Coded Voyages: Modelling Genre and Travel Writing in the Eighteenth Century

Catherine Nygren
Department of English
McGill University, Montreal
May 2024

A thesis submitted to McGill University in partial fulfillment
of the requirements of the degree of Doctor of Philosophy

Table of Contents

TABLE OF CONTENTS 2

ABSTRACT 6

ACKNOWLEDGEMENTS 8

LIST OF FIGURES AND TABLES 10

LETTERS OF INTRODUCTION 13

The Question of Genre 14

Unpacking the Genre of Travel 16
Travel Writing and the Novel 21

The History of British Travel 23

Consumers of Travel 26

Writers of Travel 30

Computational Approaches to Genre 34

Dissertation Map 36

CHAPTER 1: CREATING A CORPUS OF TRAVEL WRITING 39

Building a Corpus, Step One: A Bibliography of Titles 43

Overview of Consulted Sources: Bibliographies, Anthologies, Archive Catalogues, and Critical Works 45
Bibliographical Sources of the TWDB 51

Building a Corpus, Step Two: Combining Bibliographic Sources 54
Processing TWDB-origbib in Theory, or, How "C1ean" Should Data Be? 54
Processing TWDB-origbib in Practice 58

Building a Corpus, Step Three: Data Analysis of My Bibliographical Sources 60
Publication Counts: Year, Decade, and Source 61
Cross-References Between Sources 67

Conclusion 72

CHAPTER 2: SEARCHING FOR DATA, FROM PRINT TO DIGITAL 75

An Overview of Digital Databases 76

The First Challenge: Going to the Source 78
Eighteenth Century Collections Online 79
Nineteenth Century Collections Online 87
HathiTrust Digital Library and Research Center 89
Google Books 92

The Second Challenge: Supplementing Bibliographical Sources with Library of Congress Subject Headings 94
LCSH: Background 95
Searching for LCSH: The Eighteenth Century and Travel 96
Implementing LCSH in the TWDB 100

Going the Extra Mile: Encoding for Gender 102
Encoding Gender: In Theory 102
Encoding Gender: In the TWDB 107

Conclusion 108

CHAPTER 3: OBSERVATIONS AND REFLECTIONS ON THE TWDB 110

The TWDB and the Growth of Travel Writing 113
Annual and Decade Trends 113
Bibliographical and Data Source Trends 115

TWDB and LCSH: What is Travel Writing "About"? 117
Travel Writing Subjects and Trends 119

Finding Women in Travel Writing 123
Finding Women in Travel Writing: Previous Work 124
Counting and Identifying Women in the TWDB 125
Gendered Titles: Expanding the Bibliography 129
Gender on the Page: Claiming to be "By a Lady" 133
Subjects in Gendered Titles 134

Conclusion 136

CHAPTER 4: DESCRIPTIONS OF DESTINATIONS 138

Introduction: Perspectives and Modelling 138

Travel Writing: The Big Picture 144
From the Perspective of Bibliographers 149

Subject Headings 155
Armchair Travellers vs Real Travellers: Comparing Perspectives 155
Forms: Poetry 160
Forms: History 164
Locations 166

Writing "By a Lady" 170

The Changing Faces of Travel 175

Conclusion 183

CONCLUSION: FUTURE ITINERARIES 187

APPENDICES 193

Appendix 1.1: TWDB Metadata Codes 193

Appendix 1.2: Description of Sources 196
1.2.1: Bibliographies 196
1.2.2: Anthologies 211
1.2.3: Archives 215
1.2.4 Critical Works 220

Appendix 1.3: Extended Methodology of TWDB-origbib 223
1.3.1: Notes on Cox 223
1.3.2 Additional notes on transcription and deduplication 225

Appendix 1.4: Cross-References in TWDB-origbib 226
1.4.1: Table of Single and Multiple Sources 226

Appendix 2.1: Publication Numbers in ECCO 226

Appendix 2.2: LCSH on ECCO and HT 227
2.2.1 LCSH on ECCO 227
2.2.2: ECCO and HT Travel Keyword Comparisons 232

Appendix 2.3: Place of Publication on ECCO 232

Appendix 2.4: Notes on the Data Model 233
2.4.1: First Editions and Place of Publication 233
2.4.2: Matching the TWDB-origbib to HathiTrust, ECCO, NCCO, and Google 234
2.4.3: Harvesting LCSH: Notes on Technical Steps 237
2.4.4: Gender Encoding 240
2.4.5: Processing Plain Text and OCR Quality 241

Appendix 3.1: Bibliographical and Data Source Trends 245
3.1.1: Sources: Place of Publication and Holding Institutions 247

Appendix 3.2: LCSH Keywords in the Overall TWDB 249

Appendix 3.3: Subjects and Trends: Travel Keywords 250

Appendix 3.4: Subjects and Trends: No Travel Keywords 252

Appendix 3.5: Counting Gendered Titles 253

Appendix 3.6: Gendered Roles and Locations 255

Appendix 3.7: LCSH in the Gendered Corpus 256

Appendix 4.1: Models and “Accuracy” 259

Appendix 4.2: Mutual Recognition 262

Appendix 4.3: Pace of Change 266

BIBLIOGRAPHY 267

Primary Sources 267

Secondary Sources 268

Abstract

This dissertation analyzes the diversity of travel writing published in Great Britain between 1700-1830 through supervised predictive models. One of the most popular genres of the long eighteenth century, travel writing flourished in forms ranging from instructional guides to personal memoirs. This heterogeneity in form, content, and style has led to many definitions of travel writing and even disputes over whether it exists as a genre at all. *Coded Voyages* explores this contested landscape by algorithmically modelling different perspectives of travel writing. First, *Coded Voyages* describes the creation of a travel writing dataset of over 4,000 titles drawn from fourteen scholarly bibliographical sources. It then documents the theoretical and practical expansion of this Travel Writing Database through richer metadata such as subject headings and identification of women's roles as authors, editors, and translators. Finally, using the technique of perspectival modelling developed by Ted Underwood, *Coded Voyages* demonstrates that algorithmic models of travel writing can often identify travel titles nine times out of ten—despite conflicting scholarly definitions. These findings suggest that, far from being a borderless genre, travel writing does cohere around particular features. Comparisons of these models further reveal patterns in key subsets of travel writing, such as fictional voyages and women's writing. This dissertation thus offers new insights into popular genres of the eighteenth century, as well as the potential of computational analysis in literary studies.

Résumé

Cette thèse étudie, à l'aide de modèles prédictifs supervisés, la diversité des récits de voyages publiés en Grande-Bretagne entre 1700 et 1830. Le récit de voyage, qui constituait l'un des genres littéraires les plus populaires au cours du long XVIII^{e} siècle, prenait alors une multitude de formes allant du guide pédagogique jusqu'aux mémoires personnels. Cette hétérogénéité de forme, de contenu et de style a entraîné l'élaboration de plusieurs définitions contradictoires du récit de voyage, et même des débats sur sa véritable existence en tant que genre littéraire. *Voyages encodés* explore ce panorama contesté en modélisant ces différentes perspectives au sujet du récit de voyage à l'aide d'algorithmes. Premièrement, cette thèse décrit l'élaboration d'une base de données comportant plus de 4 000 récits de voyages tirés de 14 sources bibliographiques savantes. Par la suite, elle documente le développement théorique et pratique de cette base de données à l'aide de métadonnées riches, notamment les en-têtes et l'identification des rôles joués par les femmes dans la rédaction, l'édition et la traduction des récits. Enfin, *Voyages encodés* utilise les techniques de modélisation perspective développées par Ted Underwood afin de démontrer que les modèles algorithmiques sont fréquemment en mesure d'identifier les titres qui appartiennent à des récits de voyage dans neuf cas sur dix, et ce, malgré la multiplicité des définitions de ce qui constitue un récit de voyage, au juste. Ces découvertes démontrent que le récit de voyage est loin d'être un genre indéfini et qu'il converge plutôt vers un certain nombre de caractéristiques spécifiques. En comparant les modèles développés, cette thèse identifie aussi des tendances au sein de sous-ensembles importants du corpus des récits de voyages, dont les voyages fictifs et les oeuvres rédigées par des femmes. Elle propose ainsi de nouvelles approches pour l'étude des genres de la littérature populaire du XVIII^{e} siècle et, plus généralement, pour l'analyse informatique en études littéraires.

Acknowledgements

If there is one thing I have learned from this dissertation, it is that every journey—physical or figurative—depends on the generosity of one's hosts and travelling companions.

First, I acknowledge my responsibility to the lands that have shaped me, and their peoples. I completed this dissertation as a white settler on Tiohtià:ke (Montréal), which has long been a place of meeting and exchange among Indigenous peoples, including the Haudenosaunee and Anishinabeg nations, and traditionally cared for by the Kanien'kehá:ka. I grew up and completed my undergraduate and Master of Arts degrees on Treaty 4 Territory, Treaty 6 Territory, and the Homeland of the Métis. I honour and respect these many nations, the traditional stewards of the lands and waters which I love, and the lands on which I continue to live and work. Unlike many of the authors studied in this dissertation, I hope to recognize and follow through on my responsibility as a guest on these territories.

I am also grateful to my many academic communities. The English Department and the Burney Centre at McGill have been a home for over a decade, and I am indebted to the mentors, colleagues, and friends from those hallowed and tea-filled halls. The warmth and rigour of the eighteenth-century and digital humanities communities, especially in digital form (and especially the Space Cats Discord server), have greatly influenced my scholarship and pedagogy. This dissertation would also not exist without the extensive support of many librarians, programmers, and administrative folks, especially at McGill Libraries and HathiTrust. Ted Underwood has always supported this project, both with resources and enthusiasm. I also gratefully acknowledge the financial support of the Social Sciences and Humanities Research Council, the Institute for the Public Life of Arts & Ideas, NovelTM, Distant Reading Early Modernity, and the McGill English department.

For me, the arrivals and departures of mentors are the most significant landmarks, especially as a first-generation university student trying to find my way. Even before McGill, Allison Muri, my

academic parent, opened many doors for me after that first discerning question of “What major are you in?” Ray Stephanson pushed me to think more deeply about the eighteenth century, pedagogy, and making big decisions; I’ve passed on his advice several times since my Master’s. Peter Sabor has been a stalwart supporter of this PhD from the beginning, unflagging in his encouragement, eye for detail, and openness to new methods, and it is an honour to be his last graduate student. The late Stéfan Sinclair’s generosity, in both code and spirit, is woven throughout this project, just as it is through so many others; I wish you could have seen it, Stéfan. Richard Jean So stepped in for the final stretch as a co-supervisor, with all of the challenges resting therein, and my work is the better for it.

I am grateful also for companions who traveled with me on this journey. Matthew, Jordan, Katherine, Jérémie, and Amy, you are always ready with a place to rest or a push to get going. Alyssa and Hannah, you cheered me on from the first to the final steps. Anna, colleague and comrade, provided advice on long sentences and stakes and pushing through to the end. Manuel, your faith and high expectations are the best cure for impostor syndrome, both in and out of academia. Emma, you have made me a better person—don’t laugh.

At the end of this decade-long odyssey, my thoughts turn again to family, both near and far. My parents, Randy and Elaine, always caring and hard-working, encouraged me to read and to try new things, even when it took me far from home. My siblings, Trevor and Finch, remind me of a world beyond academia. And finally, Brennan—thank you for being a home I can always return to.

List of Figures and Tables

Table 1.1: Date Range of TWDB Sources 46
Figure 1.1: Date Range of Consulted Sources 47
Figure 1.2: TWDB-origbib Publication Numbers, by Year 62
Figure 1.3: TWDB-origbib Publication Numbers, by Decade 62
Table 1.2: Number of Titles From Each Bibliographic Source 63
Figure 1.4: TWDB-origbib Titles by Source 65
Figure 1.5: TWDB-origbib Titles by Source, Without Cox 66
Table 1.3: Number of Cross-References in TWDB-origbib 67
Table 1.4: Ratio of Single Citations to Total Titles 71
Table 2.1: Number of Volumes with Travel LCSH in ECCO Modules 84
Table 2.2 Most Common LCSH Subdivisions in ECCO (1700–1800) 85
Figure 3.1: TWDB Publication Counts by Decade: All Volumes 114
Figure 3.2: Title Publication Counts by Decade: TWDB vs TWDB-origbib 114
Figure 3.3: TWDB Publication Counts: Volumes vs Titles 115
Figure 3.4: TWDB Volume Counts by Decade: Search Source 116
Figure 3.5: TWDB Volume Publication Counts by Decade: File Source 117
Figure 4.1: Model of `alltravel` in the TWDB 146
Table 4.1: Thirty Most Influential Terms of the TWDB Travel Model 147
Figure 4.2: Sample Predictive Models of Cox and TEE 150
Table 4.2: Accuracy of Bibliographic Models 151
Figure 4.3: Mutual Recognition of Bibliographical Sources 153
Table 4.3: Bibliographical Mutual Recognition with McVeagh 154
Table 4.4: Accuracy and Mutual Recognition of Imaginary and Fiction Models 157
Figure 4.4: Model Predictions of Select Imaginary Voyage Titles 159
Figure 4.5: Poetry Model Prediction 163
Table 4.5: Accuracy of Location Models 167
Figure 4.6: Model Predictions of Women's Voice 172
Figure 4.7: Pace of Change in TWDB (20-Year Average) 176
Figure 4.8: Pace of Change in ECCO Travel Writing (20-Year and 60-Year Average) 178
Table 4.6: Pace of Change in HT Travel Writing: Comparing 1771–1800 and 1801–1830 178
Figure 4.9: Pace of Change in HT Travel Writing (20-Year Average) 179
Table 1.1.1: TWDB Metadata Codes 193
Table 1.2.1: Cox Statistics in TWDB-origbib 199
Figure 1.2.1: Cox Titles in TWDB-origbib 200
Table 1.2.2: McVeagh Statistics in TWDB-origbib 200
Figure 1.2.2: Cox Titles in TWDB-origbib 201
Table 1.2.3: BTW Statistics in TWDB-origbib 204
Figure 1.2.3: BTW Titles in TWDB-origbib 205
Table 1.2.4: Robinson Statistics in TWDB-origbib 206
Figure 1.2.4: Robinson Titles in TWDB-origbib 207
Table 1.2.5: Gove Statistics in TWDB-origbib 208

Figure 1.2.5: Gove Titles in TWDB-origbib 209
Table 1.2.6: Murray Statistics in TWDB-origbib 210
Figure 1.2.6: Murray Titles in TWDB-origbib 210
Table 1.2.7: Murray Statistics in TWDB-origbib 212
Figure 1.2.7: Murray Titles in TWDB-origbib 213
Table 1.2.8: TEE Statistics in TWDB-origbib 214
Figure 1.2.8: TEE Titles in TWDB-origbib 214
Table 1.2.9: NCCO-Brynmawr Statistics in TWDB-origbib 217
Figure 1.2.9: NCCO-Brynmawr Titles in TWDB-origbib 218
Table 1.2.10: NCCO-Travel Narrative Statistics in TWDB-origbib 219
Figure 1.2.10: NCCO-Travel Narrative Titles in TWDB-origbib 220
Table 1.2.11: Andrews Statistics in TWDB-origbib 221
Figure 1.2.11: Andrews Titles in TWDB-origbib 221
Table 1.2.12: Leask Statistics in TWDB-origbib 223
Figure 1.2.12: Leask Titles in TWDB-origbib 223
Table 1.3.1: TWDB-origbib excerpt, showing cross-references 224
Table 1.3.2: TWDB-origbib excerpt with consolidated cross-references 224
Figure 2.1.1: ECCO Publication Numbers, by Year 226
Figure 2.1.2: TWDB Publication Numbers, by Year 227
Table 2.2.1: Keyword Results in ECCO (Online vs Hard Drives) 228
Table 2.2.2: Counts of Volumes Holding Each LCSH Keyword in ECCO Modules 229
Table 2.2.3: Most Common Main Headings in ECCO (1700–1800) 230
Table 2.2.4: Most Common Subdivisions in ECCO (1700–1800) 230
Table 2.2.5: Most Common Main Headings in ECCO Travel Writing 231
Table 2.2.6: Most Common Subdivisions in ECCO Travel Writing (1700–1800) 231
Table 2.2.7: Raw Counts and Ratios of Keyword Results in ECCO and HT Searches 232
Figure 3.1.1: Ratio of Found to Remaining Titles in TWDB-origbib 246
Table 3.1.2: Number of Found TWDB-origbib Titles 246
Figure 3.1.3: Raw Count of Found to Remaining Titles in TWDB-origbib 247
Table 3.1.3: Place of Publication of TWDB Volumes 248
Table 3.1.4: Top Ten Institutions Contributing to the TWDB 248
Table 3.2.1: Titles with No LCSH, by TWDB fileSource 249
Table 3.2.2: Number of Titles Returned by Each Travel Keyword 249
Table 3.2.3: Titles with Travel Keywords Cross-referenced with Travel 249
Table 3.3.1: Most Common Complex LCSH (Main Heading and First Subdivision Only) 250
Table 3.3.2: Count and Frequency of All LCSH Elements (Main Headings and Subdivisions) 250
Table 3.3.3: Most Frequent LCSH Elements by Decade 251
Table 3.3.4: Top Five Most Frequent LCSH Main Headings by Decade 251
Table 3.4.1: Count and Frequency of Main Headings (No Titles with Travel Keywords) 252
Table 3.4.2: Count and Frequency of Subdivisions (No Titles with Travel Keywords) 252
Table 3.5.1: Unmatched Gendered Titles in TWDB-origbib 253
Table 3.5.2: Gendered TWDB Titles of the 1720s 254
Table 3.6.1: Ratio of Claim Locations 255

Table 3.6.2: Gendered Roles in the TWDB 255
Table 3.6.3: Location of Claim of Gendered Author Role 256
Table 3.6.4: Location of Claim of Gendered Translator Role 256
Table 3.7.1: Main Heading Frequencies in Gendered TWDB Corpus 256
Table 3.7.2: Main Heading and Subdivision Frequencies in Gendered TWDB Corpus 257
Table 3.7.3: Titles in the TWDB holding `wom` 257
Table 4.2.1: Mutual Recognition of Bibliographical Sources 262
Table 4.2.2: McVeagh's Irish Travels and Location Mutual Recognition 263
Table 4.2.3: Voyages, Imaginary: Select Titles 264
Table 4.2.4: Poetry Model Mutual Recognition 265
Table 4.2.5: History Model Accuracy 265
Table 4.2.6: History Model Mutual Recognition 265
Figure 4.3.1: Pace of Change in LCSH `Description and Travel` Model 266

Letters of Introduction

> Thus the whole circle of travellers may be reduced to the following *Heads*.
>
> Idle Travellers, Inquisitive Travellers, Lying Travellers, Proud Travellers, Vain Travellers, Splenetic Travellers.
>
> Then follow the Travellers of Necessity.
>
> The delinquent and felonious Traveller, The unfortunate and innocent Traveller, The simple Traveller, And last of all (if you please) The Sentimental Traveller (meaning thereby myself) who have travell'd, and of which I am now sitting down to give an account—as much out of *Necessity*, and the *besoin de* Voyager, as any one in the class.
>
> —Laurence Sterne, *A Sentimental Journey through France and Italy* (15)[1]

> The literature of travel is gigantic; it has a thousand forms and faces.
>
> —Percy G. Adams, *Travel Literature and the Evolution of the Novel* (282)

The eighteenth century was a critical time of development for British travel writing. Authors, including both veteran travellers and writers adapting others' texts, interpreted, criticized, and presented their journeys and destinations, explicitly and implicitly developing concepts of aesthetics, morality, identities, and nationalism. Such writing flourished in a variety of forms, including instructions for gentlemen on the Grand Tour, antiquarian accounts, adventures of buccaneers, critiques of slavery, and translations and adaptations of classical texts. Authors of domestic and foreign travel literature include Grub Street hacks, explorers of the New World,

---

[1] Sterne's "preface in the Desobligeant" is in chapter 7.

and major literary figures, such as Daniel Defoe, Lady Mary Wortley Montagu, Henry Fielding, Samuel Johnson, James Boswell, Laurence Sterne, and Mary Wollstonecraft. Increasing public demand for travel-related texts and the genre's potential for examining and experimenting with social and philosophical ideas thus elevated travel literature to one of the most important genres of the century.

This dissertation, *Coded Voyages: Modelling Genre and Travel Writing in the Eighteenth Century*, collects thousands of English-language travel writing texts published in Great Britain between 1700–1830, drawing on the work of past bibliographers, to create a Travel Writing Database (TWDB). After exploring the dataset, I analyze this huge genre by using perspectival modelling. In doing so, this project contributes to the field's ongoing effort to recognize and analyze more of travel writing's "thousand forms and faces" (Adams *Travel* 282).[2]

The Question of Genre

One of the most common divisions—whether explicitly or implicitly—in establishing a corpus or canon is the genre of the works. Genre has been a part of literary criticism since Aristotle's *Poetics*, though the discussion has moved far beyond the mode of imitation in poetry. In a general sense, genre can "refer to a distinctive category of discourse of any type, spoken or written, with or without literary aspirations" (Swales 33). In the 1970s and 1980s, however, many scholars, particularly linguists and literary theorists, attributed genre to underlying linguistic patterns and typology. For example, Douglas Biber's 1988 "A Typology of English Texts" creates a taxonomy distinguishing between a genre, decided by external, non-linguistic, "traditional" criteria, and a text type, determined by internal, linguistic criteria. Biber's article is

[2] The literature review in this introduction focuses on the genre of eighteenth-century travel writing. I review bibliography and digital humanities methods in later chapters.

still recognized as "a pioneering work and a milestone in this field of research," but his "taxonomy of text types has not gained ground among linguists and his text typology is too fuzzy and subjective to be generalized" (Santini 19).[3] Critics such as Jacques Derrida have further questioned the theoretical possibility of establishing concrete genres; in "The Law of Genre," Derrida suggests that "Every text participates in one or several genres…yet such participation never amounts to belonging" (65). Instead, "the law of the law of genre…is precisely a principle of contamination, a law of impurity, a parasitical economy" (59).[4] To add to the confusion and endless caveats, especially as new genres rise or old ones fall (or we gain different perspectives in later centuries), congeneric forms—genres appearing around the same time as each other—"cast light by virtue of their deeper similarities, independently of any putative influence" (Hulme "Colonial" 93). Genre thus remains far from a settled category.

Despite this "contamination," however, genre shapes not only the contemporary media landscape of bookstore shelves and Netflix recommendations, but more importantly for this study, literary criticism, bibliography, and algorithmic studies of genre.[5] In general, scholarly attention has shifted from theoretical arguments to sociocultural or practical debates. For most contemporary genre research, Ted Underwood suggests that "Scholars tend to treat genre as a social phenomenon rather than a symptom of deeper linguistic structures" (*DH* 8). Marina Santini, in her "State-of-the-Art on Automatic Genre Identification," suggests that "It is worth noting that most projects on automatic genre identification/classification do not bother very much with these theoretical issues: in many cases what they aim to achieve is a classification of

---

[3] For further discussion, see also David Y. W. Lee, "Genres, Registers, Text Types, Domains and Styles."
[4] Derrida playfully opens his "Law" with repetitions of "Genres are not to be mixed. I will not mix genres" (55).
[5] The main journal for academic studies of genre, *Genre: Forms of Discourse and Culture,* is "devoted to the study of the codes, conventions, and histories of generic forms in the context of their cultural manifestations and effects" ("About the Journal").

documents not based on the 'content', but on other features" (3). Thus, even though genre is a common term, it "remains a fuzzy concept, a somewhat loose term of art" (Swales 33) that is defined according to particular fields, scholars, and even individual projects. This is especially evident in the conception of travel writing as a genre.

Unpacking the Genre of Travel[6]

Defining travel writing has long been a point of contention in modern criticism. Can a vade mecum on travel practices be in the same genre as Sterne's experimental *Sentimental Journey*, satires involving a trip to the moon, or memoirs of a woman in the Highlands of Scotland?[7] Sometimes, critics—both those contemporary with the authors or looking back with modern hindsight—focus on whether the travel writing was true: that is, had the authors actually seen what they were describing, or were they simply writing from their Grub Street garret?[8] In 1770, the *Critical Review* laments that "Because there have been lying travellers, the veracity of almost every traveller is suspected" (qtd. in Batten 21). A work may still use the conventions of "real" travel writing even if the travel did not actually happen; as Charles Batten argues, reviews, prefaces, and essays show that authors of travel writing were "Clearly aware of the literary demands of their age, [and] they wrote in a firmly detailed tradition" (3). Some of these conventions and expectations, however, change during the 130 years under examination in this dissertation.[9]

---

[6] A play upon Barbara Korte's first chapter, "Charting the Genre," in *English Travel Writing from Pilgrimages to Postcolonial Explorations*.
[7] See chapter 4 for a discussion of *Letters from the Mountains* (1806) by Anne MacVicar Grant.
[8] See Adams' 1962 monograph, *Travelers and Travel Liars, 1660–1800* and Innes M. Keighren et al. (11–8).
[9] See Batten for several examples, including how "evaluations of [Joseph] Addison's [*Remarks on Italy*] progressively became more and more severe" as the century progressed (18–9).

The perceived literary value or cultural influence of a travel text may also affect its categorization. As Barbara Korte points out in her survey of the field, some definitions have excluded a considerable number of travel texts because of "the view that travel writing is only literature if couched in 'literary' language or 'poetic prose'" (Segeborg qtd. in Korte 14). Jean Viviès, in *English Travel Narratives in the Eighteenth Century* (2002), for example, examines six "case studies" of Boswell, Johnson, Sterne, and Tobias Smollett because the authors' "literary reputations were not determined by their travel writings" (24).[10] Even current discussions about travel writing subgenres are rife with prioritizing literary elements; for example, as Nigel Leask discusses, the purported division of subjective and objective travel texts elevates one form of travel writing over the others (*Curiosity* 7–8).

In general, however, the field of eighteenth-century travel writing is expanding its borders rather than narrowing them. Instead of trying to detect whether the author wrote the truth, Korte characterizes travel texts as narratives, usually (but not exclusively) in prose, that "claim—and [make] their readers believe—that the journey recorded actually took place, and that it is presented by the traveller him or herself" (1). Benjamin Colbert similarly defines travel writing based on genre and witness, focusing on nonfiction texts (but not paratexts) about travel that actually happened—including children's storybooks based on real travel, but not certain subgenres such as shipwreck literature.[11] Others also define travel writing through exclusion: in the conclusion of *Travel Literature and the Evolution of the Novel*, Adams defines a "récit de voyage" by what it is not: it is not

---

[10] Reading Viviès during my compulsory research project on Daniel Defoe's *Tour thro' the Whole Island of Great Britain* was part of what inspired this dissertation project.
[11] For a full description, see Colbert's entries in appendix 1.2.1.

> just a first-person journal or diary…[not] necessarily a story with a simple, uncontrived plot…not just a set of notes jotted down each day or whenever the traveler has time…not just an objective report…not a branch of history any more than it is of geography…not just an exploration report…not just a complete record of a journey. (280–2)

Adams concludes his definition by negatives by arguing that travel writing "cannot be a literary genre with a fixed definition" because "it includes so many types both by form and by content" (282).[12] Jan Borm draws similar attention to variety through his list of "terms in use" for travel writing, noting that "their sheer abundance raises the question of what we actually mean by the travel book and travel writing" (13).[13] Borm uses this list to set up his own definition, which echoes Adams': it is "not a genre but a variety of texts, both predominantly fictional and non-fictional, whose main theme is travel" (13). Similarly, Mary Baine Campbell describes travel writing as "a genre composed of other genres" (6).[14] More recently, Katrina O'Loughlin suggests that "In its distinctive narrative heterogeneity, eighteenth-century travel writing might therefore be best understood less as a unified 'genre' than a powerful register of simultaneity, cultural comparison, and critique: a dialectical impulse at the heart of early modern sociality" (6–7).[15] Tim Youngs goes one step further, arguing that "Travel writing feeds from and back into other forms of literature. To try to identify boundaries between various forms would be impossible and

---

[12] Adams argues that this variety of form and content also makes it impossible to define the novel.

[13] Borm's list includes "'travel book', 'travel narrative', 'journeywork', 'travel memoir', 'travel story', 'travelogue', 'metatravelogue', 'traveller's tale', 'travel journal', or simply 'travels' (*The Travels of Sir John Mandeville*), and, in a different vein, 'travel writing', 'travel literature', 'the literature of travel' and 'the travel genre'" (13).

[14] Grzegorz Moroz suggests that Campbell's theoretical positions are, in general, "strongly anti-generic" (22). See Moroz's overview of the field in "Travel Book as a Genre in the Anglophone Literary Tradition," Campbell's chapter "Travel Writing and Its Theory," and Katherine Turner's summary of critical approaches (4–11). Adams' "Travel Literature of the Seventeenth and Eighteenth Centuries: A Review of Recent Approaches" offers a useful time capsule of the field in 1978.

[15] Korte suggests that "More acutely than any other genre, then, travel writing is defined by the interaction of the human subject with the world" (5).

I would be deeply suspicious of any attempt at the task" (8). Dennis Porter is similarly wary of "recent efforts to isolate 'literary travel' from other kinds, to define the characteristics of the genre, and to formulate a poetics" (3). These "hazardous" definitions—focusing on exclusion or the impossibility thereof—speak to the inherent constraints of generic categorization.[16] However, they also emphasize the especially challenging nature of assigning labels to travel writing, a provocation which this dissertation embraces by accepting multiple, sometimes conflicting definitions.

One challenge in applying labels to past genres is both a blessing and a curse: hindsight. Campbell suggests that that "It is perhaps only from the armchair of the postcolonial twentieth century that these works can be seen as bearing a close enough family resemblance to constitute a genre" (*Witness* 5).[17] Leask points out the dangers of anachronistic divisions, critiquing the supposed "disjuncture between scientific and literary travel writing in the decades after 1790–1820 (decades which saw the retrospective construction of 'romantic ideology'), rather than seeing it as essentially constitutive of the genre in the period itself" (*Curiosity* 7). He continues, arguing that "To generalize travel writing in the romantic period as merely 'subjective' is to ignore not only the majority of travelogues produced during the period but also the testimony of contemporary commentators" (*Curiosity* 7–8). Hindsight can also be an advantage, however. In a later overview of the field, Campbell points to feminism and postcolonialism as ongoing reformations of conceptions of travel writing. As in other historical studies of literature, feminist

---

[16] Definitions were of concern in the eighteenth century as well. In 1755, Johnson in *The Rambler* notes that "definitions are hazardous. Things modified by human understandings, subject to varieties of complication, and changeable as experience…are scarcely to be included in any standing form of expression, because they are always suffering some alteration of their state. Definition is, indeed, not the province of man" (*Selected Works* 79). Johnson's *Dictionary of the English Language* was published this same year.

[17] Campbell focuses on texts written before the seventeenth century and how travel writing is a "slow assembling of the features that now identify a work as 'travel literature'" (5).

scholarship has resulted in travel texts by women writers being "discovered, rediscovered, or revisited" ("Travel Writing" 264).[18] Similarly, Edward Said's *Orientalism* created an "epistemological shift" in understanding questions of empire, the Other, and the travelling, or stationary, self (Campbell "Travel Writing" 265).[19] These new theoretical and methodological approaches continue to develop our understanding of the genre, and as I discuss in chapter 4, they can provide ways of comparing these multiple historical perspectives.

Without a strict definition, or even a common one, it may come as no surprise that there is hardly a canon of travel writing.[20] Charles Grivel describes travel writing as "a neglected literature; compared to the samples of canonical genres, it can hardly offer something like 'works'" (qtd. in Korte 269). Korte suggests, however, that "The absence of a canon is not, in itself, problematic. Quite the contrary: travel writing offers the reader literary ground which is previously untrodden and unmapped, and in which there is a lot to discover for oneself," arguing that scholars should "counteract canon formation" (17).[21] Travel writing may lend itself to the task naturally, since, as Steve Clark suggests, the genre is "collective and incremental rather than singular and aesthetic" (1). Certainly, authors of travel writing known for their literature, such as those studied by Viviès—male and white, with perhaps Wollstonecraft as a token woman—may

---

[18] See Colbert, "British Women's Travel Writing, 1780–1840: Bibliographical Reflections." See also Yoojung Choi's 2020 dissertation, *Women's Mobility, Travel, and Literary Representations in the Long Eighteenth Century* and O'Loughlin's monograph, *Women, Writing, and Travel in the Eighteenth Century.*

[19] For a further overview of the theory of travel writing and how it connects with other theoretical developments and schools, including psychoanalysis, linguistics, and ethnography, see "Travel writing and its theory" by Campbell in *The Cambridge Companion to Travel Writing* (261–78). See also Claire Lindsay's overview of "Travel Writing and Postcolonial Studies" and Hector Roddan on *Orientalism* and early modern historiography.

[20] Katherine Turner's *British Travel Writers in Europe 1750–1800* opens with how travel writing, "although culturally preeminent in its day, has fallen victim to the vagaries of canon-building" (1).

[21] Korte addresses the "problem of availability," relying on texts accessible in print and in libraries as her primary examples, supplemented by "lesser-known examples" (17). Turner similarly notes how "hundreds of European travelogues…remain out of sight in most libraries, and have generally been out of print since the eighteenth century" (1).

be the first to come to mind.[22] Turner and Korte both agree, however, that such a focus on 'literary' texts limits our recognition of how authors use genre conventions, as well as ignoring more popular or influential travel titles, or simply those that contribute to the heterogeneity of the form.[23]

Shifting attention from texts with "literary" merit or those that are considered "true" encourages us to look at the thousands of other travel texts of the long eighteenth century, and the many styles, topics, and interventions that they offer. This dissertation and its accompanying dataset embrace Derrida's genre "contamination," or, as Jonathan Raban describes travel writing, this "notoriously raffish open house where very different genres are likely to end up in the same bed" (9). At the same time, I also agree with Batten's assertion that "[literary] conventions often govern a travel writer's actions and descriptions" (4). This project also takes Korte's position that travel writing's "hybridity in terms of text modes and style" (15) is also influenced by the author's "particular strategies—including specific artistic principles and designs"—in participating in a "particular form of writing" (2–3). These last observations echo Underwood's description of modern genre studies as focusing on genre as a "social phenomenon" ("Life Cycles" 6). As the rest of this introduction shows, elements as abstract as aesthetic movements and as practical as armed conflict affect authors' engagement with this complex genre.

Travel Writing and the Novel

---

[22] Few studies, especially overviews, focus on British texts by writers of colour. Few were published, and fewer still have modern editions. See modern Broadview editions of *The Letters of the Late Ignatius Sancho, an African* (1784), *Thomas Clarkson and Ottobah Cugoano: Essays on the Slavery and Commerce of the Human Species* (1786), *The Interesting Narrative of the Life of Olaudah Equiano* (1789), and *Woman of Colour: A Tale* (1808). See also *Unchained Voices: An Anthology of Black Authors in the English-Speaking World of the Eighteenth Century* and related sections in "Travel Writing in a Global Context" in *The Cambridge History of Travel Writing*.

[23] Turner argues that "we lack a coherent picture of [popular 'literary'] travelogues context, and a proper sense of how representative or innovative they really are" (4). For Korte, "focusing on only a select group of texts makes the scholar blind to important historical developments and general generic features" (15).

In studies of travel writing, the most studied "cross-contamination" is with the novel, a congeneric form studied much more than travel writing.[24] Their histories deeply intertwine, so much so that Adams wrote a whole book on the subject.[25] Travel is a staple of fictional plots, including Defoe's adventurous shipwreck *Robinson Crusoe* (sometimes considered the first novel), Samuel Richardson's psychological *Clarissa*, and Henry Fielding's picaresque *Tom Jones*.[26] As Bohls notes, "It is scarcely possible to discuss the eighteenth-century novel without speaking of travel" ("Age" 97). Although novels would, by the end of the century, supersede both sermons and travel texts to become the most popular genre, initially the title pages of novels "tried to capitalize on the contemporary popularity of travel books by suggesting the similarity of their wares" (Hunter *Before Novels* 353). As Turner points out, however, this focus risks being myopic:

> [fixating on travel writing's] influence on other literary forms, predominantly the novel, which has appeared in recent (and not-so-recent decades) . . . effectively relegate[s] travel literature to the status of slave to the master discourse of imaginative prose—a procedure curiously at odds with eighteenth-century reading and critical practices. (8)

This trend has continued to some degree with modern criticism: in the *Cambridge Introduction to Travel Writing*, the opening paragraphs of "Travel Writing in the Eighteenth Century" focus on the relationship between the novel and travel writing, even positioning modern travel writing

---

[24] Clark also describes the "impurity of the form" of travel writing (2).
[25] See *Travel Literature and the Evolution of the Novel*, which Elizabeth A. Bohls calls "a magisterial study" ("Age" 97).
[26] All of these works feature prominently in Ian Watt's influential (and often criticized) *The Rise of the Novel: Studies in Defoe, Richardson, and Fielding* (1964).

as relying more on the eighteenth-century novel than on the travel writing of that period.[27] As O'Loughlin recently suggests, the "critical suspicion" and "marginalisation of travel writing has, however, clearly shifted in the last two decades," suggesting that "Leask rightly identifies the field of travel writing studies as 'one of the major achievements of interdisciplinary scholarship' in recent years" (Leask qtd. in O'Loughlin 7). Indeed, scholars of travel writing come from many fields such as history, economics, and anthropology, demonstrating travel writing's "raffish" nature and relevance to multiple disciplines, and any survey of the field requires attention to these many cultural contexts.[28]

The History of British Travel

In 1764, Boswell, future author of his own *Account of Corsica* (1768) and *The Journal of a Tour to the Hebrides with Samuel Johnson* (1785), wrote in his Grand Tour journal that "I am made for travelling" (135). He was not part of a new phenomenon: people had been travelling for thousands of years, for business, family, pilgrimages, and pleasure, as well as reasons of escape and enslavement.[29] While travelling was certainly not a uniquely British phenomenon, the English had a reputation for being particularly prone to travel, even in medieval times.[30] By the

---

[27] Tim Youngs opens the chapter with the assertion that "The narrators of many of today's travel books owe something of their character to the humorous protagonists and comic adventures of the eighteenth-century novel, as well as to the introspection of the literature of the later decades of that century" (38). For a recent nuanced reading of novels and travel writing, see Bohls' "Age of Peregrination: Travel Writing and the Eighteenth-Century Novel." For travel writing and fiction beyond just the eighteenth century, see Janicke Stensvaag Kaasa, "Travel and Fiction."

[28] *Studies in Travel Writing*, the main journal of the discipline, is open to all fields and periods and was started by Tim Youngs in 1997. See also Keighren et al. for an overview of interdisciplinary approaches (6–11).

[29] Histories of travel often gesture to Pausanius' guidebook to ancient Greece in the second century CE as an important model (Batten 42; Bohls and Duncan xiii-iv). See Maria Pretzler's *Pausanias: Travel Writing in Ancient Greece*.

[30] Korte refers to *The Travels of Sir John Mandeville*, where the narrator proclaims that "We are in a climate under the rule of the moon, which is a planet that moves quickly—the traveller's planet" (2). In 1766, the *Monthly Review* suggests that "The English are beyond all doubt the greatest travellers in the world" (qtd. in Turner 2). I focus primarily on British travellers below, since these authors comprise the majority of this dissertation's dataset. Any titles published in Britain, regardless of nationality of the author, would probably account for this cultural context.

early modern period, travellers were expanding on the horizons of their medieval counterparts. Explorers were not just on pilgrimages to Canterbury or the Holy Land, but "discovering" islands on the other sides of the world. The Grand Tour, where wealthy young men and their tutors would spend months or years travelling the Continent to learn manners, languages, and history, was an established institution by the 1700s. These adventures provided material not only for letters home, but also for satirists, who critiqued the newly cultured young men's foreign manners and sexual diseases.[31] Instructions, such as James Howell's *Instructions for Forreine Travel* (1642) and Vicesimus Knox's essay "On the Manner of Writing Voyages and Travels" (1778) attempted to encourage not just appropriate social behaviour while abroad, but also one's responsibilities. The concept of *utile dulce*, or pleasurable instruction, was both explicit and implicit in travel writing intended for men, women, and children alike.[32] What began as a practice of upper-class, classically trained gentlemen at the beginning of the century had shifted, by 1830, into a practice accessible to much of the middle-class, and one reflecting many of their sensibilities.[33]

Travel became significantly easier as transportation improved, especially with roads and later, in the 1830s, steamship and train travel. In the early eighteenth century, "the roads of Britain had been in a severe state of dilapidation: travelling for reasons other than necessity was

---

[31] For example, in Alexander Pope's *Dunciad in Four Books*, "I see advance / Whore, Pupil, and lac'd Governor from France" (4: 271–2; see also 272n.). William Combe and Thomas Rowlandson's series of poems about Doctor Syntax, beginning with *Doctor Syntax in Search of the Picturesque* (1812) satirized William Gilpin, which the *Athenaeum* called "the most popular work of the first ten years of this century" (qtd. in Gray 232). Erik Gray notes that "Today, however, the poem has all but disappeared from view" (232).

[32] See Batten, *Pleasurable Instruction*, especially 24–9. Other guides to travel include Josiah Tucker's *Instructions for Travellers* (1757) and Leopold von Berchtold's *Essay to direct and extend the inquiries of patriotic travellers* (1789). Colbert describes J.W. Cunningham's *Cautions to Continental Travellers* (1818) as an "anxious response" to "the sign of mass travel and mass culture" (*Shelley's Eye* 3).

[33] James Buzard outlines how the long eighteenth century, roughly 1660–1840, "is marked by the emergence of this new paradigm for travelling—that of the 'Grand Tour'—and concludes with the first glimmerings of another paradigm that absorbed and superseded it: that of mass tourism" ("The Grand Tour and After" 38).

only for the hardy" (Kinsley 5). With the rise of turnpikes, "the improvements in the national road network, and the increased mobility and improved travelling conditions that resulted, began to have a fundamental effect on the leisure pursuits of the population" (Turner 5).[34] These advances were not limited to Britain, either; Europe's stagecoach networks following the Napoleonic wars made travelling the Continent significantly easier and cheaper (Korte 43), as did technologies for travel at sea, including both navigation and cartography.[35] But even if travellers knew their routes, travel could still be dangerous. In addition to natural risks, there were those provided by the citizens and governments hostile to the British.[36] This was particularly evident with travel to France; the Treaty of Paris in 1763 "made Continental travel physically more feasible, [and] victory rendered it patriotically enjoyable" but later conflicts, such as the French Revolution and Napoleonic Wars, made travelling in Europe unpredictable (Turner 25).[37] This danger encouraged domestic tourism, as well as travel to other European locations off the beaten track, such as Wollstonecraft to Scandinavia.[38] British travel was not

---

[34] For example, in Defoe's *Tour Thro' the Whole Island of Great Britain* (1724), he advocates for a turnpike system, which would "see the Roads all over England restor'd in their Time to such a Perfection, that Travelling and Carriage of Goods will be much more easy both to Man and Horse, than ever it was since the Romans" (2: 191–2). These developments had dramatic effects: the "London-York run [went] from five days in the seventeenth century to four days in 1706 and to thirty-one hours in 1790" (Ousby qtd. in Rogers 784).

[35] The British Parliament published "An Act for providing a public Reward for such Person or Persons as shall discover the Longitude at Sea" in 1714. See also Richard Sorrenson's "The State's Demand for Accurate Astronomical and Navigational Instruments in Eighteenth-Century Britain." For advancements in surveying and maps, see Mary Sponberg Pedley (19–34).

