

## From Archive to Database: Using Crowdsourcing, TEI, and Collaborative Labor to Construct the *Maria Edgeworth Letters Project*

Hilary Havens <hhavens1\_at\_utm\_dot\_edu>, University of Tennessee, Knoxville  <https://orcid.org/0000-0002-8632-5847>

Eliza Alexander Wilcox <swilcox7\_at\_vols\_dot\_utm\_dot\_edu>, University of Tennessee, Knoxville

Meredith L. Hale <mhale16\_at\_utm\_dot\_edu>, University of Tennessee, Knoxville  <https://orcid.org/0000-0001-9740-7437>

Jamie Kramer <jkrame21\_at\_vols\_dot\_utm\_dot\_edu>, University of Tennessee, Knoxville

### Abstract

This article unpacks the archival, textual, and encoded layers that comprise the *Maria Edgeworth Letters Project (MELP)*, an open-access digital archive containing the correspondence of the Anglo-Irish Regency author Maria Edgeworth and her circle. These layers reveal the impossibility of flattening or standardizing our work and instead advocate for a more inclusive and collaborative digital humanities model that accommodates both institutional and volunteer labor. Just as different methods were used to approach each archive and manage our project across multiple institutions, each transcription requires a different level of care, especially as various notes and collaborators are cited in the final project. Through the use of TEI, we can flexibly represent diverse aspects of each letter while still maintaining a database-readable structure. We endeavor to connect each person, place, or work identified in Edgeworth's letters and our database to a larger network of linked data in order to place our project in conversation with other archival resources. For entities that are unidentified or unknown, we create new name authority files or produce internal data files that can be viewed by our collaborators and users. *MELP's* flexible structure thus allows it to strive for interoperability while refusing to efface the individual traces of its collaborators, entities, and material artifacts.

### Introduction

The Anglo-Irish Regency author Maria Edgeworth (1768-1859) was “the most commercially successful novelist of her age” [McCormack 2004], yet her important, sprawling correspondence has only been published in excerpts and partial editions up to this point.<sup>[1]</sup> The primary goal of the *Maria Edgeworth Letters Project (MELP)* is to produce a searchable corpus of Maria Edgeworth's correspondence, which contains over ten thousand sheets spread across more than forty institutions. In addition to the infrastructural obstacles to accessing and collating Edgeworth's letters is the reality that a physical edition would be unwieldy as a research tool and unlikely to be completed. *MELP's* choice to produce a digital archive was also influenced by changing institutional funding structures. Researchers continue to face difficult choices when doing archival research, whether as a result of budget cuts, climate-conscious travel decisions, or lack of time [Wright 2023]. Thus, *MELP* has developed a hybrid and sustainable digital workflow incorporating both institutional and volunteer collaborators that flexibly supports an increasing corpus over multiple stages of transcription and participation.

By *hybrid*, *flexible*, and *sustainable*, we mean to gesture to the uneasy balance between physical archives and digital databases and between individual contributions and “cleaned” large-scale data. These tensions underpin *MELP's* digital workflow, as it transforms Maria Edgeworth's large and dispersed physical archive into a searchable database. Ed Folsom, one of the directors of the open-access *Walt Whitman Archive*, notes, “Often we will hear *archive* and *database* conflated, as if the two terms signified the same imagined or idealized fullness of evidence” [Folsom 2007, 1575]. However, despite the shared “desire for completeness” in both archives and databases, archival attention to physical preservation does not automatically dovetail with the detail and information that databases categorize. Even if

1

2

Edgeworth's letters were stored in a single physical institution, users would access and use those physical letters in ways different from how those same letters would be structured and presented in a database. Folsom's metaphor of database as "rhizome" is apt: "the subterranean stem that grows every which way and represents the nomadic multiplicity of identity — no central root but an intertwined web of roots" [Folsom 2007, 1573]. The iterative XML tagging process that *MELP* uses to create entities contributes to the larger "web" of tagged persons, places, and works within the archive, rather than focusing on the "central" or chronological effort of quickly uploading digital reproductions of the letters themselves.

Folsom's metaphor that values the "intertwined web" over the "central root" can also be translated to *MELP*'s data processing workflows. *MELP* approaches collecting, transcribing, and encoding through a lens that prioritizes editorial choices to retain as much of Edgeworth's correspondence "as-is" rather than organizing or transforming that data into a "cleaned" version. By "as-is", we mean creating a digital corpus that represents Edgeworth's original letters in a form as close to the originals as possible, even if that means leaving in unclear word choices, unknown subject entities, and confusing or contradictory text. The transcription checking and encoding process involves encoders making editorial choices rather than encoders functioning as data "cleaners" who have to fix or correct data and then submit it to Edgeworth subject experts. Moreover, these encoders use a flexible tagging structure that either links to or creates authority files for persons, places, and works, depending on whether those entities appear in external authority services like VIAF, GeoNames, or Wikidata. Our process, then, views "cleaning" data not as a singular process, but rather, as a complex, labor-intensive set of practices that align with the argument Katie Rawson and Trevor Muñoz make in "Against Cleaning" [Rawson and Muñoz 2019, 280]. Rawson and Muñoz challenge the assumption of an underlying order to data that renders human connection as depressingly two-dimensional, which can be extrapolated to the complicated relationship connecting digital humanities and long eighteenth-century studies with data in archives.

*MELP* thus prioritizes data and its importance for archival study over narratorial unity. Viewing the humanist goals of a digital humanities research project as separate from its data collection methodologies is a mistake that some humanist researchers have made that ultimately undermines the "powerful critiques of the existing systems of data analysis" [Rawson and Muñoz 2019, 281]. In eighteenth-century studies, there have been many critiques against adopting digital methodologies without critically understanding those methods. For example, Cassidy Holahan positions *Eighteenth Century Collections Online (ECCO)*, the largest digital database containing searchable facsimiles of eighteenth-century texts, as an "opaque archive" [Holahan 2021, 804]. Holahan defines "opaqueness" through the way scholars have approached *ECCO* as a natural resource without interrogating its "scope, biases, and [the] limitations of its contents" [Holahan 2021, 804]. Tracking and making visible the forces, whether physical or ideological, that shaped *ECCO* helps dispel the "sense of comprehensiveness, authority, and neutrality that hide the underlying history and decisions that have shaped the collection" [Holahan 2021, 824]. In a similar fashion, our approach to Maria Edgeworth's correspondence recognizes that our editorial voice might obscure the relationships and networks in the corpus and seeks to address those implicit biases by giving a transparent account of our process and choices in this essay. *MELP* views each step of the corpora building process as an opportunity to reconceptualize how data can be described and organized to embrace the "messiness" of the corpus itself. Our approach to Edgeworth's correspondence is to capture, organize, and encode data without placing it into a larger narrative. In digitizing each letter, uploading it for crowdsourced transcription, and then checking and encoding that transcription, the data's messiness is treated as inseparable from its meaning.

The three different layers of our project — archive, transcription, and encoding — reveal the challenges inherent in uniting texts, people, and institutions from different sources. While other archives seek to efface the distinctions between the disparate entities that comprise them, *MELP* embraces a transcription and encoding structure that promotes exchange with other sources while preserving unique aspects of our data: while we do reformat images and some metadata from the more than thirty archives represented in *MELP*, we are still able to preserve unique archival and transcription aspects through the flexibility inherent in the Text Encoding Initiative (TEI) schema and eXtensible Markup Language (XML) as data format. As "extensible" suggests, XML was created with the intention of the data format being stretched to meet new use cases despite some limitations due to its hierarchical nature. TEI, as a schema that is manifested in XML and that was created to deal with the intricacies of text, also supports *MELP*'s aims. Similarly,

in the lists of people, places, and works that undergird our project, we have prioritized linking to a series of named authorities to place our project in conversation with others, leaving room for lesser-known entities that we establish ourselves — either through our own creation of name authorities or internal project documentation. There is also an unwieldiness in the combination of institutional and volunteer labor that powers our project: the project is managed and coordinated across five institutions located in four different states; the transcriptions have been largely supplied by volunteers through the Zooniverse platform; and though institutionally-funded research assistants have standardized the transcriptions, the Talk boards and planned annotations allow individual contributors to preserve their own voices. Through our descriptions of the different layers of *MELP*, we hope to model the ways in which a digital archive can flexibly embrace the diverse array of contributions and contributors without flattening its data.<sup>[2]</sup>

## Digital Editions of Correspondence

*MELP*'s attention to Maria Edgeworth is part of a digital humanities lineage of projects that seek to compile and publish manuscript correspondence of a single author or literary circle, which often reveals unknown, undertheorized, or misunderstood aspects of the authors' lives and the world in which they lived. The *Shelley-Godwin Archive*, the *Carlyle Letters Online*, *The Walt Whitman Archive*, and *Digital Mitford* exemplify the kind of work already done in this area of digital humanities, and each project contributes to the broader understanding of how text encoding might be used to reshape the canonical and historical understanding of their central figures.<sup>[3]</sup> These previous projects have provided a template for *MELP*'s sustainable process suitable for encoding over ten thousand pages of correspondence while recognizing institutional collaborators (paid from the institutions hosting the project) and volunteer collaborators (unpaid, often virtual, and external to the institutions). *MELP* builds upon the accomplishments of its predecessors by inviting the larger public to participate in the project and by ensuring the interoperability of its letters and database while also prioritizing a non-hierarchical view of Edgeworth's letters.

Traditional standard and scholarly editions of correspondence have generally been published in print by university presses, though the increasing turn to the digital has heralded the arrival of authoritative digital editions. Gone are the days when presses would greenlight dozens of volumes dedicated to a single correspondence, such as W.S. Lewis's 48-volume scholarly edition of Horace Walpole's correspondence published with Yale University Press. While the material heft of the print editions is attractive, their cost — usually at least \$100 a volume — is generally prohibitive to those unaffiliated with a large university library. The accessibility and reach of digital editions is connected to a shift towards a greater democratization of knowledge and open access. Moreover, traditional print editions prioritize a chronological presentation of correspondence, which can flatten the connections between letters written at different times. A primary aim of *MELP* is to create a digital database capable of capturing the inherent complexity and “messiness” of a network as large as Edgeworth's across a wide range of dates and recipients. On a related note, one of the great strengths of digital editions of correspondence is that the text can easily be searched using the original letter images, and digital encoding is occasionally available in parallel. There are several projects, such as the *Carlyle Letters Online*, *The Walt Whitman Archive*, and the *Shelley-Godwin Archive* that unite letter photographs, transcriptions, and encodings in a pop-up or diplomatic presentation that is available to the public.<sup>[4]</sup> By encoding the transcriptions, these archives signal their interoperability — the ability of these archives to exchange information with other resources and each other, especially if sustainable file formats are used [Muñoz and Viglianti 2015].<sup>[5]</sup>

Much of the labor for these digital archives is funded and completed by institutions, though there is a turn towards pedagogy and volunteerism that has influenced *MELP*'s structure and workflows. The *Carlyle Letters*, for instance, outsourced their encoding to DNC Data Systems of Mumbai, India, which was then checked by a team of copyeditors. The *Shelley-Godwin Archive*'s encoding has largely been performed by its editorial team, though they signal their desire for the archive to become a collaborative project. The project that has served as a collaboratively encoded model for *MELP* since its inception and that has influenced our coding guidelines is *Digital Mitford*.<sup>[6]</sup> *Digital Mitford* is similarly concerned with interoperability through its use of TEI and through its recruiting and training of volunteer coders.<sup>[7]</sup> To these ends, the editors offer a regular summer coding school that trains researchers in TEI through collaborating with their project. There is also an application process to join the project as a scholarly editor. This project is more open to volunteers than most and contains excellent documentation: in their description of *Digital Mitford*, Elisa Beshero-Bondar

6

7

8

and Kellie Donovan-Condrón detail exactly how information between their project and others will be exchanged within the “network of linked data, a digital database from which we can extract and study information we are collecting about people and texts of the nineteenth century” [Beshero-Bondar and Donovan-Condrón 2017, 140]. They highlight the “systematic and transferrable methods of editing and text encoding” that they teach as well as careful forms of documentation as ways of making their project both a collaborative and pedagogical enterprise [Beshero-Bondar and Donovan-Condrón 2017, 192].

