanueva, F., Gaskell, G., Tornese, P., Vila, J., Gómez, Y., Allen, A., Veltri, G., & Codagnone, C. (2018). "Behavioural study on the transparency of online platforms." https://ictlogy.net/bibliography/reports/projects.php?idp=3535

Luu, C. (2019, September 11). "The Life Changing Linguistics of Nigerian Scam Emails." JSTOR Daily. https://daily.jstor.org/the-life-changing-linguistics-of-nigerian-scam-emails/

Marsh, S., & Dibben, M. R. (2003). "The Role of Trust in Information Science and Technology." *Annual Review of Information Science and Technology*, *37*(1), 465–498. https://doi.org/10.1002/aris.1440370111

Miralis, N. G. (n.d.). "Cyber laundering: What are some key challenges in detection and regulation?" *Lexology*. https://www.lexology.com/library/detail.aspx?g=98410b41-b400-4ab4-9575-29184ab74fe2

Noble, S. U. (2018). Introduction: The Power of Algorithms. In *Algorithms of Oppression* (pp. 1–14). NYU Press; JSTOR. https://doi.org/10.2307/j.ctt1pwt9w5.4

Reporting a Campaign. (n.d.). GoFundMe Help Center. http://support.gofundme.com/hc/en-us/articles/203604694

Roberts, S. T. (2017). "Content Moderation." *Encyclopedia of Big Data*. https://escholarship.org/uc/item/7371c1hf

Shankar, V., Urban, G. L., & Sultan, F. (2002). "Online trust: A stakeholder perspective, concepts, implications, and future directions." *The Journal of Strategic Information Systems*, *11*(3), 325–344. https://doi.org/10.1016/S0963-8687(02)00022-7

The GoFundMe Guarantee. (n.d.). GoFundMe. https://www.gofundme.com/c/safety/gofundme-guarantee

Thorpe, D. (n.d.). "Worries About Fraud Top List Of Crowdfunding Concerns." Forbes. https://www.forbes.com/sites/devinthorpe/2017/11/29/worries-about-fraud-top-list-of-crowdfunding-concerns/#6beff58648f0

Victor, D. (n.d.). "Woman and Homeless Man Plead Guilty in $400,000 GoFundMe Scam." The New York Times. https://www.nytimes.com/2019/03/07/us/gofundme-homeless-scam-guilty.html

Weisbaum, H. (2018). "Trust in Facebook has dropped by 66 percent since the Cambridge Analytica scandal." NBC News. https://www.nbcnews.com/business/consumer/trust-facebook-has-dropped-51-percent-cambridge-analytica-scandal-n867011

ACCESSIBILITY STATEMENT

I used the Web Content Accessibility Guidelines (WCAG) 2.2 published by the World Wide Web Consortium (W3C) on February 27, 2020 to inform design decisions regarding the accessibility of my site. This included, but was not limited to, the use of large-scale text, headings to organize content, color changes to indicate focus, and color contrast for ease of reading.

I have provided a PDF version of my portfolio to supplement the website.