[36] After Captain James Cook's crew's abuse of the Indigenous people of Hawa'ii, the people finally killed several sailors, including the Captain himself. See William Frame and Laura Walker (206–13).

[37] Novelist Frances Burney was famously exiled in France during the Napoleonic wars. See Margaret Anne Doody's chapter "*The Wanderer; or, Female Difficulties*: Revolution, the Rights of Woman, and 'The Wild Edifice'" in *Frances Burney: The Life in the Works* (313–8).

[38] Turner notes that "After 1815, more exotic destinations, such as Russia, Egypt, and India were both safer and fashionable, and increasingly accessible to the wives of those merchants, diplomats, and military men who were sowing the seeds of empire overseas" (49).

limited to Europe, however, and explorations for the sake of curiousity, science, and colonization continued to shift, often funded by governments or private clubs.[39]

Consumers of Travel

A final major shift in the eighteenth century is that of class, as travel became accessible to "a far more diverse body of travellers," especially those of the "middling sort" (Turner 25). Susan Lamb troubles the "powerful myths that the eighteenth-century tourist was an aristocratic youth who traveled to complete his education, or that, in any case, all eighteenth-century tourists were aristocratic men," arguing that "contemporaries were well aware of the diversity of touring populations on both the Continent and in Britain" (16, 99).[40] Bohls and Ian Duncan suggest that a "conspicuous trend of leisured tourism emerged in the 1770s" available to the middle classes, especially connected to domestic travel. By 1815, John Scott, an editor and publisher, was joking that "our book-shelves groan with the travels of persons who have suddenly arisen from almost every class and profession of life" (qtd. in Keighren et al. 4). A reviewer of Mary Shelley and Percy Shelley's anonymous *History of a Six Weeks' Tour* (1817) was more harsh, complaining that "The dashing *milords* of the last age are now succeeded by a host of *roturiers* [commoners], who expatriate themselves for the sake of economy" (qtd. in Colbert "European Tour" 6).[41] These discourses shifted the expectations for and discourse of travel writing, ranging from

---

[39] The Association for Promoting the Discovery of the Interior Parts of Africa (commonly known as the African Association) formed in 1788. The Admiralty (the commanding authority of the British Royal Navy) and the Board of Longitude influenced and commissioned Arctic and African exploration. See Keighren et al. for the influence of such organizations on Murray publishing house (1–33).

[40] Note that Lamb is unusual in her broad uses of "tourist" and "tourism," which eighteenth-century scholarship typically avoids to distinguish it from the mass tourism that began post-1830.

[41] Buzard discusses the shift to an idea of "universal access ('all sorts and conditions'), not in terms limited to one privileged class" (101–2). He pays particular attention to Samuel Rogers's poem *Italy* (1822).

guidebooks targeting those travelling with wives rather than tutors, to texts written by and for both amateur and professional travellers.

Consuming travel writing also became increasingly popular. In the 1795 introduction to the translation of the *Travels of Carl Philipp Moritz in England in 1782*, C.G. Worde suggests that "One of the most distinguishing features in the literary history of our age and century, is the passion of the public for voyages and travels. Of the books that have lately been published, there are none, novels alone excepted, that, in point of number, bear any proportion to them" (qtd. in Rogers 781).[42] Contemporary discourse around travel writing is supported by analysis of publication numbers and library circulation records. Batten's 1978 study, *Pleasurable Instruction*, opens by describing the "unparallelled popularity" of travel writing, much of which he notes may "ring hyperbolic" to modern ears (1). He suggests that nonfiction travel "accounts won a readership second only to novels by the end of the century," an assertion that Leask describes as "credible" and Turner characterizes as now "commonplace in critical work" (*Curiosity* 11; 10n26).[43] Quantitative analyses of publication numbers vary depending on definition. Edward Godfrey Cox, with a very broad definition ranging from agricultural treatises to antiquarian descriptions, lists over 4,000 titles. The *New Cambridge Bibliography of English Literature* lists just over 1,500 travel items written or published between 1695 and 1800 (qtd. in Rogers 785). William Matthews counts at least 300 titles on travel within Great Britain and Ireland between 1695 and 1830 (qtd. in Rogers 784). Small timeframes provide useful perspectives: Simon Eliot identifies an increase in travel writing in the first decades of the

---

[42] Batten also quotes the introduction to Moritz's *Travels*, as well as Robert Gray's introduction to his *Letters During the Course of a tour through Germany, Switzerland and Italy* with "no taste is more prevailing than that for books of travels" (2).

[43] For religious texts, Batten points to J.H. Plumb, who asserts that "'the output of travel books' during the 1720s and 1730s 'was second only to theology'" (qtd. in Batten 122n5).

nineteenth century (qtd. in Keighren et al. 4)[44] and Colbert's analysis shows that between 1814 and 1818, not only were there more (nonfiction) travel texts than novels published, but that several of the novels use the tropes of travel writing ("Bibliography" 13).[45] Relying on first printings of a title limits the perspective of the popularity, however. Some travel texts went through multiple editions very quickly, with estimates of reissues of between 10–15% (Rogers qtd. in Day 2). Titles were also abridged or adapted to different audiences, especially for children, and anthologies were both popular and profitable.[46]

Library lending numbers also support the "flood-tide proportions" of publications (Marshall and Williams 45). Paul Kaufman's analysis of the Bristol Library from 1773 to 1784 found that the two most popular titles were Patrick Brydone's *Tour through Sicily and Malta* (1773) and Hawkesworth's *Account of the Voyages…in the Southern Hemisphere* (1773) (qtd. in Batten 11).[47] In fact, Batten uses this data to show that, on average, travel books circulated at least twice as often compared to their non-travel counterparts (11n4).[48] In another of Kaufman's survey of eighteenth-century English libraries, an average of one in ten books were travel texts (Kaufman qtd. in Rogers 787). Similarly, travel and its close sibling history were the second

---

[44] Keighren et al. support travel's position as "second-largest genre" of the eighteenth century, behind either religious material or fiction. They point again to Eliot: "Notwithstanding the inconsistent categorization of travel texts in contemporary catalogs and indices, Eliot's statistical survey has shown that, between 1814 and 1846 at least, the portmanteau category of geography, travel, biography and history accounted for 17.3 percent of British book production (of which approximately three quarters were texts of travel), narrowly trailing religion at 20.3 percent" (5).

[45] Colbert contrasts the Centre for Editorial and Intertextual Research database of British fiction, which records an average of 58.4 new novels per year between 1814–1818. For the same time span, his data counts 98.8 travel titles.

[46] Turner estimates that about 25 collections were published in the eighteenth century alone, while Rogers documents 85 in the slightly larger time frame of 1695–1830—many of which were multiple volumes, sometimes up to 76 volumes (23; 786). Rogers bases his assessments on R. M. Wiles, *Serial publication in England before 1750*. For more on the size, cost, and profit of print runs, see Rogers (786).

[47] Brydone's *Tour* was loaned 192 times, Hawkesworth's *Voyages* 201.

[48] Specifically, on average, non-travel books circulated slightly less than 14 times, and travel books slightly more than 32 times.

most popular category at the Leeds subscription library from 1768 to 1809 (Beckwith qtd. in Rogers 788).[49] Periodicals also made travel writing more accessible, disseminating not just criticisms of texts, but content as well. Turner discusses the importance of the *Monthly Review*, the *Critical Review*, and the *Analytical Review*, which "extensively reviewed" and excerpted travel writing, exposing a "far wider readership" than of just monographs, since "perhaps as much as one-sixth of the reading public had access to one or more review journals on a regular basis" (12).[50] In Elizabeth Hagglund's study, "Reviews of Travel in the *Monthly Review*," she found that the journal reviewed an average of 39 "travel-related books" annually from 1749 to 1758, with a slow shift in content from summaries and extracts to more commentary (5).[51] These reviews thus "create[d] a taste that explorers learnt to supply" (Fulford and Kitson xxii).

Reading such texts was expected of learned men, as it could give both practical and theoretical perspectives on topics as varied as agriculture, philosophy, classical history, and forms of government; indeed, reading travel writing was almost a patriotic responsibility. In Bishop Richard Hurd's *On the Uses of Foreign Travel*, he advises (in the voice of John Locke) that "If you think I impose too great a task on your inquisitive traveller, my next advice is, That he stay at home: read Europe in the mirror of his own country…and, for the rest, take up with the best information he can get from the books and narratives of the best voyagers" (qtd. in Batten [xiii]). Many thinkers of the day seemed to agree, based on their personal libraries: Batten and Rogers point to the libraries of figures such as John Locke, Adam Smith, Samuel Johnson, David

---

[49] The most popular category was "belles lettres." Rogers estimates that "A survey of eighteenth-century English libraries found that on average travel books made up 12 per cent of holdings" (787).
[50] Wallace Cable Brown calculated "46 extensive reviews of Near East travel books, some of them continuing through several issues…between 1805 and 1825" (qtd. in Rogers 785).
[51] See Hagglund for further detail on the style of these reviews. For further examples from the *Monthly Review* et al., see Batten 1n1 for an extensive list.

Hume, and Thomas Jefferson (Batten 2n5; Rogers 789).[52] But travel writing was not just read by the stereotypical "great men": it was also considered appropriate, and even desired, for women and for children of both sexes.[53] With reductions in print costs and the increase of circulation, subscription, and public libraries, travel writing also became accessible to those with lesser economic means.[54] These factors lead Shef Rogers to argue that "travel writing was probably the most self-consciously print-informed genre of the period," with "travellers often conceiv[ing] of their travels as print-structured narratives" (784).

Writers of Travel

Though the tradition of the Grand Tour was influential and the subject of many parodies, contemporary reviews and modern scholarship demonstrate that few of the participants published travelogues.[55] Publishing was seen as low-class and for the "middling sort," and, as Bruce Redford argues, "letters and journals of the Grand Tour contributed little to this crucial staging of authority, for they tended to remain within the private sphere of the family" (81).[56] Words were still important—Redford points out the significance of more public "sartorial and linguistic

---

[52] The library at Godmersham Park, frequented by Jane Austen, contains dozens of travel titles. See *Reading with Austen*, directed by Peter Sabor. I was the project manager during the creation of the dataset and website.
[53] See Shef Rogers' "Enlarging the Prospects of Happiness: Travel Reading and Travel Writing."
[54] In her comparison of wages and prices of travel books, Hagglund notes that although "the audience for many of the books…was limited to the relatively well-off," the "preponderance of 6d and 1s pamphlets indicates the interest in travel from all strata of society" (6). G. R. Crone and R. A. Skelton further suggest that "by the middle of the 18th century the vogue for travel collections…had spread to the large middle-class reading public at a lower economic level" (qtd. in Leask 11).
[55] In 1766, for example, the *Monthly Review* noted that considering the numbers of young Englishmen on the Grand Tour, they write surprisingly few texts. The reviewer suggests that "The reason is plain: our travellers are in general young men of fortune, and are led by their tutors; and both of them, from the youth of one and the narrow education of the other, are as incapable of observation as if they were conducted through France and Italy blindfold." The reviewer continues, noting that "For want of that knowlege, steadiness, sagacity, and penetration, which can be only founded on study, and ripened by experience, they traverse the continent in a continued mist, gaping, staring, blundering along, and viewing every object in a false light" (34: 420). See also Turner (17–8).
[56] Rather, extravagant "Grand Tour portraits," "positioned carefully within the theatrical decor of the town or country house . . . Played the part of advertisements and icons" of class and authority, to the extent that "Italy-based artists such as Canaletto, Piranesi, and Rosalba Carriera made careers by painting largely for the tourist market" (Redford 81; Lamb 112).

performance," of the type that "in its most flamboyant form led to a variety of satirical attacks"—but published writing was not prioritized when claiming the authority of a Grand Tour alumnus. Instead, just like the "diverse body of travellers" travel writers included novelists, merchants, clergymen, explorers, diplomat's wives, and Grub Street hacks who plagiarized from all of the above (Turner 25). This growing middle class, according to Turner, "not only dominated the realms of published literature, but claimed most insistently to embody Englishness or Britishness, in contrast to the unpatriotic cosmopolitanism of the aristocracy" (17).[57] As Clark notes, "Travel writing has taken a mixed and middlebrow form throughout its history…Anyone can have a go, and usually does" (1). Authorship was also only one of many important roles. Translators could be anonymous or famed enough to have their names on the title page.[58] Similarly, compilers and editors could make significant sums: John Hawkesworth famously earned the unprecedented sum £6,000 for his editorial work on *An account of the voyages…[of] Captain Cook* (1773), one of the most popular travel books of the eighteenth century.[59] As with other forms of writing, authorship was not monolithic; composition and revision were influenced at multiple steps, from notanda to published work.[60]

Women Travel Writers

---

[57] Turner notes the tensions among British, English, Scottish, and Welsh identities. I follow her treatment of the issue: "In general within the discourses of eighteenth-century travel, 'English' and 'British' are deployed synonymously: or, indeed, 'English' . . . may be applied to a Welsh or Scots person, and notions of English eccentricity or liberty silently appropriated" (18). When considering other studies, Turner suggests that "It is therefore surprising how many otherwise excellent studies of travel writing still tend to overemphasize the Grand Tour component" (17). For example, she suggests that Buzard "assumes a pre-1800 landscape of travel writing defined solely by the Grand Tour and its 'overt class and gender prerogatives,'" and others, such as Dennis Porter, focus on Tour narratives without considering other forms (17).

[58] For example, on Anne Plumptre as translator, see A.E. Martin, Susan Pickford, and Glenn Hooper (*Anne Plumptre*).

[59] For comparisons of other editorial and authorial advances, see Rogers (786–7). Tim Youngs quotes Richard White in asserting that *An account of the voyages* was "the most popular travel book of the century" (50), and the title was the most borrowed in the Bristol Library from 1773 to 1784 (Kaufman 122).

[60] See Keighren et al. for several examples (18–21).

Gender was another significant factor affecting authorship in the long eighteenth century, especially as "Women did not fit the traveler's image as heroic explorer, scientist, or authoritative cultural interpreter" (Bohls *Women* 17). Cultural expectations limited women's activities in the public sphere, including travelling (especially without a husband or guardian), and travel was not cheap.[61] Rather than a monolithic group, however, women writers (or those claiming their voices) interacted with and interpreted locations, classes, and racism in varying ways—and many did so as professional writers.[62] Rather than dominant cultural forms like aesthetic treatises or philosophical enquiries, however, women writers often "chose genres more accessible to women, like the novel and travel writing" (Bohls *Women* 3). These authors had to navigate cultural movements such as sensibility with care, and they strategically used their feminine perspectives as justification for travel, writing, and publication.[63] As Bohls describes in her extensive study of how women authors subversively employ the male-dominated language of aesthetics, "Both tourism and writing for publication took women into the public realm in potentially transgressive ways" (*Women* 103).[64]

Travel writing by women has thus faced a double marginalization in criticism, leading to

---

[61] Increasing rates of literacy among women later in the century "extended their participation in aesthetic practices to reading and writing sentimental fiction and scenic tours" (Bohls *Women* 102). The resources necessary to travel and participate in the aesthetic discourses of travel writing, however, were largely restricted to higher-class women. Although "ranks of middle-class tourists had certainly swelled by the 1780s and 1790s . . . tourism's indispensable prerequisites of means, leisure, and education still restricted it to a comparative elite" (Bohls *Women* 90).

[62] A common example is Lady Mary Wortley Montagu's access to the woman-only bathhouses in Turkey. See Bohls (*Women* 27–38). Colbert notes that "For many female authors, the travel book was a first, and sometimes last, foray into print…many others had established or were establishing reputations as novelists, poets, and scientific and miscellaneous writers" ("Bibliographical Reflections" 9). Colbert also describes "less contentious authorial roles," such as editor or co-author, as part of the "evasive action" that female authors used when appearing in print (8). For professional travel writers, including novelists, in the early eighteenth century, see Choi's dissertation, *Women's Mobility, Travel, and Literary Representations in the Long Eighteenth Century*.

[63] Some authors, such as Mary Ann Parker and Lady Elizabeth Craven, "used their status as mothers to justify the publication of their travels…employ[ing] the rhetoric of sensibility and nationalism" and "the language of empire…with the language of family" (Zold 325).

[64] See Turner on the "Rise of the Woman Travel Writer" (127–180).

some critics still disregarding such texts in favour of male authors writing in more dominant genres.[65] Many male theorists, including Adams and Batten, "fail to theorize a place for women as traveling subject" (Lawrence 2).[66] Sometimes, these omissions arise from strict limits of what constitutes travel: for example, Fussell asserts that "to constitute real travel, movement from one place to another should manifest some impulse of non-utilitarian travel," a definition which Karen Lawrence asserts "too strictly polices the borders between self-motivated and forced movement, and between use and pleasure" (*Norton Book of Travel* 21; 21).[67] Travel texts by women are also more likely to be disregarded as "life-writing," or critical attention may focus on male co-authors or subjects.[68] Critics who do consider women travel writers sometimes treat them as exceptional, favouring narratives that show women as feminists without examining how their writings also uphold institutions such as colonialism, slavery, and class divisions. Sara Mills' seminal text, *Discourses of Difference: An Analysis of Women's Travel Writing and Colonialism*, does much to dispel these assumptions, but modern scholars must still be wary of criticism that situates women authors as somehow outside of the pressures, expectations, and benefits of colonial, racial, class-based, or patriarchal discourse.[69] Additionally, assumptions that

---

[65] O'Loughlin observes that "Literary historiography has tended to privilege the novel over other formal developments and has only relatively recently begun to carefully consider female-authored texts in any genre" (7).
[66] In addition to Batten and Fussell (1980), Foster and Mills also criticize Buzard's *The Beaten Track* (1993) and Youngs's *Travellers in Africa: British Travelogues 1850–1900* for their lack of reference to women travel writers, and note that even Mary Louise Pratt "treat[s] women's writing as a necessarily different and implicitly subordinate sub-genre" (5).
[67] Bohls and Duncan note of Fussell's comment that his "attitude emerges from a passing remark in his introduction to what he calls 'the heyday of travel,' the later nineteenth and early twentieth centuries" (xvi). Fussell has been critiqued in much of the contemporary scholarship for including no female authors in his study of inter-war travellers (*Abroad*).
[68] John Dussinger, for example, dismisses Hester Thrale Piozzi's travel writing, contextualizing her travel writing with the "presence" and travel texts of her estranged friend Samuel Johnson, with whom, Dussinger claims, she "seems to be carrying on a conversation…looking over her shoulder" (47). Countess Evelyn Martinengo Cesaresco's recommendation of *Observations* over all others as coming "nearest to an appreciation and understanding of the life and manners of the countries through which she passed" is brushed off by Dussinger as merely Piozzi's success at not "offend[ing] those Italian individuals who had been warm hosts" (52).
[69] For an overview of the practices of past criticism, see Mills (27–63).

women's travel writing "is essentially different from men's have been pervasive in criticism," with the "most prevalent" being "that women never obtain an objective perspective, are always interested in the domestic, and, therefore, appropriately feminine realms of the places they visit, and always identify with the objectified 'Other,' the native of the foreign land" (Kinsley 7). Thus, throughout this dissertation, I pay particular attention to how travel writing by women conforms to, reinforces, and subverts our expectations about this important subset of the genre.

Computational Approaches to Genre

Despite the acknowledged popularity of travel writing and the modern broadening of its definitions, studies of the genre still tend to focus on a series of close readings, often of significant authors. Attention to titles outside of those by famous men (and women) of letters has expanded our perspective somewhat, as has attention to reviews, personal journals or manuscripts, and paratexts, but no large-scale textual analysis of eighteenth-century travel writing has yet been completed.[70] Yet if, as has been established, travel writing is one of the most frequently published genres of the eighteenth century, then studying the entire genre through close reading is an impossible project. Thus, for this dissertation, I turn to digital humanities

[70] In a 2023 overview of digital humanities approaches to travel writing, Kevin James and Gavin Hughes do not list any projects doing textual or modelling analysis, although they note the potential of distant reading (6). The most common digital approaches to travel writing are textual encoding and geospatial visualizations. For example, see *Mapping the Lakes: A Literary GIS* and *Geospatial Innovation: A Deep Map of the Lake District*, directed by Ian Gregory, and *The Grand Tour Project*, *A World Made by Travel*, and accompanying *Grand Tour Explorer*, directed by Giovanna Ceserani. Kirstin Belgum et al. in "Mapping Travel Writing: A Digital Humanities Project to Visualise Change in Nineteenth-Century Published Travel Texts" digitize and encode the *Global Odyssey* bibliography to make entries organized and browsable by location, rather than author name. The closest digital project to this dissertation is Ryan Heuser's dissertation on abstract language in the eighteenth century, *Abstraction: A Literary History*. See also his visual summary of the chapter on fiction ("Abstraction").

approaches, especially those related to textual analysis and predictive modelling.[71]

Quantitative textual analysis, such as concordances, have a long history that predates computers. In the 1960s, as Susan Hockey notes in her "History of Humanities Computing," scholars' interest in concordances led to computational approaches such as authorship-attribution studies and stylometry, and eventually branching to other categories.[72] One early computational project is Franco Moretti's work charting the genre of novels published from 1740–1900 in Britain.[73] He identifies 44 genres and adopts the term of "generations" for the 25–30 year spans in which these genres appear and disappear (18–22). More recent research, especially by Underwood, has not found these generational patterns, however. Instead, Underwood's work, which I apply in chapter 4 of this dissertation, draws attention to cohesion and "fuzziness" on longer time scales.[74] Underwood's many works addressing genre indicate that the "real value of quantitative methods could be that they allow scholars to coordinate textual and social approaches to genre" ("Life Cycles").[75] In particular, the data that my project works with is bibliographic metadata (highly structured) and textual data (unstructured and with OCR errors).

---

[71] For recent monograph overviews of computational approaches to eighteenth-century studies, see *Digitizing Enlightenment: Digital Humanities and The Transformation of Eighteenth-Century Studies*, as well as *Data Visualization in Enlightenment Literature and Culture*, especially the work of Mikko Tolonen et al., John Regan, and Billy Hall. See also the work of the Helsinki Computational History Group (including Tolonen) on the English Short Title Catalogue and ECCO. See also chapter 4 and related appendices.

[72] For more on the history of quantitative and stylistic humanities computing, see Shawn Graham et al. in *Exploring Big Historical Data* (1–34) and Matthew G. Kirschenbaum's "What Is Digital Humanities and What's It Doing in English Departments?" in *Debates in the Digital Humanities* (2012).

[73] Early in *Graphs, Maps, Trees*, Moretti points to the book historians and bibliographers whose work he builds on, noting that "I mention these names right away because quantitative work is truly *cooperation*: not only in the pragmatic sense that it takes forever to gather the data, but because such data are ideally independent from any individual researcher, and can thus be shared by others, and combined in more than one way" (5). See also Lauren F. Klein on "Distant Reading After Moretti" regarding the field's use of the term "distant reading" (coined by Moretti) following allegations against Moretti of sexual assault in 2017.

[74] See also the *Journal of Cultural Analytics* issue on genre (2017) and projects by the Stanford Literary Lab.

[75] Much of this work is coauthored. Of particular importance to this dissertation are "Understanding Genre in a Collection of a Million Volumes" (2014); "The Life Cycle of Genres" (2017); "NovelTM Datasets for English-Language Fiction, 1700–2009" (2020); "Machine Learning and Human Perspective" (2020); and *Distant Horizons: Digital Evidence and Literary Change* (2019), especially in chapter 4.

Although giving a computer "bags-of-words" and asking for a pattern may seem counter-intuitive to current humanist methodologies, the "high-dimensional space defined by thousands of [words] gives us room to trace complex literary boundaries that don't line up with any single term," instead "register[ing] genre, topic, tone, and even . . . the social context of writing" ("High-Dimensional" 2). At the same time, these methods can contribute to close reading and other modes of analysis such as historiography.[76]

Dissertation Map

Chapter 1 describes the foundations of the Travel Writing Database (TWDB), a dataset that collates travel writing titles listed in fourteen bibliographies, anthologies, archives, and critical works.[77] Drawing on book history and bibliography studies, the chapter focuses on the selection criteria of these sources and how their definitions conceptualize travel writing. This iteration of the TWDB, called TWDB-origbib ("original bibliography"), lists 4,772 short titles, along with their authors, year of publication, and bibliographical sources. An analysis of the dataset reveals the impacts of these sources, especially on the distribution of publication dates and the cross-references between bibliographies. This chapter demonstrates that choice of bibliography significantly impacts the perspective of travel writing, especially because the vast majority of titles are listed in only one bibliographical source.

Chapter 2 documents the theoretical and practical augmentation of the TWDB-origbib to include richer metadata, including connections to machine-readable text files. First, the chapter

---

[76] Tools such as Stéfan Sinclair and Geoffrey Rockwell's Voyant Tools or Jonathan Armoza's Topic Words in Context encourage users to combine distant and close reading. See also Hoyt Long and Richard Jean So in "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning" and Sinclair and Rockwell's monograph *Hermeneutica*.

[77] As Ceserani notes in her project history for the *Grand Tour Explorer*, "Telling the story of the project itself has become a genre of its own in the digital humanities."

establishes the institutional histories and biases of its main data sources: Eighteenth Century Collections Online, Nineteenth Century Collections Online, HathiTrust, and Google Books. Then, the chapter turns to discussing the impact, bias, and application of Library of Congress Subject Headings (LCSH). Drawing on studies of LCSH and information retrieval, I suggest that LCSH are critical to study of literature of previous centuries, especially when confronted with periodization in both scholarship and archives. Finally, the chapter describes the TWDB's critical intervention of encoding women's voices in the dataset, with particular attention to how each title claims and displays women's roles.

Chapter 3 analyzes the TWDB's 4,259 titles and 5,506 volumes to provide suggestions about patterns of travel writing and provide a foundation for future use of the TWDB and its texts. Approximately 60% of the titles from the TWDB-origbib's bibliographical sources have matches in the TWDB, a total of 2,830 titles. LCSH also add 1,429 travel writing titles not listed in any bibliographical source. These LCSH reveal patterns of "aboutness" in travel writing in both the overall corpus and several key subsets and suggest the affordances and constraints of using such data in this dissertation's chapter 4, as well as future applications. The chapter also considers the 179 titles of the gendered subset of women's voices, considering the significant impact of broadening definitions beyond "real" travel.

Chapter 4 applies perspectival modelling to the TWDB, testing both the cohesiveness of travel writing as a genre and the viability of applying Underwood's methodology to different corpora.[78] The resulting models of travel writing can often predict travel writing titles at rates above 90%, indicating that there are features that are consistent within travel writing compared

[78] See chapter 4 and appendix 4.1 for further documentation of the methodology. Chapter 4 draws significantly from Underwood's method, which relies on supervised predictive modelling.

to random samples of eighteenth-century texts. The models also reveal important divisions and connections among subgenres, locations, and authorship. Chronologically across this study's 130 years of analysis, however, the analysis shows that travel writing remains relatively stable, recognizable to the decades before and after, disputing received literary histories of rapid change within the genre.

Altogether, like travel writing itself, *Coded Voyages* and the accompanying TWDB dataset document multiple, interdisciplinary, steps, arriving at new perspectives on this overlooked genre and gesturing to avenues for future studies of eighteenth-century literature and culture.

Chapter 1: Creating a Corpus of Travel Writing

> The use of travelling is to regulate imagination by reality, and instead of thinking how things may be, to see them as they are.
>
> —Samuel Johnson to Hester Thrale Piozzi (September 15–21, 1773; 2: 359)

> In our experience, the tasks of exploratory data mining and data cleaning constitute 80% of the effort that determines 80% of the value of the ultimate data mining results.
>
> — Tamraparni Dasu and Theodore Johnson, *Exploratory Data Mining and Data Cleaning* (ix)

Deciding on a corpus—regardless of media or methodology—is one of the most critical parts of any scholarly project. The process of gathering (or scraping, harvesting, munging) data, or, to use Johanna Drucker's concept, "capta" or taking and constructing data, informs and is informed by the data model, the amounts and kinds of necessary resources (ranging from time through expertise to finances), and the researchers' possible questions and answers. This process involves both practical and ethical concerns: data are not inherently objective or free of bias but are created and influenced by past and current power structures. Katherine Bode, for example, criticizes Franco Moretti's and Matthew L. Jockers' tendency to present literary data as objective facts, free of interpretation, where visualizations are "a set of two-dimensional signs…that can be grasped at a single glance" (Moretti 2013: 211, qtd in Bode "Equivalence" 20–1). Jockers describes interpretation as "flawed…anecdotal and speculative" in comparison to "comprehensive and definitive" "big data" (Jockers 2013 6; 31; qtd. in Bode *World* 22). Instead, Bode argues,

> The meaning derived from a literary-historical dataset—like the interpretation of a literary work—is shaped, profoundly, by the methodological and critical frameworks through which it is approached, and by the selections and amplifications those frameworks produce. Two scholars can read the same dataset—like the same literary work—and derive different meanings. While an independent observer may be more or less convinced by the different arguments, deciding between them depends upon access to the object on which they are based. (*World* 25)

From this perspective, the work remains deeply interpretative even at the level of dataset construction.

Bode, along with other scholars focusing on frameworks such as feminism, postcolonialism, and Marxism, thus calls for a renewed attention to methodology long before the programming begins. Although the application of methods borrowed from other disciplines such as statistics and genomics is an opportunity for digital humanities (or cultural analytics, or computational literary studies, or any of a variety of sub/fields using algorithms to analyze data) to ask new questions, this can lead to the sublimation of (equally) important parts of the process, especially that of data modelling and corpus creation. According to Bode, methodology should be foregrounded, rather than "relegated to footnotes or 'methodological caveats' (Underwood and Sellers 2015) as if they qualified rather than constituted the basis of the arguments offered" ("Equivalence" 98). Creating a data model, collecting data, processing or "cleaning" the data, and, finally, analyzing the results requires decision-making and "tweaking" in an iterative process that is an opportunity for crucial digital humanities work. As Laura Mandell argues, "acknowledging that numbers have to be *read*—that is, resisting the view that statistical results are meaningful in any self-evident way—takes place by close reading computational results and

often manipulating parameters, adjusting algorithms, and rerunning the data" ("Gender" 5). As Bode points out in her extended engagement with Moretti's and Jockers' work,

> In the absence of data publication, distant reading and macroanalysis are analogous to a scholar finding a set of documents in an archive or archives, transcribing them, analyzing those transcriptions, publishing the findings, and asserting that they demonstrate a definitive new perspective on the literary field, without enabling anyone to read the transcriptions (*World* 25).

Doing the work but not sharing the data or explaining each step not only restricts the accessibility of the work to be used in the future, but also creates challenges in evaluating the final results and interpretations of the data.

Another facet of "showing your work" in digital projects is to foreground the ethical nature of the data and labour. While many words related to creating data (gathering, scraping, harvesting, munging, mining, extracting) imply that information already exists and merely needs to be collected and analyzed, both past and current historical and cultural paradigms influence the creation and use of data.[1] The structure of this dissertation follows the advice of scholars such as Bode, Mandell, Klein, and D'Ignazio in foregrounding my data, methodology, and labour. This concern is not limited to digital humanities projects, either: Kate Ozment's recent call for a "feminist bibliography" describes her "search for women's bibliographic labor," noting that "To label [women working outside canonical bibliography] "bibliographers" would erase a gendered history of bibliographic labor" (167).[2] By presenting my own labour as an integral part

---

[1] For numerous examples both in literary studies and other fields, see Catherine D'Ignazio and Lauren F. Klein's *Data Feminism*.

[2] Ozment thanks Christine Nelson at the Morgan Library for the phrase "women's bibliographic labor."

of this project, rather than an afterthought, I situate myself and my work within the larger ecosystem of graduate student labour, eighteenth-century studies, and digital humanities.[3]

Overview of My Data Creation and Preparation

In the following pages, I detail the construction of the Travel Writing Database (TWDB): a dataset of travel writing published in English in Great Britain between 1700 and 1830. By documenting the TWDB's creation and revision, I take on Bode's challenge of creating a "scholarly edition of a literary system," which includes analysis of "the selections, assumptions, and uncertainties about evidence" that I rely on when making arguments about travel writing in the long eighteenth century ("Why" 99). In broad terms, my methodology follows this itinerary:

1. After settling on my research goal of exploring travel writing as a genre, I constructed a **bibliography** of primary source materials (including only title, author, and publication year). I developed this list, which I called TWDB-origbib, from **sources**: bibliographies, anthologies, archives, and critical works that focus on eighteenth-century travel writing. The traditional form of a bibliography, with its fields for author, title, and publication date, informed this initial construction of my data model, as did the respective bibliographical sources' definitions of travel writing.
2. To enable computational analysis of these titles, I searched for digitized editions by examining three digital repositories: HathiTrust Digital Library and Eighteenth Century Collections Online, using code that "matched" digital texts with entries in TWDB-

[3] My former co-supervisor, the late Stéfan Sinclair, aided in many of the computational approaches to the work, especially in creating the code for matching bibliographical metadata to files on ECCO and HathiTrust. For a summary of digital humanities scholarship and Donna Haraway's "Situated Knowledges," see Shawna Ross's "Toward a feminist modernist digital humanities."

origbib,[4] and then a manual integration of Nineteenth Century Collections Online, supplemented the dataset. This collection of metadata became the working version of the final TWDB.

3. To combine and contrast bibliographers' definitions of travel with those of cataloguers and large-scale datasets, I gathered titles with Library of Congress Subject Headings (LCSH) related to travel but missed or ignored by my original sources.[5]
4. I focused on a particular gap in my data: titles that were authored, translated, or edited by women. I did more customized searches in repositories for these titles and added a new repository, Google Books.

The TWDB therefore has two versions, which can answer different questions and supplement each others' answers about travel writing in the long eighteenth century. The first iteration of the TWDB, the "original bibliography" or TWDB-origbib, focuses on metadata from bibliographical sources. The final iteration of the TWDB, which chapters 3 and 4 rely on to do analysis of metadata and predictive modelling of textual data, lists titles and their full metadata, as well as linking each volume of a title to a digital file. These datasets do not align completely; instead, users can compare gaps to examine bias and other patterns, as I do in chapter 3.

Building a Corpus, Step One: A Bibliography of Titles

As with many genres of literature (and especially those of the eighteenth century), there is no ready-made corpus of digitized eighteenth-century travel writing for the aspiring text miner. Instead, with my questions in mind, I had to construct a list of texts (or a **bibliography**) and then

[4] This code matched metadata based on title, author, and publication date. I wrote it in collaboration with Stéfan Sinclair.

[5] This process involved reviewing titles missed in my original targeted search (that is, they were in TWDB-origbib, but not matched) or in the working version of the TWDB (as a duplicate, reprint, or some other form).

collect their accompanying files. My central questions of "What are the genre markers of British travel writing between 1700 and 1830?" and "How did those genre markers change in that time?" are specific enough to establish initial limitations: a publication year between 1700 to 1830, published in Great Britain, and about travel.

This last constraint, "about travel," is the most flexible, and therefore the most challenging. Because of the many existing and conflicting definitions of travel writing, I did not want to create my own. This abstention is not just because of the theoretical and practical challenges. Instead, these many perspectives of travel writing offer an opportunity to compare received literary histories and conceptions of travel writing. Like other (digital) humanists before me, I wanted to start with what Ted Underwood calls the "social evidence," or how "historically grounded interpretive communities" define the genre (*DH* xi). Either explicitly (as in the geographic or subject categories in Edward Godfrey Cox's comprehensive *Reference)* or implicitly (as in the works that Elizabeth A. Bohls and Ian Duncan include in their teaching anthology), the different sources let us explore aspects of travel writing ranging from canonicity to location. Each bibliographical source provides a different conception of travel writing, not just in definitions and theory, but in practice. These different perspectives are "a way to establish comparative relationships between different parts of the historical record," and to explore how our understandings of travel writing have been shaped by the practice and politics of citation (Underwood xii).[6]

---

[6] For examples of criticism of the canon related to the long eighteenth century, see Lora Edmister Geriguis, "Transplanting the Duchess: Margaret Cavendish and the 'Chronic Dilemmas' of Literary Anthology Construction"; Michelle Levy and Mark Perry, "Distantly Reading the Romantic Canon: Quantifying Gender in Current Anthologies"; and Mandell, "Canons Die Hard: A Review of the New Romantic Anthologies." This issue includes drama as well; Willow White focused on the neglect of eighteenth-century women playwrights in her 2019 presentation "Tracing a Literary Tradition of Women in Comedy from Aphra Behn to Tina Fey" (ACCUTE conference).

Overview of Consulted Sources: Bibliographies, Anthologies, Archive Catalogues, and Critical Works

To address questions about travel writing as a changing genre, the TWDB includes fourteen bibliographies, anthologies, archival catalogues, and critical works (my "sources"), resulting in a wide range of definitions and purposes. The construction of the TWDB began with the largest bibliographical source: Edward Godfrey Cox's three-volume *Reference Guide to the Literature of Travel*. While invaluable for sheer quantity, Cox's bibliography ignores travels after 1800, meaning that examination of movements such as Romanticism or Britain's expanding empire requires sources that favour forms of organization other than the century. My sources also allow examination of different definitions or subsets of travel writing. Some, such as Gove's imaginary voyages, focus on a genre of travel writing, while others focus on a particular publisher (Keighren, Withers, and Bell on Murray's publishing house), location (McVeagh on Ireland), archive (BrynMawr/*NCCO*), theme (trade and exploration in *TEE*), or creator gender (Robinson's *Wayward Women* and Colbert's British Travel Writing–Women's Excerpt). I include a brief description of each source in appendix 1.2, as well as observations on how their goals, gaps, and constraints shape my findings about genre throughout the long eighteenth century.

In addition to definitions and subgenres of travel writing, many sources rely on different windows of time, though many share a general interest in the end of the eighteenth century. The years from 1786 to 1800 have the most coverage; all of my sources, except for Colbert's European Tour excerpt, begin their catalogues by 1786, and none end before 1800. These divisions indicate the prevailing scholarly interest in the end of the century, whether in ideological movements such as Romanticism or the more practical impacts of faster trains and

steamships.[7] As we will see in later chapters, this overlap affects our distribution of titles' publication dates and cross-references.

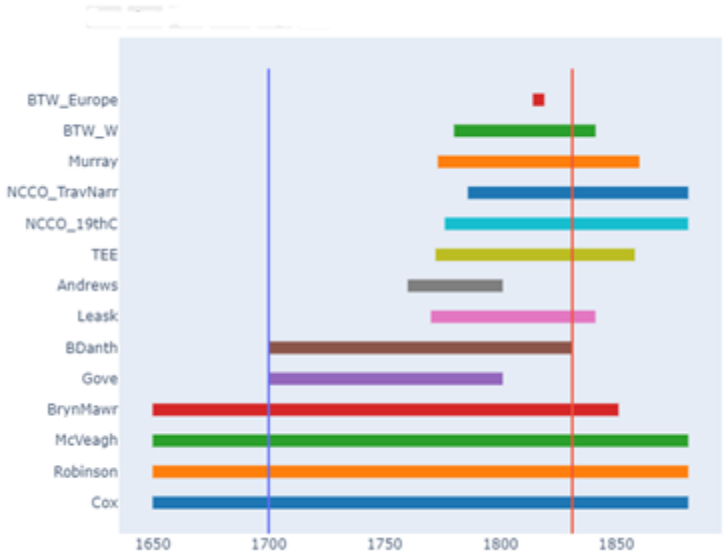
Table 1.1: Date Range of TWDB Sources

Source Name[8]	Source Published	Source Begins	Source Ends
Cox's *Reference* (cox)	1935, 1938, 1949	Earliest[9]	1800
McVeagh's Irish Travels (irishmcveagh)	1996	Earliest	1996
British Travel Writing–Women's Excerpt (btw_w)	2016-ongoing[10]	1780	1840
British Travel Writing–European Excerpt (btw_euro)	2004	1814	1818
Gove's Imaginary Travels (gove)	1941	1700	1800
Robinson (robinson_w)	1991	Earliest	1900
Bohls and Duncan anthology (bdanth)	2015	1700	1830
Travels, Explorations and Empires (tee)	2001	1772	1857
Andrews' picturesque (andrews)	1989	1760	1800
Murray Publishing House (murray)	2015	1773	1859
Leask (leask)	2002	1770	1840
Bryn Mawr's 19th Century Travel Literature (brynmawr)	Unknown-2018 (last update)	1486	1850

[7] George Dekker argues that "Besides profoundly affecting the speed, ease, safety, volume, and demographics of tourist travel, rail and steamship transportation so transformed commerce, industry, warfare, and the very face of the island that a Victorian tourist was in many respects a different person from his or her forebears" (22).
[8] For a description of TWDB metadata codes, see appendix 1.1.
[9] "Earliest" means that the source records all documents from earliest possible publication, rather than a set start date.
[10] The project formally concluded in May 2018, though work on the database continues. See "9 May 2018: Summary Report for Women Travel Writers, 1780–1840."

Source Name[8]	Source Published	Source Begins	Source Ends
NCCO–19th Century Travel Literature[11] (ncco_c19trav)	2014[12]	1776	1926
NCCO–Travel Narratives and Personal Reminiscences[13] (ncco_travnarr)	2014	1786	1921

Figure 1.1: Date Range of Consulted Sources



Even combined, these sources do not present a complete list of all travel writing titles in the long eighteenth century. Scholars have been discussing these differences—between what is written, published, preserved, catalogued, canonized—for decades. The factors affecting digital remediation are no different; although technology has allowed for digitization of some of Moretti's "slaughterhouse of literature" (2000), the resources governing the physical world—

[11] The "Collection Facts" on NCCO describe the date range for 19th Century Travel Literature as 1800–1899 (accessed 16 April 2019). Note that for search results, I record searches conducted close to when I collected my data rather than new versions of interfaces or data.
[12] The NCCO collections were released in June 2014 and upgraded to "an enhanced user interface" in December 2019. Thanks to Lonnie Weatherby for helping me find this information.
[13] The collection also includes a copy of Mandeville's *Itinerarius* published by Richard Pynson in about 1496 (STC 17246) (accessed 16 April 2019).

time, labour, cultural and financial capital—and their availability along racialized and gendered lines still influence what and how textual artifacts are being digitally created, preserved, and made accessible. These disparities mean that even more attention needs to be paid to what the authors of the Stanford Literary Lab's "Canon/Archive" describe as the differences among "the published" (all published works), "the archive" (published work that has been preserved, initially physically, and now, increasingly, digitally) and "the corpus" (the part of the archive selected for a research question or project), though their proposal that soon these three layers will "converge…into one; into that 'total history of literature', to borrow an expression of the Annales, that used to be a mirage, and may soon be reality," will remain an impossibility (2). Similarly, much documentation of the labour required to create these resources has been lost. Many archivists, cataloguers, and librarians who are critical to such work—at the stage of accession, conservation, and discovery—are often elided (by authors, corporations, or other entities) from this critical process.[14]

When Enough is Enough: Ending My Data Collection

One of the most significant challenges in this project was resisting its expansive nature. Although comprehensive bibliographical sources form the foundation of the TWDB, my dissertation is not an exhaustive bibliography like Cox's *Reference* or Colbert's BTW bibliography; rather, my needs for my data were twofold. First, I needed enough bibliographical sources to provide a varied perspective on travel writing as a genre. Second, I required enough

[14] See, for example, discussion surrounding the "Caswell Test," as coined by Bridget Wheatley, or the criticism of scholars who "discover" documents in archives and receive media coverage which often ignores the work of those who acquired, accessioned, and produced finding aids for the material.

digitized travel titles with good enough OCR to do some sort of computational analysis.[15] With this approach in mind, I stopped collecting data when I had over 4,000 titles spread across 130 years, with at least six titles per year.[16] Finally, I relied on LCSH to supplement both numbers and definitions of travel writing. This openness—to both data and methodology—remained an advantage and a challenge over the years spent on this project, but like innumerable past explorers, I set out in a direction hoping to find something interesting—or, at the very least, to survive.