Learning TEI encoding, even with the help of a coding school, is a substantial investment that is also a barrier to participation, which is why *MELP* instead has invited volunteers to assist with initial letter transcription through the Zooniverse platform. Zooniverse is “the world’s largest and most popular platform for people-powered research...made possible by volunteers...who come together to assist professional researchers”.<sup>[8]</sup> Zooniverse is a crowdsourcing platform that enables contributions to academic research without requiring specialized expertise. There is a wide range of projects: many of them are in the sciences, and some of the most popular projects are related to identifying astronomical features and classifying nature. Volunteers are able to contribute to these research projects through Zooniverse’s user-friendly interface by answering straightforward questions, taking an image survey, drawing shapes and figures, and transcribing text, the latter of which is central to *MELP*. Through our work with the NEH Institute for Advanced Topics in the Digital Humanities, “Building Capable Communities for Crowdsourced Transcription”, we spent an eighteen-month period building and establishing *MELP*’s Zooniverse website<sup>[9]</sup>, which was launched in spring 2022 and is in continuous use as we add more letters from archives to be transcribed.

9

There have been numerous other scholarly editing projects similarly reliant on volunteer labor that are hosted through the Zooniverse platform. The *Davy Notebooks Project*, which aims to transcribe all 75 notebooks of Sir Humphry Davy (1778-1829), arguably the most famous nineteenth-century chemist, will eventually be published online in Lancaster Digital Collections. The 29 volumes in *The Diaries of Michael Field* were also transcribed, at least in part, by Zooniverse volunteers, and the eventual goal is to encode the letters in TEI “within a network of linked resources” to support interoperability. *MELP* is most similar to *The Diaries of Michael Field*, as both projects combine volunteer transcription work with institutionally-funded encoding; the interconnectedness of volunteer and institutional labor is manifested across *MELP*’s multiple layers and the flexible structure of the project management, metadata, transcriptions, and encoding.

10

## Data Organization and Project Management

The central and initial challenges that *MELP* and comparable digital archives have faced can be addressed with strong project management: finding, securing, organizing, and getting permissions for the material that comprises the foundation of the archive, as well as constructing the project team. Even though project management is “primarily used in business”, according to Lynne Siemens, for large-scale projects like *MELP*, there is a “growing need for public accountability by funding agencies and others which requires, even demands, successful project completion facilitated by detailed and realistic planning”, connected to the “coordination of people, financial resources, and tasks” [Siemens 2021].<sup>[10]</sup> Collaborative digital work often goes underappreciated in the humanities because many college promotion and tenure guidelines prioritize traditional publications, though they should, as Kathleen Fitzpatrick argues, “rethink the ways that we give credit for such projects” [Fitzpatrick 2011].<sup>[11]</sup> Because collaborative work incorporates the efforts of more individuals, the scope, speed, and quality of such projects generally exceed the capabilities of a single individual, with the work and administrative burdens shared [Siemens 2015, 358]. Large-scale collaborative projects, however, need a project manager in order to be successful: according to Erik Ernø-Kjølhede, “The central task of any project manager regardless of her field is to navigate between the conflicting demands of time, cost and performance. The project manager constantly has to weigh these demands against each other and trade off one against the other” [Ernø-Kjølhede 1999, 13]. Project management makes it possible to negotiate the rhizomatic structure of *MELP*, including project deliverables and deadlines, personnel tasks and training, and the system of workflows developed generally by the librarians who have worked on *MELP*.<sup>[12]</sup> *MELP*’s workflows contain an “‘authoritative’ record of roles, tasks, outcomes and relationships” so that our work can be understood and reproduced by others [Siemens 2015, 358]. Furthermore, good project management benefits our student assistants: we aim, as Siemens does, for our students to

11

see “themselves as collaborators with few barriers between themselves and their supervisors” and learn important skills in project management and collaboration by working on interdisciplinary teams [Siemens 2015, 357, 362]. Above all, we agree with Jason Boyd, who views project management as “a key scholarly practice in the digital humanities and in the humanities more broadly” that makes “open to scrutiny many tacitly understood practices” [Boyd 2022].<sup>[13]</sup>

Due to the dispersal of our project team — across five institutions in four states — project management has played an important role in communication and in negotiation with archives regarding their Edgeworth holdings. *MELP* began in 2017, but it was only in 2020, with the improvement of digital meeting technologies like Zoom and the more widespread digitization of library manuscript holdings that project team meetings and *MELP*'s virtual archive became, respectively, more regular and larger. Like the *Walt Whitman Archive* and *Digital Mitford*<sup>[14]</sup>, *MELP* has compiled and united material from over 30 archives and universities in North America and Europe that was previously dispersed and, in some cases, unknown.<sup>[15]</sup> The hitherto gradual shift to digital resource sharing accelerated during the pandemic, which meant that, to improve accessibility and promote scholarship under lockdown conditions, many of the institutions that held pieces of Edgeworth's correspondence were not only willing to digitize their materials, but also provided them for free and occasionally published them on their own collections' websites<sup>[16]</sup>, especially the libraries located at large North American universities (such as Harvard University's Houghton Library and Yale University's Beinecke Library) and libraries that were large research institutions in and of themselves (the Huntington Library and the New York Public Library). Bringing this correspondence together digitally is a challenge, even after the pandemic, and requires the core workflow of the project to be agile in proportion to the responsiveness of each institution. Communication with these institutions was the primary difficulty, especially early in the pandemic when many other scholars were also requesting digitized material. The large amount of digitized manuscript material we were able to secure from these libraries has been an important contribution to open-access scholarship, though the open, free, and collaborative resource sharing that marked the early months of the pandemic is reverting in some cases to pre-2020 standards.<sup>[17]</sup>

Different holding institutions have different copyright policies, especially institutions in the United Kingdom that operate under the Copyright, Designs and Patents Act of 1988. According to this legislation, unpublished literary works, which include letters, were granted an automatic copyright until 2039. We consulted with two copyright librarians — Molly Keener (Wake Forest University) and Peter Hirtle (Berkman Center for Internet and Society at Harvard University) — who determined that British archives do have the right under a copyright exemption to share Maria Edgeworth's letters for publication<sup>[18]</sup>, which applies when the work is available to the public in an archive or similar institution, the work is at least 100 years old, the work's author has been dead for at least 50 years, and the present copyright owner is unknown to the publisher. Edgeworth died in 1849, and after so much time, the present copyright owner is unknown and unknowable. By this reasoning, we were able to convince English institutions to grant us access. Similarly, we had protracted negotiations with the Bibliothèque de Genève about access and copyright, which were conducted in French. The Bibliothèque de Genève holds Edgeworth's interesting correspondence with the philosopher Etienne Dumont, and after purchasing and receiving the images, we had to send a physical letter in French to request exemption from the several thousand dollar publication fees to display the images since our archive is non-profit and will be open-access. Our negotiations with the various libraries are a testament to the necessity of having a flexible approach when interacting with other institutions.

We also needed to be flexible in our use of metadata and our image processing decisions, which corresponded with the uneven levels of labor in preparing letters for transcription and encoding. Different institutions provided different amounts of metadata about their letters: sometimes we received no information besides Edgeworth's authorship, while in other cases, dates, locations, and people (connected with their name authorities) were provided. In all cases, we had to review and confirm or generate metadata associated with these letters. It was rare for holding institutions to provide information beyond the author, recipient, and date. Occasionally the date was incorrect because of Edgeworth's difficult handwriting, or the author or recipient was unidentified or misidentified. These details are important because one of the major contributions of *MELP* is to provide a comprehensive global list of Edgeworth's correspondence. Metadata is necessary to provide an accurate accounting of letter location, author, recipient, date, and permissions. We have tracked all of this information in a large spreadsheet (Figure 1), which will be available on our beta website soon.

12

13

14

Local file name (include extension)	Title	Creator	Date of Original	
File Name	dc:title	dc.creator	dc.created	dc.rights
filename	title	Author of Letter	Original creation date (date letter was written, if a date [creation date of letter] licensing	
Edgeworth_Bodleian_696_1818-01-04_page01.jpg	Letter from Black Castle to Mrs. Edgeworth, January 4th, 1818	Edgeworth, Maria	01/04/1818	
Edgeworth_Bodleian_696_1818-01-04_page02.jpg	Letter from Black Castle to Mrs. Edgeworth, January 4th, 1818	Edgeworth, Maria	01/04/1818	
Edgeworth_Bodleian_696_1818-01-04_page03.jpg	Letter from Black Castle to Mrs. Edgeworth, January 4th, 1818	Edgeworth, Maria	01/04/1818	
Edgeworth_Bodleian_696_1818-01-04_page04.jpg	Letter from Black Castle to Mrs. Edgeworth, January 4th, 1818	Edgeworth, Maria	01/04/1818	
Edgeworth_Bodleian_696_1818-01-04_page05.jpg	Letter from Black Castle to Mrs. Edgeworth, January 4th, 1818	Edgeworth, Maria	01/04/1818	
Edgeworth_Bodleian_696_1818-01-04_page06.jpg	Letter from Black Castle to Mrs. Edgeworth, January 4th, 1818	Edgeworth, Maria	01/04/1818	
Edgeworth_Bodleian_696_1818-03-03_page01.jpg	Letter from [Maria Edgeworth] to Mrs. Edgeworth, March 3rd, 1818	Edgeworth, Maria	3/15/1818	
Edgeworth_Bodleian_696_1818-03-03_page02.jpg	Letter from [Maria Edgeworth] to Mrs. Edgeworth, March 3rd, 1818	Edgeworth, Maria	3/15/1818	
Edgeworth_Bodleian_696_1818-03-03_page03.jpg	Letter from [Maria Edgeworth] to Mrs. Edgeworth, March 3rd, 1818	Edgeworth, Maria	3/15/1818	
Edgeworth_Bodleian_696_1818-03-03_page04.jpg	Letter from [Maria Edgeworth] to Mrs. Edgeworth, March 3rd, 1818	Edgeworth, Maria	3/15/1818	
Edgeworth_Bodleian_696_1818-03-03_page05.jpg	Letter from [Maria Edgeworth] to Mrs. Edgeworth, March 3rd, 1818	Edgeworth, Maria	3/15/1818	
Edgeworth_Bodleian_696_1818-03-03_page06.jpg	Letter from [Maria Edgeworth] to Mrs. Edgeworth, March 3rd, 1818	Edgeworth, Maria	3/15/1818	
Edgeworth_Bodleian_696_1818-05-01_page01.jpg	Letter from Edgeworthstown to Mrs. Edgeworth, May 1st, 1818	Edgeworth, Maria	5/1/1818	
Edgeworth_Bodleian_696_1818-05-01_page02.jpg	Letter from Edgeworthstown to Mrs. Edgeworth, May 1st, 1818	Edgeworth, Maria	5/1/1818	
Edgeworth_Bodleian_696_1818-05-01_page03.jpg	Letter from Edgeworthstown to Mrs. Edgeworth, May 1st, 1818	Edgeworth, Maria	5/1/1818	
Edgeworth_Bodleian_696_1818-05-01_page04.jpg	Letter from Edgeworthstown to Mrs. Edgeworth, May 1st, 1818	Edgeworth, Maria	5/1/1818	
Edgeworth_Bodleian_696_1818-05-16_page01.jpg	Letter from [Maria Edgeworth] to Mrs. Edgeworth, May 16th, 1818	Edgeworth, Maria	5/16/1818	
Edgeworth_Bodleian_696_1818-05-16_page02.jpg	Letter from [Maria Edgeworth] to Mrs. Edgeworth, May 16th, 1818	Edgeworth, Maria	5/16/1818	
Edgeworth_Bodleian_696_1818-05-16_page03.jpg	Letter from [Maria Edgeworth] to Mrs. Edgeworth, May 16th, 1818	Edgeworth, Maria	5/16/1818	
Edgeworth_Bodleian_696_1818-05-16_page04.jpg	Letter from [Maria Edgeworth] to Mrs. Edgeworth, May 16th, 1818	Edgeworth, Maria	5/16/1818	

**Figure 1.** List of Edgeworth's letters and metadata. The current sheet displays letters held at the Bodleian Library at the University of Oxford.

By compiling and gathering this data, it will be possible to attribute and acknowledge our letters correctly, and it will also be possible for users to locate the physical letters themselves. Images correspondingly have been delivered in various file formats and sizes in line with the internal guidelines of our partner institutions. While we have retained the original files and will display them on our archive, before they can be transcribed, they need to be reduced to 1 MB to be uploaded on the Zooniverse platform. Handling image processing and metadata from different institutions while preserving the original library content and reformatting it are other balancing acts our project performs in order to prepare letter images for transcription on the Zooniverse interface.

15

Besides Zooniverse, the applications the MELP team uses most frequently to coordinate the project's work are Google Drive, GitHub, and Zoom. Given the ubiquity of Google Drive, it has proved to be an effective tool for MELP's inter-institutional collaboration. MELP's Google Drive currently holds data (e.g., transcribed letters and image files), documentation on processes, and completed scholarship on the project. While Google Drive has many strengths, the MELP team decided that some content, particularly XML documents, would be more effectively shared, edited, and reviewed in a GitHub repository. The MELP GitHub and its associated repositories were created to meet this need. This decision did increase the barrier to participation, as few contributors had previously used GitHub, but the benefits were worth the effort. In particular, the MELP team found that it was hard to keep track of the authoritative version of XML files, like the TEI template, on Google Drive. It also was difficult to suggest edits to particular lines of code. By using GitHub, the project team was able to address these issues. We conducted two training sessions, one simply introducing the platform and another hands-on workshop for committing and pushing content, to help institutional contributors gain familiarity with GitHub. MELP uses GitHub exclusively for XML documents and processes since Google Drive is sufficient for storing text files and tabular metadata. And while Google Drive and GitHub effectively support asynchronous work, Zoom meetings have been essential to collaboratively planning, managing, and working on MELP. Given the inter-institutional nature of the project and physical separation of the project team, having reliable video conferencing software like Zoom has allowed MELP to build momentum over the past three years through monthly meetings.

16

## Transcription

MELP's commitment to a collaborative, inter-institutional approach is essential to the project beyond the initial data management and project organization stage. Once the archival data has been gathered and organized, transcription begins, which can take many forms depending on a project's desired outcomes. Other digital humanities archives have employed a variety of different labor practices; these are largely institutional labor practices that rely on labor outsourced to computers, underpaid digital laborers, and students. Using computers to do the majority of the transcription work would have been a feasible alternative to crowdsourcing. Handwritten Text Recognition (HTR) engines have benefited from the recent and substantial developments in AI technology, and the appearance of HTR

17

technology in published research is “international and rapidly growing” [Nockels et al. 2022, 367]. Transkribus, in particular, is “the most commonly used HTR tool in the cultural heritage space”, with approximately 1700 monthly users. Yet an HTR engine like Transkribus would be in opposition to *MELP*'s collaborative goals, and relying on AI to transcribe letters moves us away from the most important dimension of the humanities: the human. Involving volunteers and the public in larger research-based projects promotes “citizen science”, as *MELP* is built on “researchers interact[ing] with the public to achieve a collective goal” that furthers the humanities [Nockels et al. 2022, 377]. These “citizen scholars” are essential to how we define “collaboration” in the digital humanities, and they can more easily access and engage with digital humanities projects through online crowdsourcing than through more “traditional”, or offline approaches [Arbuckle 2019, 293–294]. [Brown 2016, 49].

Since *MELP*'s commitments to open-access and collaboration with the public are antithetical to using machine labor for transcription, other alternatives include student labor or other outsourced, cheap labor. This “effacement of student labor”, whether student or outsourced, frequently occurs in digital projects, as Spencer Keralis argues [Keralis 2018, 278]. Students are institutionally used for labor through unpaid student internships, minimum wage student employee positions, or faculty who integrate digital humanities (DH) project work into their syllabi under the auspices of gaining translatable and collaborative skills. Keralis states that this practice is “naturalized into the fabric of digital pedagogy”, yet “student labor in the classroom is never not coerced” since “students will feel coerced to participate in the professor's project ... even if an alternative assignment is offered” [Keralis 2018, 278, 286]. In response to these labor practices, Haley Di Pressi et al. argue that students earn the right to appear as project collaborators on any project to which they have made “substantive contributions” [Di Pressi et al. 2015]. *MELP* has consistently followed the labor practices outlined in the “Student Collaborator's Bill of Rights” by Di Pressi et al., which provides guidelines to prevent the exploitation of students in any DH work setting, asserting that “[s]tudents should not perform mechanical labor ... without pay” [Di Pressi et al. 2015, ¶5]. These forms of “mechanical labor” and detailed, single-action tasks are the same piecemeal assignments that DH project editors can alternately outsource using low-cost labor platforms, such as Amazon's Mechanical Turk, and this is a frequent practice. Christlein et al. outsourced labor for a transcription project to “naïve transcribers” in Vietnam and described this labor choice as “cost-effective” and “powerful” [Christlein et al. 2018, 7]. However, Mechanical Turk presents an ethically questionable solution in its frequent outsourcing of labor to developing countries as well as to workers located in the United States. Cushing reported that Mechanical Turk workers were making an average of \$1.50 per hour; seven years later, Hara et al. reported this average as \$2 per hour [Cushing 2012, ¶22] [Hara et al. 2018, 11]. Fred Benenson's *Emoji Dick* (2019) is one such digital project that was finished through the outsourced labor of hundreds of anonymous workers employed by Mechanical Turk. Workers were paid five cents for every sentence of Herman Melville's *Moby-Dick* (1851) that they translated into emoji; they were also paid two cents for each vote on the best translation for each sentence [Benenson 2010, vii]. Although the Mechanical Turk employees who worked on *Emoji Dick* made creative decisions, both by translating sentences into emoji and by voting for the best translation, no one who participated in the creation of the text is individually credited, except for Benenson. [19]

Instead of relying on cheaply outsourced or unpaid student labor, *MELP* employs a flexible, hybrid model of volunteer labor and institutional labor.<sup>[20]</sup> Over the course of *MELP*, undergraduate and graduate student workers have been employed by Texas A&M, Tennessee, Wake Forest, and Xavier, and these students have all been paid hourly wages and are recognized as collaborators. As for all of the Zooniverse volunteers, they are individually credited for their transcriptions and annotations on the Zooniverse project page, and their usernames (and real names if they chose to share them) will also appear in the completed digital archive. Crowdsourcing through volunteer-based platforms like Zooniverse, where labor is voluntary and non-coercive, provides an ethical alternative to questionable labor practices. While many of the digital archives that have served as models for ours do not have an open mechanism for accepting volunteer contributions, one of the strengths of *MELP* is its large volunteer presence, which is evidenced by our work using the Zooniverse platform. The importance of “collective intelligence” is the throughline of James Surowiecki's *The Wisdom of Crowds*, in which he argues that “[i]f you put together a big enough and diverse enough group of people and ask them to ‘make decisions affecting matters of general interest’, that group's decisions will, over time, be ‘intellectually [superior] to the isolated individual’, no matter how smart or well-informed he is” [Surowiecki 2004, xvii]. Such “collective intelligence” can be harnessed through digital platforms such as Zooniverse. More than a million volunteers come

together “to enable research that would not be possible, or practical, otherwise”.[21] Through the Zooniverse platform, volunteers are able to transcribe letters simultaneously to achieve consensus with multiple others. The transcription task interface is logical and contains instructions, as is shown in Figure 2.

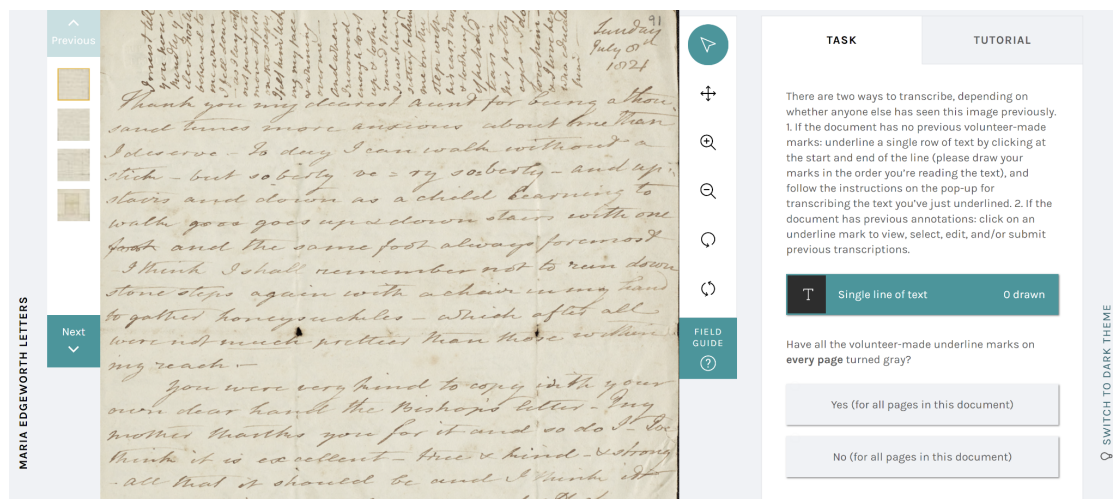


Figure 2. MELP transcription interface on Zooniverse.