Excluded Source Material

Just as important as what I include in the TWDB, however, is what I do not: these absences influence not only data sampling and results, but also the boundaries of what questions and generalizations the TWDB can support.[17] There are at least a dozen other sources not yet included in the TWDB. Many focus on European locations, echoing travel criticism's interest in the Grand Tour[18]; however, future iterations of the TWDB would benefit even more from sources focusing on travel off "the beaten track."[19] Some of these sources require wrestling with terminology, where bibliographers list "accounts" or "topographies," but not explicitly travel

---

[15] In 2015, when I began this project, my supervisors and I decided I would first collect the data and then, based on data availability and quality, decide on the appropriate computational methods. We initially hoped for at least 2,000 titles.

[16] Only three years (1808, 1809, and 1811) have fewer than ten titles.

[17] For more on generalizations in literary studies, see Andrew Piper's *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data*, especially his suggestion to explicitly "Incorporate limitations" (in both digital humanities, but also literary scholarship more generally) (59).

[18] For example, W.A.B. Coolidge's *Swiss Travel and Swiss Guide Books* (1889) or R.S. Pine-Coffin's *Bibliography of British and American Travel in Italy to 1860* (1974).

[19] For example, Thomas L. Welch's *Travel accounts and descriptions of Latin American and the Caribbean, 1800–1920: a select bibliography* (1982).

writing.[20] Others were simply unavailable or unknown during the TWDB's creation.[21] Similarly, catalogues based on libraries and archives would also further enrich the TWDB by providing more perspectives with different archival goals than the bibliographical projects listed above.[22] And another, final option is a dataset based on primary texts or contemporary reviews identifying the title as a travel text, offering a contemporaneous perspective to contrast with our modern conception of the genre.

I highlight these efforts (and in some sense, these "failures"[23]) rather than effacing them because they demonstrate several tensions with this combination of bibliographical and digital

---

[20] For example, Bernard Naylor does not define entries as travel writing or a related subgenre in *Accounts of Nineteenth-Century South America: An Annotated Checklist of Works* (1969). Similarly, see John P. Anderson's 1881 (reprinted 1976) *The Book of British Topography: A Classified Catalogue of the Topographical Works in the Library of the British Museum Relating to Great Britain and Ireland.* Some subheadings, such as "Tours," are related to travel writing, but he lists most works under counties. Anderson does not provide a clear definition of topography, though the *Oxford English Dictionary* describes "The science or practice of describing a particular place, city, town, manor, parish, or tract of land; the accurate and detailed delineation and description of any locality." Incorporating Anderson will require further deliberations about indicating these differences, or lack of, definition. Anderson's work consists of over 400 pages of two-column bibliographical entries, or 14,000 titles, which may appear multiple times, similar to Cox.

[21] Although Colbert has published two impressive excerpts of his *Database of British Travel Writing, 1780–1840* (both included in the TWDB), the bulk of the data (some 5,000 works by 3,000 creators) remains inaccessible until funding bids are finalized (Colbert in a private email, 2017). See Colbert's definition of travel writing under appendix 1.2.1. Another potential resource is *Global Odyssey: A Bibliography of Travel Literature Before 1940* (2006), by librarians Angela Courtney and Melissa (Van Vuuren) Jones. Their data shows 1364 English-language first editions published in the United Kingdom from 1700 to 1830, with 63.4% of the titles published in 1800–1830. Considering that many of the years in my original bibliography (i.e., without LCSH-sourced titles) list fewer than 20 titles, even discounting duplicates, *Global Odyssey* may significantly shift the counts for these later years, though if Courtney and Jones also relied on LCSH, the TWDB's LCSH searches may have evened out some discrepancies. Unfortunately, Courtney and Jones' work has only been cited once since its publication. I discovered this bibliography in Kirsten Belgum, Keith Handley, and Rachel Bott's 2019 article about their digital humanities project *Mapping Travel Writing*, for which they used *Global Odyssey* as a foundation. Colbert, who is among the most thorough in connecting his entries to other sources (including many that are niche or published a century earlier), does not count *Global Odyssey* among his sources. Numerous other print works, including Keighren, Withers, and Bell's John Murray bibliography (2015), do not include *Global Odyssey* in their bibliographies. It also has few search results in search engines such as Google.

[22] See, for example, Shirley H. Weber's two volumes of *Voyages and Travels in Greece, the Near East and Adjacent Regions* focused on pre-1801 and nineteenth-century travel works in the Gennadius Library in Athens, as well as the travel writing subset of the Corvey Collection (https://extra.shu.ac.uk/corvey/twfset.htm), though more information on the Corvey definitions of travel and cataloguing is needed, especially considering discrepancies between the Sheffield Hallam University catalogues and other sources like Gale's Corvey Collection module.

[23] For more discussions of failure in digital humanities, see Quinn Dombrowski's "Towards a Taxonomy of Failure" and Shawn Graham's *Failing Gloriously and Other Essays*.

humanities work. If a resource is unknown or is in a form that requires extensive resources to access (either in subscription fees, complex parsing, institutional affiliation, or just hours of labour), scholars may not—or simply cannot—include them in their work. Another similarity of bibliographical and digital projects is their propensity for sinking further into back catalogues, website archives, and general obscurity.[24]

Bibliographical Sources of the TWDB

The TWDB's bibliographical sources fall into four general categories: bibliographies, anthologies, archives, and critical works (that is, scholarly criticism in the form of books and journal articles). Below, I discuss reasons and caveats for including such categories. For more details on the editorial goals of each source, its role in the TWDB, and the number of titles it contributes, see appendix 1.2.

Bibliographies

Bibliographies, with their definitions and subsequent lists of titles, form the core of the TWDB. As D.F. Mackenzie describes it, "bibliography is the discipline that studies texts as recorded forms, and the processes of their transmission, including their production and reception" (12). Bibliography and book history, like digital humanities, are shifting fields; recent scholarship by Ozment, especially her "Rationale for Feminist Bibliography," has drawn attention to the formation of "white, male, print history" and "the thread of bibliography established by [Tanselle, Mackenzie, and McGann]" which "has, effectively, become canon" (157; 159). In particular, Ozment encourages a focus on "women's bibliographic labor" rather than strictly

[24] This is particularly important for authors who are not housed in a prestigious department, or even in a "department" at all. In addition to discussions of the "Caswell Test," see Ozment's suggestion for shifting to a "feminist bibliography" that recognizes and cites women's contributions in collecting, library studies, and so on ("Rationale for Feminist Bibliography"), as well as Mandell and Amy Earhart.

"relabeling women as bibliographers and eliding their ties to fields like librarianship" (167). While I draw primarily on enumerative bibliographies, I do not treat them as authoritative, or even comprehensive, lists of titles. Instead, I expect and investigate gaps in their knowledge. In doing so, I follow Paul Eggert in treating bibliographies as "embodied arguments about textual transmission" (177), and Bode, who suggests that "adapting bibliographical and editorial approaches to mass-digitized collections is intended to position any resulting data set as an argument about the relationship between literary phenomena and data, where the available evidence is, by definition, contested" ("Why" 100). These works range from the very broad (Cox's *A Reference Guide to the Literature of Travel*) to enumerative lists focusing on geographic areas, subgenres, and author identities. My use of these bibliographies both acknowledges and resists D.W. Krummel's wry suggestion that "bibliographies work best when they can be taken for granted" (238).

Anthologies

Anthologies have a different goal than bibliographies: editors choose a limited number of works they consider the most "important" to a subject, or "representative" of a style, at the expense of excluding others. Anthologies have a long history, arising from traditions of miscellanies in the sixteenth century and becoming similar to our current conception of anthologies in the late eighteenth century.[25] Today, text anthologies are still sold to the general public (including formats such as thematic collections or "poem-a-day" calendars), but they are most common in undergraduate classrooms.[26] They have become a critical pedagogical tool, especially for

---

[25] Mandell disagrees with Barbara Benedict's blending of miscellany and anthology, pointing to contemporary comments on anthologies to show that the new form wanted to "represent the canon of works of allegedly timeless appeal and universal importance produced throughout the history of the British literary tradition by great men (and I do mean people of the male gender)" ("Canons Die Hard").

[26] The term "anthology" has also been used for other media, ranging from radio to television and video games.

instructors teaching survey classes and introductions to different periods, concepts, and genres. Through these resources, the canon—which authors are valued, or even known—continues to influence future generations of readers, critics, and scholars[27]; as Lora Edmister Geriguis highlights, "The frequency of an author's works being taught, and the way that author is taught…are materially influenced by the configuration of the author's presence in a widely adopted anthology" (897). Similarly, Leah Price describes how anthologies "are more than a referendum. They determine not simply who gets published or what gets read, but who reads, and how…At once the voice of authority and a challenge to prevailing models of authorship" (3).

Anthologies are also not a singular entity, with shifting formats, focuses, and styles. Scholars point out that anthologies may remove important social and historical context (Benedict 7), or alternatively, they may provide too much framing for readers, prescribing a type of reading that is often focused on the figure of the author.[28] The internal organization of these documents can also influence readers' assumptions: Mandell draws attention to how in the *Norton Anthology*, the "revolutionary potential" of the new forty female authors to a new edition is effectively "effaced" by their placement.[29] Similarly, Michelle Levy and Mark Perry, in their quantitative analysis of several Romantic anthologies, demonstrate that any raw count increase in number of women in an anthology needs to be considered alongside their allocated page space, especially in comparison to the "Big Six" (Blake, Byron, Coleridge, Keats, Shelley, and Wordsworth) ("Distantly Reading"). In the TWDB's use of anthologies, I obfuscate some of

---

[27] See also John Guillory, *Cultural Capital: The Problem of Literary Canon Formation*.
[28] See Jerome McGann's description of his editorial work on *The New Oxford Book of Romantic Period Verse* in "Letters from the Editors of the New Romantic Anthologies," also qtd. in Levy and Perry (151).
[29] Mandell focuses on the sixth edition, 1993. These authors are relegated to a section entitled "Romantic Lyric Poets" rather than in the main "Romantic Period" section where Wordsworth and Coleridge—both lyric poets themselves—appear. Rather than "Romantic Lyric Poets," Mandell suggests, the section may as well be titled "minor" or "other" ("Canons Die Hard").

these important elements. For example, I do not limit my text to that reproduced in the anthology. However, the two anthologies in the TWDB—one designed for the classroom, one academic—still present evidence related to the project of canonization discussed above.

Archives and Critical Works

Finding aids and archival catalogues provide another map to travel writing. Some library holdings, such as Bryn Mawr below and the Corvey Collection, have created travel-specific catalogues; others rely on LCSH or keywords in larger library catalogues. While more work on individual archive collections may reveal patterns that vary by institution, nation, and locale, I include the lists below to highlight their unique connections and use as a tool for scholarly research and in the classroom. Similarly, critical works—books, articles, and other such texts—are primarily intended for scholarly research and citation. They are not designed to be a comprehensive list, but I include two such works as cross-references in the TWDB for their arguments about travel writing and their focus on particular locales, allowing me to test their theories against my larger corpus.[30]

Building a Corpus, Step Two: Combining Bibliographic Sources

Processing TWDB-origbib in Theory, or, How "C1ean" Should Data Be?

The concept of "cleaning" data is as fraught as deciding on data sources. This process of cleaning, wrangling, or otherwise processing data is necessary for many digital humanities goals; even handling unstructured plain text may require removing or transforming characters.[31] For projects that rely upon semi- or structured data, this step may balloon to consume a significant

---

[30] Citation analysis of a larger body of criticism would offer another way to examine the impact of canonization and travel titles. Such a discussion would require attention to the pressures of the publishing industry and academic market, such as monographs compared to journal articles and connections to tenure.

[31] Characters as in written letters, numbers, punctuation, and so on.

amount of resources. Hadley Wickham's observation, drawn from Dasu and Johnson, that "it is often said that 80% of data analysis is spent on the process of cleaning and preparing the data," emphasizes the daunting challenge for projects seeking to document all the steps of processing their data, especially when this process of "cleaning" is iterative, as it was for the TWDB (1).[32] Depending on the project, this 80% of cleaning and preparing often requires more than one person: complex projects such as eMOP, *Transcribe Bentham*, the *Grub Street Project*, and *Reading with Austen* require the knowledge and time of several people, including experienced programmers, volunteers, and project managers.[33] And projects never truly end: even if it is no longer active (whether because the goal was achieved or the resources ran out), someone still has to keep the hard drives running. These disappearances disproportionately affect women and marginalized communities; Mandell lists numerous institutional websites as well as smaller scholarly projects that are no longer accessible (for free, or at all) or have not been updated or completed ("Gendering" 513–4), as does Earhart, who notes that "Digital humanists are fond of talking about sustainability as a problem for current and future works, but it is clear that we already have sustained a good deal of loss within the broadly defined digital canon."[34] Long-term viability is required for projects requiring significant external funding, especially considering increasing guidelines from major funding bodies.[35] Ensuring that data can be used long-term—

---

[32] Searching for "80% data analysis cleaning" brings up hundreds of results quoting the 80/20 rule, including headlines from Forbes and IBM as well as articles targeting hopeful data scientists.
[33] I was a research assistant for the *Grub Street Project* and the Project Coordinator for *Reading with Austen.*
[34] Earhart credits granting models with reinforcing attention on famous authors: in her example of the National Endowment of Humanities awards from 2007 through 2010, she critiques the focus on "innovation" rather than recovery, which continues the historical attention to white male authors. The most well-known and stable digital humanities projects focused on women—the *Women Writers Project* and the *Orlando* project—receive funding not for content creation, but for their creation of tools, encoding practices, and interfaces (Wernimont "Whence Feminism"). For a history of projects focused on a single text, see Elyse Graham's "Joyce and the Graveyard of Digital Empires," *Debates in the Digital Humanities* (2019).
[35] See, for example, NEH, "Data Management Plans for NEH Office of Digital Humanities Proposals and Awards" or SSHRC, "Tri-Agency Research Data Management Policy."

one of my goals with the TWDB—can add to these required resources.

But even if a project (whether a multi-person, multi-grant team, or a single doctoral candidate) has the resources to engage in extensive data preparation, as Katie Rawson and Trevor Muñoz describe in "Against Cleaning," the expectations and standards for preparing data are still developing in the humanities. This has resulted in terms like "data cleaning" being used as shorthand, but, unlike in other disciplines where "data cleaning" has clear rules, boundaries, and margins of error, in the humanities, "there is no one single understanding of what data cleaning means." Instead of clear (if tedious) descriptions of workflows, "the specifics of data cleaning…reside in the general professional practices, materials, personal histories, and tools of the researchers," not least because many digital humanities practitioners (myself included) are self-taught to some degree (279). Rawson and Muñoz do not castigate digital humanities practitioners for these shortcomings, however:

> Rather, the collective acceptance of a connotative term, "cleaning," suggests two assumptions: first, that researchers in many domains consider the consequences of whatever is done during this little-discussed (80 percent) part of the process as sufficiently limited or bounded so as not to threaten the value of any findings; and second, relatedly, that there is little to be gained from a more precise description of those elements of the research process that currently fall under the rubric of cleaning. (279)

The many disciplines, methodologies, and goals in this "big tent" of digital humanities must therefore regularly address how data is created, handled, and funded. They have also led to debates around "What is digital humanities?", hack vs yack, the neoliberal digital humanities, and critiques and connected forums such as Nan Z. Da's "The Computational Case against

Computational Literary Studies," among others.[36] As the field continues to develop, methods may become not only more standardized, but more contentious as well.

For the TWDB, I organized the data from the fourteen bibliographical sources to create a flexible, accessible, open-source product: a simple table in CSV format, where each row is a different identifier with associated metadata.[37] The silver lining of this lengthy process is my familiarity with the TWDB data and a better understanding of the outlines of my corpus. Dasu and Johnson describe how these "seemingly unrelated topics" of data exploration and data quality management "are inseparable. The exploratory phase of any data analysis project inevitably involves sorting out data quality problems, and any data quality improvement project inevitably involves data exploration" (x). I also follow in the path of other solo graduate students: Megan Peiser, for her *The Novels Reviewed Database, 1790–1820*, notes that her transcription of reviews by hand, rather than OCR, helped her find "patterns [that] would not have been possible to find, even with text mining" (*British Women* 2.18), and François Dominic Laramée, in his apt description of "How to Extract Good Knowledge from Bad Data: An Experiment with Eighteenth Century French Texts," identifies an "intimate knowledge of the corpus" as key to extracting useful data and knowing what questions can be asked of it. This echoes Dasu and Johnson's argument that this process "sheds light on appropriate analytic strategies" (x).

---

[36] See, for example, the 2012, 2016, 2019, and 2023 editions of *Debates in the Digital Humanities* and the responses to Da in *Critical Inquiry* and *Cultural Analytics.*

[37] CSV stands for "comma-separated values," where each value is separated by a comma and each line is a different data record (in my case, a different volume). CSV is a common file format that many different programs and databases can import and export.

Processing TWDB-origbib in Practice

The concept of a spreadsheet may seem simple, but my documentation below and in the appendix shows the critical editorial decisions behind the TWDB's hundreds of rows and columns. I outline this narrative chronologically, in part to show its iterative nature, but also to demonstrate the causal relationships of my work and how many small methodological decisions slowly built the dataset. For an extended description, see appendix 1.3.

I began with the most obvious and necessary candidate: Cox's three-volume *Reference Guide to the Literature of Travel*, which is the largest bibliography and has the most generous definition of travel.[38] The PDF facsimile on HathiTrust, with its differentiated styling of publication dates, authors, titles, edition information, and cross-references, is relatively clear for human comprehension, but also required manual transcription, foreshadowing how I would record the bulk of my sources. The sheer number of Cox's entries—over 5,000 entries across hundreds of pages—quickly affected my workflow: early in the process, I ignored capitalization, and I shortened titles to the first several useful words.[39] Cox's categories also required revision of the TWDB's original data fields, ranging from extra authors to cross-references to different categories.[40] While time-consuming, recording these cross-references provided an extra layer of metadata, a shortcut to comparing different locations, even if some critics think the "regional

---

[38] I began collecting bibliographical sources in 2016.
[39] For an example of Cox's methods of recording cross-references, publication year, notes, and later editions, see volume 1 of the *Reference* (197).
[40] For example, see the fields for "author_2" and "xref" in the TWDB-origbib. The "notes" field mixes my own editorial notes and those from other sources, which would have been more useful to split into multiple columns of types and sources of annotations.

categories are often too inclusive for specialised use" (Colbert "Bibliography" 7).[41] At the end of this process, I had condensed Cox's 5,000 entries to a mere 4,072. Because Cox ends at 1800, however—and because of new approaches to travel writing, both in criticism and cataloging—over the next three years, I continued adding other sources.[42] Some sources, especially from digital resources with an established data model, such as the BTW excerpts, the Murray bibliography, and the Bryn Mawr Guide, were easy to process.[43] The majority of my sources, however, required a manual transcription (from a print resource, including McVeagh, Gove, Robinson, *TEE*, Andrews, Leask) or copy and pasting individual fields from a facsimile or digital resource (Bohls and Duncan, the NCCO collections).[44]

In some ways, however, integrating these sources into one master bibliography—or, as with Cox and his cross-references, even internal integration—proved to be the more significant challenge. Imagine cross-referencing a hundred titles to see if they are already in a spreadsheet, then repeating the step multiple times, when you—or your sources—may record titles or authors differently (a situation familiar to many, not just digital humanists). In the end, I relied on a mixture of manual review and digital methods.[45] Because I worked iteratively (adding a source

---

[41] If a title was cross-referenced, I marked it in an "xref" column, which I later combined into a single row. For many entries, this was a simple process, as I could select all the "xref" entries and sort by author or title. However, because of differences in my transcriptions or in Cox's styling of author names, or because of missing cross-references, I also used Microsoft's Fuzzy Search application. I did not use Cox's location tags in the final dataset, relying on LCSH instead.

[42] The most recent addition was Robinson's *Wayward Women* in 2019.

[43] Conventional bibliographical formats meant that dividing the text at markers such as periods created consistent, structured data that corresponded to the TWDB fields.

[44] Commenting on bibliographies of novels, Emily Friedman reflects on how "These works of enormous intellectual labor are preserved and stabilized by their instantiation in print, but to use them as something other than a reference requires laborious transformation, which due to copyright regulations cannot be then shared publicly. I often wonder how many times researchers (or their student assistants) have privately done this kind of transformative work that allows for dynamic analysis and visualization of these bibliographies, and how many sub rosa exchanges of that information occur" (356). I wonder this, as well.

[45] See appendix 1.3.2 for more details.

in steps, rather than just combining everything and then deduplicating) and because so much of the work, from transcription to deduplication, was manual, I magnified my potential for human error. However, this prolonged (and, because of other pressures of graduate school, necessary) approach did give me a better sense of how each source combined with the others. At the end of this process, I had a spreadsheet of 31 columns and over 4,000 rows.

Building a Corpus, Step Three: Data Analysis of My Bibliographical Sources

I could have ended this chapter with the following summary—"Following the iterative process of cleaning my data, my TWDB-origbib bibliography consisted of 4,772 short titles, along with their authors, year of publication, and bibliographical source"—and then promptly moved on to the next step of creating my more ostensibly complete and final data set. Doing so, however, would have ignored two important elements. First, most of my future findings rely on the data in this TWDB-origbib. Understanding the shape of the data—which sources contribute the most titles, the diachronic influence of publication numbers, and so on—is necessary to understand the impacts and limitations of my work in chapter 4 and research by others that may rely on this data. Secondly, even without matching these titles to digital files, the TWDB-origbib offers an opportunity to explore publication rates and cross-references of travel writing. As other bibliographic projects and meta-analyses have shown, exploring our received literary histories may not only challenge our expectations, but also reveal biases of past—and current—scholarship.[46] In a field such as travel writing that is understudied in favour of (supposedly) more literary productions, these gaps take on increased importance.

---

[46] For example, see the work of Mandell, Levy, and Perry, as well as Peiser. For specific discussions of travel writing bibliography, see Colbert's published excerpts of his *British Travel Writing Database* and the *Women's Print History Project*'s genre of "Travel/Tourism/Topography."

In the following sections, I discuss some of the most important elements revealed by the TWDB's origbib iteration. This combination of 14 bibliographies consists of 4,772 short titles, along with their authors and year of publication.[47] These sources shape the distribution of publication dates and the most common cross-referenced titles. My final finding, that most titles are listed only in one bibliography, suggests that choice of bibliography may significantly impact one's evidence and all subsequent analysis of travel writing, digital or otherwise.

Publication Counts: Year, Decade, and Source

The TWDB-origbib suggests that at least 4,772 travel writing titles were published between 1700 and 1830. Although this number averages to 36 titles a year, as figure 1.3 shows, the distribution is not uniform. Instead, the number of titles slowly climbs as 1800 approaches, in keeping with patterns suggested by histories of printing and ECCO.[48] The shocking plunge immediately following the turn of the century, however, points us toward the impact of bibliographical sources and their gaps. In particular, the impact of Cox, with its end point of 1800 and the lack of any nineteenth century counterpart, is reflected in both figures 1.3 and 1.4 below.

---

[47] Multivolume works combined into one entry. Note that in the charts that follow, I delete titles that are dated 17-- (a total of 5 titles). The "pubDate_mod" field in TWDB-origbib is drawn from the first four characters of the "pubDate" column, meaning that works with a date range, such as "buck's antiquities or venerable remains" by Samuel Buck, published 1727–1740 (according to Cox), is simplified to 1727. For many multi-volume titles—especially anthologies, such as John Senex's *Modern Geography*, which can be published over several years or even decades—this decision creates a necessary distortion of the publication numbers for just over 200 titles whose publication ranged over more than one year.

[48] See appendix 1.4.1. Despite ECCO's documented faults as a 1:1 proxy for eighteenth-century publishing, the online database is still the most comprehensive source of machine-readable texts for the eighteenth century.

Figure 1.2: TWDB-origbib Publication Numbers, by Year

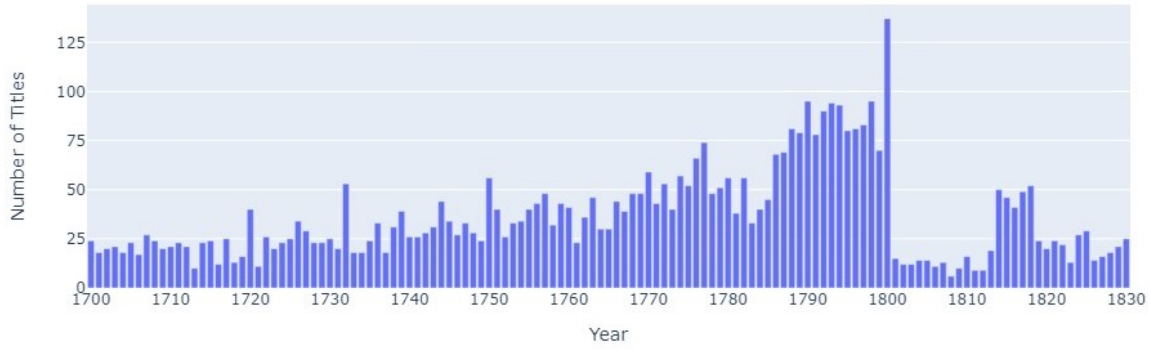
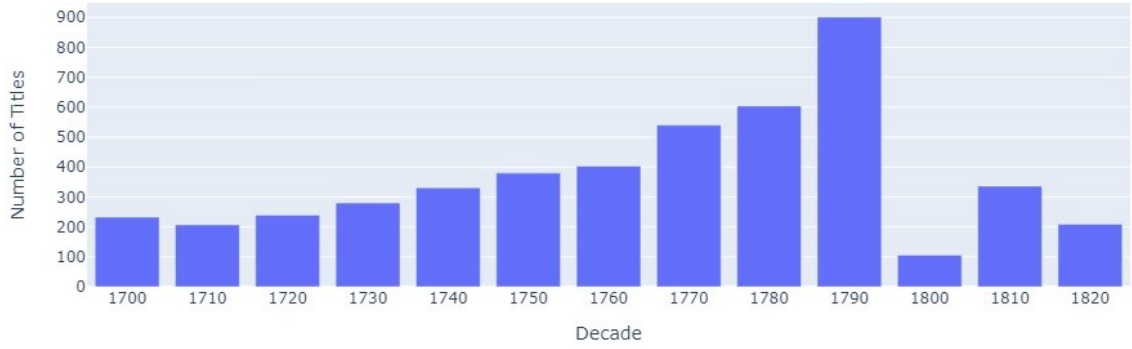


Figure 1.3: TWDB-origbib Publication Numbers, by Decade



The 14 sources in the TWDB range from comprehensive, enumerative bibliographies to specialized anthologies.[49] The number of titles from each source varies significantly based not just on time frame, but on the project's goals as well.

[49] For the full description of each source, see appendix 1.2.

Table 1.2: Number of Titles From Each Bibliographic Source[50]

Source	Number of Titles
cox	4069
brynmawr	200
btw_europe	169
btw_w	134
murray	118
ncco_c19trav	111
irishmcveagh	94
gove	77
tee	66
andrews	65
leask	60
bdanth	55
robinson_w	40
ncco_travelnarr	10

Cox's scope is certainly the largest, but the 3,800-title difference between his list and the next closest source, the Bryn Mawr catalogue, is still startling; in fact, relying on Cox alone would still catch over 85% of the titles in the corpus, though these titles would, admittedly, be published mostly before 1800. A narrow time frame, however, does not necessarily mean low numbers: for example, Colbert's British Travel Writing–Europe excerpt (btw_europe), which focuses on travel on the continent between 1814–1818, is still the third-highest contributor. Thus, when choosing sources to inform a project, chronological scope alone cannot predict the number of titles. In fact, the thorough nature of Colbert's btw_europe draws attention to the more mixed coverage of my own bibliography; if his "reasonably complete listing" is at all indicative of titles published per year (Colbert 34), then we might expect closer to 1,054 titles from the beginning of 1800 to the end of 1830, rather than the 651 currently in the TWDB-origbib.

[50] Titles may be listed in more than one source.

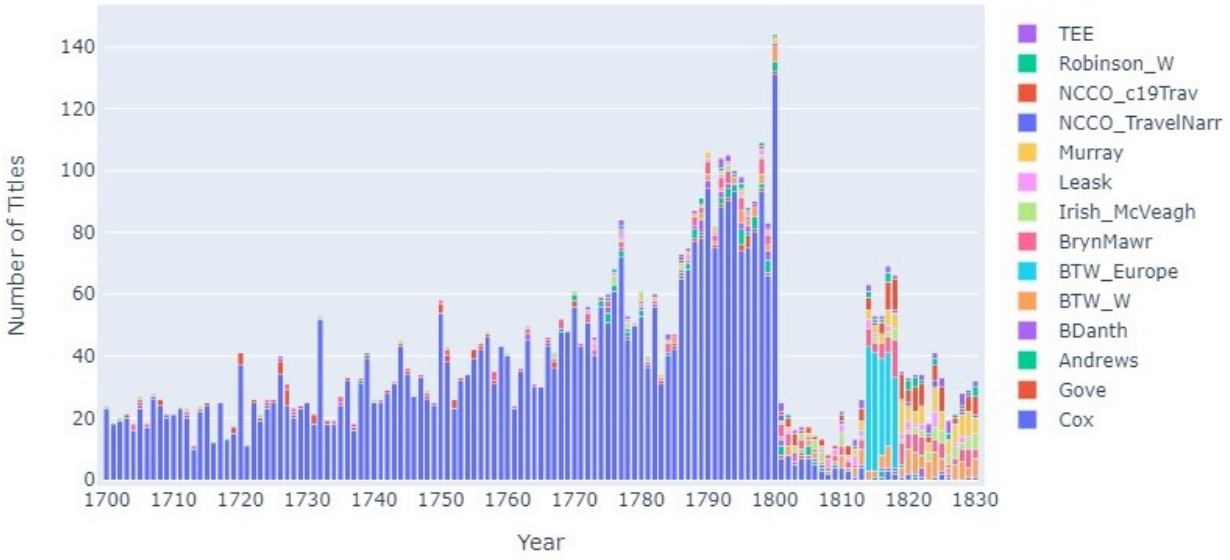
But what about the other sources in the TWDB? Even if their contributions number in the few dozen, some significant patterns offer insight into particular subgenres of travel writing, especially as they influence some of our findings in chapter 4. For example, consider Murray's potential impact: even though Keighren, Withers, and Bell focus on non-European travel titles for their bibliography, this single publishing house contributes more titles than either McVeagh's catalogue of Irish travels, or Gove's list of imaginary travels. Meanwhile, the anthologies and critical works—the Bohls and Duncan teaching anthology (bdanth), the *Travels, Explorations and Empires, 1770–1835* (tee), Leask, and Andrews—contribute fewer titles in general, as we might expect from works that do not aim at the cataloguing function of bibliographies. The source divisions and diachronic publication numbers also require contextualization. Of the 14 TWDB sources, only 5 cover the entire 130 years; instead, most focus on smaller windows of time, and especially later in the century.[51] The early eighteenth century has the most limited coverage, drawing from only a few sources.[52] In contrast, the period with the most titles is the 1790s, a decade commonly included by both sources that rely on the end of the century and those that are interested in other patterns of travel, such as the Romantic period or empire-building. Shifting the above visualization to include the annual counts thus reshapes our understanding of this influence once again. Cox is a wall of blue, before dropping off after 1800.

---

[51] See table 1.1 and figure 1.1 in "Overview of Consulted Sources: Bibliographies, Anthologies, Archive Catalogues, and Critical Works" earlier in this chapter.

[52] Sources covering the first half of the eighteenth century include the Bohls and Duncan anthology, Gove, McVeagh, Robinson, and Cox's bibliographies, and the Bryn Mawr catalogue.

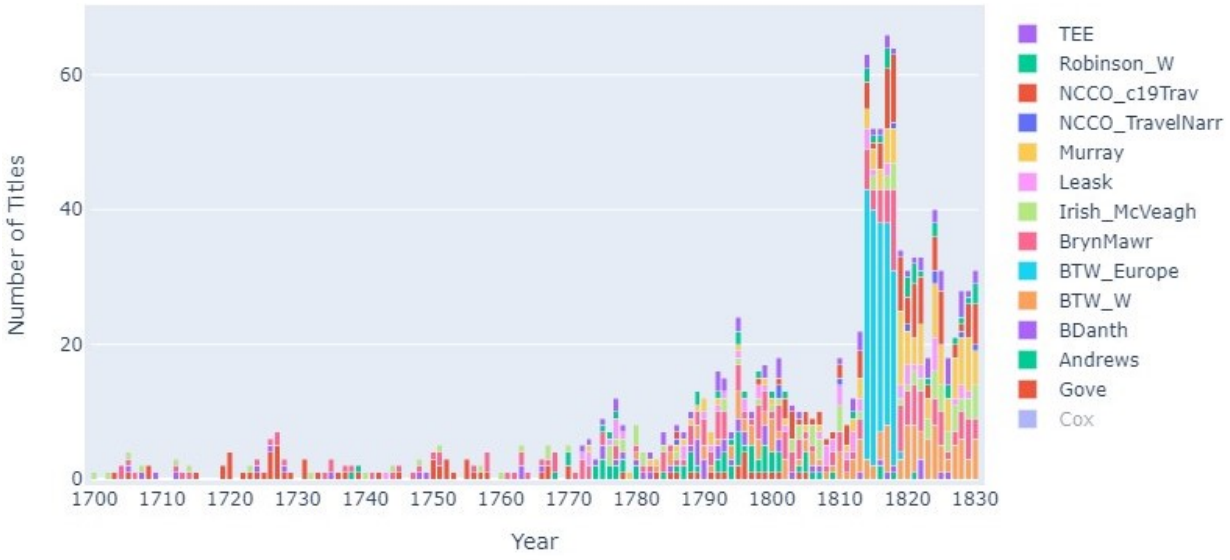
Figure 1.4: TWDB-origbib Titles by Source



If we remove Cox from the visualization, his importance becomes especially clear: many of the years prior to the 1770s have only a couple of titles, and Cox provides the only titles for 13 of the years prior to 1770.[53] In particular, figure 1.5 draws our attention to the aforementioned btw_europe, and increased presence of Murray's publishing house and women authors in the 1820s.

[53] 1701, 1710, 1711, 1716, 1717, 1718, 1721, 1730, 1746, 1754, 1759, 1765, and 1769.

Figure 1.5: TWDB-origbib Titles by Source, Without Cox



Most years have at least two sources contributing titles, while from 1780 onwards, the majority have at least five, as scholarly interest in different aspects of travel writing proliferates. Some years, however, stand out as unique: for example, 1705 is the only year prior to the 1770s to have five contributing sources. As one might expect, some sources—Cox and the Bohls and Duncan anthology—list the significant *Remarks on Several Parts of Italy* by Joseph Addison. However, three other titles and sources also contribute to this year: Gove points to Daniel Defoe's *The Consolidator; or, Memoirs of Sundry Transactions from the World in the Moon*, Bryn Mawr holds a translation of William Bosman's *Description of the Coast of Guinea* (a best-seller in his native Dutch), and McVeagh, in his Irish bibliography, lists John Dunton's *Life and Errors of John Dunton*. On the other hand, sometimes citations concentrate on a single title, such as in 1786. Of the nine sources listing titles for this year, four list William Gilpin's

famous *Observations Relative Chiefly to Picturesque Beauty*.[54] These cross-references can thus draw our attention to patterns not just in travel writing, but also in travel writing scholarship and archival practices.

Cross-References Between Sources

These observations about 1705 and 1786 may imply that cross-references are common in this dataset. In fact, most titles—4,409, or 92.5%—are listed only in one source. That leaves 360 titles with cross-references, and of those, most (260) have a single cross-reference (that is, two sources record them).

Table 1.3: Number of Cross-References in TWDB-origbib

Number of Sources Listing a Title	Number of Titles
1	4409
2	260
3	69
4	24
5	6
6	1

Once again, the predominance of Cox is clear: of these 360 cross-referenced titles, Cox contributes 226 (62.8%), demonstrating his connections with other sources. Many bibliographies' shared criteria cause the remaining connections, such as the women-created texts of BTW–Women's excerpt and Robinson's list of women travel writers, or NCCO's source material being the Bryn Mawr catalogue.[55]

---

[54] Andrews, NCCO's "19th Century Travel Literature" (ncco_c19trav), and the Bryn Mawr archival catalogue only list *Observations* for 1768. Cox lists another 59 titles.

[55] Of the 35 cross-referenced titles in robinson_w, 31 are also listed in btw_w. ncco_c19trav and brynmawr share 68 titles. Some sources, such as the btw_w, lists other bibliographical sources where they found titles, but that is the exception rather than the rule.

This may make some comparisons feel redundant; if less than 15% of the texts in the Bohls and Duncan anthology or Robinson's *Wayward Women* are unique, then why include them? Or, similarly, if a source shares no cross-references with another source, then what is the use of comparison? Where sources come together and diverge, however, are not just limitations on further questions, but also fertile ground for further speculation. I demonstrate this with perspectival modelling in chapter 4, but even with the more limited TWDB-origbib, we can ask questions important to travel writing scholars. For example, what are the "exemplary" titles in travel writing? If we consider these cross-references—created by a mixture of literary scholars, archivists, bibliographers, and organizations—what can we discern about recognition and influence?

Titles with Multiple Sources

To focus on the most common titles, let us ignore the runners-up who have only two or three citations. Instead, focusing on the 24 titles listed in at least four sources reveals the influence of the TWDB's later sources, such as the *TEE* anthology, Leask, Murray, and the BTW–Women's excerpt. Although the earliest title is John Hawkesworth's *An Account of the Voyages…in the Southern Hemisphere*, published in 1773, less than half of these titles are published before 1800, with an average publication date of 1804. The six titles with five citations also average an 1804 publication date—but even more striking, for a moment, there is a rare gender parity: Hester Lynch Piozzi's *Observations and reflections made . . . through France, Italy, and Germany* (1789), Mary Wollstonecraft's *Letters Written During a Short Residence in Sweden, Norway, and Denmark* (1796), and Anne Carter's *Letters from a Lady to Her Sister, During a Tour to Paris* (1814) are all bolstered by cross-references in women-focused bibliographies (robinson_w and btw_w). Wollstonecraft and Piozzi are also well-known eighteenth-century

figures outside of both travel writing and women's writing. In contrast to the European focus of the women authors, the titles by men explore farther afield, consisting of Mungo Park's famous *Travels in the Interior Districts of Africa* (1799), Edward Daniel Clarke's popular collection of *Travels in Various Countries of Europe, Asia, and Africa* (1810), and James Hingston Tuckey's *Narrative of an expedition to explore the river Zaire* (1818), which get a boost from anthologies (tee and bdanth) and connections between brynmawr and ncco_c19trav.[56] Few of the above titles are unknown to scholars of (long) eighteenth-century travel writing. However, just as importantly, consider those who are absent, such as Defoe, Lady Mary Wortley Montagu, Laurence Sterne, and James Boswell. With many TWDB sources beginning after 1770, these earlier authors were never a candidate for most TWDB citations, even if they are among the most cited within travel writing scholarship more broadly.

Instead, our most-cited title comes at the turn of century: Ann Radcliffe's *A Journey Made in the Summer of 1794* (1795) is recorded in the women-focused sources of robinson_w and btw_w, the anthology bdanth, the picturesque-oriented andrews, the all-consuming cox, and the brynmawr catalogue, for a total of six cross-references. But why might this title be the most listed piece of travel writing? *A Journey* is not an obscure text, but neither is it known for influence in the same way as Sterne's *Sentimental Journey* decades before. Instead, I suggest that Radcliffe's prominence is a combination of both historical circumstance and contemporary interest—a valuable case study in how data sources can influence findings. First, the publication year (1795) of *A Journey Made in the Summer of 1794* is covered by many of the TWDB sources, increasing opportunities for cross-references among these later bibliographies.[57]

---

[56] NCCO's two collections list later editions of Clarke's *Travels in Various Countries.*
[57] The *Lady's Magazine* published four extracts from *A Journey*, three in 1795 and one in 1797.

Radcliffe is also a famous author, albeit best-known for her Gothic fiction; *The Mysteries of Udolpho* (1794, the year before *A Journey*) and *The Italian* (1797) were extremely popular during her lifetime, and critics situate her as one of the most important influences on the Gothic novel.[58] Indeed, Radcliffe's popularity is a result of a confluence of social factors influencing expectations for women's participation in travel and publishing, both historically and in modern feminism's attempts to recover women's writing. As Bohls traces in her discussion of Radcliffe's *Udolpho*, "Both tourism and writing for publication took women into the public realm in potentially transgressive ways. Sensibility and taste helped legitimize a woman like Radcliffe in her pursuit of these dubious endeavors" (*Women* 103). Radcliffe's publisher for *Udolpho* and *A Journey*, the radical George Robinson, was best known not for his fiction, but for his travel narratives, translations, and periodicals, "position[ing] Radcliffe and her work outside of the insular worlds of the British novel and the fashionable milieu of the circulating library, and mak[ing] her a significant part of early conversations about human rights and even what we might today call global citizenship" (287). Thus, even without *A Journey*, Radcliffe's oeuvre still influences discussions of landscape aesthetics (note her presence in Andrews), the picturesque, and the Gothic. There is still more room for scholarship of *A Journey,* however; for example, while there are accessible scholarly editions of most of her novels, there is still no modern edition of *A Journey*—which, at least by TWDB standards, is one of the most bibliographically recognized travel texts of the long eighteenth century.[59] Even if citation does

---

[58] For an overview, see the collection *Ann Radcliffe, Romanticism and the Gothic.*

[59] For example, Oxford has published editions of *Udolpho*, *A Sicilian Romance*, *The Romance of the Forest*, and *The Italian*. Broadview Press has accepted a proposal and introduction for *A Journey* from Frances Chiu, originally forthcoming in 2021 but still unpublished in March 2024. A complete edition of Radcliffe's works from Cambridge University Press, edited by Angela Wright and Michael Gamer, is forthcoming, beginning with *Udolpho* and *The Italian* in 2024 (Wright).

not indicate influence, it does indicate recognition—and yet, *A Journey* remains, as Leah Orr points out, "often-overlooked" ("Property" abstract).[60]

Titles Listed in Only One Bibliography

But what, then, about these under-recognized titles in the dataset? Most of the titles in TWDB-origbib are listed in only one source. In general, these titles tend to be from bibliographical sources, with most listed in Cox, the BTW excerpts, McVeagh, Gove, and Murray. Predictably, Cox, our most prolific bibliographer in TWDB-origbib overall at 4,072 titles, also lists the most single-citation titles, at 3,846. However, as before, the btw_europe dataset indicates the importance of focused bibliographies for expanding historical knowledge: nearly 90% of its titles are not listed in any of my other sources. Similarly, McVeagh's focus on Ireland appears to be outside the scope or notice of our other sources, many of which were either more attentive to travel that was more domestic or farther afield than just across the Irish sea.[61]

Table 1.4: Ratio of Single Citations to Total Titles

Source	Titles in This Source Only	Total Title Count	Ratio
cox	3843	4069	0.94
btw_europe	152	169	0.90
irishmcveagh	81	94	0.86
gove	49	77	0.64
btw_w	74	134	0.55
murray	63	118	0.53
ncco_travelnarr	5	10	0.50
leask	24	60	0.40
ncco_c19trav	36	111	0.32
brynmawr	43	200	0.22
andrews	13	65	0.20

[60] Orr's "Property, Money, and Benevolence in *The Mysteries of Udolpho*" uses *Journey*'s attention to financial costs to contextualize the "economic concerns" in Radcliffe's novels (87). For a discussion of models of writing by women and these titles in particular, see chapter 4.