Zooniverse projects are required to conform to set templates that permit a limited amount of information directly in the interface, so we also included a Field Guide and Tutorial (visible in Figure 2) to share additional information. In fact, many of the entries in the Field Guide were suggested by volunteers, who had questions about the ampersand, the Edgeworth family, how to handle numbers and tables, common but difficult to read names within the correspondence, and how to approach stray pink underline marks. Some of the volunteers even generated language for the Field Guide, including the entry on the different colors of underline marks.[22]

The Talk boards on Zooniverse are another way for volunteers to interact directly with researchers by posting a comment related to a letter after finishing a transcription or on the general message boards. There have been hundreds of posts on these message boards, which is a testament to the dedication of our volunteers, who tend to be, as Melissa Terras writes about crowdsourcing volunteers, “highly motivated and skilled individuals” who are “committed to the project for the long term, appreciate that it is a learning experience, which gives them purpose and is personally rewarding, perhaps because they are interested in it, or see it as a good cause” [Terras 2015, 424, 427]. One of the earliest contributions to MELP’s Talk boards is a handwriting guide called “Tips for Transcribing Maria Edgeworth’s Handwriting” that began as an undergraduate research assistantship project to advise transcribers on reading Romantic era handwriting and tricky letters. The guide now has substantial contributions from volunteers, including entries addressing particular words, letters, and names, as well as connections to other Zooniverse projects.[23] As of August 2023, there are 203 discussions and 845 comments on individual letters on Zooniverse that have helped us figure out the meaning of names and terms like “vandyke”, which is a type of lace[24], as well as identify connections to other projects and archives, such as the *Davy Notebooks Project* on Zooniverse and the Beddoes archive.[25] The interactions between volunteers and researchers on the Talk boards also serve a pedagogical function, as researchers share information on Edgeworth and the historical, cultural, and social contexts in which she lived. Recent discussions have touched on women’s rights during the Regency period and plays from the Romantic period by Sir Walter Scott and Joanna Baillie.[26] We are currently compiling and aggregating volunteer comments so that their work can be cited within our digital archive[27], which will incorporate annotation functionality so that later viewers can also contribute comments. The various formats and contributors of these annotations comprise another layer of MELP’s messiness, but they allow contributors to engage and be recognized for their work in the humanities, in turn encouraging more engagement on our project [Terras 2015, 435].

While the volunteer collaborators provide initial manuscript transcriptions, notes, and, eventually, annotations, the



institutional collaborators generally check the transcriptions and encode them in TEI. Once a letter has been completed on Zooniverse, research assistants on the *MELP* project team can view the transcription marks through Zooniverse's Aggregate Line Inspector/Collaborative Editor (ALI/CE), which aggregates all volunteer transcriptions of each line and statistically chooses the most likely correct transcription, in order to correct and process the transcription of the entire letter. Although one transcription is ultimately selected or composed by research assistants, Figure 3 provides an example of the multiple contributions we receive for a single line within a letter.

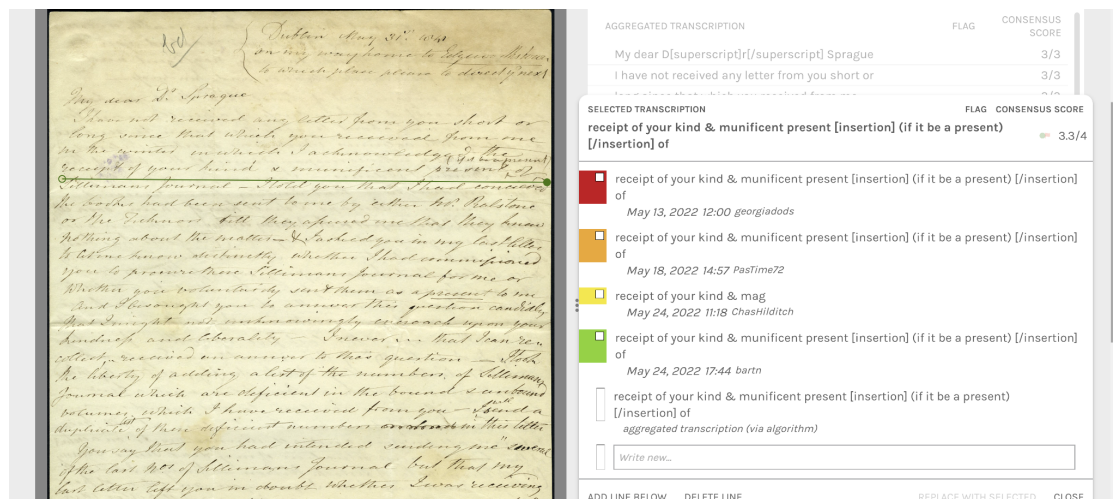


Figure 3. MELP ALI/CE interface on Zooniverse.

This process of reviewing and reconciling volunteer transcriptions through ALI/CE is a large part of the transcription workflow, or the process by which a transcription is reviewed and processed in preparation for TEI encoding, the standard for the representation of texts in digital form. Line breaks, page breaks, and tags for underlined, superscript, added, and deleted text, all noted by volunteers during transcriptions, are preserved in this process. However, before texts can be encoded, the people, places, and works mentioned in the letters need to be accounted for so that they can be reconciled with existing name authorities (i.e., files that establish standardized spellings and forms for each entity) and made discoverable in our database. From checking the transcription to amending the “ography” files, the process of selecting which files to work on often happens in back-and-forth stages. This process can lead to unexpected connections, corrections, and editorial decisions. For example, research assistants working on the letters may be unfamiliar with aspects of Edgeworth's life that are well known to the project editors, who are Edgeworth scholars. Working together over Zoom or exchanging emails allows the process to be collaborative, opening up opportunities to revise a given letter and then apply the new information to other letters. The final part of our workflow extends this flexible and interoperable process by transforming the corrected transcriptions and associated entities into digitally encoded TEI files.

23

## Encoding

After letters have been transcribed and converted to plain text, *MELP* researchers encode the files in TEI (Text Encoding Initiative), an application of XML (eXtensible Markup Language) and a standard of representation that has been used by humanists since 1994 to create machine-readable digital texts. Many museums, archives, and digital projects like ours use TEI because of its interoperability with other projects and because it is flexible enough to grow with the corpus as more materials are added regardless of the type of database chosen. *MELP*'s use of TEI is most visible in the creation of two types of documents: 1) encoded letters, and 2) indices of named entities (“ography” files).

24

[28] The suffix “ography” is typically associated with the “names of descriptive sciences”, such as research that involves classifying and categorizing [Ography, n. 2023]. We thus use the term “ography”, which derives from the standard TEI *prosopography*, to refer to our entity naming and encoding processes for people (personography), places (placeography), and textual works (workography). In selecting controlled names for entities within the “ography” files (where matches were possible), three authorities were privileged: the Virtual International Authority File (VIAF) for

persons, GeoNames for places, and Wikidata for works.<sup>[29]</sup> These authorities were chosen because they are widely used, match the expected domains of knowledge, can be edited by team members, and support URIs (Uniform Resource Identifiers) that are dereferenceable and content negotiable. Wikidata and GeoNames are completely open to public editing while additions to VIAF, an aggregator of authorities, can be made through the Library of Congress Name Authority Cooperative (NACO). Using these authorities allows *MELP* to create a searchable structure without imposing a definitive hierarchy or interpretation of the materials that is predicated on a specific narrative of Edgeworth's life. Each person, place, and title is given the same weight in the linked open data, which translates in the database as a “choose your own adventure”; users can look for specific terms rather than through a traditional linear arrangement of the correspondence. Additionally, by dividing the encoding workflow into the expansion of “ography” files, which can be completed on a spreadsheet, and iterative TEI creation, which requires XML editing software, *MELP* enables a diverse labor structure in which non-TEI experts can contribute and cross-institutional collaboration is streamlined, as our encoding and linked data practices encourage information sharing with other institutions and entities.

Many other projects construct their data internally without consideration of external name authorities or integration with other digital projects. When citing data, most projects have custom mechanisms for authoring biographies and creating annotations. In the *Willa Cather Archive*, for example, the entry for Jane Austen reads:

25

Born in Steventon Parish, Hampshire, England, Jane Austen began writing at an early age, and much of her precociously satiric juvenilia has been preserved. By the mid-1790s she had written early versions of *Sense and Sensibility* (1811) and *Pride and Prejudice* (1813). These were followed by *Mansfield Park* (1814) and *Emma* (1815), all published anonymously. *Northanger Abbey* and *Persuasion* were published after her death. Austen's reputation increased during the nineteenth century as she became a part of the canon of the English novel. Willa Cather acknowledged Austen's importance but does not discuss her.<sup>[30]</sup>

It's a detailed prosopography, or description of Austen's life and works within the context of the larger project, here Cather's letters. Most of the entry is a biography that replicates biographical information that could be found elsewhere on the web (though these are more useful for figures who are not easily discoverable), with one sentence connecting her to the Cather archive. This practice of creating prosopographies is a central task of many digital archives, including *The Letters of Charlotte Mary Yonge*, *The Olive Schreiner Letters Online*, and the *Vincent Van Gogh Letters*.<sup>[31]</sup> These entries can be detailed and time-consuming due to the “inclusive and iterative research process” they require, “combining fragments from public and academic information resources” [Fukushima, Bourrier, and Parker 2022, ¶29].<sup>[32]</sup> While Maria Edgeworth's correspondence is rich in potentially prosopographical subjects, prosopography is rarely accessible or reviewable beyond the constraints of the project, which works directly against *MELP*'s goals.