[61] See William H. A. Williams on Irish travel, including *Creating Irish Tourism: The First Century, 1750–1850* and "The Irish Tour, 1800–50."

Source	Titles in This Source Only	Total Title Count	Ratio
tee	13	66	0.20
bdanth	8	55	0.15

Considering how anthologies are attuned to canon formation, it is unsurprising to see the two anthologies, the Bohls and Duncan and the scholarly *Travels, Explorations and Empires, 1770–1835*, with some of the lowest rates of unique titles. Even so, a general teaching anthology such as Bohls and Duncan having nearly 15% unique titles, or 3 out of every 20, emphasizes the editors' goal of "encourag[ing] readers to reconsider and broaden their understanding of what constitutes travel" (xvii). But it may also recall Kate Durie's criticism of the anthology, that "If the label 'travel writing' is stretched too far, it becomes almost meaningless" (927), a consideration which Bohls and Duncan address directly; for example, one unique title is John Newton's *Thoughts upon the African Slave Trade* (1788), which they note is "not strictly classifiable as travel writing" (191).

Trawling through the rest of this "great unread" (Cohen 23) shows a rich variety of genres, locations, and authors, some of which appear multiple times, and for others, only once. Margaret Cohen's term takes on extra weight as the TWDB moves to its next step: for many of these texts, these few words of the title, this brief encounter of transcription, are all that I will ever read.

Conclusion

In the epigraph that opens this chapter, I quote Samuel Johnson's letter to Hester Thrale: "The use of travelling is to regulate imagination by reality, and instead of thinking how things may be, to see them as they are" (September 15–21, 1773 (2: 360)). For Johnson, after decades of reading widely yet rarely leaving London, his Scottish travels confronted him with differences between his travel books and his travel experience. Disheartened by the "toilsome drudgery, of wandering

in Skie," he warns Piozzi that "You will now expect that I should give you some account of the Isle of Skie, of which though I have been twelve days upon it, I have little to say" about the island's "barren desolation" (2: 359).[62] Susan Lamb suggests that, for Johnson as well as other eighteenth-century travellers, such a journey "was an opportunity to synthesize and confirm or challenge book-learning and the spatialized knowledge already available in printed form," as well as "a way of testing expectations and versions of reality" (22).[63] Similarly, the TWDB and its combination of bibliographic and algorithmic methodologies can "synthesize and confirm or challenge" our received literary histories and scholarly assumptions, adding nuance to our understanding of one of the most-published genres in the eighteenth century.

Instead of Johnson's "confine[ment]" in the "naked desart" of Skye, the TWDB's initial dataset encourages future exploration, both of the TWDB-origbib metadata and the textual data I describe in chapters 2 and 3. The TWDB lists twice as many titles as I originally set out to record, but those numbers dramatically drop at the turn of the century: the TWDB would be very different without the efforts of Cox, who contributes not only the most titles, but also the most cross-references.[64] Indeed, most travel writing titles are only listed in one bibliographical source. For those seeking a particular thematic reading list, such as imaginary travels, then Gove's

---

[62] Following his observation about the imagination, Johnson continues his description of Skye with "Here are mountains which I should once have climbed, but to climb steeps is now very laborious, and to descend them dangerous, and I am now content with knowing that by a scrambling up a rock, I shall only see other rocks, and a wider circuit of barren desolation. Of streams we have here a sufficient number, but they murmur not upon pebbles but upon rocks; of flowers, if Chloris herself were here, I could present her only with the bloom of Heath. Of Lawns and Thickets, he must read, that would know them, for here is little sun and no shade. On the sea I look from my window, but am not much tempted to the shore for since I came to this Island, almost every Breath of air has been a storm, and what is worse, a storm with all its severity, but without its magnificence, for the sea is here so broken into channels, that there is not a sufficient volume of water either for lofty surges, or loud roar" (2: 359–60).

[63] Andrew McKendry points to Bertrand Bronson's similar suggestion that Johnson sought "to consolidate his opinions, test his earlier conjectures, and formulate judgment by firsthand observation of the merits and demerits of a system of life which he had long idealized" (qtd. in McKendry 12). See McKendry, "The Haphazard Journey of A Mind: Experience and Reflection In Samuel Johnson's *Journey To The Western Islands Of Scotland*."

[64] Cox's bibliography is connected to half of all cross-referenced titles.

bibliography may suffice; for those looking for the most cross-referenced works of travel writing, arguably the most representative of the genre, then they may turn to works by women.[65] But I argue that understanding the boundaries of travel writing and its subgenres requires not just definitions and theory, but tests of the evidence and data. This dissertation thus begins with lists of titles, continues building networks of connections, and, as will appear in later chapters, finally tests "the selections and amplifications those frameworks produce" (Bode *World* 25). Of note is that combining and comparing these sources does not create a holistic, simplified conception of the genre. Rather, much like travel writing's own variety of topics, styles, and voices, these multiple perspectives create a more complex, rich, and sometimes conflicting map of the genre's borders. Our construction and documentation of the TWDB thus implicitly follows Leopold von Berchtold's argument in his 1789 *Essay to direct and extend the inquiries of patriotic travellers*:

> All those gentlemen who flatter themselves to have travelled usefully, will agree, that a traveller ought to know what to look for, in order to discover the object of his pursuit before he comes on the spot; otherwise it is too late: the principal work is to see how far the ideas we had formed of an object were founded on reason during anticipation" (17).

In the upcoming chapters, as I further enrich, analyze, and transform TWDB data, this initial "object" of the TWDB-origbib does not fade away, but rather gains importance as the foundation for the next step of the journey.

---

[65] Radcliffe's *Journey* is the most-cited title in the TWDB. See earlier in this chapter.

Chapter 2: Searching for Data, From Print to Digital

> JOHNSON. "[B]ooks of travels will be good in proportion to what a man has previously in his mind; his knowing what to observe; his power of contrasting one mode of life with another. As the Spanish proverb says, 'He, who would bring home the wealth of the Indies, must carry the wealth of the Indies with him.' So it is in travelling; a man must carry knowledge with him, if he would bring home knowledge."
>
> BOSWELL. "The proverb, I suppose, Sir, means, he must carry a large stock with him to trade with."
>
> JOHNSON. "Yes, Sir."
>
> —James Boswell, *Life of Johnson* (3: 301–2)

The "large stock" of the TWDB's original bibliography consists of short titles, authors, publication dates, and bibliographical references—a resource that combines decades of scholarship. The final form of the TWDB, however enhances this initial metadata with full publication data, subject headings, and links to digitized editions.[1] While this compilation may seem mechanical—and indeed, it is often repetitive, iterative, and even tedious—the editorial decisions range from handling limited corporate databases to encoding complex concepts such as gender. By documenting this next iteration of the TWDB and analyzing the results, I shed light on the differences between "the published" and "the archive" ("Canon/Archive"). Travel writing was one of the most popular genres in the eighteenth century, but as I show in this chapter, much

[1] File IDs connect to plain text and frequency files for each volume.

has been lost in the transitions from original printing, to physical archive, to digital collection and cataloguing. Even with these gaps, however, the TWDB can let us ask questions about our conceptions of travel writing, as well as how these conceptions now mediate our access to titles and files both familiar and foreign.

As this chapter demonstrates, the editorial process makes the question of "what is travel writing" both more concrete and abstract, as our definitions intersect in dozens of libraries, hundreds of subjects, and thousands of texts. First, I describe the specific histories and biases that factor into relying on eighteenth-century databases. Second, I outline how Library of Congress Subject Headings can supplement bibliographical sources, attending to this approach's limitations. Finally, I address the challenge of encoding gender beyond a simple binary, adding to ongoing efforts to reclaim and theorize the role of women in travel writing. For a more thorough description of methodology, see appendix 2.4: Description of the Data Model.[2] The documentation in both the appendix and the chapter proper function not just as a practical (and, I argue, necessary) tool for anyone seeking to use the TWDB or verify my findings, but also as "an extended historical argument in and of itself" (Bode *World of Fiction* 3).

An Overview of Digital Databases

The transition from bibliographical sources to digital files changes what titles are known and accessible to contemporary scholarship, as well as how those titles are organized, framed, and perceived. These practical and theoretical concerns are especially relevant when making

---

[2] I take Katherine Bode's argument seriously when she critiques how "'the stylistic protocols of literary criticism' mean that issues deemed methodological are often relegated to footnotes or 'methodological caveats'—as if they qualified rather than constituted the basis of the arguments offered" (54, quoting Underwood and Sellers 2015). Within this static dissertation, I focus on my "engagement with disciplinary infrastructure" (54) in the main body of the text, and describe the step-by-step approach in related appendices. In doing so, I balance the requirements of a dissertation, while linking arguments, both theoretical and practical, as much as possible.

sweeping statements about genres such as travel writing. Both bibliographers and digital book repositories rely on what the Stanford Literary Lab's "Canon/Archive" calls the "archive," or "that portion of published literature that has been preserved—in libraries and elsewhere—and that is now being increasingly digitized" (2). As Katherine Bode notes, "both digitized collections and analog bibliographies are derived from…predominantly the collections of major (usually American or British) university libraries" (*World* 24).[3] This remediation into a digital format, however, relies on the further "acts of production and reception, critical and technical, that produce digital collections" (*World* 44). [4] Ironically, as James Mussell argues, "a successful digital resource is often one that masks its workings and appears to be something that it is not" (5). Therefore, just like the TWDB's bibliographical sources, and the TWDB itself, my digital sources have histories (and futures) that dictate the content, form, and display of the collection.[5] When approaching any kind of database, archive, or collection, scholars should always question not just the content, but the background as well.[6] These interrogations can begin to resemble a primary school list of "who, what, where, when, and why," but they support diverse lines of inquiry into every aspect of a database's construction and use. During my development of the TWDB, I often reflected on the following questions, though I did not always find an answer: What was digitized (ephemeral material? fiction? English language?); by whom (academic

---

[3] Depending on the bibliographer, some may accept listing in a publisher's catalogue or advertisement as evidence of publication; others require evidence of a physical copy.

[4] In his discussion of digitized newspaper collections, Mussell similarly argues that "Every digitization project is also an editorial project and all editorial projects must define in some way what it is they edit. Editing is thus an interpretive exercise and all editions make some sort of argument about whatever it is that they publish" (4).

[5] Library science regularly addresses these issues, both in their conceptual and practical nature. See, for example, the roundtable discussion from the 2015 Charleston Library Conference, "Text and Data Mining Contracts: The Issues and Needs."

[6] Even the terms used for different sources provide a framework for understanding a collection. See Kenneth M. Price's "Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?" I expect scholars to be the primary users of the TWDB. For additional context connecting to art, the public, and corporate influence, see Catherine D'Ignazio and Lauren F. Klein's *Data Feminism* and Safiya Noble's *Algorithms of Oppression*.

librarians? Private corporations? North America, Europe, Global South?); using what technological or editorial priorities; how is it stored, accessed, searched, and displayed (paywalls or public domain? open-source data or no download option?); and what are the plans for future sustainability?[7] And even if we know these answers on one day, the digital archive may be a different iteration the next, with updated content or search algorithms. With accurate, explicit documentation on the part of both the digital source and the scholar, these changes can be a strength; without, it quickly becomes impossible to maintain clarity on what version of evidence is supporting an argument.

To support future use of the TWDB and to explore how the TWBD's different file sources—Gale's ECCO and NCCO, the HathiTrust Digital Library, and Google Books—may have impacted this dissertation's findings, this section describes the aspects of each collection's history, metadata practices, and accessibility.[8] These questions are even more important as the TWDB combines these different corpora into a single dataset, and then treats them as contiguous or interlocking—despite the TWDB including multiple literary movements and periods, as well as filetypes. These sources' practices, ranging from Optical Character Recognition (OCR) to metadata structure, have the potential to affect TWDB data and findings even more than transcribed bibliographical sources.

The First Challenge: Going to the Source

In some ways, finding eighteenth-century texts is not difficult. An online search can often bring up a digital facsimile on Google Books or a print-on-demand edition on Amazon. For scholars,

---

[7] For a similar discussion of the differences between analogue and digital projects, see Bode in *World of Fiction*, (44–6). For a discussion of projects that are no longer accessible or updated, see Laura Mandell, "Gendering Literary History."

[8] See Bonnie Mak's "Archaeology of a Digitization."

specialized digital archival products such as Gale's Eighteenth and Nineteenth Century Collections Online (ECCO and NCCO) are also longtime standard sources of primary material.[9] Trying to access and use these thousands of texts, however, is challenging. Private corporate resources such as ECCO or NCCO require expensive institutional subscriptions, and although reading images on the privately owned Google Books or the public HathiTrust is comparatively simple, accessing computer-readable, rather than human-readable, text files on a mass scale is not. Even searching metadata may be impractical without tools, technical knowledge, or authorization. And these problems go beyond the practical to inform the TWDB's central questions: ECCO, NCCO, HathiTrust, and Google Books, and their source libraries, all vary in their purposes, collection and digitization guidelines, and time frame.[10] For any project like the TWDB, combining multiple databases becomes an exercise not unlike relying on pieces from different puzzles that were never designed to fit together.

Eighteenth Century Collections Online

ECCO is the "the primary digital source for the study of printed eighteenth-century texts in the English language" and has had "a profound impact on how researchers conduct scholarship of the period" (Tolonen et al. "Anatomy" 95; Gregg *Old Books* 1).[11] It provides the majority of the TWDB data, but the archive's priorities, construction, and timeframe of 1701–1800 create limitations for its use and exemplify some of the challenges of relying on corporate databases.

---

[9] Early English Books Online, or EEBO, hosted by ProQuest, is the early modern companion to these resources.
[10] Lawrence Evalyn similarly examines the histories and "lacunae" of the English Short Title Catalogue (ESTC), Eighteenth Century Collections Online (ECCO), Eighteenth Century Collections Online Text Creation partnership (ECCO-TCP), Google Books, and HathiTrust. See Evalyn's section on "Material Histories of Eighteenth-Century Databases" in *Database Representations of English Literature, 1789–99* (41–64).
[11] For a comprehensive history of ECCO, see Stephen Gregg's *Old Books and Digital Publishing: Eighteenth-Century Collections Online*. Gregg especially analyzes how ECCO communicates the "bookishness" of its data. I'm grateful to Gregg for sharing this research prior to publication.

This commercial resource, sold by Gale, contains over 180,000 titles (200,000 volumes or 32 million pages) printed between 1701 and 1800 and published anywhere in the world. These works are available to members of subscribing institutions in several formats.[12] Despite ECCO's seemingly comprehensive nature, however, the process of ECCO's digitization, cataloguing, and interface involve many important caveats, particularly as many aspects are not published in an obvious place on their interface—or, indeed, anywhere at all on the platform.

ECCO's creation, especially as a product balancing corporate, public, and scholarly demands, "challenges the notion of a linear, progressive history" (Gregg *Old Books* 1). It begins with the English Short Title Catalogue (ESTC) and the microfilming of texts, especially in the British Library, in the late 1970s, a project from which modern ECCO still inherits bibliographic metadata, and thus errors.[13] The project expanded from its original goal of "first and significant editions of each title" (except for all editions by "major authors"—all white men) to include "all distinct editions of a work by the late 1980s."[14] Digitization and OCR were outsourced to various companies, with the first iteration of ECCO published in 2003, and ECCO II in 2009.[15] However, because of the lack of public documentation regarding editorial boards and

---

[12] According to Gregg's communications with Gale, "In 2020 2,092 institutions and consortia in forty-two countries subscribe; in 2019 around 7.7 million search results, images, or texts were retrieved worldwide" (1). All file formats derive from the original microfilm photographs of the physical pages. In addition to online viewing, users can download PDFs, page images, and plain text of the texts through ECCO-Gale Primary Sources (ECCO-GPS) interface (original ECCO only provided PDFs and images); and XML and TIFF formats distributed on hard drives to subscribing libraries. As a response to the COVID-19 pandemic, the American Society for Eighteenth-Century Studies (ASECS) partnered with Gale to provide access to ECCO for ASECS members beginning in July 2020, which now provides perpetual access for members based in North America and the Caribbean. Jisc's Historical Texts interface, another important access point, was retired in July 2024 ("Frequently Asked Questions").

[13] For several detailed examples, including authorship attribution and format errors, see Jim May's "Some Problems in ECCO (and ESTC)."

[14] For a full list of the 28 authors, see Gregg (23). The latter "distinct editions" do not include texts such the Bible with their considerable number of reprints. See Mikko Tolonen et al. in "The Anatomy of Eighteenth Century Collections Online (ECCO)" for a quantitative analysis of significant editions and titles.

[15] For full details, see Gregg (48–73).

consultations, the influences on and of some of these editorial decisions remain unknown. This fault is one of the most significant to Jim May, who suggests that

> These big commercial products offer profusion at a cost to precision and might be excused as being 'too big' for editorial supervision. No scholars' intellectual reputations stand behind them. More scholarly rigor was no doubt needed when the filming by Gale and its predecessors was done to decide which copy should be filmed. (28)

May also points out that "ECCO and the film series underlying it are necessarily slanted toward works owned by the libraries where filming was done," a critique which Patrick Spedding builds on to suggest that, because of the focus on particular, well-endowed libraries, the "microfilm series is not a random—and therefore randomly representative—selection of items from ESTC" (441). ECCO continues to develop: in 2020, Gale—the parent company holding ECCO—began digitizing another 90,000 titles, and in 2023, they re-ran OCR on several segments of the dataset.[16] Hard drives with image and text files have been available to subscribers for a decade, and the new Primary Sources interface allows access to each page's underlying OCR. These improvements point to ECCO's ongoing importance to eighteenth-century research, especially as interest in computational methods increases, ranging from basic keyword searches to data visualization.[17]

As computational work continues to gain momentum for both scholars and companies, however, so too does research into ECCO's role as a representative resource.[18] In Mikko

---

[16] According to a Q&A with Gale senior product manager Megan Sullivan, "For ECCO, we reran all documents written by women and BIPOC authors as well as any material written by prominent eighteenth-century authors as defined by the original ECCO selection criteria. We also reran anything identified as a dictionary, encyclopedia, or a periodical" ("Optical Character Recognition (OCR) Improvement").
[17] For an analysis of recent use of ECCO, including citations in recent publications, see Cassidy Holahan.
[18] Spedding lists several examples, noting that "The literature on ECCO is dominated by reviews and review-essays" (n1).

Tolonen et al.'s thorough "account of the anatomy of ECCO: what is missing, what are the imbalances, and how representative it is with respect to its source catalog," they point in particular to "publication format, reprinting phenomena, the overrepresentation of popular authors, and the underrepresentation of certain publication places in ECCO, as well as how many of these are conditioned on the year of publication of the work" (95).[19] Although the TWDB's approach and dataset creation avoids some of these issues, random samples from ECCO may have more reprints and popular authors than expected.[20] Even more than ECCO's composition, however, OCR quality is "the most well-known topic in discussions of ECCO" (Gregg *Old Books* 63), impacting everything from keyword searches to my data modelling in chapter 4.[21] Mark Hill and Simon Hengchen's comparison of ECCO's OCR to the manual transcriptions of the ECCO-TCP project shows a 77.4% accuracy rate (828), and the eMOP project calculates accuracy levels of 86% to 89% (6:21)[22]—both of which are much lower than Gale's estimates of "in the low to mid 90%'s," numbers which are now visible on the GPS interface for each page, rather than an invisible assumption (ECCO "FAQ" qtd. in Gregg *Old Books* 65).[23] A further complication is that different OCR engines were used for ECCO Part I and Part II, and probably for the upcoming Part III.[24] Because of such "crucial limitations," Tolonen et al. note that Gale's most recent addition to its digital humanities suite, the *Digital Scholar Lab*, is "virtually unusable for many research tasks" ("Corpus" 33). Even with these challenges, however, ECCO remains

---

[19] Similarly, Evalyn "conclude[s] that mass digitization has not selected against works attributed to women, but instead selected against 'authorless' works," including both ECCO and HT initiatives (134–5).
[20] For example, reprints are largely avoided by algorithmic and manual deduplication.
[21] For more details on my methodology regarding OCR accuracy, see appendix 2.2.5.
[22] Christy et al. compare ECCO's commercially created OCR transcription accuracy (89%) to eMOP's more open-source versions (86%). In tests on some of the most inaccurate pages, the eMOP workflow resulted in higher accuracy than Gale or Readex (71% to 64%) (6:21).
[23] Holahan suggests that this is the "most significant[]" aspect of the GPS platform (820). See also Paddy Bullard's accuracy estimations based on 15 random pages from five books (755–6).
[24] See Gregg (*Old Books* 59–60); see also Matthew Christy et al. for a technical description of OCR and eMOP.

the predominant access point for digital remediations of eighteenth-century primary sources, and the only one to provide such OCR predictions (and have them scrutinized so closely).[25] These factors are especially significant as ECCO supplies the most titles to the TWDB.

ECCO and Travel Writing

If ECCO is the most "comprehensive" access point to eighteenth-century documents, just how much travel writing does it contain? Again, ECCO's organization reflects definitions of travel writing, with their attention to heterogenous topics. For example, on the original ECCO interface and the hard drives, ECCO lists eight different modules or subject areas.[26] Following the traditionally close relationship of history, geography, and travel writing, the description of "History and Geography" explicitly identifies travel writing as a topic, noting that the module "feature[s] travel accounts, pilgrimages, topographical histories, and gazetteers…[a] variety of travel accounts, from explorations of the British Isles to adventure travel…[and] Translations of travel accounts, especially by French travelers" ("Overview"). This module also includes histories both ancient and modern, titles on heraldry and genealogy of the peerage, and biographies. This framing and organization thus reinforces travel writing's position as a nonfiction genre with closer connections with history rather than literature.

In fact, because of travel writing's genre flexibility, instead of relying on modules, the more accurate way to count travel writing on ECCO is to examine the Library of Congress

---

[25] Tolonen et al. call ECCO the "most comprehensive dataset available in machine-readable form for eighteenth-century printed texts. It plays a crucial role in studies of eighteenth-century language and has vast potential for corpus linguistics" ("Corpus" 19). The Helsinki Computational History Group (including Tolonen et al.) are conducting most current large-scale examinations of ECCO and the ESTC.

[26] History and Geography; Social Sciences; Fine Arts; Science, Technology, and Medicine; Literature and Language; Philosophy and Religion; Law; and General Reference. The Gale hard drive metadata use the tag "module" and the ECCO interface uses the term "Subject Area." These categories often overlap in subject matter, but each volume is only in one module. Hypothetically, volumes of the same title could be split across different modules, but I have not conducted a review. See Tolonen et al. for a discussion of "anomalies" in the composition of these categories ("Corpus").

subject headings, which indicate the "aboutness" of a work.[27] Counting the 4,782 volumes[28] that contain LCSH with at least one of the TWDB's LCSH keywords—`travel, discover, explor, voyage, guidebook,` and `antiquit`—shows that Gale classifies over 60% of all these volumes as part of the History and Geography module, but every module, even Law, holds some travel writing, highlighting the heterogeneous nature of the genre.[29]

Table 2.1: Number of Volumes with Travel LCSH in ECCO Modules[30]

Module	Raw Count	Ratio of Travel Titles in Each Module
History and Geography	3095	0.647219
Social Sciences	780	0.163112
Literature and Language	479	0.100167
Religion and Philosophy	150	0.031368
Fine Arts	117	0.024467
Medicine, Science and Technology	81	0.016939
General Reference	61	0.012756
Law	19	0.003973
Total	4782	

In fact, counting LCSH on ECCO can support one common scholarly claim about travel writing: that it was one of the most popular genres of the century. Table 2.2 shows that the LCSH subdivision `description and travel` is the eleventh most common subdivision in ECCO, following another notably popular genre, sermons and religious writing. Novels, the most

[27] There are 257 titles with relevant LCSH in other languages. Also note that because I selected using LCSH, all these volumes have LCSH; however, 1,668 volumes of the 207,614 on the ECCO hard drives do not have LCSH (less than 1%). For a description of the methodology and structure of LCSH, see appendix 2.2.3.
[28] I count volumes rather than titles. If counting only volumes with `currentVolume` "0" or "Volume 1", the number is reduced to 3,800.
[29] Note that these keywords sometimes pick up LCSH unrelated to travel, such as `Travell, Thomas, Sir`. Based on my later explorations, when I use LCSH to add titles to my corpus, these titles make up only a few volumes of the total.
[30] This same calculation with a single title instead of multiple volumes results in a 0.609 ratio for History and Geography, with other modules staying largely the same.

popular genre later in the century, do not make an appearance, but this may be due to the challenge of cataloguing fiction.[31]

Table 2.2 Most Common LCSH Subdivisions in ECCO (1700–1800)

Subdivision	**Raw Count**	**Ratio to Overall LCSH Subdivisions**
early works to 1800	135919	0.317965
18th century	29388	0.068749
great Britain	21811	0.051024
History	18756	0.043877
Sermons	15847	0.037072
England	11527	0.026966
politics and government	10083	0.023588
n.t.	6027	0.014099
o.t.	5541	0.012962
Periodicals	5467	0.012789
description and travel	3957	0.009257
Ireland	3617	0.008462

Knowing the data exists and then accessing it, however, are two distinct challenges. And the solutions to these problems are not stable: for example, over the course of this dissertation, Gale integrated the standalone ECCO platform into their Primary Sources (GPS) interface, which in turn changed search algorithms, file accessibility, and subject cataloging.[32] Most users access ECCO through the online GPS interface, where they can search for texts using various metadata fields, download PDFs and plain-text of titles, and do "light" digital humanities methods such as

[31] See Gregg, "1748: 'Fiction' in the Database."
[32] Gale retired the original ECCO interface on December 18, 2020. Similarly, the original interface for Early English Books Online, the early modern equivalent to ECCO, was retired on July 7, 2020, with users redirected to the new ProQuest interface (which still maintains LCSH integration).

Term Frequency.[33] Doing so, however, makes the user vulnerable to the shifting methods of the online interface. For example, when searching for travel writing, one might search for `description and travel`, the most common Library of Congress subject heading in the TWDB. The original ECCO interface returns 3,024 results. GPS returns 0.[34] This is because GPS requires consistent subject indexing across all its content, so ECCO's subjects now rely on machine-aided indexing.[35] Although Gale representatives confirm the original LCSH will be re-integrated in the future, the project continues to see delays.[36] Unfortunately for explorers of travel writing, this means that discovering travel texts has become more difficult: for example, *The history. Of the discovery and conquest of the Canary Islands* holds, on ECCO, the LCSH of `Canary Islands--Description and travel`. On GPS, the "Related Subject" is `Hostages`.[37] Meanwhile, *The history of a voyage to the Malouine (or Falkland) Islands*, with its multiple LCSH on ECCO, has no GPS "Related Subjects" at all.[38] Nowhere on the interface

---

[33] The former ECCO interface limited number of PDF pages to 250, requiring multiple downloads for longer titles. GPS allows downloading of the entire volume, as well as the plain text, though as recently as 2017, this only included what Gale metadata had designated as the Main Body of the text, as opposed to the Front Matter (title page, preface, dedications) and Back Matter (indexes, advertisements, and so on). At some point between 2017 and 2020, downloading the "full text" began to include these divisions—another undocumented change.

[34] Searches conducted July 17, 2020, and April 4, 2024.

[35] While GPS's MAI may be consistent for subject indexing across their different databases, this translates into a significant depletion of resources for both this particular instance and for other similar, subject-oriented queries. While Gale is increasing its access to DH tools through the *Digital Scholar Lab* (which requires its own subscription), it draws from the same metadata as the rest of the GPS ecosystem. Thus, a download of the metadata via the *Digital Scholar Lab* of the three titles listed above (*Three Tracts*, *The history. Of the…Canary Islands,* and *The history of a voyage*) does not list anything in the "Subject" column. Note that the downloaded metadata also relies on the Gale Document Number as an identifier and does not include the ESTC number or edition field, limiting interoperability with other, richer metadata such as the ESTC.

[36] Gale has been planning to reintegrate LCSH for several years (Holahan n28), but LCSH are still missing as of 2024. Gale confirms that the project will be resumed in 2025 (Sullivan).

[37] Gale ID CW0100336746.

[38] Gale ID CW0100958323. *History of a voyage* on ECCO included `Patagonians--Early works to 1800; Natural history--Falkland Islands--Early works to 1800; Magellan, Strait of (Chile and Argentina)--Early works to 1800; Patagonia (Argentina and Chile)--Description and travel--Early works to 1800.`

are these subjects explained.[39] This technological change, intended to increase interoperability, has disabled a critical method of finding one of the most popular genres of the century.[40]

For the TWDB, however, I detour around these public interfaces. Instead, I use the XML files on hard drives from ECCO, which include full metadata (including LCSH) as well as the text. This way, I can rely on my own searches, rather than the hidden algorithms of an online interface. Of course, this seemingly standard route is a challenge for future pathfinders wanting to reproduce my steps: even if their institution can afford the hard drives, do they have the same version?[41] Thus, though ECCO supplies the most volumes and metadata for the TWDB, relying naively on it as a transparent source will actually result in missing many of the resources—both hidden and expected—along the way.

Nineteenth Century Collections Online

NCCO, another Gale resource, provides only a handful of files for the TWDB, and if numbers were my only concern, I would not have included NCCO. However, like ECCO, NCCO is a prominent resource in the field (especially prior to the rise of HathiTrust and Google Books), and thus it warrants discussion. NCCO differs from ECCO in more than just period; instead of relying on the Nineteenth Century Short Title Catalogue (NSTC),[42] Gale instead created a series of "modular, subject-specific 'archives'" that "explore the themes and movements most

---

[39] The document accessed by clicking "Help" at the bottom of the page simply notes that "Related Subjects displays the subjects used for indexing an article. Clicking on a subject delivers additional articles indexed with the same subject."
[40] For more discrepancies between the original ECCO interface and GPS, see appendices 2.2 and 2.4.5.
[41] For example, even the total number of volumes differs depending on interface: the McGill ECCO hard drives hold 207,614 XML volumes, the online ECCO interface holds 207,627, and the eMOP ECCO metadata holds 207,662. Following the upcoming updates to ECCO, Gale notes that "the original OCR will be archived and available to be shipped to users on a physical drive" ("Optical Character Recognition (OCR) Improvement").
[42] See Frank Robinson for a description of the creation of the NSTC. ProQuest acquired the NSTC in 2005.

requested by scholars and researchers" (Product Highlights: NCCO).[43] Like ECCO, the editorial guidelines for these modules are unclear,[44] and Gale has folded NCCO's independent interface into GPS, with similar challenges for searching by subject.[45] For example, the XML of *Letters from the Caucasus and Georgia* holds LCSH such as `Women Travelers, Voyages and Travels`, and `Iran—Description and Travels`; online, it has only the MAI-created Related Subject of `Nadir Shah, Shah of Iran.`[46] While *Letters from the Caucasus* can still be found relatively easily because of its inclusion in NCCO's "Mapping the World: Maps

[43] Gale representatives describe different reasons for avoiding the NSTC. Seth Cayley suggests that "finding the guiding concept for NCCO was more challenging than for ECCO, because there was no trusted equivalent of the ESTC for the nineteenth century." Meanwhile, in addressing NCCO's "substantial differen[ce]" from ECCO, Ray Abruzzi from Gale states that "In planning *NCCO,* Gale did consider following the *Nineteenth Century Short Title Catalogue* (*NSTC*) in an attempt to achieve a similarly comprehensive experience, but we came to realize that such an approach would cause *NCCO* to fall short of its objectives…within the *NCCO* vision, the *NSTC* could only play a partial—though important—role in documenting the 19th century. Why is this the case? Principally because *NCCO* is not limited to books (though books are an important part of the resource). Rather, *NCCO* places strong emphasis on manuscript materials; ephemera; newspapers; government documents; personal archives; illustrations; photographs; and other content types, too. Most crucially, *NCCO*'s vision is a global one, covering many regions of the world across the 'long' 19th century, not strictly the period 1801–1900" (McWilliam). The modular setup also allowed Gale to sell individual archive access, rather than all of NCCO.

[44] To choose these collections, Gale relied on NCCO's International Advisory Board, whose "role is to suggest topics, themes, concepts, and regions of study, and to recommend the institutions, associations, and scholars needed to provide expert input and identify specific collections"; the Board also "helped conceptualize and commission new content [and] advised on the creation of headnotes that provide context for exploration of the individual collections within the archives," in addition to providing feedback about design and user experience ("About Nineteenth Century Collections Online"). Depending on the webpage and the date of publication, the list of members on the advisory board changes and may have conflicting information. For example, a list in 2018 lists 7 scholars; a more recent list (July 15, 2021) posted on the GPS-NCCO website lists 10 (some, but not all, of which, are the same). See, for example, "NCCO Advisory Board" and "Advisory Board Members."

[45] NCCO's original interface—more modern than ECCO and closer to GPS—did not contain LCSH either. Currently, GPS still enables browsing by module and has replicated the "About" text for each module.

[46] Gale ID CWVQGJ422652176.

and Travel Literature" archive[47] (which contains a variety of "collections"[48]), much other travel writing is in other archives such as "Science, Technology, and Medicine: 1780–1925."[49] Thus, because of NCCO's modular, rather than comprehensive, goals, as well as the lack of LCSH, projects cannot just use NCCO where ECCO stops, and even relying on it as a varied source of travel writing is much more fraught. In recognition of NCCO's influence on eighteenth-century studies, I do include NCCO as both a bibliographical source and file source. However, because of NCCO's limitations, the TWDB must primarily rely on other digitization projects for the nineteenth century.[50]

HathiTrust Digital Library and Research Center

If NCCO is a particularly imperfect source, then, what is the alternative—or at least, another piece to try to fill the gaps? For the TWDB—as with other projects exploring the nineteenth century—the HathiTrust Digital Library and Research Center (HT) is the primary option. HT has become increasingly popular for cultural analytics projects because of its comprehensive (rather

---

[47] According to a Gale Fact Sheet from 2014, Dr. Jordana Dym in the history department at Skidmore College curates the archive, though it is unclear exactly what "curation" means in this context; for individual items in these collections, Gale relied presumably on a mixture of the library's catalogue divisions, local archivists and scholars, and their own criteria ("Mapping the World: Maps and Travel Literature Factsheet"). Terminology such as "Selected Journals and Papers" and discrepancies between a library archive's catalogue and what NCCO has published online obfuscate the editorial decisions influencing the content of these online resources. However, because the timeline of NCCO's digitization and curation efforts is not published, it is unknown whether these differences are because of library acquisitions or other constraints.

[48] These collections focus mostly on maps, including collections such as the National Archives: Selected Maps Representing the Long 19th Century, the King George III Topographical Collection, and Travels and Travelers in the 19th Century: Selected Journals and Papers [from the British Library].

[49] For example, a general online keyword search of `voy*` in the "Science, Technology, and Medicine: 1780–1925, Part II" archive results in several travel-writing titles, such as Basil Hall's *Voyage to Corea, and the Island of Loo-Choo*, William Parry's *Journal of a Voyage for the Discovery of a north-west Passage*, and *Travels through the Southern Provinces of the Russian Empire*; however, only *Voyage to Corea* had any terms in the Related Subjects field (`Korea`; `Ryukyu Islands`). Reviewed on April 22, 2024. At McGill, each module from NCCO is on its own hard drive.

[50] Many smaller digitization projects focus on particular figures or collections. *Networked Infrastructure for Nineteenth-Century Electronic Scholarship* allows for interoperable searching of many of these datasets.

than modular) attempts and supportive infrastructure. As with ECCO, as use of the dataset has increased, scholars and HT itself have pointed out cautions with using the collection. Even so, HT has more coverage for later centuries, and, usefully for the TWDB, LCSH, which makes HT the closest thing to an interoperable dataset with ECCO. Where ECCO stops in the TWDB, then, HT picks up the slack.

HT is a "not-for-profit collaborative of academic and research libraries," the main purpose of which is to maintain a repository to archive and share their digitized collections ("Welcome"). Established in 2008, in 2024, it holds over 18 million total volumes.[51] These titles consist of digitized texts ingested from Google and the Internet Archive, as well as from numerous public and university libraries, most of which are American, with digitization and OCR completed by partner institutions or other entities like Google.[52] Most works were published after 1900; only 1.1% of the collection, or 92,193 works, are dated 1700–1799. From 1800 through 1829, on the other hand, there are roughly 125,106 works ("HathiTrust Dates"), already drawing close to ECCO's 180,000 titles.[53]

HT's limitations include an American focus, uneven chronological coverage, and a tendency toward monographs prioritized by academic libraries. As Ted Underwood explains in his work on identifying fiction in HT, "In its raw form, HathiTrust probably comes closest to answering [what got bought by libraries]," but that may leave out ephemeral works, juvenile fiction, dime novels, and other works not commonly collected by academic libraries ("Understanding Genre" 36). In their more recent datasets of fiction in HathiTrust from 1700 to

---

[51] Of these, 6,739,307, or approximately 39%, were in the public domain.
[52] This aspect is particularly important because of how different methods handle the long "S" of eighteenth-century printing.
[53] Search conducted June 20, 2020. I did not capture the number of titles when I did my initial searches and data harvesting two years prior.

2009, Underwood, Patrick Kimutis, and Jessica Witte further note that "While academic libraries collect works by famous writers around the world, coverage of popular culture and juvenile fiction is weaker, especially outside an Anglo-American context" ("NovelTM Datasets" 4). Even within the American publishing sphere, HT is limited: Underwood et al. estimate that "HathiTrust contains a little more than half of the nineteenth-century fiction titles mentioned in *Publishers Weekly*" (4–5).[54] In the shorter, detailed lists of the NovelTM dataset, the authors eschew the eighteenth century entirely, as the "coverage in HathiTrust is uneven, and the amount of fiction published in the century is small enough that it would be possible to start with a bibliography rather than a sample" (9). While the TWDB ignores divisions between fiction and nonfiction, some of Underwood et al.'s observations about fiction—its ephemerality, its sheer bulk, its coverage in academic libraries—certainly apply to travel writing as well.

The accessibility of HT is also a significant advantage; like ECCO and NCCO, HT has an online interface for document viewing and downloading files in various formats, but its real strength is in its APIs, tutorials, work environments, and personal support that allow for mass downloading or manipulation of HT data.[55] Another advantage is HT's LCSH, which are incorporated both into their backend metadata schema and publicly accessibly information. Even though, as Eamon P. Duffy points out, for HT's Bibliographical Metadata Specifications for contributing organizations, "subject headings are neither required nor 'strongly preferred' elements of acceptable records" (10), in 2012, he found that roughly 75% of HT's MARC

---

[54] They also note that "In the twentieth century, that ratio drops to less than a quarter" (5).

[55] Certain resources, such as mass plain text downloads or access to items in copyright, are only available to member-affiliated researchers, but others are accessible to anyone with a non-profit or research institution affiliation—or even just to the general public. For example, the Extracted Features Dataset, with its 17+ million volumes of fiction, drama, and poetry, or the HathiTrust+Bookworm application, are open for public use ("Troubleshooting and FAQs"). HT released the Extracted Features Dataset, currently in version 2.0, under a Creative Commons Attribution 4.0 International License.

records contained LCSH.[56] Additionally, my own searches show that HT's proportions of travel keyword LCSH are similar to those of ECCO.[57] The keyword `travel` (connected to the `description and travel` subdivision) is again most common, and even the ratios for other keywords vary by, at most, 4%.[58] These similar ratios are despite the different time period for each search (1700–1800 for ECCO, and 1700–1830 for HT). While HT is not suitable as a complete substitute for ECCO, then, it is a crucial supplement when crossing the divide at 1800, especially since the LCSH make ECCO and HT the closest thing to an interoperable subject database across the two platforms—just as LCSH are intended to work.

Google Books

At this point, why add more complexity—especially a source as opaque as Google Books?[59] For some titles, however, Google Books held the only digital facsimile, especially when I was searching for titles by women. In reviews of the TWDB-origbib after the initial matches of metadata to digital files, 66 of the women-authored titles remained unmatched.[60] While manual searching resulted in matches for several of these titles in ECCO, NCCO, and HathiTrust, 20 of

---

[56] In Duffy's study, "any record containing at least one instance of a 600, 610, 611, 630, 650, 651 or 655 field with a second indicator value of "0" was deemed as having a subject headings field containing LCSH" (10).
[57] In 2017, I was using the now-defunct SolR Proxy HathiTrust API; numbers would have been slightly lower than when I completed the following searches on the HT Digital Library interface in the summer of 2020.
[58] For a table comparing search results, see appendix 2.2.2.
[59] See James Somers' history of Google Books in "Torching the Modern-Day Library of Alexandria."
[60] This lack is due to the additional pressures of the print market on women; for example, smaller or private print runs decrease the likelihood that their titles are in multiple archives or archives that can afford digitization. For others, however, manual searches showed that the lack of results were because of combinations of discrepancies in the metadata, such as in the title (mucrus vs Muckrus) or the author (for example, "a lady" on the title page, married name in bibliographical source, and birth name in digital metadata).

them—nearly a third of the missing 66—were only available on Google Books.[61]

While adding another database for an additional 20 titles may seem superfluous, especially compared to the hundreds of titles from other TWDB sources, avoiding this extra work would also mean encouraging the obfuscation of these authors. While these few titles are unlikely to affect studies relying on random samples of the thousands of volumes in the TWDB, I also wanted to enable smaller studies focusing on the gendered corpus, whether by myself or future users of the dataset.

Summary of Sources

A Library of Babel-style archive containing all texts of the long eighteenth century is impossible, as pages have been lost to events dramatic or mundane, to "the flames and the privy" (Friedman 353). A collection of titles that perfectly represents the eighteenth century is similarly unachievable, in theory as well as practice. All the digital sources above are far from either of these goals, despite claims to the contrary.[62] In combining these sources, the TWDB takes on their shortcomings, especially when using their data as a nontravel comparison. But because of the TWDB's focus on travel writing, especially through bibliography, we can begin to analyze some of the blank spaces in the map instead of treating these gaps as only natural or accidental—or ignoring them altogether.

---

[61] Benjamin Colbert's BTW–Women's excerpt was particularly useful for finding these works, as it links to many digital first editions, including titles that were bound together (see the works of Sophie Dixon, *A Journal of 10 Days Excursion* and *A Journal of Eighteen Days Excursion*, TWDB: d_4GAAAAQAAJ). Others I found by searching for the title and author in the Google Books web interface. Once I identified these 20 titles, I downloaded the accompanying EPUB file (several more titles were available as PDFs, but I did not include them because of the additional hurdle of performing OCR). I then transformed the EPUB file into a plain text file using the open-source e-book software, Calibre, and removed the Google copyright information at the beginning and end of each file so that it did not affect predictive modelling.

[62] As Bode observes, for "[p]roprietary mass-digitized collections such as Google Books, Early English Books Online, and The British Newspaper Archive," "the commercial imperatives of these enterprises arguably depend on them presenting these collections as comprehensive" (*World* 47). See also Somers.