26

Instead of following the prosopography model of these other projects, *MELP* has focused on links between it and other authorities as a way of contributing to open linked data. For example, *MELP*'s entry for Jane Austen merely consists of the title of her name authority file (Austen, Jane, 1775-1817) and the link to her Virtual International Authority File (<http://viaf.org/viaf/102333412>), which contains related names and works in a series of links instead of a prose paragraph. *MELP*'s use of open linked data is based on *Digital Mitford*'s model, which uses “a growing network of linked data, a digital database from which we can extract and study information we are collecting about people and texts of the nineteenth century” [Beshero-Bondar and Donovan-Condrón 2017, 141]. Such a network “make[s] available hitherto unknown data about publishers of periodicals, theatre managers and actors, poets, artists, as well as politicians and educators — an extensive network bonded by mutual influence and support” [Beshero-Bondar and Donovan-Condrón 2017, 141]. Following in this tradition of linked data, *MELP* uses the “ography” files to create an iterative tagging process that draws upon the full titles of personography, placeography, and workography. We identified people, places, and works as the three most important elements to our current project, allowing us to track the individuals within Edgeworth's community of correspondence, the places she traveled to or mentioned, and the works she read and discussed. These three categories, we felt, would address some of the most pressing research questions that future database users would have. Each entity is defined through a unique `@xml:id` value, which is used to link mentions within letters to the project's authoritative textual form of the name as well as external authority URIs where applicable.

27

Personographies and placeographies, like *MELP*'s, are common in the TEI landscape<sup>[34]</sup>, though project-specific decisions needed to be made. For example, the team decided not to include some attributes like gender or sex in the personography, in alignment with the Program for Cooperative Cataloging's practice for Library of Congress authority records [Billey et al. 2022].<sup>[35]</sup> Rather than using all of the elements and attributes present in TEI to help describe a person, such as birth, nationality, and occupation, *MELP* instead has focused on finding existing authority URIs for these individuals that typically include this information externally. Because the URIs for the authorities *MELP* selected are content negotiable, this specific information can be theoretically retrieved as RDF (Resource Description Framework) simply by using the URI. For instance, the URI associated with the place "Edgeworthstown" (<http://sws.geonames.org/2964434>) returns information on alternate names, population, latitude, and longitude (Figure 4).

```
@prefix cc: <http://creativecommons.org/ns#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix gn: <http://www.geonames.org/ontology#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix wgs84_pos: <http://www.w3.org/2003/01/geo/wgs84_pos#> .

<https://sws.geonames.org/2964434/> a gn:Feature ;
  gn:alternateName "Meathas Truim",
    "Meathus Truim",
    "Mostrim",
    "Edgeworthstown"@en,
    "Meathas Troim"@ga ;
  gn:countryCode "IE" ;
  gn:featureClass <https://www.geonames.org/ontology#P> ;
  gn:featureCode <https://www.geonames.org/ontology#P.PPL> ;
  gn:locationMap <https://www.geonames.org/2964434/edgeworthstown.html> ;
  gn:name "Edgeworthstown" ;
  gn:nearbyFeatures <https://sws.geonames.org/2964434/nearby.rdf> ;
  gn:parentADM1 <https://sws.geonames.org/7521314/> ;
  gn:parentADM2 <https://sws.geonames.org/2962839/> ;
  gn:parentCountry <https://sws.geonames.org/2963597/> ;
  gn:parentFeature <https://sws.geonames.org/2962839/> ;
  gn:population "2072" ;
  gn:wikipediaArticle <https://en.wikipedia.org/wiki/Edgeworthstown> ;
  rdfs:isDefinedBy <https://sws.geonames.org/2964434/about.rdf> ;
  rdfs:seeAlso <https://dbpedia.org/resource/Edgeworthstown> ;
  wgs84_pos:lat "53.7" ;
  wgs84_pos:long "-7.61667" .

<https://sws.geonames.org/2964434/about.rdf> a foaf:Document ;
  cc:attributionName "GeoNames"^^<https://www.w3.org/2001/XMLSchema#string> ;
  cc:attributionURL <https://www.geonames.org> ;
  cc:license <https://creativecommons.org/licenses/by/4.0/> ;
  dcterms:created "2006-01-15"^^<https://www.w3.org/2001/XMLSchema#date> ;
  dcterms:modified "2022-01-16"^^<https://www.w3.org/2001/XMLSchema#date> ;
  foaf:primaryTopic <https://sws.geonames.org/2964434/> .
```

**Figure 4.** RDF returned for the URI for "Edgeworthstown". This RDF is formatted as turtle, or Terse RDF Triple Language.

This information is communicated in semantic triples, or sets of three entities, which include a subject, predicate, and object. Because of this, *MELP* did not create exhaustive entries for the persons and places it lists in the personography and placeography. For entities that do not have established name authorities, we explore newspaper, scholarly, and genealogical databases for information, similar to the creators of *Digital Dinah Craik*. Thus, within our personography, we either embed key entities within a larger network of linked data or, when necessary, create them ourselves, making our project a flexible resource in dialogue with other networks of projects and individuals that allows those with content expertise to participate without additional barriers. *MELP*'s third "ography" file, the workography, is — to our knowledge

— unique in its content and scope since there are no existing workography examples, though it resembles the personography and placeography files. Edgeworth's letters include numerous references to her own novels and short stories as well as to the works of her contemporaries, such as Walter Scott and Frances Burney, so documenting these connections in a separate workography file is essential for researchers. Because of this, *MELP* needed to devise the XML structure of the file to meet the project's needs. The project team contacted the TEI listserv on 4 November 2022 to get input on this dilemma. None of the responses we received directly addressed our use case. The elements `listBibl` and `bibl` form the backbone of this "ography" file. The TEI was flexible enough to also allow *MELP* to model works within works, which is especially important for referencing Edgeworth's short stories (see Figure 5).

```
<bibl xml:id="Moral1">
  <title>Moral Tales for Young People</title>
  <idno>https://www.wikidata.org/wiki/Q80225023</idno>
  <author>
    <persName>Edgeworth, Maria, 1768-1849</persName>
    <idno type="VIAF">http://viaf.org/viaf/315684946</idno>
  </author>
  <ref corresp="Moral1">
    <bibl xml:id="Knaps1">
      <title>The Knapsack</title>
      <idno>https://www.wikidata.org/wiki/Q80031478</idno>
    </bibl>
  </ref>
  <ref corresp="Moral1">
    <bibl xml:id="Fores1">
      <title>Forester</title>
    </bibl>
  </ref>
  <ref corresp="Moral1">
    <bibl xml:id="Pruss1">
      <title>The Prussian Vase</title>
    </bibl>
  </ref>
</bibl>
```

Figure 5. Works within works present in the workography. `@corresp` is used to link short stories, essays, etc. to their parent works.

We adapted `@corresp` to indicate how works within the workography are related to one another, which is particularly helpful for tracking works within other works, such as Edgeworth's individual short fictions in her larger collections of tales. This "ography" file will allow scholars to track allusions to literary and non-literary texts across Edgeworth's entire correspondence. Instead of charting loose affinities based on similar publishing periods or identifying one instance of Edgeworth's direct commentary, users will be able to see all appearances of a text within the archive, creating new critical pathways. Thus, by creating the three "ography" files, *MELP* has been able to generate its own linked data system that can be represented in standardized TEI tags within the letters and in the entities listed in our Corpora-based searchable database.

The "ography" files, alongside other typographical features, provide the basis for encoding transcribed letters in *MELP* into a template following the *TEI Guidelines for Electronic Text Encoding and Interchange*.<sup>[36]</sup> After a letter has been transcribed on Zooniverse, there is a multi-step process that we have developed for encoding the letter and defining entities in the appropriate "ography" file.<sup>[37]</sup> Using metadata either directly from the institution the letter is held in or metadata gathered after examining the letter itself, the encoder records in the TEI header the author, the recipient, their locations, the date on which the letter was sent, and the institution in which it is housed. The component parts of each letter (opener, salute, body, closer, postscript, and address) are encoded according to the correspondence structure

from the TEI guidelines, retaining line breaks, paragraph breaks, and page breaks. During this encoding process, which follows standard TEI guidelines, researchers tag each `persName`, `placeName`, and `title` with a corresponding `@ref`, linking back to the local name authority files created through the “ography” process. Given the detailed TEI workflow that all members of the project team can access, the process of encoding each letter does not require extensive coding experience. By separating the processes for tracking down named entities and encoding the letters, *MELP*'s approach allows a wider variety of collaborators to participate in the project, from tenured faculty and librarians to graduate and undergraduate researchers, including a recent group of the latter who were paid to learn TEI in a summer 2023 undergraduate institute co-hosted by Wake Forest University and Xavier University of Louisiana.

The wide variety of contributors to the project necessitates detailed instructions, both for creating encoded letters and “ography” files, so that contributors can complete the work consistently and ensure that future contributors can be onboarded in a timely fashion. Starting from step one through the final step, if a person has a basic working knowledge of TEI, they can encode for *MELP*. We believe that decentralizing the knowledge required to contribute to *MELP* and specifically tailoring our TEI template and “ography” files to be replicable without long-term training are key to producing a project that has long-term sustainability. As people leave and join the project, this relatively low barrier to entry minimizes the risk that the project will not be completed or maintained; distributing the knowledge and creating the “ography” files has helped *MELP* grow quickly and deeply. The “ography” files, which establish `xml:id` values for entities mentioned in the letters are a good example of this as they require limited technical knowledge. Each “ography” file is built in a spreadsheet that creates associations between the internal `xml:id` for each person, place, or work and how that person, place, or work will appear in our Corpora database. *MELP* decided to use spreadsheets as the foundation for the “ography” files so that any person who can input data into a spreadsheet will be able to contribute to the project's growing dataset.

Term (Lname, Firstname Mname, Birth date - Death date)	ID	VIAF URI
Achard, Madame	AchaX1	
Adams, Abigail, 1744-1818	AdamA1	<a href="http://viaf.org/viaf/32015075/">http://viaf.org/viaf/32015075/</a>
Adams, John, 1735-1826	AdamJ1	<a href="http://viaf.org/viaf/95296943">http://viaf.org/viaf/95296943</a>
Addison, Joseph, 1672-1719	AddiJ1	<a href="http://viaf.org/viaf/7413288">http://viaf.org/viaf/7413288</a>
Agnew, Elizabeth	AgneE1	

OpenRefine Export Template  
(GREL)