The Second Challenge: Supplementing Bibliographical Sources with Library of Congress Subject Headings

We can also understand, or even fill in, these blank spaces, by examining titles identified via the Library of Congress Subject Heading controlled vocabulary (LCSH). Relying on a mixture of bibliographical sources and LCSH creates a corpus based on scholars' and cataloguers' combined estimations of "aboutness," rather than using the presence of a few words in a title or text to determine the topic, theme, or genre.[63] This intersection of categories and definitions, and their mutability and instability, reflect these tensions in the related literary, historical, and library science traditions. Rather than treating either a single bibliography or a metadata system as authoritative, then, this is an opportunity to compare decisions, the "conflicting practices" of different users, cataloguers, and scholars (Underwood "Historical Significance" 68). As Underwood summarizes in his analysis of similarities in LCSH genre and topics, subject headings are "typical of the categories that interest humanists: they are tangled up with time, for reasons that have as much to do with the history of observers as with the history of the object" ("Historical Significance" 61). As I will show, in practice, LCSH generally treat travel as more of a topic than as a genre, placing the metadata structure in conversation with travel writing definitions such as Jan Borm's, where travel writing "is not a genre, but a collective term for a variety of texts both predominantly fictional and non-fictional whose main theme is travel" (13). Thus, while LCSH may not be suitable for explorations of all genres or topics, for travel

[63] For a discussion of genre markers on title pages in eighteenth-century fiction (including terms such as "history," "letters," and "account"), see Leah Orr, "Genre Labels on the Title Pages of English Fiction, 1660–1800."

writing—a genre already focused on a particular topic—subject headings are a powerful tool for gathering data and a rich feature for analysis.[64]

LCSH: Background

Since the TWDB's construction and analysis relies so much on LCSH, I outline some of its key elements below. The LCSH controlled vocabulary is produced from the subject authority file maintained by the United States Library of Congress. As of March 2020, it held 348,246 authority records, with single words or phrases (and combinations thereof) representing topics, persons, corporate bodies, geographical areas, and other phenomena (LC, "Introduction to Library of Congress Subject Headings"). Libraries, archives, and other institutions use this vocabulary for indexing, cataloging, and accessing records in physical and digital library catalogues.[65] The LCSH vocabulary is the most widely used such thesaurus in the world, in part due to the "administrative and managerial machinery of [the Library of Congress]" and thus has been translated into many languages (*LS & IT* "Library"). The LCSH vocabulary dates back to 1898, when the LC converted its catalog and relied on the American Library Association's *List of Subject Headings for Use in Dictionary Catalogs* (1st ed., 1895; 2nd ed., 1898), adding new categories as needed (Stone 2). These iterations have continued to the modern day: as terminology and scholarship change in the face of shifting cultural and social justice issues, terms are added (approximately 5,000 a year), revised, and removed (LC, "Introduction"). As

---

[64] In contrast, for other research, LCSH may be problematic or even unsuitable. For example, see Underwood's categorization of "fiction" in HathiTrust ("Understanding Genre") and Gregg's comparison of "fiction" tags in ECCO and ESTC ("'Fiction' in the Database").

[65] Unlike most of the TWDB bibliographers, these cataloguers remain unnamed. Gregg's history of ECCO and LCSH describes the "idiosyncratic" subject headings of ECCO, which derive from a "a combination of human interpretation and computer-aided harvesting of data," including from "the Northwestern University experiment in 2005," ESTC's contracts with "HTC Global in 2007," and "'existing' library records" (71–2).

Emily Drabinksi has pointed out, library classification schemes like the LCSH are not objective, but are "products of human labor that carry traces of all the intentional and unintentional racism, sexism, and classism of the workers who create them" (198).[66]

Two elements of LCSH are especially important to the TWDB: "aboutness" and structure.[67] When determining the "aboutness" of a work and applying subject headings, cataloguers assign headings for topics only if they consist of 20% of the work: thus, if a title has the LCSH `Italy--description and travel`, at least a fifth of the book should, ostensibly, be about travel in Italy. LCSH are also structural, with a main heading (`Italy`) optionally followed by subdivisions (`description and travel`). As we will see in chapters three and four, these rules create a degree of consistency that make LCSH a valuable resource not just for finding titles, but also for analyzing them.

Searching for LCSH: The Eighteenth Century and Travel

The first challenge in finding travel writing via LCSH is knowing which terms to search for. The LCSH database contains some candidates: `Travel in literature`,[68] `Travel`

---

[66] See also Steven A. Knowlton's "Three Decades Since Prejudices and Antipathies: A Study of Changes in the Library of Congress Subject Headings." Cataloguers have also made a renewed effort to handle the "subject/form confusion…a not-the-same, not-different state of incomplete fusion" of earlier cataloguing to enable distinctions between works *about* a subject or *examples* of a form (Miller 171). This discussion of subject versus form culminated in the Library of Congress establishing the Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT), which has developed since initial discussions in 2007 to include literature in 2015 ("LoC Genre/Form Terms PDF Files"). The MARC field 655 holds these terms, but it is unclear how many institutions have updated their records with LCGFT, especially since some LCGFT overlap with previously established and still valid LCSH. MARC 655 may also hold terms from other thesauri and controlled vocabularies. For example, HT titles in my corpus included "Travel Literature" in MARC 655 via the Getty's *Art and Architecture Thesaurus* (aat) or the *Genre Terms: A Thesaurus for Use in Rare Book and Special Collections Cataloguing from the Association of College and Research Libraries* (rgbenr). As a result, although LCGFT includes both "travel writing" and "guidebooks" as authorized terms, MARC 655 and LCGFT were not part of my analysis.

[67] For a full description and examples of LCSH structure, see appendix 2.4.3.

[68] Created in 1986, for "works on travel as a theme in literature" ("Travel in literature").

writing,[69] and Travelers' writings, [nationality].[70] However, searching these subjects on ECCO and HT usually returns zero results.[71] Instead, travel writing is easier to identify in terms that indicate topic, rather than genre, and in subdivisions. For example, the LCSH main headings of voyages and travels, voyages, imaginary, voyages around the world, and voyages all return results, but shifting beyond just these main headings is even more productive. Duffy, in his case study searching for primary documents related to Peruvian bat guano mining in the 1800s, found that

> paradoxically…the main headings are somewhat less important…Rather, the key is to use certain free-floating subdivisions that follow the main heading and identify several document types that often prove valuable as primary sources, . . . [including] Description and travel (for travel narratives) (6).

And this subdivision is not static. As Duffy notes, "*description and travel* now encompasses all travel writing, while before 1991 works describing cities or colonies were assigned the now defunct free-floating subdivision, *description*" (10).[72] Other potential terms for eighteenth-century primary sources, however, return little to no relevant results on ECCO or HT: "Journeys," "Pilgrims and pilgrimages," and "Scientific expeditions"

---

[69] Created in 1991, for "works on the authorship of writings by travelers that are often presented in narrative form or as memoirs" ("Travel writing").

[70] Created in 1992, consisting of "Collections of works written by travelers from a specific country" ("Travelers' writings, English"). See also "Travel writing." Examining travel-related genre subdivisions of the LCSH database by searching under the "GenreForm" tag returns the subdivisions of aerial photographs, tours, pictorial works, gazetteers, aerial views, and guidebooks, all of which are associated with the description and travel topic as a subdivision or earlier established form. See the main Library of Congress Subject Headings page.

[71] "Travel writing" on the now defunct ECCO interface returns 25 results, mostly periodicals, while the other terms return no results on both platforms. Search conducted July 16, 2020.

[72] ECCO (hard drive and defunct online interface) only holds one title with the description subdivision: *An account of Tangier*, which also holds description and travel (Gale ID CW103939768). HT holds 20 English-language titles with descriptions (rather than the singular description).

all returned less than a dozen volumes,[73] while "`Grand tours (Education)`" returned no results at all.[74] Cataloguing of eighteenth-century travel writing thus often tends toward general main headings and subdivisions about travel, rather than specific types of journey.

For historical research, these LCSH take on a particular importance: as Jeffrey Garrett reports in the 2005 experiment applying LCSH to ECCO files at Northwestern University, he found that subject headings significantly increased the number of search results.[75] He describes how

> For a number of reasons, some having to do with changes in the lexicon, some with a century-specific perceived need for circumlocution, words such as 'hygiene' and 'prostitution' occurred far less frequently in the eighteenth century than they do today—not to mention the often disastrous effects of pre-1800 orthography on modern-day keyword searches. (70)

With the advent of full text searching, including tables of contents and indexing, however, searching by subject has fallen out of favour; instead, "keyword searching has become the most often used, and, in fact, the preferred method of conducting a search in any online system" (Gross et al. 4). Numerous studies and discussions have addressed the feasibility and usefulness of assigning LCSH, questioning whether LSCH, with their high cost of human labour, should be abandoned or modified.[76] The preponderance of studies comparing different search methods,

---

[73] Furthermore, some of these documents were reprints, later editions, etc.
[74] By "not relevant," I mean not published between 1700–1830, in English, in Britain. Search conducted April 14, 2020 on the defunct ECCO interface.
[75] This was before ESTC and ECCO metadata were enhanced with subject headings in 2009. See Gregg in *Old Books* (70–9).
[76] Other options include of keyword searching, folksonomies, query expansion, or new prototypes and experimental interfaces that adapt to both user needs and the information system. For an overview of various studies, see Tina Gross et al., "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching."

however, show that controlled vocabularies still greatly enhance results by reducing the number of hits to a manageable level or by increasing coverage of results that would otherwise be missed (Gross et al.). Even with full text available, Garrett demonstrates that

> keyword searching in full-text databases is no substitute for searches run against . . . bibliographic files with ample descriptors and subject headings. . .The demonstrable fact is that full-text searching of eighteenth-century texts often does not retrieve examples of terms that describe the work as a whole or even important topics or aspects of the work, especially as we might describe them today. (75)

Tina Gross et al.'s 2015 study comparing keyword, title, and subject searches in a contemporary academic library database conclusively supports this position: they found that an average of 27% of keyword search results would be lost without the subject headings.

If we test this approach on travel writing, the results are even more dire, missing nearly half of the travel titles in the TWDB. In the TWDB 2,816 of its 4,259 titles do not have any of the TWDB travel keywords[77]; even if expanding the list to include tokens that often appear in travel writing titles (`excursion, journey, tour, trip, expedition, visit`), the title search misses 2,262 titles, or nearly half of all titles. Searching full-text and all metadata for the travel keywords on the online ECCO or HT interfaces returns too many results to review (over 100,000). Searching only titles is more manageable (7,586 volumes for ECCO; 3,705 titles for HT), but such searches will ignore subgenres that use other title terms. The TWDB titles with no travel keywords do often contain other terms, such as journal, memoir, history, narrative, account, or letters, but these search terms are also more likely to return titles that are not travel

---

[77] Travel keywords consist of `travel* OR voyage* OR guidebook* OR discover* OR antiquit* OR explore*`.

writing. LCSH let us address travel writing in many forms, going beyond what is marketed in a title.

Implementing LCSH in the TWDB

With these factors in mind, I chose LCSH based on metadata I already had from an initial scrape of HT matches to TWDB-origbib titles.[78] The final iteration of the TWDB relies on six terms in LCSH: `travel*` OR `voyage*` OR `discover*` OR `explor*` OR `guidebook*` OR `antiquit*`.[79] Travel writing scholars may immediately call foul: what about history, especially if the TWDB terms include antiquities?[80] After all, even eighteenth-century travel writers such as Henry Fielding called travel writing a branch of history, and Cox's bibliography has sections for both "History and Chronicle" (3.372) and "History and Antiquities" (3.404).[81] Here, however, is where the intersection of travel writing definitions, LCSH terms, and "aboutness" becomes important. The LCSH `history` gathers titles with history as a subdivision under not only geographical names, but also corporate bodies, sacred works, ethnic groups, and other topical

---

[78] All queries of the HT LCSH metadata relied on the now-defunct SolR Proxy HT API.

[79] I include `antiquit*` based on the focus of Cox and Leask (see also M.O. Grenby and Rosemary Sweet). Another candidate for future iterations of the TWDB is `shipwrecks`, though some scholars (such as Colbert) do not include this subgenre as travel writing. Still, `shipwrecks` as a subject (without any of the other travel keywords) returns 144 volumes on ECCO (60% of which include "Crusoe" in the title). Similarly, on HT, there are 40 titles with the `shipwrecks` LCSH, and 4 with "Crusoe" in the title. Some of these volumes and titles are later editions or reprints, and some (at least eight with just the `shipwrecks` LCSH and no travel keywords) are already in the TWDB through TWDB-origbib metadata, so it is not a significant difference in overall numbers of the corpus. in ECCO XML, the subject headings are in the `locSubject` tag. Using longer tokens can reduce the labour to review the results but can also sometimes miss some useful terms: for example, `guide` rather than `guidebook` returns `pilot guides` (31 volumes), but also `color guides` and `life skills guides`. Searches conducted June 29, 2020 on the defunct ECCO interface.

[80] Percy G. Adams takes the opposite approach, arguing that "Travel writing, in spite of what librarians and historians have often said, is not a branch of history any more than it is of geography" (280).

[81] Fielding writes in his preface to the 1755 *Journal of a Voyage to Lisbon* that "there is no other branch of history (for this is history) which hath not exercised the greatest pens" (547). See chapter 4's discussion of travel writing and history.

headings.[82] Even Cox is careful to limit his inclusion of history, examining only particular historical subgenres.[83] Other related subdivisions, such as `economic conditions`, `politics and government`, or `social life and customs`, also remain impractical to implement, though with potential as a future angle of inquiry.[84] The diversity of LCSH attached to even the six TWDB keywords provides an expansive sense of the many applicable LCSH, but also means removing titles with similar but undesirable terms.[85] For travel scholars working on a larger corpus, or a corpus in a different time period, more caution may be needed: the LCSH vocabulary contains dozens of topics including `travel`, such as `Travel with donkeys`, `Used travel trailers`, `Income tax deductions for travel and entertainment expenses`, and `Seismic traveltime inversion`. The TWDB keywords—`travel`, `voyage`, `discover`, `explor`,

---

[82] On HT, for example, a subject search for history with TWDB limits resulted in more than 5,000 titles (search conducted April 15, 2020); see the LCSH page for history for more information on the subdivision ("History").

[83] Cox writes in his explanatory note for "History and Chronicle" that he focused on "regional histories, such as those of counties and parishes, [which are] more concerned with antiquities, genealogies of prominent families, records, and the like, and less with recounting transactions of the passing hour," especially because "As a rule, [they] are more likely to have occasioned a certain amount of travel on the part of the writer, though they exhibit little evidence of personal reactions" (3: 372).

[84] Creating a list of locations from the LCSH authority file, a corresponding list of keywords for subdivisions, and then either accepting all results or using personal judgment based on metadata made this unfeasible. Examples of titles and LCSH include *A complete history of Algiers* (1728), with the LCSH `Africa, North---History_Algeria---History` (HT mdp.39015073766605), or *Memoirs of the life and travels of the late Charles Macpherson* (1800) with `Slavery---West Indies_West Indies---Social life and customs` (HT nyp.33433082344973).

[85] The list of removed terms includes `Travell, Thomas, Sir`, `Period of discoveries, 1385-1580`, `discoveries and cautions from the streets of zion`, `plain discovery`, `act for indemnifying such persons as shall upon examination make discoveries touching the disposition of publick money`, `church of england---prayers and devotions---guidebooks---early works to 1800`, `finance, personal---great britain---guidebooks`, `history of the colonization of the free states of antiquity`, `antiquitates judaicae`, `review of the doctrine of the eucharist, as laid down in Scripture and antiquity`, and `antiquity and duration of the world`. In ECCO, titles were sometimes mistakenly included as a LCSH. For example, `Duty of consulting a spiritual guide, considered` (ESTC: T124153), also in the ESTC data.

`guidebook`, `antiquit`—do not capture every travel writing text, as our contrast with bibliographies in the next chapter demonstrates. However, LCSH do allow us to engage with the diversity of travel writing far beyond a title or full-text search, both enriching bibliography and enabling comparisons of texts based on what they are "about," as I show in chapter 4.

Going the Extra Mile: Encoding for Gender

Encoding Gender: In Theory

In Yoojung Choi's study of women travel writers, she critiques how "The current narrative of the eighteenth-century history of travel writing centers around a few canonical male writers," which "reifies the myth that only male fiction writers…joined in the fad of travel writing" (3). In the TWDB, I engage with this "myth," adding to Choi and others who analyze how women (or their imitators) participated in the genre of travel writing as authors, editors, and translators. The TWDB lists only printed work, and attempts the first printed incarnation of a single title rather than prepublication drafts, multiple editions, reviews, or reader reception, areas with further rich avenues for research.[86] Even with this limitation, as with other projects, I come up against the challenge—or, as Laura Mandell describes, the feminist opportunity—of structuring TWDB data to identify and examine women-authored texts.[87] Scholars and activists have troubled the definitions of gender for decades, and these tensions have necessarily carried over to digital

---

[86] Margaret Ezell has written extensively on women's manuscript writing in relationship to print culture. See, for example, "Editing Early Modern Women's Manuscripts: Theory, Electronic Editions, and the Accidental Copy-Text." Choi also points out the limitations of focusing on published travel writing over manuscript writing (see especially 10–2).

[87] My approach to gender, book history, and digital humanities was greatly influenced by the attendees of the *Women in Book History Symposium, 1660–1836* in 2018 in Vancouver, many of whom I cite throughout this dissertation.

infrastructures.[88] Miriam Posner, in her discussion of "What's Next: The Radical, Unrealized Potential of Digital Humanities," identifies handling "categories like gender that are not binary or one-dimensional or stable"—that is, "develop[ing] models of the world that have any relevance to people's lived experience"—as "the most complicated, challenging computing problem I can imagine, and DH hasn't even begun yet to take it on" (2016). Instead, as Mandell discusses in "Gender and Cultural Analytics: Finding or Making Stereotypes?", most contemporary studies that measure a "gender" signal use the "M/F" division. She notes that "In the fields of cultural analytics, computational linguistics, and quantitative sociology, such work typically appears in articles with 'gender' in the title—paradoxically because 'male' and 'female' are biological sex terms rather than gender terms" (3). Instead, although gender is not entirely distinct from sex, it is a cultural, rather than a biological term.[89] As such, gender is also an intersectional category[90]: Posner discusses how gender data models should include the person's descriptor, additional descriptor(s), associated pronouns, time, place, imputed gender, race, and class.[91] Forcing a M/F binary on the data ensures that the model will only produce results fitting that model, rather than exploring how "these structures themselves constitute data" (Posner "What's Next"). Based on this long history, it is unsurprising that the first principle of Catherine D'Ignazio and Lauren F. Klein's "Feminist Data Visualization" is to "Rethink Binaries," which the TWDB embraces through encoding authorial gender in multiple ways (2).

---

[88] For a useful overview that connects feminist identity politics directly to computational methods, see Susan Brown and Laura Mandell's introduction to "The Identity Issue" in *Journal of Cultural Analytics*. For an overview of feminism in data visualization, science and technology studies, human computer interaction, digital humanities, and critical cartography and geographic information systems, see D'Ignazio and Klein, "Feminist Data Visualization."
[89] Mandell refers to Eloy LaBrada on how terms like cisgender, transgender, and genderqueer designate the relationship a person has to their assigned gender.
[90] Kimberlé Crenshaw theorizes the term intersectional "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color," which appeared in the *Stanford Law Review* in 1991.
[91] For one example of gender encoding over time, see Pamela L. Caughie et al.'s discussion of data models and gender ontologies as applied to *Man into Woman* (1933), the posthumously published life narrative of Lili Elbe.

In early years of feminist work on the eighteenth century, these goals were often constructed and framed as recovery projects: increasing attention and accessibility to texts by women, especially as printed scholarly editions, would enable their accessibility in both research and the classroom. This method has continued into the digital age: large projects such as the *Orlando Project* and the *Women Writers Project* were conceived as a counter to the marginalization and lack of availability of women's writing (Wernimont). Jacqueline Wernimont, in "Whence Feminism? Assessing Feminist Interventions in Digital Literary Archives," draws attention to the "exuberant celebrations of abundance and presence" in commentary on these projects. For her, this "emphasis on familiar patriarchal tropes of size, mastery, and comprehensive collection" leads to the assumption that more (or all) women will be present in the archive. However, this "additive approach" (Rooney qtd. in Wernimont)—does not mean that "women's work [will become] a visible, central part of literary history" (Wernimont).[92] In the TWDB, I include women's work, but I also purposefully feature it in my analysis in order to make women's travels, writings, and voices more prominent.

Two projects that influence the TWDB's conception of authorial gender are Benjamin Colbert's *Women's Travel Writing, 1780–1840* (one of the TWDB's bibliographical sources, btw_w) and Megan Peiser's *Novels Reviewed Database, 1790–1840.* Colbert records gender in a "M/F/Unknown" format on *Women's Travel Writing, 1780–1840*, but does not distinguish whether they are identified as such on the title page, by personal research, or other means. His categories include principal author, contributor, editors, translators, and illustrators, "so long as

---

[92] Mandell notes that "Many digital recovery projects of women's writing have…never realized their ambitions," and, in the current academic atmosphere, "are not in great demand" (513–5). See also Amy E. Earhart, "Can information be unfettered? Race and the new digital humanities canon."

contributions to the book are not peripheral" ("About").[93] Peiser, who focuses on novels, includes three fields in her model that can describe an author's gender.[94] First, she records the gender as represented by the review's bibliographical heading (names [including pseudonyms], titles, or phrases ("Jane," "Countess," or "By a Lady") as "F" for female, authorial gender (*British* 2.10). Secondly, she records the reviewer's use of pronouns in the body of the review. Finally, she relies on research regarding the author's gender, ranging from Ralph Griffith's personal notes in the *Monthly Review* to modern editions of the work, recording where she finds each piece of information.[95] Other ongoing bibliographical projects include the *Women's Print History Project, 1750–1830*, which also lists figures such as compositors, engravers, and publishers (also encoded as "Female, Male, or Unknown"), while the *Women in Book History Bibliography* instead focuses on secondary sources on women's writing and labor in the eighteenth century and beyond.[96] These projects, Kate Ozment's recent "Rationale for Feminist Bibliography," and the TWDB indicate that gender remains an ongoing concern in bibliography, book history, and the disciplines' remediation into the digital realm.

Other studies focusing on gender sometimes rely on computational approaches rather than bibliographical. In *Distant Horizons*, Underwood uses "names recorded in the US census as

---

[93] Users can search via author, contributor, editor, or translator, but not illustrators, though illustrators are included as a browsable category (last accessed April 22, 2024). Colbert also includes biographical profiles on each figure; as of the end of the project's funding in May 2018, 45 of the 91 creators in his database had little to no information in extant biographical sources, creating a valuable resource that Peiser highlights in her review of the project ("Review").

[94] Peiser also includes reviews of translations of novels, citing the English translator as the author (women often translated foreign works, taking advantage of publishing rates that treated new and translated prose fiction as the same). Including these women also "acknowledges the invisible labor of many women writers who are often ignored in histories of the novel, especially in quantitative studies" (*British* 2.11–2). See also Peiser's "Reviews as Database."

[95] Many thanks to Megan Peiser for providing a pre-publication copy of her dissertation and database.

[96] In Evalyn's dissertation, he includes categories that "register the extent to which the author of a work in question can be understood through the lens of gender" (74).

a guide" for the gender of names in his HT dataset, though he notes the shortcomings when identifying "ambiguous names or non-European names or pseudonyms or multiple authors" (135). [97] Underwood and Sabrina Lee also construct a smaller corpus based on *Publishers Weekly* that they manually review for authorial gender; Underwood then uses these datasets in his examination of the proportion of both women's authorship and women characters. Other scholars commonly focus on differing features, especially vocabulary, to identify differences in style or predict gender identities of authors in texts ranging from novels to speeches and tweets.[98] However, without enough nuance, as Mandell points out in her critique of Jan Rybicki's work on novels at Chawton House, these studies fall into stereotype (Rybicki "*Vive la Difference*"; Mandell "Finding").[99] Instead of accurately measuring some sort of "gender signal", Rybicki's model is "guessing about what [Mandell] would call 'textual gender,' the stylistic and textual features associated with gendered genres," as women's increased participation in genres like the sentimental novel dictate their topics (15). She suggests including more categories, ranging from "men writing as men" to "men writing as unspecified (anonymous byline) in the voice of a woman," in order to take into consideration other influences on the style of a text (15). Similarly, in Bode's data for *World of Fiction*, she has seventeen different author name categories, including variations on attribution, honorifics, initials, and pseudonyms (88–9). Examining peri- or epitextual materials—title pages, prefaces, authors' letters—gives insight into the system that the book was inhabiting. This is especially relevant when some authors, editors, and translators may remain unknown. Others have been "discovered"—or disproven—over the years (though

---

[97] Evalyn found the "gender" package for R, the "current standard tool for large-scale gender identification[,]…to be inadequate for the eighteenth century" (71).
[98] For summaries and overviews, see Rybicki, as well as Sean G. Weidman and James O'Sullivan, "The limits of distinctive words: Re-evaluating literature's gender marker debate."
[99] See "Finding" for Mandell's use of "stereotype" (5).

sometimes library catalogues take longer to be updated).[100] The TWDB's documentation of multiple perspectives of "women's writing" enables not just inclusion of "feminine" voices, but analysis of their nuance as well.

Encoding Gender: In the TWDB

Creating this kind of thick data by examining paratextual materials may be possible with a smaller dataset of a few dozen, or even a few hundred, titles, given the appropriate resources, but for the thousands of titles in the TWDB, examining even just the facsimile title page of each work is unfeasible. How, then, to identify titles written, translated, or edited by women?[101] Some of this data comes from previous studies, such as the British Travel Writing–Women's excerpt (btw_w) or Jane Robinson's *Wayward Women* bibliography (robinson_w). But these bibliographies focus on a narrow definition of travel or of "authored by a woman"—definitions that are not necessarily inaccurate, but without the richness that the TWDB's comparisons enable.[102] Therefore, the TWDB includes tags for works that claim creation by women or that have been assigned such by scholars or institutions, relying on names as well as keywords such

---

[100] See, for example, Colbert's post on "Arrivals and Departures" from the BTW–Women's Travel Writing database. See also Elizabeth Hagglund's discussion on the disputed authorship of *Journey into the Highlands of Scotland*, for which many library catalogues (including the ESTC, ECCO, and HT) still list Mary Ann(e) Hanway as the author despite critics' evidence otherwise (*Tourists* 155–6).

[101] The only known women-illustrated title in the TWDB is *Tunbridge Wells, and Its Neighbourhood* by Paul Amsick (1810), which includes etchings by Letitia Byrne.

[102] Choi is careful to point to this issue, arguing that "For example, scholars have prioritized…factual travel accounts over fictional ones. Women's travel writing studies have limited their scope to a female traveler's published writing of her authentic travels, despite the fact that many women did not have the means to document their journeys, or did not have a chance or intention for publication" (7).

as "`lady`," "`Mrs`," "`gentlewoman`," and "`wife`."[103] This approach has numerous caveats: most obviously, women's work—as author, editor, or translator—may be incorrectly (or not at all) acknowledged on title pages and thus in institutional metadata.[104] And the inverse may be true: works that claim to tell a story "by a Lady" may be authored by men.[105] I return to this argument for broadening our understanding of "women's travel writing" in chapters 3 and 4 to demonstrate how such titles let us examine how authors resist or reinforce expectations for a feminine travelling voice. In particular, in chapter 4, the models trained on these examples can discern which titles belong to the "gendered" corpus and which are from a random sample of the eighteenth century, indicating that these authors do rely on "particular strategies—including specific artistic principles and designs"—as they engage with the particular genre demands of travel writing (Korte 2). This assessment, however, is only possible through the extra labour of encoding gender—and encoding it beyond a simple "biological sex of the author" binary.

Conclusion

The TWDB relies on practices that are decades, if not centuries, old, ranging from patriarchal printing traditions to the constantly revised LCSH vocabulary. Facing the complexity of these

---

[103] To further nuance this metadata, especially for future users, the TWDB also records the role and where in the text the claim was made by examining facsimile editions of the titles. For most of these texts, especially authorship, the title page proclaimed either a woman's identity by a gender identifier (*A sentimental tour through Newcastle; By a young lady.*, TWDB 0646501600) or by name (the famous Helen Maria Williams). Other titles identified their authors or editors as women not on the title page, but in dedications, prefaces (especially for women editing their husband's work), or, more rarely, signatures at the very end of the work (Jane Squire's *A proposal to determine our longitude*, TWDB 0206300500). The TWDB does not indicate whether the gendered claim is associated with a name or if it is anonymous. The final category of data regarding the inscribed gender is whether it is only claimed somewhere in the text or if scholars have confirmed the veracity of the claim. For full details, see appendix 2.2.4.

[104] For example, in early HT scrapes, the metadata included author and related names divided by gender, including errors such as Camille de Roquefeuil or ED Keaton (since updated to be E[dward] Keaton), as well as paratextual material such as book inscriptions (Catherine Vaslet's name, for example, inscribed on the second volume of *A new account of the East Indies* from Universidad Complutense de Madrid on HT).

[105] James Raven lists several examples of male novelists using the "by a Lady" pseudonym in the late eighteenth century (145). For early modern traditions of "by a Lady," see Ezell, "'By a Lady': The Mask of the Feminine in Restoration, Early Eighteenth-Century Print Culture."

data, systems, and interfaces, both print and digital, can be overwhelming. Emily Friedman notes the temptation "to throw up our hands, turn our backs on large-scale collaborative projects, and cultivate our small, impressionistic gardens" (358). ECCO has incomplete and biased coverage and NCCO is more specialized, even if a scholar's institution can afford access. Meanwhile, HT has little coverage before 1800, and Google Books is essentially impossible to use on a large scale. All are plagued with OCR errors, some of which are documented, others obscured. Like Friedman, however, I argue that "the perfect cannot be the enemy of the good" when creating or using such data (358).[106] The TWDB, with its accompanying documentation, is a robust dataset that allows for comparing different data sources both printed and digital, filtering for OCR quality, and exploring travel writing through subjects obvious or esoteric. My editorial decisions are just some of many influenced by academic practices, financial markets, and available technology. And these artifacts are not static: many will continue to change, some of them daily. In this discussion of the construction of the TWDB, then, I provide not just the background for any future findings arising from the project, but also a model for other projects seeking to do similar historical literary research. This second iteration of the TWDB continues the "data cleaning" of chapter 1, embracing and extending Katie Rawson and Trevor Muñoz's "invitation to scrutinize, perhaps reimagine, and almost certainly rename this part of our practice."[107]

---

[106] At the same time, as this dissertation demonstrates, I am wary of a desire for perfection. Mak argues that "the database is a performative space in which readers enact desires for its completeness, and that such desires have been carefully and strategically fostered" (1518).

[107] In addition to Rawson and Muñoz, see Bode in "From World to *Trove* to Data: Tracing a History of Transmission" (*World* 59–81).

Chapter 3: Observations and Reflections on the TWDB

> Here the pitiless reader, sitting quietly at his table with his map before him, will say to the poor, hungry, harassed traveller, exposed to all the trouble of war: "I see no account of Apheoditopolis, Crocodilopolis, Ptolemais—what is become of all these towns? What had you to do there, if you could not give any account of them? . . . have I not relied upon you to give me some information on all these subjects?"
>
> —Vivant Denon, trans. Arthur Aikin, *Travels in Upper and Lower Egypt* (1803, 2: 22)

> All of this [travel scholarship] is heady stuff that needs, as with the novel, to be grounded on the reading and analysis of thousands of books, not just of exploration reports but of all travel literature.
>
> —Percy G. Adams, *Travel Literature and the Evolution of the Novel* (1983, 162)

The Travel Writing Database (TWDB) contains 5,506 volumes of travel writing, on subjects as varied as the health benefits of spa water, tours to the island of love, memoirs of foreign courts, and the abolishment of slavery. But there are also hundreds that are missing, either from our bibliographical sources or online datasets. In this chapter, I explore what the TWDB metadata reveals about travel writing, even as I address what Lawrence Evalyn calls "the fantasy of

completeness" (64).[1] Margaret Ezell similarly warns of how "comprehensive" datasets encourage the "belief that if you don't find something…it did not exist" ("Big Books" 9). Instead, what the TWDB can and cannot include becomes yet another argument about our received histories of the genre—received not just from scholars, but from institutions. As Katherine Bode argues in her documentation of the *Trove* project, researchers should "expose the constructed and conditional nature of the curated dataset: not in order to discount but to establish its capacity to stand as a reliable foundation for [our] own literary-historical arguments, and for those of others" (81).[2] If we are to use the TWDB, we must ask how representative the TWDB is of travel writing, or at least, travel writing as created by bibliographers, libraries, and datasets. As this chapter demonstrates, travel writing tends to remain focused on "real journeys," with the most cited titles also being the most digitized, even as subject headings point out the significant diversity of the field.

The first questions that the TWDB can answer are quantitative. If, as scholars suggest, travel writing was one of the most published genres in the century, then how do the TWDB records represent travel writing's growth from 1700 to 1830? Even though 41% of the titles from the TWDB-origbib are still missing, the TWDB contains at least 150 titles every decade, enabling comparisons ranging from bibliographies to Library of Congress Subject Headings

---

[1] Bonnie Mak also works with the term "fantasy," drawing from Sherry Turkle who "characterizes the environment of digitizations thus: 'Computer precision is wrongly taken for perfection. The fantasy, visceral in nature, is that computers serve as a guarantor'...This fantasy harbors the implicit assumption that all valuable information is online, and that such information, apparently having been certified by computational processes, is necessarily complete, comprehensive, and accurate" (1520).

[2] For Bode's description of *Trove*'s constraints and affordances, see 59–81.

(LCSH).[3] Without Eighteenth Century Collections Online (ECCO), the early eighteenth century is scarce; without HathiTrust (HT) after 1800, the TWDB would replicate the troubled periodization of other projects. LCSH further mitigates this century division, pointing to the critical nature of interoperable data structures. The TWDB's final combination of sources opens new paths for exploring travel writing across this time period, enabling Percy G. Adams' call for the theorization of travel writing to rely on "thousands of books, not just of exploration reports but of all travel literature" (*Travel Literature* 162).

Secondly, the TWDB data gestures toward trends in the "aboutness" of travel writing in the eighteenth century. The relative stability of the most common subjects, such as `description and travel`, contrasts with the otherwise flexible and encompassing nature of the genre. Although one might expect Grand Tour destinations or other foreign locales to be most common, the TWDB actually holds more titles on domestic travel than any single foreign destination. LCSH also reveal how the genre relies on subgenres and responds to world events and popular trends, both through the 2,655 titles with travel writing LCSH and the 1,604 titles without.[4] Finally, expanding definitions of travel writing beyond a documented "real journey" makes room for authors, especially women, who are engaging with the genre through forms as varied as novels, guidebooks, and juvenile literature. Thus, through numbers, subjects, and definitions, the TWDB contributes to a more intersectional perspective of travel writing, addressing Ezell's question of how "the digital…[can] help us to imagine new forms and

---

[3] The final TWDB dataset consists of 4,259 titles or 5,506 volumes. Of these volumes, the majority are single volumes (3,506 volumes or 63.7%), while the remainder are multi-volume works (2,000 volumes, 36.3%). The designation of single or multi-volume works is based on the metadata from the data source rather than the actual number of volumes in the TWDB. Some works may have multiple volumes listed in the metadata, but the data source may hold only one of the volumes; the TWDB data records such a title as a multi-volume work.

[4] If relying only on travel-related LCSH, the TWDB would miss 37.7% of its current titles.

functions for literary histories" ("Big Books" 9).[5] I argue that—just like travel writing itself—the TWDB must consider these different genres, sources, and arguments to understand how these texts, and the eighteenth century more broadly, conceptualize, organize, and understand the world.

The TWDB and the Growth of Travel Writing

Annual and Decade Trends

The diachronic publication trends of the TWDB are similar to the TWDB-origbib data from chapter 2: the number of titles peaks in 1791–1800, with a drop in 1801.[6] In the TWDB, this high point is less prominent: compared to the TWDB-origbib, the TWDB has fewer titles before 1800 and more titles afterward (figure 3.1). This difference demonstrates the significant impact of the HT LCSH harvest on the TWDB, simultaneously filling and thus emphasizing the gap left by the lack of a comprehensive bibliography of travel writing in the early 1800s.[7] Still, even considering only single titles rather than volumes, the TWDB holds more than 150 titles every decade. After the 1710s, this includes more than 10 titles a year, and after 1770, at least a couple dozen. In direct comparisons with the number of novels published during this time, Benjamin Colbert points to "[Charles] Batten's impression that travel writing 'won a readership second only to novels' at the end of the eighteenth century" (13). Colbert's data suggests that "by the

---

[5] Kristi Siegel states that "Without sufficient attention to determinants such as race, class, location, historical circumstance, and power—to name just a few—any conclusions drawn about women's travel become meaningless" (1). See also Katherine Turner, who sees "the discourse of individuality and eccentricity" in eighteenth-century travel writing as "largely a middle-class phenomenon" (21).

[6] When comparing TWDB and TWDB-origbib, I rely on titles, as TWDB-origbib does note record volumes.

[7] Ten titles have more than 15 volumes; the vast majority of multi-volume works have three volumes or less. Multi-volume works are most common after 1760, with peaks in 1790, 1820, and 1824. 1790 holds 36 volumes of *English botany*, 1820, 26 volumes of *Topography of Great Britain*, and 1824, 30 volumes of *The modern traveller*. The TWDB condenses date ranges into publication dates of the first volume, skewing the representation of when a volume was actually printed.

early nineteenth century this trend was reversed," with nonfiction travel titles overtaking novels (13).[8] Depending on one's questions, this dataset may be too big or too small, but for the methodology of perspectival modelling in chapter 4, these titles offer a healthy number from which to draw various subsets and samples.

Figure 3.1: TWDB Publication Counts by Decade: All Volumes

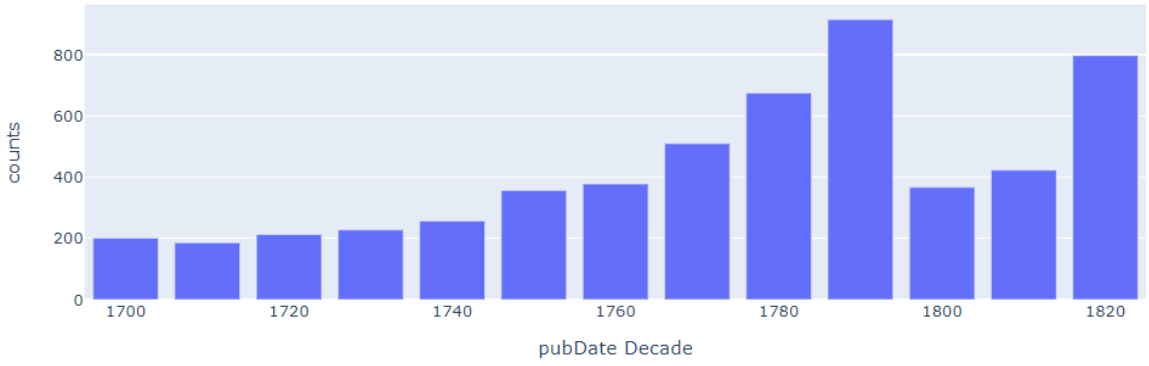
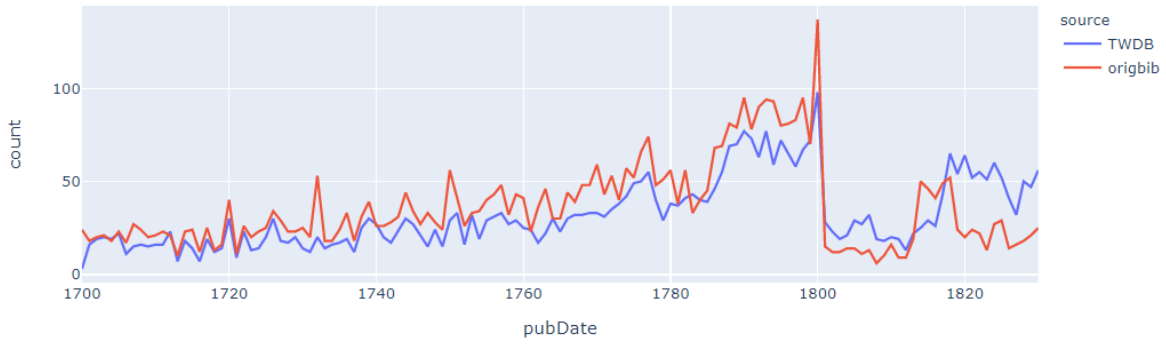
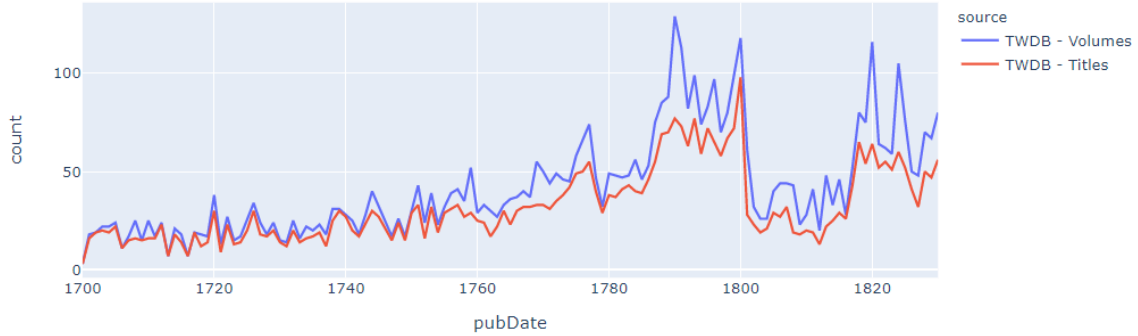


Figure 3.2: Title Publication Counts by Decade: TWDB vs TWDB-origbib



[8] Furthermore, Colbert observes, "several [novels] in this period imitate the travel genre, particularly satires and comic novels" (13). See the introduction for a literature review of publication numbers.

Figure 3.3: TWDB Publication Counts: Volumes vs Titles



Bibliographical and Data Source Trends

Charting the different TWDB sources demonstrates their significant impact on the dataset. Nearly 60% of the TWDB-origbib entries connect to a digital file, and this ratio is relatively stable diachronically, meaning that just because a title is published later in the century does not guarantee that it is digitized or discoverable.[9] Instead, the more significant correlation is cross-references: the more bibliographies that cite a title, the more likely it is to have a digital counterpart.[10] All titles with more than four citations are in the TWDB. Of the titles with two to three citations, 90% are listed. However, the vast majority of titles (4,409) have only one citation, and of these, the TWDB only links 56.6% with a digitized edition. This means that titles already documented in multiple bibliographies are over-represented in the TWDB. Furthermore, since many of these bibliographical sources focus on "real" travel, the diversity of travel writing may be under-represented.

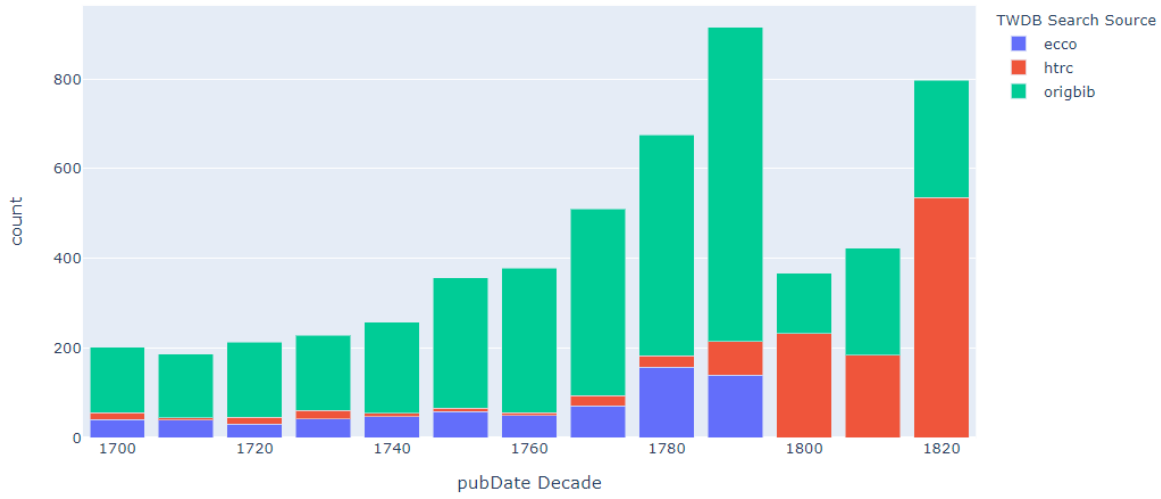
---

[9] The average number of TWDB-origbib titles found is closer to 50% in the 1700s, 1730s, 1740s, and 1810s, and about 70% in the 1800s and 1820s. See appendix 3.1.

[10] Manual searching also influences these ratios: as will be discussed further in the overview of the gendered titles, 84.6% of gendered titles in TWDB-origbib have a match in the TWDB, compared to 58.4% of ungendered titles.

The TWDB's collection of titles by LCSH supplements these bibliographical listings and becomes especially influential after 1800. As figure 3.4 shows, after the turn of the century, LCSH contribute more volumes to the TWDB than all the bibliographies combined. Doing quantitative comparisons across the century divide without these titles skews the results. Instead of discussing travel writing more generally, such analysis would instead compare large enumerative bibliographies like Edward Godfrey Cox's *Reference* to more specialized sources focused on a single area.[11] As I discuss later in this chapter, LCSH come with their own biases, but for questions of travel writing as a genre spanning decades, continents and styles, bibliography alone is inefficient.

Figure 3.4: TWDB Volume Counts by Decade: Search Source

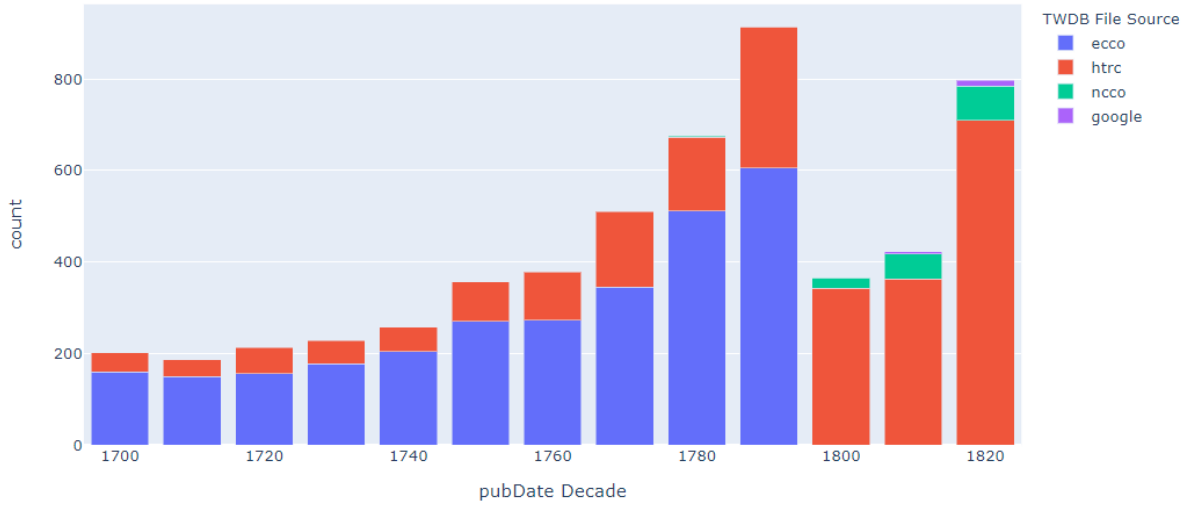


Examining which database provides files to the TWDB in figure 3.5 also shows the effects of periodization. ECCO and HT dominate, switching prominence at the turn of the

[11] Further integration with Benjamin Colbert's full dataset and the *Global Odyssey* bibliography could add further variety, though Colbert still focuses on "real" travel.

century, with several dozen files from NCCO and Google Books scattered throughout the latter parts of the dataset. The lack of early travel writing on HT echoes the findings of Underwood's suggestion to consider sources other than HT for the eighteenth century, as the "coverage . . . is uneven" ("NovelTM Datasets" 9). Even though I searched HT for TWDB-origbib files before ECCO, HT contributes less than 60 volumes a decade before 1751. These numbers emphasize ECCO's "crucial role" as a dataset for the eighteenth century (Tolonen et al. "Corpus" 19).

Figure 3.5: TWDB Volume Publication Counts by Decade: File Source



TWDB and LCSH: What is Travel Writing "About"?

LCSH metadata provides connections not just between HathiTrust and ECCO, but also within the TWDB corpus and to other genres of the eighteenth century.[12] Nearly every title in the TWDB contains at least one LCSH; only 87 do not, and most of those come from Google Books,

[12] See chapter 4 for comparisons of models based on LCSH including fiction, history, and different locations.

which has no public LCSH.[13] This thorough coverage makes LCSH useful not just for methods of harvesting and sampling, but also for analysis. Furthermore, the TWDB allows comparisons between three relatively equal categories of the corpus: (1) TWDB-origbib entries *with at least one* LCSH travel keyword (1,226 titles); (2) TWDB-origbib entries *with no* LCSH travel keywords (1,604 titles), and (3) titles with *no* TWDB-origbib connection, but *with at least one* LCSH keyword (1,429).[14] The vast majority of the LCSH titles are captured by the word `travel`, but just as the TWDB considers definitions beyond "real" travel, so too can we consider LCSH beyond such obvious terms.[15]

In fact, even though one's first thought of travel might be international, the most popular LCSH suggest a domestic focus. In the top 20 complex LCSH alone, Great Britain or a location within its borders make up nearly half of the entries.[16] Notably, `great britain` appears four times, combined with the subdivisions `description and travel`, `history`, `colonies`, and `antiquities`, suggesting concerns not just for the history of the growing

---

[13] Of the titles in the TWDB, 87 have no LCSH, 1,720 have only one complex LCSH, and the remaining 2,452 titles have more than one. 87 titles, or 2% of the TWDB, may seem small, but since all of the Google titles are related to the gendered corpus, any LCSH analysis or sampling based on LCSH leaves out 16.8% of the gendered titles. The TWDB has better coverage than HT in general, as is shown by Duffy's 2012 analysis where 25% of public-domain HT MARC records did not have LCSH. See also HT's 2015 account of MARC coverage for each field. The field with the highest coverage—655, topical term—is at nearly 65% (Organisciak "MARC Coverage"). Eleven TWDB titles have more than the maximum ten recommended by the Library of Congress; one title, a report on Acts of Parliament, holds 26 complex LCSH (mdp.35112204864187). See appendix 3.2.

[14] The keywords are `travel`, `voyage`, `antiquit`, `discover`, `guidebook`, and `explor`.

[15] These analyses come with caveats. For example, the MARC records for a title in the TWDB may vary between institutions (sometimes even between different editions of the same work in the same database). Unless all catalogues draw from a common database such as the ESTC, the cataloguers may assign different LCSH, despite the guidelines set out by the Library of Congress. Basing these decisions on title pages, tables of contents, and other paratextual materials can be a challenge, as both the LCSH documentation and travel writing bibliographers suggest. Therefore, prioritizing files from different sources, even if using the same "first edition title" may create different results. See appendix 3.2.

[16] The LCSH of `england`, `london (england)`, `scotland`, and `wales` with `description and travel` are also in the top 20, along with the current or former colonies of the `united states`, `america`, and `canada`. Within main headings, the most common collocations are nearly all with `great britain`, especially `great britain` with itself (for example, `great britain---antiquities`, `great britain---description and travel`). See appendix 3.3.

empire, but also its future. In fact, in a count of all LCSH, without disambiguating between main headings, subdivisions, or their combinations, `great britain` and `england` are once again in the top five most common, right after the more general `description and travel`, `early works to 1800`, and `history`.[17] In contrast, nearly 2,000 LCSH only appear once, including terms as varied as `fish ponds`; `generative organs, female`; `linen industry`; `dreams`; and `dwarfs`. In fact, of the 3,179 unique LCSH elements in our corpus, 2,993 of them, or 94.2%, appear 10 times or less. These terms may be less obviously useful for methods such as sampling, but they do identify the "most unique" (in LCSH terms) titles in the TWDB.

Travel Writing Subjects and Trends

As one might expect, travel writing's most common LCSH of `early works to 1800` and `description and travel` are relatively stable diachronically, and especially in titles published before 1800.[18] Occasionally, the conventional destinations of France and Italy, or the more revolutionary United States, rise into the most common elements, but the collected LCSH of the TWDB tend towards domestic interests.[19] Calculating only the most common main headings, rather than including subdivisions as above, reveals more interesting diversity, but

---

[17] See appendix 3.3.

[18] Prior to 1800, `early works to 1800` is between 12.5 and 16.6% of all LCSH each decade. Its successor to the top position, `description and travel`, is always one of the top two terms from 1700–1830, but from 1801 onwards, it makes up over 20% of all LCSH, peaking at 26% in the 1800s. This rise may be in part to the increase of titles harvested by LCSH rather than by bibliography. The geographic locales most often connected to the `description and travel` subdivision are Great Britain, Europe, England, Italy, and the United States. See the discussion of cross-references below.

[19] In the 1810s and 1820s, `history`, consistently in the top 5 across all decades, moves to second most common. The other terms appearing in the top five include `great britain` (12 decades), `england` (10 decades), `antiquities` and `france` (2 decades), and `18th century`, `united states`, and `italy` (1 decade).

must be considered with caution.[20] These subjects point to a travel writing market sensitive to publication trends, current events, and public tastes. In the 1710s, for example, `roads` is one of the most popular terms.[21] The popularity of `voyages, imaginary` in the 1720s is due to responses (especially "keys") to Jonathan Swift's *Gulliver's Travels* in 1726. In the 1730s, interest in the new settlement in Georgia lifts the state to the top five, alongside `jews` via the Jewish histories of Flavius Josephus.[22] In the 1820s, the appearance of `greece` in the top five terms reflects the trend of philhellenism and public interest in the Greek War of Independence and the death of George Gordon, Lord Byron.[23] Perhaps most interesting is the decrease of `great britain` throughout the eighteenth century, finally disappearing from the top five main headings in the 1820s, suggesting a potential shift in attention beyond the island's borders.[24] These findings are certainly exploratory rather than foundational, but these patterns in "aboutness" encourage scholarly attention to how travel writing topics within the TWDB and in travel writing more generally vary from decade to decade.[25]

Because the TWDB harvests titles based on LCSH travel keywords, it makes sense that LCSH such as `description and travel` are overrepresented. Bibliographical sources,

---

[20] Some decades have less titles, meaning repetitions of main headings may affect them more. Similarly, `scotland` is a different main heading from `orkney islands (scotland)`.

[21] This is because of only seven main headings across five titles. See Christine M. Petto for the popularity of road maps in the seventeenth and eighteenth centuries.

[22] For more on eighteenth-century translations of Josephus, see Gohei Hata, "Robert Traill: The First Irish Critic of William Whiston's Translation of Josephus" in *Making History: Josephus and Historical Method*.

[23] See William St. Clair in *That Greece Might Still Be Free: The Philhellenes in the War of Independence*.

[24] The main heading of `great britain` peaks at 11.3% in the 1700s to nearly 5% in the 1770s and 1800s. In the 1810s, it drops to third most frequent (2.4%) before disappearing from the top terms in the 1820s. This decrease may also affected by Cox's British volume ending at 1800. Other common terms include `voyages and travels` in the top 5 for 12 decades (except the 1760s) and the remainder ranging from eight decades (`united states`) to one, including geographical terms such as `spain`, `greece`, `scotland`, and `india` and topical terms such as `jews`, `roads`, `indians of north america`, and `voyages, imaginary`.

[25] See the introduction for an overview of the main research interests in travel writing scholarship.

however, offer a counterpoint useful for both sampling and analysis. In general, bibliographers and cataloguers tend to agree on which texts belong to the genre, since over 80% of TWDB titles listed in a bibliography also hold one of the LCSH travel keywords. This alignment, however, relies on definitions in both bibliography and cataloguing of "real travel," with sources focusing on authentic journeys having more travel keywords in their titles' subject headings.[26] But this comparative approach reveals that `voyages, imaginary` are a more difficult concept to catalogue. In Philip Babcock Gove's bibliography of the *Imaginary Voyage in Prose Fiction*, less than 60% of the titles have a travel keyword, with the most common LCSH of `voyages, imaginary`, `18th century`, `english fiction`, and `fiction`.[27] The difference is even more stark for Cox, whose wide variety of categories means that only 38.4% of Cox's matched digital files have a travel keyword. Analyzing Gove or Cox, then, will rely on titles with subjects other than stereotypical travel. Specialization can also occur beyond the travel keywords. For example, Nigel Leask's focus on "an antique land" is clear in the frequent main headings of `mexico`, `india`, `egypt`, and `ethiopia`. For the sources focusing on women's writing, Helen Maria Williams's works contribute to `france` as the most frequent main heading. The list from the Murray publishing house includes the rare `arctic regions`, demonstrating the publisher's connections with official Arctic explorations.[28] Finally, for titles sourced from John McVeagh's *Irish Travel Writing*, the most frequent main heading is, unsurprisingly, `ireland`. However, `great britain` is also among the most common,

---

[26] The most frequent LCSH within the bibliographical sources follow the trends of the larger corpus described above, especially for `description and travel`.
[27] If focusing on main headings, Gove's focus on `voyages, imaginary` expands to topics such as `utopias` and `interplanetary voyages`.
[28] See Keighren et al. (45–51).

pointing to their close proximity, both geographical and conceptual. In short, the LCSH can also provide insight into the selections of bibliographers and cataloguers, directly impacting what questions one might ask of the TWDB data.[29]

Bibliographical sources are also critical for capturing travel writing's heterogeneity beyond LCSH travel keywords. There are 1,604 titles with no travel keywords; without these works, the TWDB would not only be smaller, but also much more focused on traditional or easily recognized conceptions of travel. Most significantly, this corpus nearly doubles the number of titles at the intersection of travel writing and fiction, poetry, and drama.[30] This convergence of genre and form has long been of significant interest to travel writing scholarship, and it is critical for chapter 4's comparison of fictional travel and fiction more generally.[31] Other unique additions to the corpus through these titles include `agriculture`, `gardening`, and `botany`, which are largely (or in the case of `gardening`, entirely) in Cox.[32] Similarly, via Cox's "Spas" section, the corpus carries all of the `mineral waters` and `hydrotherapy` tags. This heterogenous combination of travel writing supports Adams' extended description of what travel writing *is not*, which ends by suggesting that it "cannot be a literary genre with a fixed definition any more than the novel is; it is not even sui generis since it includes so many

---

[29] For example, for work on Ireland, one may want to ignore any titles that also have an LCSH relating to Great Britain or to include any titles with the subject of Ireland, even if not listed in McVeagh.

[30] This corpus contributes 79 instances of the following LCSH elements across 73 titles: `english poetry`, `poetry, epistolary poetry, english`, `humorous poetry, english`, `scottish gaelic poetry`, `narrative poetry`, `english fiction`, `fiction, epistolary fiction, english`, `biographical fiction, english`, `english literature`, `juvenile literature`, `controversial literature`, `erotic literature`, `english drama`, `english drama (comedy), drama`, and `theater`. The rest of the TWDB, with the travel keywords, carries 105 LCSH across 88 titles.

[31] See "Travel and the Novel" in this dissertation's introduction.

[32] Cox includes these titles in his sections focusing on "Natural History" and "Agriculture, husbandry, and gardening" (both in volume 3, focused on Great Britain).

types both by form and by content" (*Travel Literature* 282).[33] It is only through both bibliography and subject headings that the TWDB can both list these titles, and then—as chapter 4 demonstrates—bring them into view.

Finding Women in Travel Writing

As the explorations above demonstrate, expanding the definition of travel writing from beyond "authentic narratives" is crucial for understanding the scope of travel writing in general. However, I argue that this approach is also needed for studying travel writing by women.[34] Current women-focused bibliographies, such as the BTW–Women's excerpt and Jane Robinson's, focus on "authentic" travel or titles by "real" women, but the TWDB's encoding of multiple aspects of women's authorship allows for "a more complex model of gender and the way it interacts with other factors" (Foster and Mills 3). For example, while women (or their claimed voices) make up a relatively small part of the TWDB, most assert their influence publicly on the title page, claiming a space in the public sphere. The LCSH for these titles both participate in the trends of travel writing more generally, but also reflect the locations easier for women to visit and topics often connected with women's writing in general. Katrina O'Loughlin draws attention to how this "significant corpus of eighteenth-century travel writing remaining to us directly challenges preconceptions about women's use of the genre," and as will be seen in the TWDB below and in chapter 4, this "use of the genre" includes not just personal narratives, but texts encouraging everything from religious reflection to nation-building to travel itself (8).

---

[33] Adams asserts, for example, that travel writing is not just "in prose…a set of notes… an objective report… an exploration report" (280–1).

[34] This applies to eighteenth-century women's writing more generally. David Mazella et al. "also argue for the critical value of recognizing variation, intermixing, and elaboration of existing genres in the newly expanded digital corpus that is now coming into view" (45).

Finding Women in Travel Writing: Previous Work

While some bibliographers, such as Robinson, have focused on women who write about travel, the only scholar to track more detailed statistics of women travel writers and their works is Benjamin Colbert in his *British Travel Writing* database (btw_w). He notes the following about the publication of travel books by women:

> In the years 1780 to 1840, around 5000 travel books were published by around 3000 authors, but, of those, only 204 can be identified as produced by 146 women. Of these titles, 168 can be classified as narratives, as opposed to collections, guidebooks, letterpress plate books and topographical descriptions. Whether we consider narratives specifically, or travel writings in the aggregate, the fact remains that women accounted for only around 5% of travel books published in Britain and Ireland during this period (a 20 to 1 ratio of men to women). ("Bibliographical Reflections" 6)

In fact, Colbert identifies only nine travel books written by women between 1700 and 1780 ("Bibliographical Reflections 165–6n11).[35] All of these titles but one are in the TWDB: Lady Margaret Pennyman's *Miscellanies in prose and verse*, published in 1740 by Edmund Curll.[36]

---

[35] Colbert lists the following titles in "British Women's Travel Writing, 1780–1840: Bibliographical Reflections," but his database begins at 1780, meaning that these titles are not flagged with btw_w in the TWDB: Marie-Catherine d'Aulnoy, The Ingenious and Diverting Letters of the Lady——T*ravels into Spain* (1691); Elizabeth Justice, *A Voyage to Russia* (1739); Lady Margaret Pennyman, *Miscellanies in Prose and Verse, by the Honourable Lady Margaret Pennyman. Containing, I. Her Late Journey to Paris* (1740); Lady Mary Wortley Montagu, *Letters of the Right Honourable Lady M—y W—y M——e* (1763); Marie Anne (LePage) Fiquet DuBoccage, *Letters Concerning England, Holland and Italy* (1770); Elizabeth Percy, Duchess of Northumberland, *A Short Tour Made in the Year One Thousand Seven Hundred and Seventy One* (1775); Anne Miller, *Letters from Italy* (1776); Jane Vigor, *Letters from a Lady, Who Resided Some Years in Russia* (1776); Mary Ann Hanway, *A Journey to the Highlands of Scotland* (1776) (see Elizabeth Hagglund on Hanway's disputed authorship [155–6]); Jemima Kindersley, *Letters from the Island of Teneriffe, Brazil, the Cape of Good Hope, and the East Indies* (1777).

[36] Of the remaining eight titles, seven were listed in different bibliographical sources (Cox, Elizabeth A. Bohls and Ian Duncan's *Anthology*, Robinson's women-focused bibliography, the Bryn Mawr collection, and Andrews' picturesque bibliography). The Duchess of Northumberland's Short Tour (1775) was added by the LCSH `description and travel`. Colbert also lists Madame d'Aulnoy's *Travels into Spain* (1691).

This absence demonstrates the vagaries of bibliography, digitization, and cataloguing: despite being held in prominent libraries, none of the TWDB's bibliographical sources, not even the expansive Cox, include this title.[37] Cataloguers have not described it as travel writing either, despite the explicit description of travel in the title.[38] This lack of attention also applies to criticism: as Yoojung Choi points out, scholars often focus on Elizabeth Justice's *A Voyage to Russia* in 1739 or Lady Mary Wortley Montagu's *Embassy Letters* in 1763 as early examples. Choi posits that "Pennyman's account of her journey to Paris has been forgotten because it is but one section of her miscellaneous works, and therefore has been categorized as miscellany rather than proper travel writing" (*Women's Mobility* 17). Extending the genre to include sections of miscellanies is only the beginning; as the TWDB shows below, expanding the definitions of travel writing, and of "by a woman," offers even greater nuance for how conceptions of travel writing circulated in both broader culture and among different audiences.

Counting and Identifying Women in the TWDB

To "count" travel writing "by" a "woman" involves criteria beyond the simple definition of travel writing. I highlight these words to emphasize how each creates important questions of context, counting, and ontology. First, the TWDB's labels still create a binary, though not the stereotypical M/F divide: instead, the TWDB categorizes titles as "by a woman" or "by unknown/unlabelled gender" and lists the roles that women held in these works, as well as those

---

[37] *Miscellanies* is held at the British Library, the Bodleian, the Houghton, and the Newberry.

[38] The full title is *Miscellanies in prose and verse, by the Honourable Lady Margaret Pennyman. Containing, I. Her late journey to Paris, giving an Account of the present State of the Court of France, and of all that is curious and remarkable in that famous City. II. Poems on several occasions, with Familiar Letters to a Friend.* The only subject heading connected to the copy on ECCO (pre-GPS) is "Pope, Alexander (1688–1744)" and, on ESTC, "Pope, Alexander, 1688–1744. Essay on man."

roles' public nature.[39] In total, the TWDB lists 179 titles, or 4% of the total, with a woman in some role. Depending on one's perspective, 179 titles may seem high, especially in comparison to Colbert's more selective 145.[40] At the same time, this number may seem comically low, especially when looking at other features: searching for authors with John in their name, for example, returns more results than the gendered corpus.[41]

Names are one of the main challenges of counting women, as women's names and influence were often elided from printed title pages (and therefore, institutional metadata). Colbert describes how, while some women such as Helen Maria Williams and Lady Morgan were "unapologetic and . . . adversarial" in their gendered authorial voices, "the majority of women represented in the database were conscious of the pitfalls of appearing too boldly in print and many . . . took evasive action in prefaces, introductions, notes to readers and advertisements, or by other paratextual means" (157; 158). Another popular prose genre of the period, the novel, provides additional context: Leah Orr, in her study of fiction from 1660 to 1750, found that 50% of the title pages list no author, and another 20% list only a tagline or a pseudonym, and as such are "functionally anonymous" ("Genre" 80). James Raven's findings are similar, where "over 80 percent of all new novel titles published between 1750 and 1790 were published anonymously,"

---

[39] See chapter 2 for encoding gender. In short, roles may include an author (full or in part), translator, or editor, and where their role is claimed, such as on the title page, in a peritext, in external research, and so on.

[40] Colbert counts 204 titles published from 1780 to 1840. Of those, 74 were published in the 1830s, with 6 published in 1830 (combining the online BTW-W (as of July 1, 2020) with his 2016 article, "British Women's Travel Writing, 1780–1840: Bibliographical Reflections"). Nine additional titles were published from 1700 to 1780. There are too many differences between the TWDB and BTW-W criteria to do a flat comparison; while we consulted many of the same sources, his definition of travel writing is relatively narrow (see appendix 1.2.1). Still, despite the different criteria of the TWDB and Colbert, the ratios of gendered titles are relatively comparable in the scope of the databases' larger corpora: Colbert found that "women accounted for only around 5% of travel books published in Britain and Ireland during [1780–1840]", and in the TWDB, the percentage is similar, at 5.9% during the same time frame ("Bibliographical Reflections 156). While most gendered TWDB titles, 159, came from bibliographical sources, another 20 came from the LCSH search, resulting in an additional 12 titles from HT and 8 from ECCO.

[41] This is also the case for William, Thomas, or James.

with rates in 1790 dropping to 62% anonymous, the 1810s to less than half, but then "returning by the late 1820s to about 80% of all annual new novel titles" ("Anonymous" 143).[42] These high levels of anonymity emphasize the importance of the methods of book history and bibliography, where scholars often go beyond the evidence presented on a title page. Some "anonymous" authors "were an open secret or very easily discoverable" for contemporary audiences, and modern metadata structures have often silently erased anonymity when an author is known (Raven "Anonymous" 145).[43] Works by minor authors or those of less interest to wider scholarship, however, may remain anonymous far longer. In other words, there are certainly more titles authored, edited, or translated by women within the TWDB, but their complex provenance remains unknown.

Within the TWDB, however, the database holds digital equivalents for 86.4% of the gendered titles in the TWDB-origbib. This rate is about 25% higher than the unknown gender corpus, of which only approximately 60% of titles have a digital match, indicating the impact of doing custom manual searches rather than generic scripted searches. Even with custom searching and using LCSH, there are 29 titles that I could not find. These titles may exist in other repositories, they may have different metadata than my search queries, or they may simply not be digitized.[44] For example, *The life and suprizing adventures of Friga Reveep* is held by only a few

---

[42] Other checklists of the eighteenth-century novel include (as cited in Raven "Anonymous") W H. McBurney, *A Check List of English Prose Fiction, 1700–39*; C. Beasley, *The Novels of the 1740s*; and Raven, *British Fiction 1750–1770: A Chronological Check-List of Prose Fiction Printed in Britain and Ireland*. See also Emily Friedman (356).

[43] Silent inclusions of names, though with some benefits, make searching for pseudo- or other forms of anonymity challenging.

[44] For example, the "letters of dr. johnson + piozzi, hester," which Cox dates as 1788, is probably the *Letters to and from the late Samuel Johnson, LL.D. . . .Published from the original Mss. in her possession*, by Hester Lynch Piozzi (where Piozzi is identified only in the title and not in the author field), but could also be the 1786 *Anecdotes of the late Samuel Johnson*, which does list Piozzi as the author.

libraries.[45] Other missing titles, such as Maria Riddell's 1792 *Voyages to the Madeira*, are cross-referenced by multiple bibliographical sources and held by several libraries, but even these are no guarantee of online accessibility.[46]

Similar to the larger TWDB corpus, the gendered corpus peaks in 1800, followed by a steep decline, and no titles at all published in 1803. The years preceding 1790 have low publication numbers, with generally less than ten titles published a decade, except for the 1720s, which features fiction by novelists such as Penelope Aubin, Eliza Haywood, and Delariviere Manley.[47] Of all the years in the TWDB, 1820 holds the most titles, at nine. This rise in publication numbers echoes the increase found by Colbert, despite the different method of identification (table 2, "Bibliographical Reflections" 160). These numbers may seem low in the context of "big data," with under a dozen titles per year compared to the sometimes several dozens of the TWDB more generally. However, both contemporary reviewers and modern scholars point to their significance. Katherine Turner states that "between 1770 and 1800 almost twenty travelogues by women were published, generally to critical acclaim," a "tiny. . . [yet] still a significant figure" (53), and she quotes the *Critical Review*'s observation in 1777 that "Letters of female travellers are now become not unusual productions" (qtd. in Turner 128).[48] What, then, can we learn by looking at the variety of these "productions"—both "unusual" and not?

---

[45] No scholarly criticism focuses on this Robinsonade. The ESTC lists the publication date as 1755, but Raven suggests an amendment to 1770 ("Publication" 38n13). Martha Pike Conant calls *Reveep* "a feeble imitation of *Robinson Crusoe*…with some resemblance to an oriental tale" (48).

[46] See appendix 3.5.

[47] See appendix 3.5. Orr also points to a small increase in travel fiction in the 1720s. She suggests that the lack of a sharp rise following the publication of *Robinson Crusoe* "indicates that it was perhaps not as widely imitated as some twentieth-century scholars have assumed" (*Novel Ventures* 189–90). See also Choi's discussions of these authors in *Women's Mobility*.

[48] Turner does not list which texts she includes.

Gendered Titles: Expanding the Bibliography

To demonstrate how this sample can add to our understanding of gender and travel writing, we can consider Colbert's list of only nine travel books by women between 1700 and 1780. The TWDB holds 37 books for the same time frame: 27 from bibliographical sources and 10 from LCSH searches. Most of the titles from bibliographical sources are fictional to some degree, and therefore not listed by Colbert or Robinson.[49] Instead, Gove and Cox, with their inclusion of imaginary or fictional travels, contribute them, expanding the "women of travel writing" to include Aubin, Manley, Haywood, Sarah Fielding, and Sarah Scott, all familiar names to those working with eighteenth-century novels. In fact, as Choi demonstrates, Aubin "created her own unique career path as a female author by specializing in 'lady's travel' in the 1720s: writing stories of global travel with strong female characters" (*Women's Mobility* 71).[50] These novelists use travel writing as a "niche marketing strategy" at a time when most travel writing was "exclusively about male travelers" (71).[51] Similarly, Jasmine Proteau argues that "women-authored guidebooks…have been historically overlooked in scholarship" and pose challenges for categorization.[52] It is especially significant, then, that the TWDB includes two guides by women, both listed by Cox. Mary Chandler's popular *A description of Bath: a poem* (1769501300) reached eight editions by 1767 and, according to Elizabeth Child, "helped to inaugurate the local

---

[49] Most of these titles also lack any travel-related LCSH.
[50] For more on Manley, Aubin, and other early eighteenth-century women who blended travel and fiction, see Choi's dissertation, *Women's Mobility, Travel, and Literary Representations in the Long Eighteenth Century*, and article, "'Every Jolt Will Squash Their Guts': Women's Stagecoach Travel in Delarivier Manley's *Letters*."
[51] Elizabeth A. Bohls similarly notes how women authors strategically contribute to aesthetic theory not through "the discourse, treatise, or inquiry" but "instead…genres more accessible to women, travel writing and the novel" (*Women* 3).
[52] James Buzard and Pieter François suggest that the divisions between travel guide and travel book begin to concretize after the rise of the Murray and Baedeker guidebooks of the 1830s (67; 73). See Betty Hagglund in "Gendering the Scottish Guidebook: Sarah Murray's *Companion and Useful Guide*" for context regarding guidebooks in the late eighteenth century.

demand for place-specific literary commodities" (158).[53] The privately printed *A short account, Of the principal seats and gardens, In and about Twickenham*, by Jael Henrietta Pye, was Pye's first publication before she moved on to other genres (0079200300).[54] Another notable addition to the TWDB, again via Cox, is Jane Squire's *A proposal to determine our longitude* (0206300500), a scientific pamphlet by the only woman to participate openly in debates surrounding the 1714 Longitude Act, though her work did not win an award.[55] By expanding the borders of travel writing beyond an "authentic" narrative of personal travel, the field broadens to consider how women were using the framework and discourses of travel within their own writing, even if their prose is about a fictional journey, or lacking a journey at all.

Searches by subject heading contribute another ten titles to the gendered corpus.[56] *A short tour* (1775, njp.32101073814558) by Elizabeth Percy, Duchess of Northumberland, is listed by no bibliographers except by Colbert, despite its title.[57] The TWDB also features an anonymous translation of the anonymously authored *The antidote; or an enquiry into the merits of a book, entitled A journey into Siberia . . . Translated into English by a Lady* (1772, 0180500700), a text which critiques Jean-Baptiste Chappe d'Auteroche's 1768 work and his unfavourable descriptions of the region.[58] Eliza Haywood's titles, however, do not feature her own journeys: *The dumb projector: being a surprizing account of a trip to Holland made by Mr. Duncan*

---

[53] See Child for further discussion of women authors in and around Bath, especially Sarah Fielding and Sarah Scott. David Shuttleton analyzes the poem as "feminocentric civic mapping" (447).
[54] ECCO spells her name as Joel-Henrietta Pye. As Colbert notes, "For many female authors, the travel book was a first, and sometimes last, foray into print" ("Bibliographical Reflections" 159).
[55] Elizabeth Johnson (1721–1800) also submitted an anonymous pamphlet, *The geography and astronomy of the created world, and of course the longitude: being the fourth book by the author of the Explanation of the vision to Ezekiel* (1785), to the Board of Longitude competition. She was unsuccessful.
[56] If the TWDB only used LCSH subheadings, eschewing bibliographical sources, the TWDB would hold only 98 of the 179 titles.
[57] Colbert lists the text in a footnote listing women's travel writing before 1780. *A short tour* is not in the BTW–Women's excerpt. LCSH: `description and travel`.
[58] LCSH: `description and travel`.

Campbell (1725, 0497000100) draws on the popular story of the deaf Scottish soothsayer.[59] *Memoirs of the court of Lilliput* (1727, mdp.39015078568626) details amorous activities "missing" from *Gulliver's Travels.*[60] Finally, *A letter from H---- G----g, Esq; One of the Gentlemen of the Bed-Chamber to the Young Chevalier* (1750, 0115501200) is a political parody for which Haywood was arrested for seditious libel.[61] Another piece of fiction is *The travels and surprising adventures of Mademoiselle De Leurich* (1316400701), published in 1751. This abridged piracy of *The entertaining travels and adventures of Mademoiselle de Richelieu* (1740) was gathered by the LCSH of `voyages, imaginary` and `travelers`; *Richelieu*, on the other hand, is tagged with `English fiction---18th century` on ESTC/ECCO. Moving away from novels, *A warning against the Quakers* (1708, 0278001200), a translation of the writings of Antoinette Bourignon (1616–1680), focuses on religion and the infamy of the French mystic's life.[62] Like many other travel writing authors, women translate, adapt, and lie about both their own travels and those of others.

Sometimes, the LCSH also add travel writing in other forms: *The Court miscellany, or, Ladies new magazine* (1765, 1671100501) is distinct from most of the TWDB because it is a periodical (and the only one attributed to a woman, though "Matilda Wentworth" is a pseudonym of the male editor, Hugh Kelly).[63] Some of its articles are travel-oriented, such as "A Letter from the celebrated Monsieur le Pays, when in England, during the Reign of Charles II. relative to

---

[59] LCSH: `description and travel`.
[60] LCSH: `voyages, imaginary`.
[61] LCSH: `description and travel`. While the ESTC, ECCO, and HT metadata for *A letter* lists Haywood as the likely author behind the pseudonym, David Brewer notes that "Few twentieth-century Haywood scholars have accepted [Alexander] Pope's attribution or even taken it particularly seriously" (217).
[62] LCSH: `antiquities`. For more on travel and travel writing by Quaker women, see Choi's overview (*Women's Mobility* 13–6).
[63] LCSH: `travel writing`. This LCSH is rare in the TWDB.

London and the Manner of Inhabitants" and "An extraordinary Instance of noble and disinterested Friendship in a British Sailor; extracted from Hughes's History of Barbados, on which coast the Accident happened." This issue points to another potential lacuna in the TWDB, as periodicals often excerpted and abridged travel accounts.[64] LCSH also add poetry to the TWDB: *An epistle from Oberea, Queen of Otaheite, to Joseph Banks* (1774, 0346100100) is one of a number of satires published about Banks' travels.[65] A more sincere lyric is *Needwood Forest* (1776, nyp.33433074860648).[66] Although this privately printed poem was advertised as written by Francis Mundy, Anna Seward wrote some of the shorter poems appended to the work, and, in her correspondence, claims to have written sections of the longer poem and contributed collaboratively to other parts (DeLucia "Local Poetry" 159). Like the larger corpus, then, the authors of this subset participate in multiple subgenres, ranging from fiction to parody to poetry.

This diversity of topic and form continues even after women begin to publish more travel writing after 1780, which is when Colbert begins his bibliography. As above, these inauthentic travels often take the form of fiction, memoirs (both fictional and real), and local histories. Of particular note is the increasing amount of juvenile and educational literature, especially in titles added via LCSH.[67] This "underdeveloped" (Colbert "Bibliographical Reflections" 164) or "almost entirely overlooked" (Grenby 171) subgenre connects travel writing with children's

---

[64] See JoEllen DeLucia's "Travel Writing and Mediation in the *Lady's Magazine*: Charting 'the meridian of female reading'" for more on the remediation of travel narratives into serial formats intended for women.
[65] LCSH: `account of the voyages` (which is an invalid LCSH in the ECCO data via McGill). While the LCSH was not displayed on the (now defunct) ECCO interface, searching for `account of the voyages` in the subject field still returned several editions of this title. For criticism and parodies of Banks, see Laura J. Rosenthal in *Infamous Commerce: Prostitution in Eighteenth-century British Literature and Culture* (193).
[66] LCSH: `description and travel`.
[67] In Colbert's BTW–Women's excerpt, "Travel novels, poetry and other forms of fiction have generally been excluded, although an exception has been made for travel storybooks for children that are based on an author's actual tour or contain abridgements of third-party tours (de facto collections)" (152–3).

writing, antiquarianism, and nation-building. Although Priscilla Wakefield's prolific catalogue has received attention, many similar authors in the TWDB, including Emily Taylor, Frances Thurtle Jamieson, and Isabella Jane Towers, remain absent not just from travel writing scholarship, but eighteenth-century studies more generally.[68] These titles and their forms, by turns or simultaneously "literary," "popular," and "heterodox," gesture to what further research can offer both to travel writing, but also studies of other genres of the long eighteenth century.

Gender on the Page: Claiming to be "By a Lady"

Teasing out the "evasive action" that Colbert describes, and how we know to count it, creates more context about how women were publicly participating in travel writing—and in fact, much of it was public. Nearly three-quarters of all the TWDB's gendered works identify the involvement of a woman on the title page in some fashion, whether by drawing on the claim to authenticity by using a name (ranging from the unknown Sophia Barnard to the famous Helen Maria Williams) or presenting a more modest anonymity such as "by a lady." These works range in genre, from abridgements for children to what scholars often classify as novels, from satire to "found" manuscripts. Some are authored by women, while others interject a digression in a woman's voice.[69] Similarly, while some translations reach for authority by avoiding any identification of the translator, others feature reputable translators, such as Anne Plumptre, on the title page.[70] Guides and topographical descriptions often attempt a more objective tone by avoiding identities, as does some juvenile literature. This anonymity makes identifying the roles

---

[68] Regarding Wakefield, see Ruth Graham, Bridget Hill, and Jacqueline M. Labbe.
[69] For example, see *The adventures of Abdalla... Intermix'd with the story of Mrs. Villars* by William Rufus Chetwood (1726, 0388100100).
[70] On Plumptre as translator, see A.E. Martin, Susan Pickford, and Glenn Hooper. For an overview of Plumptre's literary career including her novels, see Deborah McLeod's introduction to *Something New* (xii-xvi).

of women more challenging, especially as some of them require multiple tags.[71] By far the most common role in the TWDB data, however, is that of author, accounting for 77.6% of the roles.[72] And, because the TWDB indicates where this claim was made, these two features can be cross-referenced to compare the public nature of the gender claim. For authorship, then, 80% of these claims were made on the title page. In contrast, of the 17 translated titles, only 63% advertise that the translation was by a lady. Considering that even by the 1720s, "The way the publishers displayed the list[s] of books [in advertisements] indicates that the genre of 'lady's travel writing' became a recognizable category and genre identifier in the literary market," perhaps it is not surprising that those responsible for marketing texts would target this niche audience (Choi *Women's Mobility* 71n3).

Subjects in Gendered Titles

Even with caveats of the smaller size of the gendered corpus, LCSH still provide useful insight into common subjects, as well as their limitations in finding titles by and about women.[73] The most common main headings connect to geographic locations, especially those more accessible to women: France, Great Britain, Italy, and Europe. O'Loughlin suggests that attention to such locations is especially important for women's travel writing, since "the popularity of certain destinations at particular periods" reveals critical "connections—chronological or historiographical" in how women used the genre (8). Expanding to include subdivisions initially provides a familiar pattern, with `description and travel`, `early works to 1800`,

---

[71] For example, *Zelia in the desert. From the French. By the lady who translated ...* (1789, 0000300201). Sometimes, translators added their own content, as with *Translation of the letters of a Hindoo rajah ... to which is prefixed a preliminary dissertation ... by Eliza Hamilton* (1796, nyp.33433075896377)

[72] Adding partial authorship includes 82.6% of titles.

[73] The gendered subset is both a fraction of the size of the TWDB and holds fewer LCSH because of several titles sourced from Google Books. See appendix 3.7 for data on this section.

`france`, and `history` as the top four terms. Then, however, the subgenres that scholars identify as women's specialties appear, with `social life and customs` and `fiction` hovering around 2% of the overall LCSH; in the larger corpus, these LCSH are at 0.8% and 0.2%, respectively.[74] Where LCSH fail rather spectacularly, however, is in identifying texts that are about women. Several LCSH identify women directly in the topic, often designating their status as other or exceptional.[75] However, only six titles in the gendered TWDB hold the LCSH of `women`, `women travelers`, `women authors`, and `women spies`, and in the larger TWDB, only another four titles have women-related subjects.[76] Relying on LCSH, then, is not a viable way to find travel titles by or about women, at least in the eighteenth century.

This synthesis—mixing bibliographical sources, LCSH results, and book history—offers ways to extend the context of gendered travel writing. The broader the definition, of not only travel, but of "by a woman," increases the variety and number of titles, especially when considering fictional works. As Choi notes in her study of early modern and early eighteenth-century travel writing, focusing research after 1780 "establishes and perpetuates a narrative that women started to travel and consciously write about their experience only from the late eighteenth century" (*Women's Mobility* 7). Instead, these titles support Choi's thesis: that

---

[74] Ana Alacovska suggests that publishers encouraged this focus since they "approved female travel authorship only on gender-appropriate themes such as foreign customs, manners, food and dress" (134).
[75] See Elizabeth Hobart (112–3) and Kristin H. Gerhard et al. regarding "women" and LCSH.
[76] The gendered subset includes *Letters written by a Peruvian princess* (1748, 1122101500), *Facts. The female spy* (1783, 0726000300), *An account of two charity schools* (1800, 0611601200), *The traveller in Asia* (1817, uc1.31822038214573), *Narrative of the operations and recent discoveries within the pyramids* (1820, hvd.fl4sp7), and *Letters from the Caucasus and Georgia* (1823, NCCOF0257-C00000-B0123800). In the rest of the TWDB, the anonymous *The spy at Oxford and Cambridge* (1744, 0555201700) holds `women`, while three other titles hold other new terms: Richard Twiss's *A trip to Paris in July and August, 1792* (1793, nyp.33433069322661), which identifies the bookseller "Mrs. Harlow" on the title page (`women publishers`), and two ballads, *The Loyal-lovers garland* (1760, 0149102400, `women sailors`) and *The Blind beggar of Bednal Green* (1720, 1292703400, `man-woman relationships`, `young women`).

expanding the definition of travel writing is crucial for understanding not only women's travel writing, but also the genre of travel writing as a whole.[77] O'Loughlin suggests that "Travel writing of this period is distinguished by its popularity, diversity, curiosity, and experimental impulses" (7). Any broad claims about what travel writing by women is about requires attention to these "experimental impulses," whether in the recognizable form of a journey or a genre further afield. As with other writing by women that has "largely slipped outside considerations of eighteenth-century literary and cultural history" (O'Loughlin 7), more work remains to make travel writing more discoverable, accessible, and available for analysis.