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- teiHeader -->
  </teiHeader>
  <text>
    <body>
      <p>This is a personography that defines people associated with the Maria
      Edgeworth Letters project.</p>
      <listPerson>
        <person xml:id="AchaX1">
          <persName>Achard, Madame </persName>
        </person>
        <person xml:id="AdamA1">
          <persName>Adams, Abigail, 1744-1818</persName>
          <idno type="VIAF">http://viaf.org/viaf/32015075/</idno>
        </person>
        <person xml:id="AdamJ1">
          <persName>Adams, John, 1735-1826</persName>
          <idno type="VIAF">http://viaf.org/viaf/95296943</idno>
        </person>
        <person xml:id="AddiJ1">
          <persName>Addison, Joseph, 1672-1719</persName>
          <idno type="VIAF">http://viaf.org/viaf/7413288</idno>
        </person>
        <person xml:id="AgneE1">
          <persName>Agnew, Elizabeth</persName>
        </person>
        <!-- Additional person elements -->
      </listPerson>
    </body>
  </text>
</TEI>
```

Figure 6. Personography creation process showing selected portions of the initial data entry in a spreadsheet and the resulting TEI product.

The flat spreadsheet data is then transformed into TEI using an export template written in the Google Refine Expression Language (GREL) within OpenRefine (Figure 6). As data is continually added to the spreadsheets, TEI can be regenerated using the existing export template. Instructions have been written to ensure the sustainability of the project and remove barriers to others contributing. Distributing this labor has allowed *MELP* to reconcile the divide Brown, Clements, and Grundy describe as the choice between “greater complexity and a fairly tight-knit group of collaborators, and lesser complexity and a more open collaborative model” [Brown, Clements, and Grundy 2006, 324]. Our approach to Maria Edgeworth’s vast correspondence offers a model that invites greater complexity and intentional work amongst a small group of scholars (institutional collaborators) as well as pursuing an open collaborative model through using Zooniverse to transcribe Edgeworth’s letters (volunteer collaborators). A broad overview of how the different processes covered in this article work together to create a digital archive appears in Figure 7. By defining these steps, from letter acquisition to encoding, we can visually demonstrate how our labor is distributed and provide guidance for similar projects.

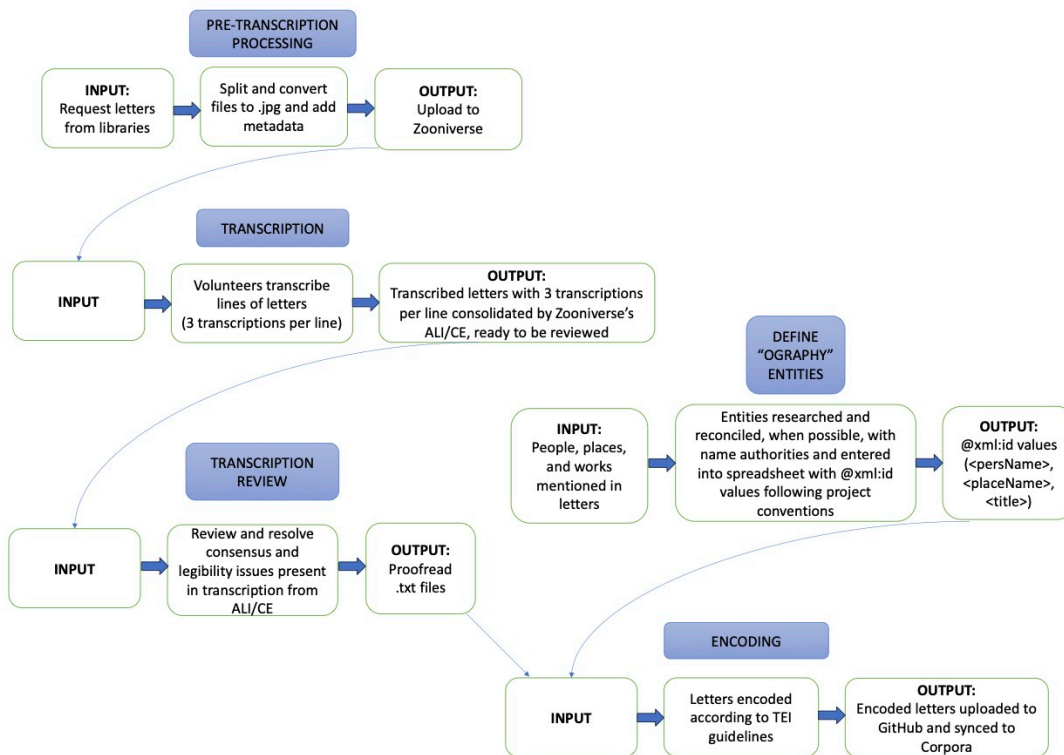



Figure 7. Visual depiction of MELP's workflows.

## Conclusion

The undergraduate encoders and Zooniverse volunteers demonstrate MELP's public scope beyond the institutionally employed professors, librarians, and graduate research assistants. Their contributions appear within a full-text and entity-searchable database that has been available on MELP's pilot website from spring 2024.<sup>[38]</sup> The database unites all of the various types of labor within MELP: the metadata collection, the crowdsourced transcription, the "ography" files, and the text encoding. MELP's database has been built in Corpora, a no-SQL database, or a database that stores data in a format other than relational tables.<sup>[39]</sup> No-SQL databases are ideal for projects like MELP that have a constantly changing number of entities and thus require structural flexibility. To maintain the larger "web" of tagged persons, places, and works within the archive and transfer it to a database structure, a No-SQL structure is essential. Corpora has successfully supported a number of digital humanities projects, including *Carlyle Letters Online* and the *New Variorum Shakespeare*. Within the database, people, places, and works are tagged within each letter and linked to all other letters that mention those entities (Figure 8).

Identifier	Edgeworth_Harvard_604742_1829-12.xml
Title	Letter from Maria Edgeworth to Mr. [James Townsend] Mackay, December 23rd,1829
Date Composed	1829-12-23
Author	Edgeworth, Maria, 1768-1849 (PERSON)
Recipient	Mackay, James Townsend, 1775-1862 (PERSON)
Images	Edgeworth_Harvard_604742_1829-12_page01.jpg Edgeworth_Harvard_604742_1829-12_page02.jpg Edgeworth_Harvard_604742_1829-12_page03.jpg
Entities Mentioned	United States of America (PLACE) Marcet, Mrs. (Jane Haldimand), 1769-1858 (PERSON) Candolle, Augustin Pyramus de 1778-1841 (PERSON) Ballsbridge, Ireland (PLACE) Bloomfield (estate), Ireland (PLACE)
HTML	 <p>Miss Edgeworth's compliments to Mr. Mackay - she has been intending ever since she has been in this neighborhood to here indulge her self in a visit to the Coll<sup>rs</sup> Botanic garden - But weather now and her ill health when weather permitted have deprived her of this pleasure She leaves Bloomfield early on Saturday morning - Her only chance of having</p>

**Figure 8.** Letter from Maria Edgeworth to Mr. [James Townsend] Mackay, 23 December 1829, on Corpora.

The output from Corpora will soon be connected to *MELP's* public WordPress website so that other researchers and members of the general public can use the database to search and explore the Edgeworth correspondence, assembled through various methodologies and by numerous contributors.

Critical to this user experience, Corpora enables the retention of messy data yet capitalizes upon existing linked data structures, like controlled vocabularies, when appropriate. It maps the “messiness” of *MELP's* corpus to established terms and URIs for people, places, and works. Corpora allows us to both show the original, sometimes familiar, references to these entities in the letters and link out to regularized names and authoritative data. Clicking on a given entity allows the user to see other letters associated with that entity along with the variety of ways it has been referred to in those sources, with external links to established vocabularies, if identified. When present, external URIs share valuable information, like birth dates for people and coordinates for places. Nevertheless, these external links are optional and only associated with an entity when this application can be done with certainty. In this way, Corpora supports two different webs users can traverse: an internal web of entities mentioned within the letters and an external web of linked data sources. Having both webs available to users allows Edgeworth's correspondence to retain the messiness inherent to a correspondence network this wide in scope, varied in content, and written over decades across the beginning and end of two centuries. Ultimately, *MELP* privileges the letters, with their inclusion of incomplete names, misspellings, and sometimes unclear text, as our final authority and the source with which most users should primarily interact.

Although Corpora and WordPress will be the project's public interfaces, we want to close by noting that while database and website technologies may change, the building blocks of our project — the plain text (.txt) and extensible markup (.xml) documents — are purposefully written in file formats known for their simplicity, longevity, and usability. This is the last and, perhaps, the most important aspect of *MELP's* flexibility. The .txt and .xml files that populate our Google Drive and GitHub repositories are key in facing *MELP's* largest challenge: preservation. Choosing to create a digital archive is itself an act of preservation in that a “suite of tools, operations, standards, and policies” are necessary to “help ensure that this investment is not squandered” [Conway 2010, 65]. Yet Drew VandeCreek's study clocks the longevity of most NEH-funded digital humanities projects at between eleven and sixteen years.<sup>[40]</sup> Funding issues and obsolete file formats or programs are often the cause of this obsolescence. Our final act of collaborative project management has been to select plain text files and eXtensible Markup Language documents, which can be edited with a variety of applications, as means to promote the longevity and interoperability of the project amidst the inevitable shifts in website and database technologies. Even if, when we reach the end of our project, we are unable to acquire, transcribe, and encode all ten thousand sheets of Edgeworth's correspondence, the material that we do assemble in our open-access digital archive should resist obsolescence and remain the authoritative source for Edgeworth's correspondence for years to come.

35

36

37



# Acknowledgements

We would like to thank our collaborators on the *Maria Edgeworth Letters Project*, including the three other editors, Susan Egenolf (Texas A&M), Jessica Richard (Wake Forest), and Robin Runia (Xavier of Louisiana), our project manager, Carrie Johnston (SMU), our former data curation librarian, Heather Barnes (Wake Forest), and our technical consultant, Bryan Tarpley (Texas A&M), as well as the other research assistants from the University of Tennessee who have contributed to the project: Katie Haire, Autumn Hall, Ivy Kiernan, and Ziona Kocher. We would also like to thank the 860 volunteers who transcribed the first 744 letters, whose usernames are listed here: <https://www.zooniverse.org/projects/mariaedgeworthletters/maria-edgeworth-letters/about/results>. Additionally, we are grateful to Cailin Roles and the anonymous reviewers at *Digital Humanities Quarterly* for their helpful feedback.

38

The *Maria Edgeworth Letters Project* has been supported by a number of internal and external grants. We have received a National Endowment for the Humanities: Humanities Collections and Reference Resources Foundations fellowship, as well as several grants from the University of Tennessee: a New Research in the Arts and Humanities award, two summer research assistantships, two English department assistantships, and a UT Humanities Center Digital Humanities Fellowship. We would like to thank Misty Anderson (UT English), Amy Elias (UT Humanities Center), and Holly Mercer (UT Libraries) for their support of the project. The *Maria Edgeworth Letters Project Zooniverse* website was initially developed as part of the 2021-22 project cohort of Building Capable Communities for Crowdsourced Transcription (<https://sites.google.com/umn.edu/atdhcrowdcohort/home>), an Institute for Advanced Topics in the Digital Humanities (HT-262556-20) generously funded by the National Endowment for the Humanities and led by Sam Blickhan, Evan Roberts, Ben Wiggins, and Trevor Winger.

39

## Notes

[1] Some letters have previously been published in [Butler 1972], [Colvin 1971], [Colvin 1979], [Edgeworth 1867], and [Pakenham 2017].