Conclusion

The Travel Writing Database consists of metadata and textual data for 4,259 titles and 5,506 volumes sourced from ECCO, HT, NCCO, and Google Books. While approximately 60% of the titles from the TWDB-origbib's bibliographical sources have matches in the TWDB (2,830 titles), the TWDB also goes beyond bibliographies to include titles based on their LCSH, which adds another 1,429 titles. The 179 titles in the gendered corpus of the TWDB offers a further microcosm of how these different features interact: the number of titles by women increases as the definition of travel expands, as is shown by titles such as those drawn by Cox, Gove, and the many cataloguers responsible for the LCSH. These features and methods are not stable, however; if I were to attempt this study a few years from now, the LCSH metadata attached to ECCO titles may be inaccessible, and institutions are adding new content to HT daily. Still, the TWDB in its current iteration—influenced by the eighteenth-century print market, historical and modern

---

[77] This applies to eighteenth-century women's writing more generally; Mazella et al. "also argue for the critical value of recognizing variation, intermixing, and elaboration of existing genres in the newly expanded digital corpus that is now coming into view" (45).

collection practices, and the costs of digitization and accessibility—represents the largest corpus of digital travel writing spanning 1700–1830 today.

Chapter 4: Descriptions of Destinations

> He travels and expatiates, as the bee
> From flow'r to flow'r, so he from land to land;
> The manners, customs, policy of all
> Pay contribution to the store he gleans,
> He sucks intelligence in ev'ry clime,
> And spreads the honey of his deep research
> At his return, a rich repast for me.
> He travels and I too. I tread his deck,
> Ascend his topmast, through his peering eyes
> Discover countries, with a kindred heart
> Suffer his woes and share in his escapes,
> While fancy, like the finger of a clock,
> Runs the great circuit, and is still at home.
>
> —William Cowper, *The Task* (1785, IV: 107–119)[1]

> [O]ne is struck by the differences between the customs of his own country and those of other nations. Then penetrated by that difference, one seeks the reason for it, and there for him is the beginning of political, commercial, and moral combination.
>
> —John Adams, introduction to *Modern Voyages: containing a variety of useful and entertaining facts…* (1790, qtd. in Adams *Travellers* 13)

Introduction: Perspectives and Modelling

Over the course of this journey thus far, we have read the narratives of travellers before us, packed our bags with the required supplies, and eyed the routes of others with appreciation,

[1] See David Higgins regarding Cowper's poem and travel writing.

skepticism, or awe. But now, as we approach the summit, we can choose what viewpoints to stop at along the way.[2] In particular, this chapter directs our gaze at five critical elements of travel writing as a genre. First, it examines our current definitions of travel writing, revealing that even with heterogenous titles and different bibliographical guides, travel writing does cohere around particular words, especially words directly related to travelling. In fact, this "travel signal" is so strong that computational models can often distinguish travel texts against random samples nine times out of ten. Travel writing is not a monolith, however, and this chapter's application of subject headings reveals divisions and connections among the significant travel subgenres of imaginary voyages, poetry, and history. Looming over this vista, however, is Great Britain's growing empire and its influence on descriptions of people and landscapes. Participating in all these subgenres are women authors, whom models suggest often write similarly to their male and ungendered colleagues. Finally, shifting back to a diachronic survey of the whole field, this chapter demonstrates that even with the "discovery" of new peoples, countries, and philosophies, the distinctive features of travel writing remain relatively stable over the long eighteenth century.

My metaphor of vantage points is, of course, not accidental; we could explore the TWDB's metadata and textual data in many ways, such as topic modelling or location mapping.[3] But to directly address the TWDB's questions about genre, definition, and the loose borders of travel writing in the long eighteenth century, the remainder of this project relies on Ted

---

[2] Guides to picturesque viewpoints proliferated in the eighteenth century. For example, Thomas West's *Guide to the Lakes* (1778) describes "all the select stations, and points of view, noticed by those who have made the tour of the lakes, verified by repeated observations, with remarks on the principal objects as they appear viewed from different stations" (3). As David Stewart notes, "When people in the Romantic period viewed a landscape, they did so in the context of debates and discussions about *how* to view land, and *which* lands were worth viewing" (160).

[3] For example, the *Grand Tour Project* at Stanford maps travels in Italy, and the *Grub Street Project* maps both printing houses and travels in literature within London. Maps—digital or otherwise—are also models. As C. Board notes, "no map can perfectly depict reality, but in not doing so it is all the more useful" (671).

Underwood's approach of "perspectival modelling," which uses predictive models to examine the mutability of different categories.[4] First, we train a statistical model on two sets of data (in our case, sets of travel writing and non-travel writing).[5] During this training, an algorithm "reads" each text, counting the frequencies of features, especially words.[6] Then, it combines this data to create a representation or "model" of travel writing. Critically, these models are not exact copies of texts or "mimetic representations of the world" (Piper "Think Small" 652).[7] Instead, models quantify "a relation between variables": a label, or predicted variable, of "travel writing," and the predictors, or "whatever linguistic features turn out to signal genre" based on the training set (Underwood *DH* 20). After training the model on hundreds of examples of travel writing, we show this predictive model a new text and the model "predicts" the likelihood of it belonging

---

[4] For more on the history and conception of models, especially in digital humanities, see Andrew Piper in *Enumerations* (9–12) and "Think Small: On Literary Modeling," and Katherine Bode in "Why You Can't Model Away Bias" (100). For a full description of Underwood's approach, see *Distant Horizons* (*DH*), especially appendices A (Data) and B (Method). For code, see the relevant GitHub repository. Reviews of *DH*, including methodology, are generally positive. See, for example, Dan Sinykin, Tess McNulty, Daniel Rosenberg, and Alison Booth, and applications of the methodologies in *DH* by Andrew Goldstone, who cheekily remarks, "I am docking [Underwood and Jordan Seller's essay "How Quickly Do Literary Standards Change?", which influenced *DH*] Reproducibility Grade down from 100 to 99.95," and Jonathan Goodwin, who extends the work on science fiction to look more closely at Darko Suvin's lists of science fiction. Critiques of *DH* include that of Nan Z. Da, whose main critique of "The Life Cycles of Genres" (the foundation for *DH* chapter 2 and therefore this chapter's methodology), is based in part on a belatedly acknowledged error regarding how the random sample handles publication dates. The best review of *DH* addressing both epistemology and methodology is Bode's "Why You Can't Model Away Bias." I am grateful to Underwood for not only sharing his code and data, but also for answering several queries and supporting this project over the years.

[5] Underwood relies primarily on regularized logistic regression. See appendix B (Methods) in *DH*.

[6] A token is "an instance of a sequence of characters," most commonly words divided by spaces (Manning et al.). In preparation for modelling, each plaintext file was run through Underwood's tokenizer script, which tokenizes the text and creates a set of normalized features. The primary features of concern are word frequencies; because of the TWDB's messy OCR, I removed punctuation, but left other elements such as capitalization and bigrams, though they rarely come up as the most influential features.

[7] Piper suggests that "The question is not whether a model perfectly represents the world, but as [philosopher of science] Gabriele Contessa has argued, the ways in which models facilitate potentially valid inferences about the world: 'Faithful epistemic representation is a matter of degree. A vehicle does not need to be a completely faithful representation of its target in order to be an epistemic representation of it'" (18). Richard Jean So similarly draws on historian Mary Morgan to observe that "statistical models are not just summaries of or reports about data, they are mechanisms with which individuals reason and think" ("All Models Are Wrong" 670). Underwood also discusses the interdisciplinary aspect of models and their history in other fields (*DH* 21–2, 143–70). See also Bode's discussion of models in "Why You Can't Model Away Bias" (99–101).

with the other travel titles.[8] By looking at the model's accuracy, or how many volumes the model can identify correctly as travel writing, we can assess whether travel writing contains distinct features that set these titles apart from other texts of the eighteenth century.[9] Instead of relying on a single variable, such as the truth of a journey, a predictive model instead considers patterns across thousands of word frequencies.

A simple list of word frequencies may not seem to be a useful representation of a single text, never mind an entire complex genre. A computational model, however, can compare hundreds of these lists to create multidimensional relationships. In one such application of predictive modelling, focusing on science fiction and fantasy, Underwood describes how "a model might simply count words, treating the relative frequency of each word as a dimension like height or width. Since the vocabulary of fiction contains thousands of words, this will produce a space with thousands of dimensions" ("Machine Learning" 98).[10] Underwood is frank about the accuracy—past, current, and future—of this approach:

> The methods I use to train models in this chapter (and throughout [*Distant Horizons*]) are commonly used for text classification: regularized logistic regression on several thousand features, mostly the frequencies of words in the texts . . . I cannot prove that no better methods will ever exist. All I can say is that lexical models capture human judgments about genre rather well (accuracy above 90% is not uncommon), and researchers have been trying for decades to find a better strategy, without much success. (43)

---

[8] A single "title" may have multiple "volumes." The models rely on these volumes, meaning that different volumes from the same title may have very different predictions.

[9] A high accuracy means that travel writing does have distinct features, since the model can identify the travel writing volume compared to other volumes. On the other hand, an accuracy of 50% means the model cannot tell the difference between the travel and non-travel texts.

[10] For an example of applying predictive models to poetry, see Hoyt Long and Richard Jean So in "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning."

These predictive models are revealing on their own, but Underwood's "perspectival" modelling then takes the approach a step farther. Instead of relying on just one predictive model, we can create as many models as we have definitions or categories, and then compare their predictions.[11] The more similar their predictions, the more similar their representations, or perspectives, of travel writing. Our assessment of travel writing thus becomes not just about travel writing itself, but also a comparison of how bibliographers, cataloguers, and other "social perspectives" have defined and collected travel writing and its many subgenres—and the "relative strengths of the boundaries" between these definitions (*DH* xvi; 43).

As an example, we can compare two perspectives on *Millenium Hall* (1762) by Sarah Scott, a novel featuring travel to a utopian society. A model trained on true accounts of journeys by women is skeptical of *Millenium Hall* fits, with predictions of under 30%. A model trained on fictional travel writing, however, is confident that the bluestocking novel fits, at over 96%.[12] In this instance, the model supports our scholarly traditions. However, for other travel novels, this situation may be reversed, complicating analyses that rely only on "real" travel writing.[13] In addition to counting and statistical analysis most easily achieved with computers, this method thus works as "a memory-wiping flashbulb that allows us to strategically erase our knowledge of the future or past as needed. The computer knows nothing about literary history," such as theories of gender, the novel, or Romanticism, but "models only the evidence we give it" (*DH* 36). Thus, the algorithms' "readings" cannot be influenced by modern arguments about literary or historical impact, such as discourse about the importance of Laurence Sterne's *Sentimental*

---

[11] In *DH*, chapter 2 explores genre, especially the Gothic, detective fiction, and science fiction, but later chapters examine prestige and character gender.
[12] Literary scholarship on *Millenium Hall* also focuses on it as a novel.
[13] See the discussion of Penelope Aubin later in this chapter.

Journey, published in 1768.[14]

In addition to developing our understanding of travel writing and its hybrid (sub)genre(s), then, this chapter also examines whether this approach, used by Underwood on genres such as detective fiction, the Gothic, science fiction, and fantasy, against other random fiction, can work on the heterogenous (both in subject and style) genre of travel.[15] The tests in this chapter demonstrate that these methods, developed for and tested on relatively contemporary genres, can be useful for older genres and texts. This chapter is a direct answer to Underwood's call for his method "to be tested from many angles. The point of sharing code and data is to help readers test those claims in a consequential way" (183). The "claims" tested in the rest of this chapter include Underwood's methods, as well as our received literary histories about travel writing, demonstrating the benefit of new angles and contexts for considering both code and conventions.

Over the course of this chapter, each section explores a recurring topic in travel writing scholarship, relying on the diverse perspectives encoded within the Travel Writing Database (TWDB). After demonstrating that travel writing does have distinctive features, I compare the fourteen definitions of travel writing in the TWDB. Then, using Library of Congress subject headings (LCSH), I reveal the complex relationships of travel writings' subgenres, especially fiction, history, poetry, and texts describing different locations. I also apply this perspectival

---

[14] We hear, for example, that "Developments in English travel writing were catalysed by [*Sentimental Journey*]" (Korte 56), that "The phrase quickly became a kind of shibboleth for authors" (Chandler 11), and that the novel "was pivotal both in the history of tourist practice and in the history of travel writing" (Lamb 27). For adaptations of *Sentimental Journey*, see "Sentimental Journeys?: Adaptations of Sterne's Travel Narratives" by Mary-Céline Newbould in her *Adaptations of Laurence Sterne's Fiction Sterneana, 1760–1840* (35–74).

[15] In addition to *DH* for detective fiction, the Gothic, and science fiction, see "Machine Learning and Human Perspective" for comparisons of science fiction and fantasy. One prominent difference is that Underwood tests his fiction categories against other fiction, but I test the TWDB against both fiction and nonfiction in Eighteenth Century Collections Online (ECCO) or HathiTrust (HT). This approach is necessary for travel writing, since it is so diverse in its definitions; a heterogenous genre, mixing fiction and nonfiction, asks for a similarly diverse contrast set. See appendix 2.2 for insight into the contents of ECCO and HT.

approach to compare chronological subsets, revealing that a model trained on texts published before 1801 can still predict texts with a similar accuracy 20 years later, despite the momentum of the Romantic movement. However, before we zoom in to subgenre, author, or decade, it may be useful to ask the most basic question: can travel writing, with all its "hybridity in terms of text modes and styles" (Korte 15) and seeming lack of borders, be discerned against other writing of the period? After all, for some critics, travel writing is "a genre composed of other genres" or "not a genre but a variety of texts, both predominantly fictional and non-fictional, whose main theme is travel" (Campbell 6; Borm 13).[16] Let us explore whether models can recognize this literature, despite (or, perhaps, because of) its "thousand forms and faces" (Batten 281).[17]

Travel Writing: The Big Picture

We can begin with the most obvious question: can predictive modelling, based on word frequencies, correctly identify travel texts from the TWDB against a random collection of fiction and nonfiction published between 1701 and 1830? In short, yes: models can predict a TWDB text correctly about 87.3% of the time.[18] If there were no difference between travel writing and a random sample, then we would expect the accuracy of the model to be around 50%, which is what happens if I train the model on a random sample of texts and then compare those "random" texts to the rest of a random sample.[19] When comparing the TWDB's travel writing corpus against random samples, however, the model finds it much easier to discern patterns. To extend Jonathan Raban's metaphor, if travel texts tend to "end up in the same bed," they do seem to

---

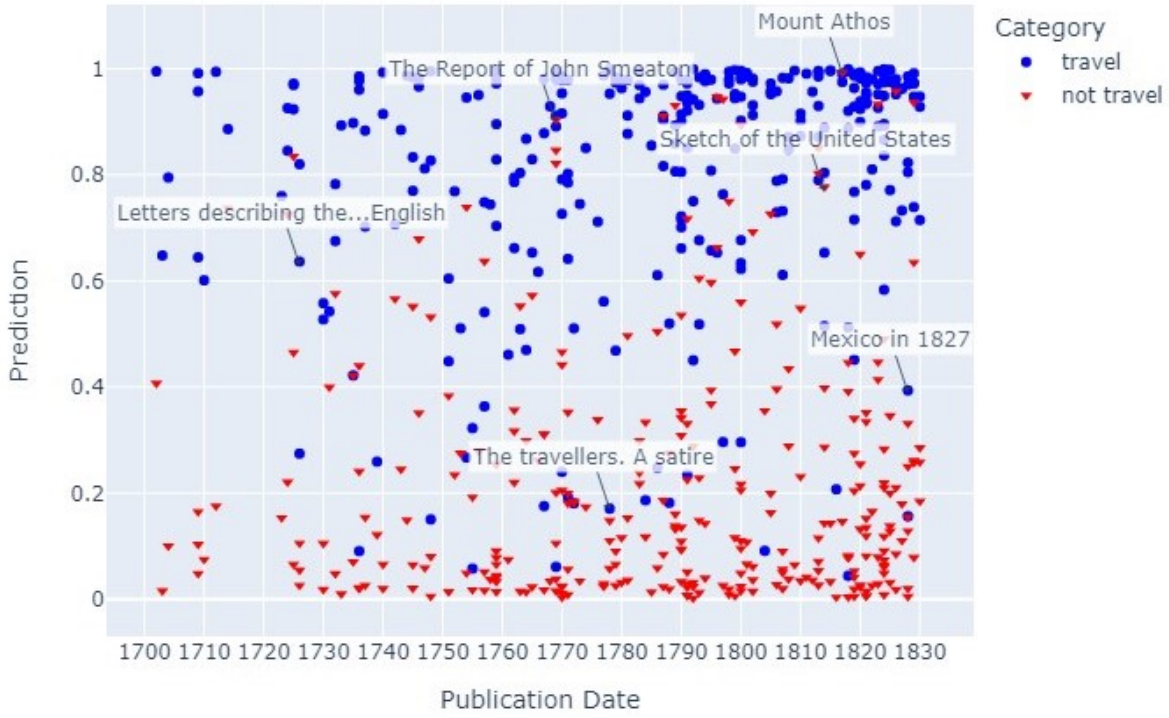
[16] Underwood quotes similar claims by Mark Bould and Sherryl Vint: "there is no such thing as SF—but instead multiple and constantly shifting ways of producing, marketing, distributing, consuming and understanding texts as SF" (qtd. in "Machine" 96–7).
[17] For a full description of travel writing, its definitions, and its history, see the introduction.
[18] Compare to Underwood's results for detective fiction at 93.4% (*DH* 49), the Gothic at 81% (57), and science fiction at 90.6% (58).
[19] For more on the meaning of accuracy, see appendix 4.1.

have a type. We may find this high number surprising; after all, we have travel titles as varied as Wollstonecraft's *Six Weeks Tour* (1817, NCCOF0257-C00000-B0037400), *An account of the culture of potatoes in Ireland* (1796, 0965600600), and Eliza Haywood's *Philidore and Placentia: or, l'amour trop delicat* (1727, 1243600302) in the same corpus.[20] Many of the random titles that end up predicted as travel writing—that is, the false positives—are also what we might consider travel writing, but were not labelled as such by the subject headings or bibliographers (see figure 4.1 below as a sample). For example, the models often expect the ostensibly non-travel titles of *The ruins of Gour described* (1817, njp.32101055309627), *The German Gil Blas; or, the adventures of Peter Claus* (1793, 0201000401), and *Mount Athos: An Account of the Monastic Institutions and the Libraries on the Holy Mountain* (1817, mdp.39015070444032), to be travel writing, and human eyes might agree, especially after comparing with the metadata of other titles in the TWDB. Of course, there are surprises in the models too: chess manuals may also be ranked highly, since their contents are mostly letters and periods (akin to tables in navigational charts or other appendices in travel writing), with few non-travel words to dissuade the model.[21]

---

[20] For titles with multiple volumes, I list the earliest one in the TWDB.
[21] Other texts with many non-Latin characters that get OCRed as single letters and periods, such as *Muntakhbāt-i-hindī or, Selections in Hindustani* (umn.319510021811833), may also rank highly.

Figure 4.1: Model of `alltravel` in the TWDB[22]



Although we must be cautious in putting any importance on single words—after all, these features are in relation in a space with more dimensions than an ordered list can indicate—there are some themes that stand out among the top influencers. As table 4.1 shows below, words directly related to travel, such as variations on visit, travel, and departure, are common at the top of the list: according to the model, travel writing is a genre that explicitly discusses its own main action, connecting with Jan Borm's suggestion that travel writing is any text with the "main theme [of] travel" (13).

[22] This sample `alltravel` model contains 300 travel and 300 nontravel volumes from 1701–1830.

Table 4.1: Thirty Most Influential Terms of the TWDB Travel Model[23]

traveller	country	climate
inhabitants	women	undoubtedly
governor	inhabited	barren
tolerable	entrance	fo
sort	english	arrival
descent	farther	priests
travelling	province	strangers
plenty	countries	island
arrive	houses	scarce
curiosity	abundance	these

We can also attend to another topic common at the top of the list: words of business and administration, such as governor, commerce, council, company, merchant, and trade. In general, these models reflect Britain's growing mercantile and expansionist interests. These interests connect not just with Britain's increasing political and economic influence in North America, the Pacific, Asia, and Africa, but also with the responsibility often implicit, if not explicit, in travel writing: in understanding how other nations and economies worked, travellers could bring resources home to the burgeoning British Empire.[24] The other end of the spectrum—the terms that predict the random sample—do not have any immediately discernible patterns, perhaps because of the variety of texts used in the sample.

The TWBD's subdivisions based on source have similar results. Models of only titles identified by bibliographers are about as accurate, at 87.7%, as the entire TWDB at 87.3%. If relying on titles with subject headings holding travel-related LCSH, however, the model can

[23] The model includes the entire TWDB, from 1701–1830. This example is from the second `alltravel` model.
[24] Samuel Johnson commented that "Every nation has something peculiar in its manufactures, its works of genius, its medicines, its agriculture, its customs and its policy. He only is a useful traveller, who brings home something by which his country may be benefited; who procures some supply of want, or some mitigation of evil, which may enable his readers to compare their condition with that of others, to improve it whenever it is worse, and whenever it is better to enjoy it" (*Idler* 97 1760, Yale *Works* 2:300).

distinguish them 90% of the time—slightly more accurate than if we only modeled titles from our original bibliographies or from the entire TWDB.[25] The LCSH model can also often recognize titles in bibliographical models, demonstrating that the definitions of bibliographers and cataloguers are much more similar than they are different—at least as far as the models are concerned.[26] For scholars looking for alternative ways to collect travel-writing titles beyond typing up bibliographies, then, these models suggest that the collection's features will be similar to those deriving from many travel bibliographies. The perspectives of bibliographers and cataloguers are broadly aligned.

These findings on their own are significant. Definitions of travel writing differ significantly, and as a genre travel writing draws from and influences a variety of others, including geographical texts, religious tracts, and novels. Nevertheless, titles identified as travel writing are still cohesive enough that a model based on these heterogenous texts can identify them nine times out of ten. As we will see below, models trained on different subsets of travel writing, organized by features as diverse as bibliographer, subgenre, claimed author gender, and location, do diverge in interesting ways when they look at each other. But even if these subsets may have similarities with, or even be considered part of, other genres of the eighteenth century, there is something distinct enough about their travel writing features that models can almost always identify them, at least 85%, if not 90%, of the time.

---

[25] The travel LCSH include (`travel*`, `voyage*`, `discover*`, `explor*`, `guidebook*`, `antiquit*`). Models of volumes harvested only via LCSH, with no listing in a bibliographical source, average 88.3%.

[26] The travel tags model's mutual recognition with other models averages 84%, which is on par or above the average mutual recognition that other bibliographical sources have for each other. In other words, on average, the LCSH tags and bibliographical sources can recognize each other more easily than the sources can recognize each other.

From the Perspective of Bibliographers

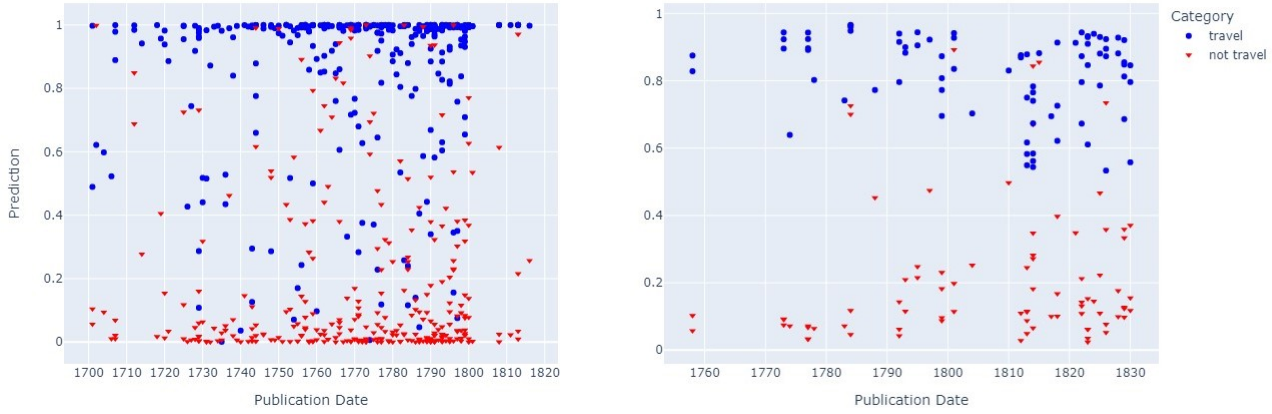
Just as we started with bibliographers for the construction of the TWDB and its definitions of travel, we can start with them for our analysis. As table 4.2 shows, models of individual bibliographical sources are generally more accurate than the overall TWDB, implying that as their definitions of travel writing get more concentrated, so too do the features that influence the models. The smallest travel writing bibliography, NCCO's "Travel Narratives and Personal Reminiscences" with only ten volumes, often has the lowest accuracy, sometimes as low as 70%; accuracy here depends greatly on what volumes end up in the non-travel sample.[27] The next step in considering accuracy is the most general TWDB designations: all the titles listed in bibliographical sources, Cox's bibliography, and everything included in the TWDB. I have made much of Cox's heterogeneity, and we see my intuitions confirmed in these results: Cox is as challenging to predict as the overall TWDB. This might not be surprising, since Cox's diverse definitions of travel writing, ranging from fen drainage to explorations of the Arctic, makes up much of our original bibliography and the TWDB in general.[28] After the more general categories, the accuracy slowly climbs into the 90s, often with TEE (*Travels, Explorations and Empires, 1770–1835*) at the top of the list.

---

[27] The related TWDB tag is `ncco_travelnarr`.

[28] In the final metadata used for modelling, there are 962 volumes cross-referenced with both Cox and a travel LCSH, and 1,280 Cox volumes with no travel LCSH (only 46 of which have no LCSH at all). For travel LCSH and no Cox, there are 1,956 volumes, but 68% of those were published after 1800, which is when Cox mostly stops cataloguing travel writing. See Cox in appendix 1.2 for more.

Figure 4.2: Sample Predictive Models of Cox and TEE



Since when "modeling a homogenous group of works, accuracy ordinarily increases as one gathers more data," these results confirm travel writing's heterogeneity (55).[29] These thematic differences are evident in the top terms for these models: from the perspective of Andrews' picturesque bibliography, for example, the view is one of stones, hills, and mountains, with a dash of pastoral families and houses, while TEE often pays attention to labour and resource extraction, commonly through slavery. Choice of travel writing bibliography, therefore, will have an impact on the user's resulting perspective of travel writing, whether that of a computational model or a human scholar.

[29] In his experiments with the Gothic, Underwood found larger models combining ghost stories, horror, and other such titles in the Gothic tradition resulted in lower accuracies. He notes that "The growing shakiness of this Gothic edifice as samples grow larger betrays a weakness somewhere in its foundation," especially compared to detective fiction's increased accuracy, where combining corpora is "a compromise that 'levels upward'" (*DH* 55; 49).

Table 4.2: Accuracy of Bibliographic Models

Model	Average Best Accuracy[30]	Total Volumes[31]
ncco_travelnarr	81.7	10
alltravel (entire TWDB)	87.3	300
cox	87.3	300
origbib (volumes from a bibliographical source)	87.7	300
lcsh_traveltag (volumes containing a travel LCSH)	90.0	300
irishmcveagh	91.3	55
btw_w	91.7	152
leask	92.7	96
btw_europe	93.0	74
andrews	93.0	63
murray	93.7	118
ncco_c19trav	93.7	180
gove	94.0	85
robinson_w	94.0	45
bdanth	94.0	60
brynmawr	94.3	243
tee	95.7	81

Bibliographical Sources' Perspectives of Each Other

The average best accuracies for each source in table 4.2 are telling insofar as they can indicate how much the models can tell the bibliography-based models apart from the random sample. But if thinking about comparisons of definitions and categories, we must also consider similarities between the different perspectives (or conceptions) of travel writing. To measure this, we

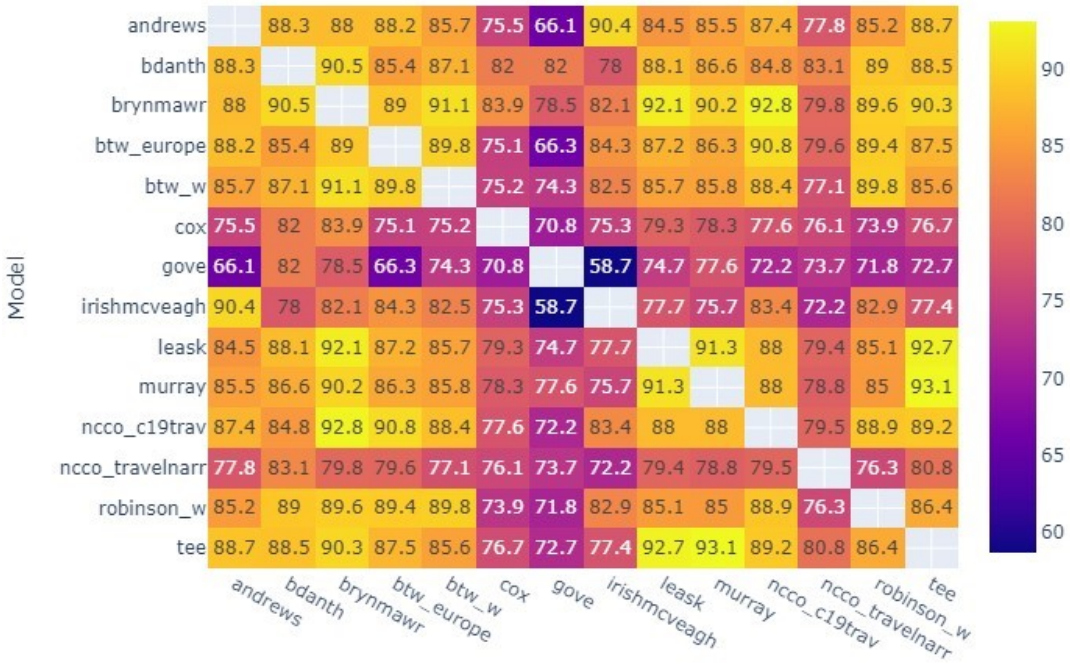
[30] The average best accuracy for individual models is always of three modelling runs.
[31] The cap on the training size of the model is 300 volumes. In the categories where there are more than 300, the model selects volumes randomly. Different models may also include the same volume. For example, Ann Radcliffe's *A Journey Made in the Summer of 1794* (1795, 0593000100) is listed in six bibliographical models: Andrews, Bohls and Duncan, the British Travel Writing–Women excerpt, Bryn Mawr, Cox, and Robinson.

continue testing Underwood's notion of perspectival modelling, shifting to his "method of mutual recognition," where we "ask whether the differences defining one genre are parallel to the differences that define another" (47). By taking Model A and asking it to predict which Model B volumes are travel writing, and vice versa, and then looking at the number of correct predictions, we can see if the models rely on the same combinations of features. The result is expressed as a percentage: if the mutual recognition is high, especially above 90%, then they have similar distinctive features. The closer the percentage gets to 50%, or only as good as a random guess, then the more divergent the models are in their different "perceptions" of what travel writing is. In the rest of this chapter, I often return to this method, not just for interpreting the differences between sources, but also across subgenres, locations, and time periods. As Underwood notes, "When I need to decide whether two models of genre are similar…this is the test I will trust" (48).

So just how similar are the models of these disparate bibliographical sources, with their different focal points ranging from gender of the author to location to the veracity of the travel? If we take an overall average of each bibliographical source's perspective on the others, and then take an average of those, the resulting 82.6% is far above a random choice. However, relying only on this average flattens out more interesting and useful comparisons, especially regarding connections between sources not just in their definitions, but also their respective focuses.[32]

[32] For full lists of mutual recognition rates, see appendix 4.2.

Figure 4.3: Mutual Recognition of Bibliographical Sources



For example, can the model of picturesque travel writing in Andrews correctly identify McVeagh's Irish travels, and vice versa? Yes, nine times out of ten—suggesting similarities in how travel writers handled the picturesque and Ireland.[33] Sometimes, high similarities may be due to volumes being in both models (and therefore influencing their perspective on what travel writing "looks like") but for McVeagh and Andrews, there are no such titles: the collections are distinct, making the high prediction rate even more notable. When one compares McVeagh's Irish travels to other perspectives, however, the mutual recognition drops, often below 80%, and, notably, to below 60% when comparing McVeagh's Irish travels and Gove's imaginary travels.

[33] Andrews can predict McVeagh's Irish bibliography 87.3% on average, and McVeagh Andrews, 93.2%.

Table 4.3: Bibliographical Mutual Recognition with McVeagh

Model	Average Mutual Recognition with McVeagh (irishmcveagh)
gove	58.7
ncco_travelnarr	72.2
cox	75.3
murray	75.7
tee	77.4
leask	77.7
bdanth	78.0
brynmawr	82.1
btw_w	82.5
robinson_w	82.9
ncco_c19trav	83.4
btw_europe	84.3
andrews	90.4

We might be reassured by the implication that Ireland is indeed a real place. But these differences between McVeagh's Irish travels, Gove's imaginary travels, and Andrews' picturesque travels also emphasize how the relationship of Great Britain and Ireland differs from those of many other locales, including those that Great Britain's empire was trying to swallow. For example, McVeagh's mutual prediction rates for `description and travel` writing about Great Britain and Ireland are above 90%, with Europe at 86.6%, North America at 80.8%, and all other locations below 77.2%.[34] Models thus draw our attention to features in Irish travel writing connected to poverty, industry, labour, and improvement, and they reinforce scholars' attention to Ireland's important connections to the development of the picturesque.[35]

[34] See appendix 4.2.
[35] For example, see Finola O'Kane's *Ireland and the Picturesque: Design, Landscape Painting and Tourism 1700–1840.*

There are interesting patterns in other sources, as well; for example, Cox's heterogeneity also results in low mutual prediction rates, often in the 70–80% range, and Cox's model is consistently better at recognizing other models compared to when these models look at Cox.[36] These comparisons, however, also offer a way to explore that assertion of some scholars: that to truly be travel writing, the first part—the travelling—must have actually taken place in person, rather than from the so-called armchair. Just how different are voyages of the imagination to voyages of the body? Gove's list of imaginary travels is the only bibliographical source focusing on such titles, and it also stands out as the most unrecognizable, with other models typically predicting Gove's titles correctly only 50–70% of the time.[37] What then, can we discover if we look more closely at the networks of sources and terms to explore this subgenre of writing?

Subject Headings

Armchair Travellers vs Real Travellers: Comparing Perspectives

To explore the imaginary, let us take a few different perspectives using our various labels. Gove's 1941 *The Imaginary Voyage in Prose Fiction: A History of its Criticism and a Guide for its Study, with an Annotated Check List of 215 Imaginary Voyages from 1700 to 1800* remains the primary work on the history of the term. Imaginary voyages are, at their most basic, "a narrative of a voyage performed in the imagination," a definition which is, in turn, "amplified and interpreted" by Gove.[38] As Gove notes, "I know of no work before the nineteenth century in any language in which an author so informs his readers by specific use on the title page of the

---

[36] For example, Cox is 83% accurate when looking at McVeagh's 110 volumes (55 travel volumes and 55 nontravel), but when McVeagh tries to identify Cox's 600 volumes (300 travel), the model is accurate only 73.9% of the time. The chart above takes the average of all the volumes added together (75.3%), rather than the average of the models' perceptions of each other (78.5%), following Underwood's method.

[37] Murray, the Bohls and Duncan anthology, and Cox models correctly predict Gove, on average, 72.2, 78.6, and 79.3% of the time.

[38] See 175–8 for his full description.

phrase imaginary voyage, although there is no mistaking the fictional implications of clearly impossible voyages and of transparent pseudonyms" (6). All of our perspectives, then, are even more interpretive than other more self-proclaimed subgenres of travel writing, such as the voyage or epistles. For example, Gove includes titles ranging from obviously false to attempting truth, such as *Iter lunare: or, a voyage to the moon* (1703, 0258401300), the satirical *Gulliver's Travels* (1726, 0647100501), and Christian Friedrich Damberger's *Travels Through the Interior of Africa* (1801, NCCOF0257-C00000-B0093800).[39] Gove thus describes the imaginary voyage as "an organic, shifting division of fiction, recognizable, but indefinable as a static, fixed, and exclusive genre" (viii). The only other bibliographical source to explicitly identify imaginary voyages is Cox, who includes a category of "Fictitious" travel writing, though he does not describe the requirements for this category.[40] However, in addition to relying on Gove and Cox, we can also turn to subject headings: a few dozen volumes carry `voyages, imaginary`, with over two hundred TWDB volumes tagged with some variation of `fiction`[41] (and some, but not all, carry both).[42] Gove expresses wariness of subject headings as a way of finding and categorizing imaginary voyages, since they only "reveal the wide range of this classification but will not help to define it, for the librarian is not concerned with establishing mutually exclusive

---

[39] *Travels Through the Interior of Africa*, although presented as true, was both extremely popular and discovered by the publisher to be false shortly after publication. For more on the history and an excerpt, see *TEE*, Part II, volume 5, 93–158.

[40] As William H. Sherman notes, Cox's "coverage of "Fictitious Travels" (particularly in plays and poems) is especially incomplete" (21). Some known fictitious titles, such as *Millenium Hall*, are in other sections. See Cox (2: 470–8).

[41] The term `literature` in subject headings catches a broad range of terms, including `controversial literature`, `juvenile literature`, `art and literature`, and so on. For this reason, I relied on the more specific category of fiction, despite its known issues (for example, see Gregg, "1748: Fiction in the Database").

[42] Only 17 volumes, including *Gulliver's Travels* and several keys to Swift's satire.

subject headings" (5).[43] By comparing these models, we can thus compare Gove's definition with those of cataloguers, as well as with bibliographers of "real" travel writing, to reveal differences between these "travellers" and "travel liars" (Adams) and those who fall in between.

The models that rely more on the `voyages, imaginary` LCSH (`lcsh_imaginary`) do tend to be less accurate, with the largest mixed bag of imaginary and fiction (`imagfiction`), which includes all of our relevant subject headings and bibliographies—162 volumes total—tending to be the least accurate. Gove's definition and tags connected to fiction, however, fare rather well when compared to the rest of the corpus.

Table 4.4: Accuracy and Mutual Recognition of Imaginary and Fiction Models

Model	Accuracy Against Random Sample	fictitious	gove	lcsh_fiction	lcsh_imaginary	nontravel_fiction	imagfiction
fictitious	89.7		85.4	81.1	84.6	76.7	85.3
gove	94.0	85.4		92.3	90.6	89.6	90.7
lcsh_fiction	91.0	81.1	92.3		88.3	88.8	90.9
lcsh_imaginary	91.0	84.6	90.6	88.3		83.9	92.9
nontravel_fiction	92.7	76.7	89.6	88.8	83.9		86.0
imagfiction	91.3	85.3	90.7	90.9	92.9	86.0	

These models are, predictably, also quite good at recognizing the titles—some of which are shared—in the other imaginary or fictional models. An exception is Cox's list of fictitious voyages and travels, whose average mutual recognition is never above 86% for the other models.

[43] Gove also cites a conversation with the Chief Cataloguer of the New York Public Library which indicates that the institution had abandoned using "voyages, imaginary" as an LCSH. However, the NYPL catalogue still has many contemporary works with this subject heading. See the Library of Congress for more information ("Voyages, Imaginary").

And some titles are often miscategorized in interesting ways: the model is uncertain about various trips to the moon, and the model often suggests that other pieces of fiction fit in with the travel fiction of the TWDB, such as *The vicar of Lansdowne; or, country quarters: a tale* (1789, 0192500202), *The astonishing history, and adventures, of Miss Betsey Warwick, the female rambler* (1795, 1030701100), and *The recluse. A Fragment* (1787, 0149300800). Earlier, we asked how models would handle Sarah Scott's *A Description of Millenium Hall and the Country Adjacent*, a 1762 novel about a female utopia which is not listed in any of the imaginary or fiction datasets listed above.[44] Once the models look at Scott's novel, however, they are very confident: Cox's fictitious model, the LCSH imaginary voyages model, and the overall combination `imagfiction` model are all over 95% certain that *Millenium Hall* fits with their models. Meanwhile, Gove and the LCSH travel fiction models drop their predictions to around 86%.[45] In contrast, other models that tend to focus on real travel or a wide breadth of travel, such as the entire `alltravel` TWDB model, the model of travel tags, or even Robinson and the BTW–Women's model, suggest that *Millenium Hall* is probably something other than their respective definitions of travel writing.[46] Supporting the suggestions of the model, and of scholars, that *Millenium Hall* is more fiction than travel writing is the high accuracy (96.9%) of the non-travel fiction model.

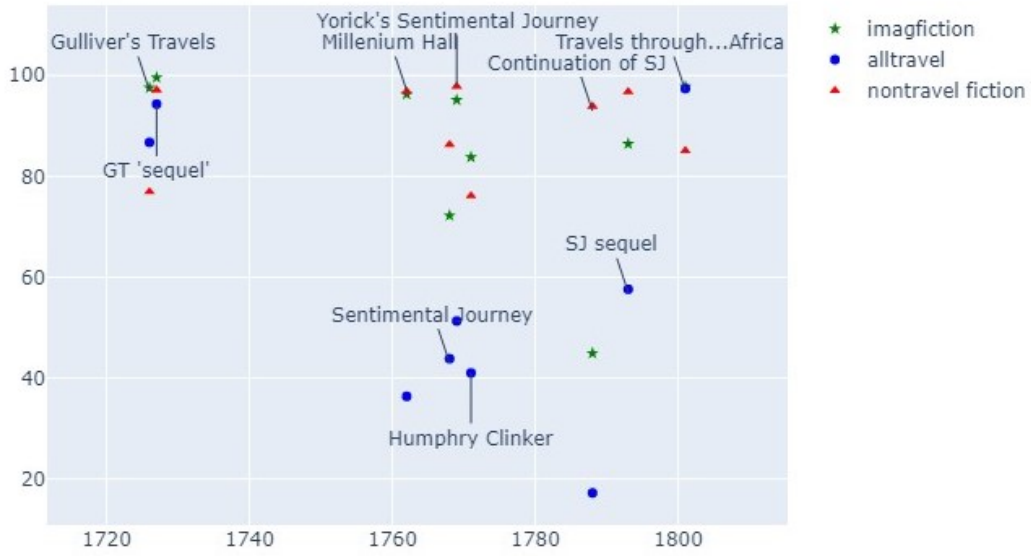
Looking at other fictional titles often deemed significant by travel scholars reveals mixed results as well: for Sterne's famous novel *A sentimental journey through France and Italy*, the imaginary voyages subject heading model is most often correct, averaging around 95.9%, but the

---

[44] The metadata source (HT) included no LCSH and neither Gove nor Cox list it.
[45] For full results for *Millenium Hall* and the titles later in this section, see appendix 4, table 4.2.3: Voyages, Imaginary: Select Titles.
[46] The prediction by the BTW–Women's excerpt is 29.2%, and Robinson's model, 23.2%.

other imaginary models never pass 78.6%, with Cox's Fictitious model averaging the lowest at 67.3%. "True" (or at least, not explicitly imaginary) travel writing models are, however, even more skeptical of *Sentimental Journey*; the other bibliographical sources' average prediction is 31.2% and the overall TWDB and travel tags average prediction is 45.4%. A few other categories are more confident—for example, travel writing poetry at 68.5%, and the women's authorship against other travel writing at 82.6%—but, again, perhaps the most interesting is the model of nontravel fiction, which claims *Sentimental Journey* against the eighteenth century sample with an prediction of 89.1%.[47]

Figure 4.4: Model Predictions of Select Imaginary Voyage Titles



According to these models, then, *Sentimental Journey* typically fits better with nontravel fiction

[47] The false continuations of Yorick's journey often follow *Sentimental Journey*'s pattern of scoring more strongly for the imaginary fiction categories, despite, as Turner suggests, that "Often, the imitation is purely stylistic." She continues by considering contemporary reviews: "The *Critical Review* notes of a 1788 *Continuation* (0761301400) that 'the only imitation of Sterne in this production, is in the breaks, and dashes, and scanty pages, in all which the imitator infinitely exceeds the original': the *Monthly* concurs, adding that 'Sterne had but one blank leaf in a volume; but this book (if you measure by meaning) is all blank, from the beginning to FINIS'" (101). For more on *Sentimental Journey*, see "Sentimental travels: 'so much the ton'" in Turner's *British Travel Writers in Europe 1750–1800* (86–120).

of the eighteenth century rather than travel writing of a historical, geographical, or even fictional nature. Our models and much contemporary scholarship agree that *Sentimental Journey* "is today regarded as a novel rather than a travel account" (Korte 56). Similarly, Tobias Smollett's *Humphry Clinker*, depending on the volume, is fairly at home in the imaginary fiction models, ranging from 73.3% (Gove) to 87.3% (the overall imaginary fiction model), and in the nontravel fiction model at 81.5%.[48] Swift's *Gulliver's Travels*, meanwhile, is always above 85% in the imaginary fiction models—but also in models trained on the entire TWDB (85.5%), Cox (91%), and the TWDB's bibliographical sources mixed together (98%), while for nontravel fiction, only 78.4%.[49] Like *Sentimental Journey*, *Gulliver's Travels* also had imitators; one "sequel," a supposed third volume published in 1727, scores much more strongly not just for other bibliographical sources, but also other locales, and especially for nontravel fiction at 97.1%.[50] So, unlike *Millenium Hall, Sentimental Journey*, and *Humphry Clinker*—and even those attempting an imitation—Swift's more generic satire of travel writing in *Gulliver's Travels* aligns more closely with a variety of models of eighteenth-century travel writing.