[2] Our long-term goal is to create linked data for other entities, such as keywords or subjects, but we identified the people, places, and works as the best starting point to intervene and transform Edgeworth's correspondence archives into a useful database.

[3] *Digital Dinah Craik* has also influenced our project. Their transcriptions and TEI letters are available through TAPAS (<https://tapasproject.org/node/443>).

[4] The *Carlyle Letters* similarly uses Texas A&M's Corpora database to structure its data. See <https://carlyleletters.dukeupress.edu/about-editorial-methods/>.

[5] File format, as Desmond Schmidt has argued, can contribute to interoperability — for example, by providing separate .txt and .xml files in a digital edition because markup standards vary by project [Schmidt 2014, 14–15].

[6] See the “About” pages and documentation for each of these projects: <https://carlyleletters.dukeupress.edu/about-project/> (*Carlyle Letters Online*), <http://shelleygodwinarchive.org/about/> (*Shelley-Godwin Archive*), and [https://digitalmitford.github.io/DM\\_documentation/](https://digitalmitford.github.io/DM_documentation/) (*Digital Mitford*).

[7] The two goals of *Digital Mitford* are “to produce the first comprehensive scholarly edition of the works and letters of Mary Russell Mitford, and to share knowledge of TEI XML and other related humanities computing practices with all serious scholars interested in contributing to the project”. See <https://digitalmitford.org/index.html>.

[8] See <https://www.zooniverse.org/about>.

[9] See <https://www.zooniverse.org/projects/mariaedgeworthletters/maria-edgeworth-letters>.

[10] The NEH, for example, requires that grant recipients complete performance and, often, financial reports. See <https://www.neh.gov/grants/manage/performance-reporting-requirement>.

[11] Fitzpatrick continues, “scholars have frequently encountered obstacles to having non-print work given appropriate credit, and many scholars in the humanities also report difficulty having their coauthored publications taken seriously as part of their record of production” [Fitzpatrick 2011, 73].

[12] Copies of *MELP*'s workflows can be found in the "Documentation" folder of our GitHub repository: <https://github.com/Maria-Edgeworth-Letters-Project/me-tei/tree/main/Documentation>.

[13] The "four key foci" Boyd identifies are "defining the bargain with team members; translating between team members; facilitating perpetual peer review; and researching scholarly exchange. Far from being an interloper into the digital humanities, project management and the project manager can be a key practice and role in realising a more fulfilling ecosystem of scholarly exchange" [Boyd 2022].

[14] The *Walt Whitman Archive* brings together material from more than 50 repositories ([https://whitmanarchive.org/manuscripts/finding\\_aids/index.html](https://whitmanarchive.org/manuscripts/finding_aids/index.html)), and there are 28 separate collections listed in *Digital Mitford* (<https://digitalmitford.org/lettersData.html>).

[15] The project is currently supported by 31 institutions, as well as private collectors. They are, in alphabetical order: Beinecke Library (Yale University), Bibliothèque de Genève, Bibliothèque Nationale de France, University of Birmingham, Bodleian Library (University of Oxford), Boston Public Library, British Library, Brown University, Chawton House Library, Claremont Colleges, Dartmouth College, Duke University, Houghton Library (Harvard University), Huntington Library, University of Iowa, King's College (University of Cambridge), University of North Carolina, Morgan Library and Museum, National Library of Ireland, National Library of Scotland, New York Public Library, Ohio University, Pennsylvania State University, Princeton University, University of Reading, Rosenbach Museum and Library, Trinity College Dublin, University College London, University of Pennsylvania, University of Virginia, and Vassar College.

[16] For other examples of increased resource digitization during the pandemic, see [Marek 2022] and [Gross 2022].

[17] Questions about the future of the expanded access brought on by the pandemic are addressed in the August 2021 webinar "Resource Sharing: What Will We Keep from the Pandemic?" See <https://exlibrisgroup.com/blog/resource-sharing-what-will-we-keep-from-the-pandemic/>.

[18] See, for instance, [The National Archives 2022, 11].

[19] *Emoji Dick* asserts that the text has been "edited and compiled by Fred Benenson", with "Translation by Amazon Mechanical Turk". The text's front matter discloses that "over eight hundred people spent approximately 3,795,980 seconds working to create this book" [Benenson 2010, i, vii].

[20] We do not use unpaid student labor in our project. Any graduate and undergraduate student contributions have been supported through funded research assistantships.

[21] More information about Zooniverse can be found here: <https://www.zooniverse.org/about>.

[22] The Field Guide can be viewed as a pull-out menu on the right edge of *MELP*'s Talk boards on Zooniverse: <https://www.zooniverse.org/projects/mariaedgeworthletters/maria-edgeworth-letters/talk>.

[23] Ivy Kiernan's work on *MELP*'s handwriting guide and the responses it generated can be found here: <https://www.zooniverse.org/projects/mariaedgeworthletters/maria-edgeworth-letters/talk/4463>.

[24] Discussions of "vandyke" lace can be found here: <https://www.zooniverse.org/projects/mariaedgeworthletters/maria-edgeworth-letters/talk/4462/2886410>.

[25] Connections with the *Davy Notebooks Project* and the Beddoes archive can be found here: <https://www.zooniverse.org/projects/mariaedgeworthletters/maria-edgeworth-letters/talk/4462/2562237>.

[26] Respectively, these discussions on women's rights during the Regency period and Romantic-era drama can be found here: <https://www.zooniverse.org/projects/mariaedgeworthletters/maria-edgeworth-letters/talk/4462/2889065> and here: <https://www.zooniverse.org/projects/mariaedgeworthletters/maria-edgeworth-letters/talk/4462/2957674>.

[27] We cite volunteers by their usernames unless they have agreed to share their actual names with us. See: <https://www.zooniverse.org/projects/mariaedgeworthletters/maria-edgeworth-letters/about/results>.

[28] Examples of both types of documents are present on the project's GitHub repository: <https://github.com/Maria-Edgeworth-Letters-Project/me-tei/tree/main>.

[29] The locations of these authorities are: <https://viaf.org/> (VIAF), <https://www.geonames.org/> (GeoNames), and

[https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page) (Wikidata).

[30] A link to Jane Austen's prosopography from the *Willa Cather Archive* can be found in Willa Cather's letter to Carrie Miner Sherwood on 28 June 1939: <https://cather.unl.edu/writings/letters/let1455>.

[31] Fukushima, Bourrier, and Parker posit that their in-progress archive, *Digital Dinah Craik*, “model[s] people as data” by combining “scholarly tools such as the TEI with para-scholarly genealogical databases and recently digitized nineteenth-century newspapers and books to give the fullest account of each person entry possible” [Fukushima, Bourrier, and Parker 2022, ¶3].

[32] As the *Syriac Persons, Events, and Relations (SPEAR)* project notes, new approaches to prosopography might focus more on the relationship editors create through “the scholar's interpretation of what a historical source asserts about a person” instead of a readily identifiable “fact” in an external authority file or database. Given the emphasis on broad network relationships in Maria Edgeworth's correspondence, *MELP* chose to focus less on deep prosopography, even though the case *SPEAR* makes for “factoid prosopography” is compelling [Schwartz, Gibson, and Torabi 2022, 5].

[33] For more information on *MELP*'s `&xml:id` conventions, see our documentation on GitHub: [https://github.com/Maria-Edgeworth-Letters-Project/me-tei/blob/main/Documentation/XML\\_ID\\_Conventions.pdf](https://github.com/Maria-Edgeworth-Letters-Project/me-tei/blob/main/Documentation/XML_ID_Conventions.pdf).

[34] Some examples of projects with placeographies that influenced our own include the *Map of Early Modern London* ([https://mapoflondon.uvic.ca/encode\\_persons.htm](https://mapoflondon.uvic.ca/encode_persons.htm)) and the *Michael Leary Letters* (<https://www.tapasproject.org/tei-learning-docs-bc/files/leary-placeography>).

[35] The report states: “Do not record the RDA gender element (MARC 375) in personal name authority records. Delete existing 375 fields when editing a record for any other reason.” Other library sources, such as the Trans Metadata Collective's “Metadata Best Practices for Trans and Gender Diverse Resources” also recommend not specifying gender unless it is essential and the individuals being described have self-identified their gender. Given that *MELP* deals exclusively with historical figures, self-identification is not possible, and attempts to interpret gender identity from other sources can easily be biased.

[36] Our TEI template is available in the *MELP* GitHub repository: <https://github.com/Maria-Edgeworth-Letters-Project/me-tei/blob/main/TEItemplate.xml>.

[37] For further information about our TEI encoding practices, please see the “Edgeworth TEI Workflow” document in *MELP*'s GitHub repository: <https://github.com/Maria-Edgeworth-Letters-Project/me-tei/blob/main/Documentation/Edgeworth%20TEI%20Workflow.pdf>.

[38] *MELP*'s pilot website is available here: <https://melp.dh.tamu.edu/> (accessed: 25 April 2024).

[39] Documentation for Corpora can be found here: <https://gitlab.dh.tamu.edu/bptarpley/corpora>.

[40] For example, funding lapsed for our sister project that is based in Oxford and Ireland, the *Digital Edgeworth Network* (<https://digital.humanities.ox.ac.uk/people/digital-edgeworth-network>) at the end of November 2021. Barone, Zeitlyn, and Mayer-Schönberger have suggested that “a sort of fatal creeping obsolescence can occur that is caused by a mix of under-funding, lack of investment in technical updating and neglect” [Barone, Zeitlyn, and Mayer-Schönberger 2015, ‘creeping’]. Lucky and Harkema have noted that “the products of grant funded digital projects have the tendency to be left to bit rot on the open Web, or disappear entirely, not long after the money runs out” [Lucky and Harkema 2018, 191]. See also [Cummings 2023].

## Works Cited

**ALI/CE Zooniverse, n.d.** ALI/CE Zooniverse Aggregate Line Inspector / Collaborative Editor [ALI/CE] Zooniverse. Available at: <https://alice.zooniverse.org/> (Accessed: 25 August 2023).

**Arbuckle 2019** Arbuckle, A. (2019) “Opportunities for social knowledge creation in the digital humanities”, in Crompton, C., Lane, R.J., and Siemens, R. (eds.) *Doing more digital humanities: Open approaches to creation, growth, and development*. London: Routledge, pp. 290–300. Available at: <https://doi-org.utk.idm.oclc.org/10.4324/9780429353048> (Accessed: 17 December 2023).

**Barone, Zeitlyn, and Mayer-Schönberger 2015** Barone, F., Zeitlyn, D., and Mayer-Schönberger, V. (2015) “Learning from failure: The case of the disappearing web site”, *First Monday*, 20(5). <https://doi.org/10.5210/fm.v20i5.5852>. (Accessed: 25 April 2024).

**Benenson 2010** Benenson, F. (ed.) (2010) *Emoji dick; Or the whale. By Herman Melville*. Translated by Amazon

Mechanical Turk. Morrisville, NC: Lulu Press.

- Beshero-Bondar and Donovan-Condron 2017** Beshero-Bondar, E. and Donovan-Condron, K. (2017) "Modelling Mary Russell Mitford's networks", in Winckles, A.O. and Rehbein, A. (eds.) *Women's literary networks and Romanticism*. Liverpool: Liverpool University Press, pp. 137–95.
- Billey et al. 2022** Billey, A. et al. (2022) *Revised report on recording gender in personal name authority records*. Available at: <https://www.loc.gov/aba/pcc/documents/gender-in-NARs-revised-report.pdf> (Accessed: 20 August 2023).
- Boyd 2022** Boyd, J. (2022) "Digital humanities project management as scholarly exchange", *IDEAH*, 2(2). Available at: <https://doi.org/10.21428/f1f23564.a4156d43> (Accessed: 6 December 2023).
- Brown 2016** Brown, S. (2016) "Towards best practices in online collaborative knowledge production", in Crompton, C., Lane, R., and Siemens, R. (eds.) *Doing digital humanities: Practice, training, research*. London: Routledge, pp. 47–64. Available at: <https://doi.org/10.4324/9781315707860> (Accessed: 17 December 2023).
- Brown, Clements, and Grundy 2006** Brown, S., Clements, P., and Grundy, I. (2006) "Sorting things in: Feminist knowledge representation and changing modes of scholarly production", *Women's Studies International Forum*, 29(3), pp. 317–25. Available at: <https://doi.org/10.1016/j.wsif.2006.04.010> (Accessed: 27 August 2023).
- Butler 1972** Butler, M. (1972) *Maria Edgeworth: A literary biography*. Oxford: Clarendon Press.
- Carlyle Letters Online, n.d.** *Carlyle Letters Online*, n.d. Available at: <https://carlyleletters.dukeupress.edu/> (Accessed: 25 August 2023).
- Christlein et al. 2018** Christlein, V. et al. (2018) "Handwritten text recognition error rate reduction in historical documents using naive transcribers". *Proceedings of the Gesellschaft für Informatik e.V. GI-workshop*. Berlin, Germany, 25 September, pp. 1-8. Available at: <https://doi.org/10.18420/infdh2018-13> (Accessed: 27 August 2023).
- Colvin 1971** Colvin, C. (1971) *Maria Edgeworth: Letters from England, 1813-1844*. Oxford: Clarendon Press.
- Colvin 1979** Colvin, C. (1979) *Maria Edgeworth in France and Switzerland: Selections from the Edgeworth family letters*. Oxford: Oxford University Press.
- Conway 2010** Conway, P. (2010) "Preservation in the age of Google: Digitization, digital preservation, and dilemmas", *The Library Quarterly: Information, Community, Policy*, 80(1), pp. 61–79.
- Copyright, Designs and Patents Act 1988** Copyright, Designs and Patents Act 1988. Available at: <https://www.legislation.gov.uk/ukpga/1988/48/contents> (Accessed: 26 August 2023).
- Cummings 2023** Cummings, J. (2023) "Academics retire and servers die: Adventures in the hosting and storage of digital humanities projects", *Digital Humanities Quarterly*, 17(1). Available at: <https://www.digitalhumanities.org/dhq/vol/17/1/000669/000669.html> (Accessed: 26 August 2023).
- Cushing 2012** Cushing, E. (2012) "Dawn of the digital sweatshop", *East Bay Express*, 1 August. Available at: <https://eastbayexpress.com/dawn-of-the-digital-sweatshop-1/> (Accessed: 23 August 2023).
- Davy Notebooks Project, n.d.** *Davy Notebooks Project*, n.d. Available at: <https://www.zooniverse.org/projects/humphrydavy/davy-notebooks-project> (Accessed: 23 August 2023).
- Di Pressi et al. 2015** Di Pressi, H., et al. (2015) "A student collaborators' bill of rights", 8 June. Available at: <https://humtech.ucla.edu/news/a-student-collaborators-bill-of-rights/> (Accessed: 26 August 2023).
- Digital Dinah Craik, n.d.** *Digital Dinah Craik*, n.d. Available at: <https://www.digitaldinahcraikproject.org> (Accessed: 25 August 2023).
- Edgeworth 1867** Edgeworth, F. (1867) *A memoir of Maria Edgeworth: With a selection from her letters*. 3 vols. London: Joseph Masters and Son.
- Ernø-Kjølhede 1999** Ernø-Kjølhede, E. (1999) *Project management theory and the management of research projects*. Department of Management, Politics, and Philosophy, CBS. MPP Working Paper No. 3/2000.
- Fitzpatrick 2011** Fitzpatrick, K. (2011) *Planned obsolescence: Publishing, technology, and the future of the academy*. New York: New York University Press.
- Folsom 2007** Folsom, E. (2007) "Database as genre: The epic transformation of archives", *PMLA*, 122(5), pp. 1571–79.
- Fukushima, Bourrier, and Parker 2022** Fukushima, K., Bourrier, K., and Parker, J. (2022) "The lives of mistresses and

- maids: Editing Victorian correspondence with genealogy, prosopography, and the TEI", *Digital Humanities Quarterly*, 16(1). Available at: <http://www.digitalhumanities.org/dhq/vol/16/1/000595/000595.html> (Accessed: 27 August 2023).
- Gross 2022** Gross, B. "Research in the time of COVID: Virtual fellowships at the Linda Hall Library", *Technology and Culture*, 63(4), pp. 1140–56.
- Hara et al. 2018** Hara, K. et al. (2018) "A data-driven analysis of workers' earnings on Amazon Mechanical Turk". *Proceedings of the 2018 CHI conference on human factors in computing systems*, Montreal, Canada, 21–26 April, pp. 1–14. Available at: <https://doi.org/10.1145/3173574.3174023> (Accessed: 23 August 2023).
- Holahan 2021** Holahan, C. (2021) "Rummaging in the dark: ECCO as opaque digital archive", *Eighteenth-Century Studies*, 54(4), pp. 803-826.
- Keralis 2018** Keralis, S.D.C. (2018) "Disrupting labor in digital humanities; or, the classroom is not your crowd", in Stommel, J. and Kim, D. (eds.) *Disrupting the digital humanities*. Goleta, CA: Punctum Books, pp. 273–94.
- Lucky and Harkema 2018** Lucky, S. and Harkema, C. (2018) "Back to basics: Supporting digital humanities and community collaboration using the core strength of the academic library", *Digital Library Perspectives*, 34(3), pp. 188–99.
- MELP TEI GitHub Repository, n.d.** *Maria Edgeworth Letters Project [MELP] TEI GitHub Repository*, n.d. Available at: <https://github.com/Maria-Edgeworth-Letters-Project/me-tei> (Accessed: 26 August 2023).
- Marek 2022** Marek, H.M. (2022) "Navigating intellectual property in the landscape of digital cultural heritage sites", *International Journal of Cultural Property*, 29, pp. 1–21.
- McCormack 2004** McCormack, W.J. (2004) "Edgeworth, Maria (1768–1849), novelist and educationist", *Oxford Dictionary of National Biography*, 23 September. Available at: <https://www.oxforddnb.com/view/10.1093/ref:odnb/9780198614128.001.0001/odnb-9780198614128-e-8476> (Accessed: 26 July 2023).
- Muñoz and Viglianti 2015** Muñoz, T. and Viglianti, R. (2015) "Texts and documents: New challenges for TEI interchange and lessons from the Shelley-Godwin archive", *Journal of the Text Encoding Initiative*, 8. Available at: <http://journals.openedition.org/jtei/1270> (Accessed: 25 August 2023).
- Nockels et al. 2022** Nockels, J. et al. (2022) "Understanding the application of handwritten text recognition technology in heritage contexts: A systematic review of Transkribus in published research", *Archival Science*, 22, pp. 367–92. Available at: <https://doi.org/10.1007/s10502-022-09397-0> (Accessed: 8 August 2023).
- Ography, n. 2023** "Ography, n." (2023), *Oxford English Dictionary*. Available at: <https://doi.org/10.1093/OED/1154538062> (Accessed: 27 August 2023).
- Pakenham 2017** Pakenham, V. (2017) *Maria Edgeworth's letters from Ireland*. Dublin: Lilliput Press.
- Rawson and Muñoz 2019** Rawson, K., and Muñoz, T. (2019) "Against cleaning", in Gold, M.K. and Klein, L. (eds.) *Debates in the digital humanities*. Minneapolis, MN: University of Minnesota Press, pp. 279–92.
- Schmidt 2014** Schmidt, D. (2014) "Towards an interoperable digital scholarly edition", *Journal of the Text Encoding Initiative*, 7, pp. 1-20. Available at: <http://journals.openedition.org/jtei/979>. (Accessed: 25 April 2024).
- Schwartz, Gibson, and Torabi 2022** Schwartz, D.L., Gibson, N.P., and Torabi, K. (2022) "Modeling a born-digital factoid prosopography using the TEI and linked data", *Journal of the Text Encoding Initiative*, pp. 1–35.
- Shelley-Godwin Archive n.d.** *Shelley-Godwin Archive*, n.d. Available at: <http://shelleygodwinarchive.org/> (Accessed: 25 August 2023).
- Siemens 2015** Siemens, L. (2015) "'More hands' means 'more ideas': Collaboration in the humanities", *Humanities*, 4, pp. 353–68. Available at: <https://doi.org/10.3390/h4030353> (Accessed: 6 December 2023).
- Siemens 2021** Siemens, L. (2021) "Introduction", *IDEAH*, 2(2). Available at: <https://doi.org/10.21428/f1f23564.f888546a> (Accessed: 6 December 2023).
- Surowiecki 2004** Surowiecki, J. (2004) *The wisdom of crowds*. New York: Anchor Books.
- Terras 2015** Terras, M. (2015) "Crowdsourcing in the digital humanities", in Schreibman, S., Siemens, R., and Unsworth, J. (eds.) *A new companion to digital humanities*. Hoboken, NJ: Wiley-Blackwell, pp. 420-39.
- Text Encoding Initiative, n.d.** *Text Encoding Initiative*, n.d. Available at: <https://tei-c.org/>. (Accessed: 20 August 2023).

**The Diaries of Michael Field, n.d.** *The Diaries of Michael Field*, n.d. Available at: <https://michaelfielddiary.dartmouth.edu/home> (Accessed: 25 August 2023).

**The National Archives 2022** The National Archives (2022) *Copyright and related rights*. Available at: <https://cdn.nationalarchives.gov.uk/documents/information-management/copyright-related-rights.pdf> (Accessed: 26 August 2023).

**The Walt Whitman Archive n.d.** *The Walt Whitman Archive*, n.d. Available at: <https://whitmanarchive.org/> (Accessed: 25 August 2023).

**VandeCreek 2022** VandeCreek, D. “Where are they now? The 2020 status of early (1996–2003) online digital humanities projects and an analysis of institutional factors correlated to their survival, preservation”, *Digital Technology & Culture*, 51(3), pp. 91–109.

**Wright 2023** Wright, D.W.M. (2023) “Travel and the climate crisis: Exploring COVID-19 impacts and the power of stories to encourage change”, *Journal of Tourism Futures*, 9(1), pp. 116-135. Available at: <https://doi.org/10.1108/JTF-03-2020-0043> (Accessed: 28 August 2023).



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.