Forms: Poetry

Another important form of travel writing is poetry, which scholars often note for both its rarity and its influence. Addison's poem *A Letter From Italy,* published in 1704 (a year before his prose *Remarks on Italy* (0168500300)), inspired responses and imitators throughout the century,

---

[48] Cox's fictitious model, however, only predicts *Humphry Clinker* at 62.5%.
[49] Of the location models, *Gulliver's Travels* scores anywhere from 30–75%—except for the model of the Middle East, which predicts *Gulliver's Travels* at 92.8%.
[50] According to Sir Walter Scott, who called this supposed third volume "the most impudent combination of piracy and forgery that ever occurred in the literary world," the text was "almost entirely stolen from an obscure French work, called 'L'Histoire des Severambes'" (2: 343).

primarily in prose but also some in verse.[51] Both of Addison's works idealize the classical world, a sentiment carried forward by numerous prose narratives (though, as Turner notes, "fewer and fewer as the century progresses") as well as "a handful" of poetic verses (13). Instead of towards Addison's much-vaunted classicism, however, Turner points to the poem's "Whiggish celebration of British liberty" as more influential, and which "bec[a]me a commonplace of travel writing" (13).[52] Turner suggests that the most significant poem of the eighteenth century, however, was Oliver Goldsmith's incredibly popular *The Traveller.*[53] First published in December 1764, four more editions were published before the end of 1765, and six more before Goldsmith died in 1774—though, surprisingly, none of the TWDB sources list it as a piece of travel writing.[54] Turner claims that this success, however, does not seem to have spurred a flurry of verse travel writing on Europe; she counts "few if any significant or widely popular poetic accounts of Continental Europe in the second half of the century, especially in comparison to the amount of topographical or picturesque verse describing the British Isles at this time" (13).[55] Turner bolsters her argument for the dominance of prose by examining reviews, concluding that "the vigorous and copious cultivation of prose travel narrative as a genre by the main literary reviews does not have a comparable parallel for the verse of the period" (13). With a similar focus on a significant poem, in C.W. Thompson's overview of what he calls the Romantic literary travel book, he claims that Byron's *Childe Harold's Pilgrimage*, published through

---

[51] Not listed by any sources; not in the TWDB.
[52] This position was handled differently in different forms: as the eighteenth century progresses, "in prose accounts it comes under increasing scrutiny, whereas verse responses to Continental travel more often adhere to the simplistic Addisonian model" (Turner 13).
[53] Goldsmith is also the reviewer who lamented, "What information can be received from hearing" that a traveller "went up such a hill, only in order to come down again" (*CR* 7: 505, 1759).
[54] Cox often mentions Goldsmith, and occasionally *Traveller*, in his notes to his bibliography, but *Traveller* never gets its own listing.
[55] See Malcolm Andrews for a summary of the history of this movement (3–23). See George Dekker's *Fictions of Romantic Tourism* for the development on sensibility and poetry and later influences on tourism (25–53).

1812–1818, "made poetry the preferred mode of expression for a [Romantic] writer's travels" (270).[56] He points to Samuel Rogers's *Italy, a Poem* (1822–28, mdp.39015025920235) as the primary example, as well as other poems—such as Rev. George Crowley's *Paris in 1815 a Poem*—as following that tradition.[57] The TWDB metadata certainly supports Turner's position, with only 69 volumes with the LCSH of poetry, and a few more added by hand through titles; in the end, only 57 are used in the poetry model.[58] Regarding Thompson's assertion, most of the poetry in the TWDB was published before 1800, at least a decade before the first cantos of *Childe Harold's Pilgrimage* in 1812. While poetry may have been important for Romantic travellers, not much of their work has ended up in the TWDB—or at least, not with metadata that is primarily poetry or "poetic-enough" for it to be clear through subject headings or titles.[59]

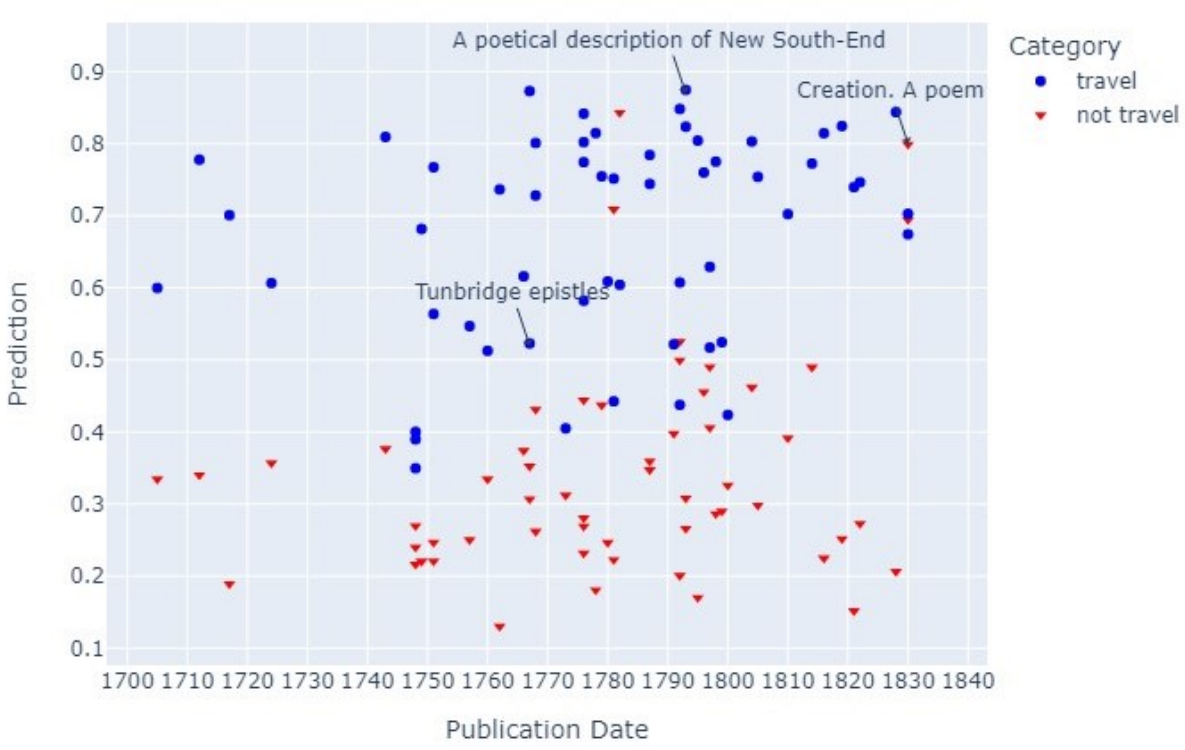
Accuracy of travel poetry models is typically lower than our other sources and tags by a small amount, hovering around 89.7%.[60] Nevertheless, although the model may accurately predict many volumes, depending on the random sample set, it is sometimes very uncertain, rarely getting over 80% certainty for any volume.[61] Travel poetry models are even worse at

---

[56] Thompson suggests that "The history of the English Romantic travelogue is distinguished by the fact that the appearance of *Childe Harold's Pilgrimage* between 1812 and 1818 seems to have brought almost to a halt the creation of such work in prose by leading authors" (270).
[57] Neither *Italy* nor *Paris* are included by my bibliographical sources; instead, *Italy* was added via LCSH.
[58] These numbers do not include the many titles that may have poetry interspersed. Some may also be questionable: the *Memoirs of Mrs. Laetitia Pilkington* (0095100801) are not entirely poetry, despite having `English poetry` as an LCSH. The tag used in the TWDB is `lcsh_title_poetry`.
[59] William St. Clair suggests that "Romantic poetry and modern travel writing went together," a "marriage [that] usually took place within the sheets of the books" as poets drew on the experiences of the prose travel writers (*Reading* 233).
[60] The TWDB's poetry model draws from 57 volumes.
[61] For example, in the second poetry model, the highest prediction was 87.5% for *A poetical description of New South-End* (1793, 0398600300).

recognizing the other models in our corpus: they hover around 68% for the more general travel models.[62]

Figure 4.5: Poetry Model Prediction



In fact, travel poetry models tend to be more similar to nontravel poetry models in the larger corpus. If we train a model on nontravel volumes with "poetry" and then compare the models' views of each other, they can predict each other 80% of the time, as opposed to the under 80% predictions common for the other travel writing sections. This suggests that the elements that

[62] One interesting exception is when the travel poetry models look at Andrews' picturesque, predicting those volumes with 82.8% accuracy (though Andrews looking at poetry is below average at 64.7%); perhaps not surprisingly, poetry and picturesque travel writing have more in common than the large rabble of travel writing generally.

distinguish travel poetry in the TWDB are more similar to poetry in general, rather than prose travel writing.

Forms: History

History is a significant category because of the genre's many connections with travel writing. Narratives discussing foreign locales as well as domestic would often include some reference to the history of the area, especially if the author took the educational responsibility of travel writing seriously. For eighteenth-century authors, the genres were related: Adams suggests that for Swift, "travels were close to history" (*Travel Literature* 143), and Henry Fielding goes even farther in his categorization (and praise) of his own writing, wondering in his Preface to the 1755 *Journal of a Voyage to Lisbon,* "why there should scarce exist a single [travel] writer of this kind worthy our regard; . . . there is no other branch of history (for this is history) which hath not exercised the greatest pens" (547). He continues, suggesting that "in reality, the *Odyssey*, the *Telemachus*, and all of that kind, are to the voyage-writing I here intend, what romance is to true history, the former being the confounder and corrupter of the latter" (548). More recently, John Tallmadge traces several different theoretical approaches to genre, noting that many of them ignore travel writing to focus on other genres, especially the novel. Simply relying on truth to categorize genres creates "a violent reduction of the texts." Instead, Tallmadge recommends that scholars should "consider literature of exploration as a true hybrid combining certain features of both 'reportage' and imaginative fiction" (2). Based on the work of theorists such as Peter Demetz, Robert Scholes and Robert Kellogg, "the literature of exploration would be considered as a type of history, adhering to the communication situation of a report but enjoying the freedom of historical rhetoric" (7). Tallmadge claims, however, that even his particular definition of travel writing—"factual accounts of voyages of discovery written by the explorers themselves

or by participants in their expeditions"—still relies on literary techniques such as compelling narrative personas, plots, themes, and motifs, troubling a spectrum that divides based on rhetoric or truthfulness (2; 10–2).[63]

This close relationship between travel writing and history is certainly demonstrated in the prominence of the TWDB's LCSH, where `history` is attached to over 1,000 volumes, with nearly 800 of those available to model.[64] A model trained only on these travel-history volumes is around 88% accurate when trained against a random sample, slightly lower than poetry, as well as the overall travel tags.[65] As before, however, we can look more closely at rates of mutual recognition: if travel writing, especially "true" travel writing, is a subset of history, we might expect travel-history models to have very high rates of mutual recognition with other models, since they would, in a sense, be setting the genre standard. Instead, the models disagree with Fielding and the scholars above: low mutual recognition rates with the bibliographical sources, most below 80%, suggest that travel-history models are relatively distinct from a variety of travel writing models, even models focusing on travel that actually happened.[66] In particular, travel-history and travel-poetry models have low mutual recognition rates of less than 60%, only slightly better than a random guess. Similarly, travel-history and the various imaginary and fiction models' mutual recognition ranges from 60–75%. Instead, travel-history models perform best on the large generic models, such as all the bibliographies together and the entire TWDB, but even then, they never reach above 86.1%. Perhaps this situation reveals where the application

---

[63] For his definition, Tallmadge also "exclude[s] second-hand accounts and journalistic treatments which, it seems to me, belong properly to the domain of history" (2).
[64] According to the Library of Congress, "history may be used under most headings to designate a historical treatment of the topic in question." See "History H 1647" for a full description.
[65] See Table 4.2.5: History Model Accuracy in appendix 4.
[66] See Table 4.2.6: History Model Mutual Recognition in appendix 4.

of LCSH diverges from theorists and eighteenth-century contemporaries; if nothing else, it suggests caution when subsuming travel writing under other genres.

Locations

Travel is always interested in some way in a location, whether that is around the corner, the continent, or one's own mind. As above, we can again use LCSH, this time to explore different conceptions of geographical locations in British travel writing. Travel writers, readers, and reviewers valued novelty of location, and modern scholars of travel writing often rely on locations as focal points for research, organizing principles in anthologies, and useful limitations for bibliographies.[67] To narrow down the subgenre of travel writing, I rely on the `description and travel` LCSH, a free-floating subdivision used "under names of places for descriptive works or accounts of travel, including the history of travel, in those places" ("Description and Travel H 1530"). I then organize the nations and, more rarely, cities and other categories into larger geographical groups.[68] This is one way of modelling how titles' perceptions of distinct geographical locations, and their cultural connotations, may be similar across oceans, even as they are the combined creation of authors, translators, editors, publishers, and larger discourses about the world beyond Great Britain's shores.

Before we examine the mutual recognition, it is useful to consider the accuracy of models combining description and travel with particular locations. Against the random sample, these location models have high accuracies similar to many other TWDB models: of the 11 models, all

---

[67] Even within the TWDB, there are several location-based bibliographies, including the British Travel Writing–Europe excerpt (btw_europe), McVeagh's Irish travels (irishmcveagh), Murray's (mostly) non-European travels (murray), and so on. Others, such as the Bohls and Duncan anthology, are organized by location (bdanth).

[68] Any of these categories will exclude LCSH of the others—for example, a volume can only be in North America, not in North America and Africa. This will affect collections, but also longer trips such as those going around the world. We lose some examples, some richness, and some nuance in order to get at some cursory results.

but 2 can be predicted above 91%. In the small collection of titles about the Caribbean, the accuracy of 78.7% is hindered by the difficult-to-predict *Whartoniana* (1727, 0595000601), with its collection of verses, and an economic tract—bearing no mention of the Caribbean beyond the South Sea bubble—entitled *The Present state of the national debt* (1740, 1447101600).[69] The smallest corpus, Oceania, despite having only 15 volumes, is still more accurate than the Caribbean, at 87%. In a sharp contrast, volumes describing Africa are easiest for models to predict, with one of the highest accuracies across all models at 97.3%.[70] European models, despite including locations ranging from Sweden to Italy, also have high accuracies, averaging 95.7%, suggesting that the features for these locations are consistently distinct compared to the random sample, making them easy for the model to predict.[71]

Table 4.5: Accuracy of Location Models

Model	Average Accuracy	Total Volumes
loc_carib	78.7	28
loc_oceania	87.0	15
loc_seasia	91.3	48
loc_northam	93.0	269
loc_meast	93.3	73
loc_ireland	93.7	30
loc_latinam	93.7	75

[69] This latter title is a useful reminder that location LCSH are as vulnerable to the subjective nature of categorization (and human error) as other LCSH, and the impacts of such errors can be more influential in smaller corpora.
[70] Across three runs, the Africa model consistently mispredicted *Narrative of a Residence in Algiers* (1818, NCCOF0257-C00000-B0161000), but few other travel titles, as non-travel; the models are more likely to predict false positives. Most titles in this model were published after the formation of the Association for Promoting the Discovery of the Interior Parts of Africa (commonly known as the African Association) in 1788.
[71] Africa is three times the size of Europe, but Eurocentrism and logistics meant that British writers (and publishers) were typically more familiar with the nuances of European cultures and geography, especially because of the tradition of the Grand Tour. More fine-tuned exploration could, for example, divide Europe into locations associated with the Grand Tour (France, Italy) and other "newly discovered" locations, such as Sweden, Portugal, Corsica, and Russia.

Model	Average Accuracy	Total Volumes
loc_easia	94.0	34
loc_gb	94.3	300
loc_europe	95.7	299
loc_africa	97.3	104

Distinct patterns appear when comparing these models.[72] Continuing our discussion of Great Britain and Ireland, the model of the descriptions and travel titles of those islands are very similar, at 91.6% mutual accuracy. The location with the highest mutual recognition with Ireland is Europe, at 87.5%; all other models are at 82.4% or lower, suggesting that the gaze of British empire—or at least, the gaze produced for British audiences—differed significantly when looking at Ireland compared to those islands further afield.[73] For example, models of Oceania are distinct from European, British, and Irish models; instead, Oceania's closest relations are Africa (86.9%) and east Asia (86.2%). These differences in accuracy mean that—at least for some models—the location-related features are generally as or more important than "universal" travel writing features. These accuracies also add context for observations about theorized parallels between locations; for example, Nigel Leask suggests that "representations of Indian or Pacific topography were influenced by the picturesque taste popularised by William Gilpin's tours of England, Wales, and Scotland, [and] 'northern' tourists frequently compared Scottish Highlanders or Irish and Norwegian peasants with Tahitians, Native Americans, and other exotic peoples" ("Eighteenth-Century Travel Writing" 96). Models offer new ways to consider features connected with these conceptions of British and Other: the word "barbarous," for example, is

[72] Note that for the location tags, no travel titles overlap in the training corpus, so the travel portions of the models are completely individual.
[73] The Caribbean (77.4%), Great Britain (79.2%), and Oceania (79.6%) have the lowest average mutual recognitions with other categories. Africa is the highest, at 86.2% average mutual recognition.

often among the most influential terms for models of all locations except for Great Britain and the Caribbean. For postcolonial scholars with knowledge of the nuance of how these locations were conceptualized for a British audience, the models offer an additional way of answering Mary Louise Pratt's question, "How has travel and exploration writing *produced* 'the rest of the world' for European readerships at particular points in Europe's [and Great Britain's] expansionist trajectory?" (4).[74] For example, in a more detailed study focused on individual nations, we could adjust our models from the larger European and Latin American categories (with their 79.2% mutual prediction rate) to focus on individual countries as a way to consider Pratt's assertion that "It is not surprising, then, to find . . . British accounts of Italy sounding like . . . British accounts of Brazil" (10).[75] The mutual recognition of models trained on descriptions of Italy and Brazil, however, is only 74.6% accurate, suggesting that, alongside close reading particular descriptions, models can offer contrasting ways of "looking at" a location. These perspectives do not have to be half a world away, either: the nearly 90% mutual recognition rate of Great Britain and Europe lends support to Chloe Chard's argument that "Like most other scholars in the field, I view the concept of the Tour as one that determines the way in which travel in Europe is envisaged and undertaken from the beginning of the seventeenth century up until 1830 or so" (11). The nuances of these models, including their origin, most influential features, and time frames, offer much to scholars trained in the details of each region, a way of

---

[74] Andrews suggests that one of the "paradoxes" of picturesque tourism is tourists "will loudly acclaim the *native* beauties of British landscape by invoking idealized *foreign* models" (3).

[75] The TWDB has 20 `description and travel` volumes about Brazil, and 128 on Italy. This quotation comes from the first edition of 1992; in the second edition in 2008, Pratt adjusts her assertion to be on a larger scale: "It is not surprising, then, to find…British accounts of the Mediterranean sounding a lot like…British accounts of South America" (12).

exploring Pratt's "contact zones" (4), Felicity Nussbaum's conceptualization and Leask's subsequent exploration of "torrid zones," and other postcolonial lenses.[76]

Writing "By a Lady"

Perhaps even more than the other "subgenres" of travel writing mentioned thus far, the category of travel writing by women is a perennial source of interest for scholars.[77] As noted in chapter 2, the TWDB encodes gender in a variety of ways, including claimed authorship, and contains two bibliographical sources that focus on women's writing: the British Travel Writing–Women's excerpt (btw_w) and Robinson's list (robinson_w). Similar findings to our previous experiments hold true: against a random sample of all eighteenth-century non-travel writing, nine times out of ten, models can discern features to identify these titles. Robinson's 45-volume model, with its "*first-hand* travel accounts *in book form*…not [including] articles, fictional or historical works, and so on," and without any domestic travel, often creates the most accurate models, averaging 94% accuracy against the random sample (ix). If we combine all of the TWDB titles with the woman-as-author tag (`gender_au`), which includes some fiction and poetry, the accuracy (90.7%) is similar to the BTW–Women's excerpt, which focuses on nonfiction (91.7%).[78] All

---

[76] See Nussbaum's *Torrid Zones: Maternity, Sexuality, and Empire in Eighteenth-Century English Narratives* and Leask's *Curiosity and the Aesthetics of Travel-Writing, 1770-1840: 'From an Antique Land.'*

[77] Recent publications focusing on women, writing, and travel include *Gender and Cultural Mediation in the Long Eighteenth Century: Women Across Borders* edited by Mónica Bolufer Peruga et al. (2024), *Political Affairs of the Heart: Female Travel Writers, the Sentimental Travelogue, and Revolution, 1775–1800* by Linda Van Netten Blimke (2022), *Taking Travel Home: The Souvenir Culture Of British Women Tourists, 1750–1830* by Emma Gleadhill (2022), *Transatlantic Women Travelers, 1688–1843* edited by Misty Krueger (2021), *Women, Writing, And Travel in the Eighteenth Century* by Katrina O'Loughlin (2018), and "Travel Writing and Mediation in the Lady's Magazine: Charting 'the meridian of female reading'" by JoEllen DeLucia (2018). Recent anthologies include the *Women's Travel Writings* series from Chawton House and Routledge, which most recently focuses on India (2020).

[78] 92% of the titles in the `gender_au` model advertise their claimed "feminine" authorship on their title page or a peritext, such as an introduction. The `gender_au` files come from a variety of locations: 92 in HT, 55 ECCO, 18 Google, 6 NCCO.

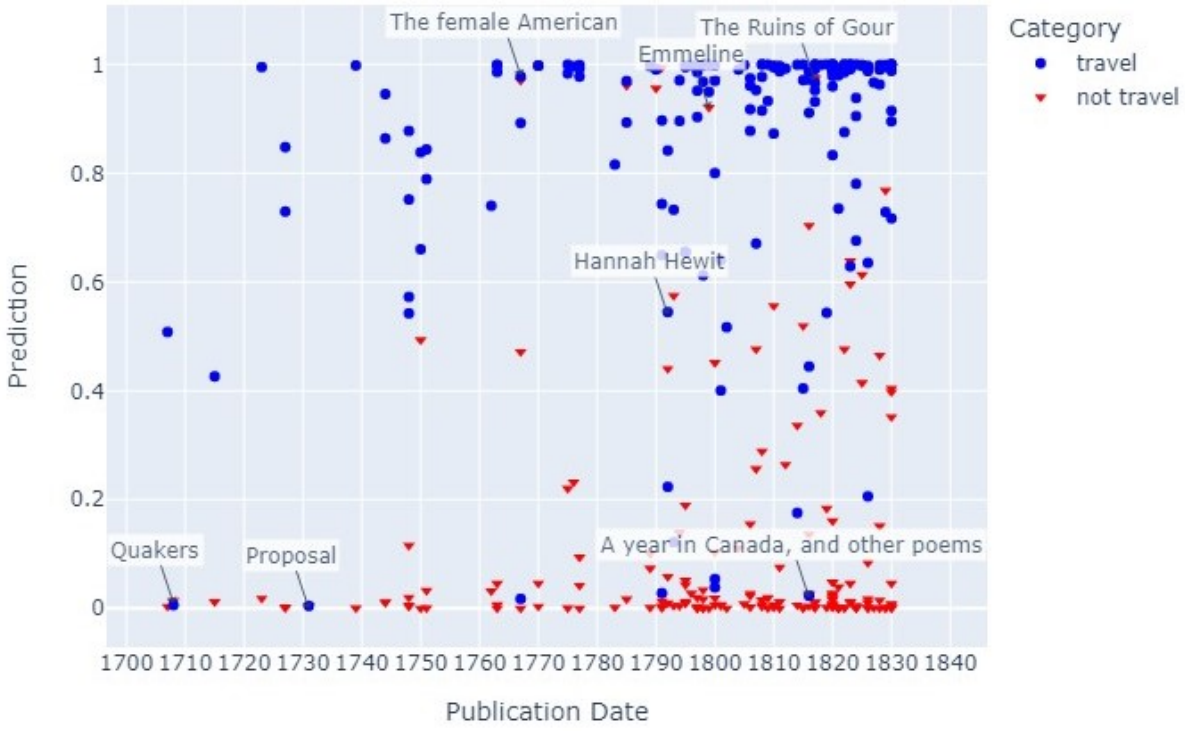
three of these models are trained against a random sample, and they are all still more accurate than the large models that combine a multiplicity of travel tags, indicating that these models of gendered writing are, in fact, more homogenous than the mass collection of bibliographical sources, titles gathered by LCSH travel tags, and the TWDB's combination of the two. But what if we train our author-gender model against not a random sample, but the rest of the TWDB? When we do so, accuracy drops, but only to 83.3%, suggesting that there is something distinctive about the features of these titles compared even to writing ostensibly in the same genre. However, this lower accuracy also means that these against-TWDB features are not *as* distinctive compared to the features that mark women's travel writing as unique from non-travel texts of the long eighteenth century.

Throughout this project, I have argued that travel writing beyond "true" travel narratives is important for exploring the genre, especially regarding women's travel writing. These models help us to build on the metadata analysis of chapter 3, suggesting particular titles that transgress boundaries and conventions. The gender-author model often incorrectly snubs a few TWDB travel titles, which may not be surprising given the titles: *A warning against the Quakers* (1708, 0278001200), *A proposal to determine our longitude* (1731, 0206300500), and *An account of two charity schools for the education of girls* (1800, 0611601200).[79] Titles in distinct forms, especially verse, such as *Tunbridge epistles* (1767, 0146303700) and *A year in Canada, and other poems* (1816, nyp.33433074899786), and the educational text *A dialogue between a lady and her pupils, describing a journey through England and Wales* (1800, 1061000100), are often

[79] *Quakers* is included by the LCSH `antiquities`, *Charity schools* and *Longitude* by Cox.

categorized as non-travel, as is—more surprisingly—*History of the Isle of Man* (1816, hvd.hxjvgw).[80]

Figure 4.6: Model Predictions of Women's Voice



The reverse—non-travel titles being predicted as fitting in with the travel—includes the usual suspects such as *The ruins of Gour described* (1817, njp.32101055309627), but also often novels, such as Charlotte Smith's first novel *Emmeline* (1799, 0485200502), noted for its landscapes.[81] The gendered model also sometimes identifies Smith's novel *Celestina* (1791, 0206900204), wherein characters undertake several journeys, including to France, Scotland, and

[80] Cox's fictitious model predicts *Isle of Man* at 85.6%, Cox at 76.6%, `lcsh_history` predicts *Isle* at 75.5%, and most other models predict below 50%.
[81] See Loraine Fletcher's introduction to *Emmeline* (23–6).

the Hebrides, as gendered travel writing, suggesting the potential for such models in examining further relationships between travel writing and novels by women. For example, an earlier novel included in the TWDB, Penelope Aubin's *The Life of Charlotta Du Pont* (1723, 0148400300), one of Aubin's several "lady's travel" novels in the 1720s, is predicted over 90% by the gendered model, as well as by most of the imaginary fiction models, and the nontravel fiction model. However, unlike several of the novels examined earlier, such as *Sentimental Journey* and *Humphry Clinker*, the model combining all bibliographies, and (surprisingly) the middle east location model, also predict *Charlotta du Pont* over 90%.[82] Even the travel tag model is moderately confident at 74%, suggesting that Aubin's novel straddles multiple genres and supporting Choi's call to reexamine these early women travel writers in particular—even if they were writing fiction.[83]

Of course, not all pieces claiming to be authored by women actually are; in these situations, we can consider how the model assesses the author's skill at matching their style or content to that of a feminine travelling voice. Two Robinsonades offer a comparison[84]: Charles Dibdin's novel *Hannah Hewit; or, the female Crusoe* (1792, 0247300201) claims on the title page *Supposed to be written by herself*, although as C.M. Owen notes, even at the time of publication,

---

[82] *The Life of Charlotta Du Pont* is included through Cox in his "Adventures, Disasters, Shipwrecks" section—not "Fictitious"—and Gove's imaginary travels. From *The Life of Charlotta Du Pont*'s title page (1723, first edition): "Giving an Account how she was trepan'd by her Stepmother to *Virginia*, how the Ship was taken by some *Madagascar* Pirates, and retaken by a *Spanish* Man of War. Of her Marriage in the *Spanish West-Indies*, and Adventures whilst she resided there, with her return to *England*. And the History of several Gentlemen and Ladys whom she met withal in her Travels; some of whom had been Slaves in *Barbary*, and others cast on Shore by Shipwreck on the barbarous Coasts up the great River *Oroonoko*: with their Escape thence, and safe Return to *France* and *Spain*."

[83] For a detailed reading of Aubin's travel titles, see Choi's chapter on Aubin (69–113).

[84] The original, *Robinson Crusoe* (0653600100), averages 84–98% for the more general travel models, 84.5–100% for the imaginary fiction models, 85% for the gendered model, and a wide range of predictions depending on the location.

reviewers knew it was Dibdin's work.[85] *Hewit* scores well for models based on imaginary travel and/or fiction, often above 90%, but only 64.7% for the gendered model, and often under 50% for models ranging from the entire TWDB to most locations. Meanwhile, most travel models predict the pseudonymous *The female American; or, the adventures of Unca Eliza Winkfield* (1767, 0028400401) at higher rates, with the gendered model at 85.5%. This is not a claim that *The female American* was actually written by a woman, or that there is something inherently "feminine" about the writing of either text, although Dibdin does clarify that "added to [Hewit's] female requisites, she had a male mind" (qtd. in Owen 199). Rather, it reveals how these titles match the model, or example, set out by our bibliographers' and cataloguers' expectations for the genre—what Mandell describes as the "'textual gender,' the stylistic and textual features associated with gendered genres" (15).

A final experiment, then: the previous chapter discussed bibliographical cross-references, where the title with the most cross-references in the entire TWDB is Ann Radcliffe's *A Journey Made in the Summer of 1794* (1795, 0593000100), with Hester Thrale Piozzi's *Observations and reflections made . . . through France, Italy, and Germany* (1789, uc2.ark+=13960=t9j38n569), Mary Wollstonecraft's *Letters Written During a Short Residence in Sweden, Norway, and Denmark* (1796, nyp.33433066613229), and Anne Carter's *Letters from a Lady to Her Sister, During a Tour to Paris* (1814, NCCOF0257-C00000-B0199300) close behind. All titles are based on actual travels, primarily in Europe, of the female author, and while they predict strongly for the author-gender model and some of the imaginary models, they are also often predicted

---

[85] Owen points to a review in *The Freemason's Magazine*, which complains that "Throughout he imitates, even to the language, De Foe; though we do not mean to say, he writes either so well, or so correctly, as that author" (qtd. on 199). The review ends thus: "We cannot close this article without advising Mr. Dibdin to be more accurate in his geography: for he has placed the Ethiopians to the north of China, and the Arabians near the Spanish settlement at Manilla!"

above 95% by many of the travel-focused models, including Cox, BTW–Women's excerpt, Robinson, Andrews, and Bryn Mawr, and location models for Europe and North America.[86] These travel models predict these titles more highly than, for example, any of the LCSH fiction models, demonstrating that the authors are aligning with travel writing's conventions. Titles by men dominate the training for most of these travel models, suggesting that these women authors may be asking similar sorts of "men's questions" to those of their masculine colleagues.[87] By the virtue of their sex (according to bibliographies) and their writing style (according to models), the four titles above, as well as many others within the TWDB, are confirmed to have the typical distinctive features of travel writing. By modelling these collections of titles and comparing them to other travel writing models, we avoid "A separatist view of women's travel writing…[which] is in danger of ignoring general generic characteristics and developments, and of reducing the travellers and their texts to the 'typically' feminine" (Korte 109); instead, we see how these authors are participating in, diverging from, and reinforcing the traditions of the genre.

The Changing Faces of Travel

A final, crucial perspective on travel writing is chronological, or examining whether models trained on one span of time can recognize those in another. Tracking changes in this mutual recognition as the century progresses may suggest shifts in the content or style of the genre. In Underwood's study of science fiction, for example, decreases in accuracy of over 10% between models trained on 30-year time frames indicate rapid changes in science fiction in the 1930s-

---

[86] None of these titles focus on North America.

[87] Wollstonecraft recollects how a host in Sweden "told me bluntly that I was a woman of observation, for I asked him *men's questions*" (qtd. in Chard 38). Chard describes how "Travellers who proclaim a female identity, moreover, intermittently define the experience of travel as one that invites an identification with various specifically masculine approaches to determining the relation between the self and the world…identifying with masculine authority…[or] masculine experience of the world" (38). Women may also strategically "resort to the superior knowledge of [their male travelling companions] where politics and economics are concerned," as Radcliffe does (Korte 113).

1950s (*DH* 63).[88] As the TWDB models below show, despite travel writing's heterogeneity and shifting cultural discourses, travel writing never changes this dramatically. The genre stays relatively stable over the course of 130 years, despite changes like the "discovery" of new continents, the growing middle class, and the rise of Romanticism.

First, we can look at the time span of decade in the overall TWDB: if we make a model of 1701–1711 and see how well it can predict the volumes in 1711–1721, and then move ahead 5 years and repeat until 1831, the loss of accuracy is almost always below 5%, except for 1716 and 1801, which still only peak at 6% loss of accuracy.

Figure 4.7: Pace of Change in TWDB (20-Year Average)

However, because we have two file sources and a significant bibliographer (Cox) that change in 1801, this peak in the TWDB models' accuracy should give us pause—and, indeed, if we compare only ECCO files, as in figure 4.8 below, there is no increased loss of accuracy leading to 1800. Instead, whether we measure by a single decade or three, the models suggest that the

[88] See also "Machine" (103–6).

mutual recognition remains approximately consistent as the century progresses.[89] Scholars, however, often point to the 1760s and 1770s as when travel writing begins to shift more rapidly, with a growing desire by readers, authors, critics, and travellers—increasingly of the middle class—for more subjective experiences and descriptions, avoiding the "disgusting repetition" of other "modern travellers" (Goldsmith 505).[90] Turner also suggests that the growing presence and influence of the literary journals, especially the *Critical Review* and *Monthly Review*, affect the "influential paradigm for travel writing and its reception in later decades" (55).[91] At our models' scale and method of comparison, however, models trained on travel writing of 1730–1760 can recognize travel writing from thirty years later with hardly any decrease in accuracy, as can be seen in figure 4.8, where the dotted 60-year average nearly aligns with the 20-year average.

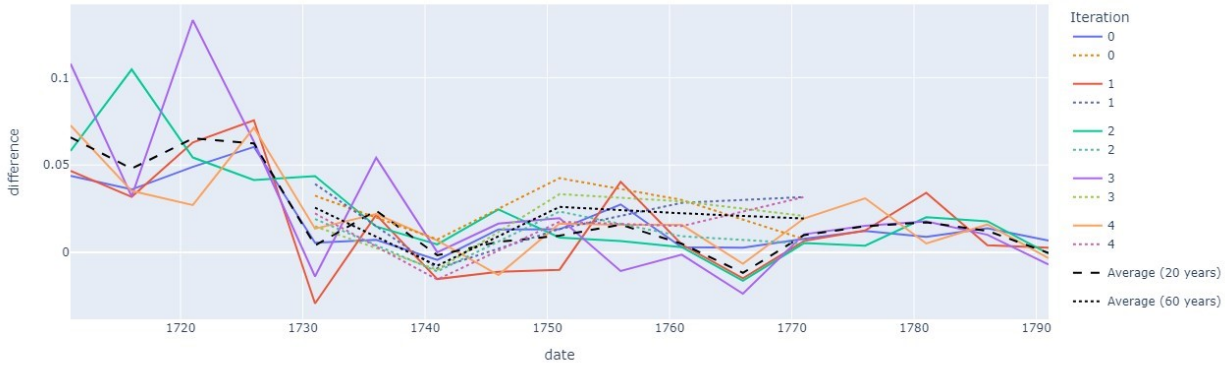
---

[89] The 30-year models step by 10 years instead of 5. The models of 1701–1730 flatten out some of the accuracy losses of the ten-year models, which is part of the puzzle of the early eighteenth century. In looking at surprising titles of this period, the models are often surprised by titles about benefits of spas, as well as all the volumes of John Stevens' English translation of Antonio de Herrera y Tordesillas' *The general history of the vast continent and islands of America, commonly call'd the West-Indies* (1725, 0098501901). Limiting the volumes to those with the `description and travel` LCSH (with no file source limits) results in similar patterns of high losses of accuracy in the early eighteenth century and around 1801 (see appendix 4.3). More work needs to be completed here, especially since the early eighteenth century has more unreliable OCR.

[90] In his 1759 review in the *Critical Review* of Aegidius Van Edmont's *Travels through Part of Europe, Asia Minor…*, Goldsmith opines that "Travels acquire one great part of their merit from being new…the reader has a right to expect recent information, that it at least excels all other accounts by giving, if not more authentic, at least more modern descriptions." Unfortunately, "In this respect, however, the purchaser of the book in question, will find himself mistaken" if that is what he expects to read (504).

[91] Scholars of travel writing often point to reviews as evidence and influence, such as the editors of *TEE* who similarly assert that at the turn of the century, "Popular reviews such as the *Edinburgh* and the *Quarterly* publicized but also criticised travellers' texts, effectively creating a taste that explorers learnt to supply. The 'foreign', as offered to the public through travellers' pages, became partially shaped by the tastes of home, in a circular process which effectively circumscribed what could be known of other places and cultures" (xxii).

Figure 4.8: Pace of Change in ECCO Travel Writing (20-Year and 60-Year Average)

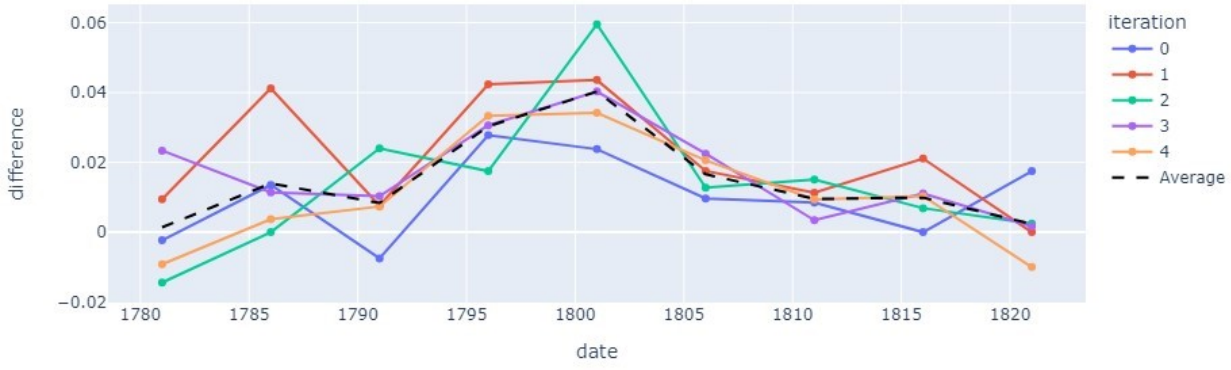


Scholars also point to other trends of the late eighteenth century and early nineteenth century, such as the picturesque, the sublime, the gothic, and the Romantic. Our ECCO files, ending in 1801, do not show dramatic changes; instead, we can bridge the century gap with our other file sources. If we model HathiTrust and Google files from 1771 through 1830—right when many of these literary movements are taking place—figure 4.9 shows that the loss of accuracy on a 20-year window is still never more than 4%, and if we use our 60-year window, looking at 1771–1800 and 1801–1830, table 4.6 shows a slightly higher loss of mutual recognition of 7.7% on average.

Table 4.6: Pace of Change in HT Travel Writing: Comparing 1771–1800 and 1801–1830

Iteration	Difference in Accuracy Percentage
0	8.7
1	5.8
2	5.9
3	7.2
4	11.0
Average	7.7

Figure 4.9: Pace of Change in HT Travel Writing (20-Year Average)



What does this mean, in practical terms? There is no agreed-upon loss of accuracy where we can say, without a doubt, "Yes, this genre is changing." In Underwood's study of science fiction, he draws attention to losses of accuracy around 10% to suggest that science fiction "change[s] rapidly" (63).[92] The average accuracy loss of the TWDB's chronological models never climbs so high. Instead, our evidence suggests that despite its heterogeneity and shifts in content, style, authorship, and even prioritized locations, travel writing on the whole changes slowly throughout the eighteenth century, with some accelerating change as the new century begins.[93]

Even if the pace of change is not as strong as in Underwood's genre studies, we can still use it to our advantage to investigate particular examples. For instance, we can examine which

---

[92] Science fiction models' "pace of change always peaks between 1930 and 1950, in a period that loosely aligns with [Gary K.] Wolfe's narrative of consolidation" of the genre during this time (*DH* 63). Models earlier and later in the time period identify losses of accuracy around 3%.

[93] These shifts do add up to larger differences: if we test the mutual recognition of a model trained on 1701–1730 to a model trained on volumes published 100 years later, the accuracy is only 69.7%. In this single experiment, the 1801–1830 model is more accurate when looking back at 1701–1730, at 76.0%, while the 1701–1730 model struggled at 64.0% to predict the volumes of the future. The different file sources and quality of OCR between these two timeframes may also influence this accuracy.

nineteenth-century predictions are "surprising" to a 1771–1800 model: that is, where the pre-1801 and the post-1801 models differ in their predictions. One perspective is to look at travel writing titles that a pre-1801 model is skeptical of, such as Walton William's *Present State of the Spanish Colonies* (1810, uc1.$b723322), James Franklin's *The Present State of Hayti* (1828, mdp.39015002621053), and Charles Mackenzie's *Notes on Haiti* (1830, nyp.33433081700399), where the pre-1801 model is surprised by language and paragraphs focusing on revolution, cruelty, and foreign relations following the Haitian Revolution.[94] For this period, however, most of the surprises, and the largest ones, are where the earlier model's predictions are higher than those of the later model—despite the predicted volumes being published in 1801–1830.[95] This means that such volumes fit better with their predecessors rather than with their contemporaries. If we look at the top 50 of these surprising volumes, another surprise awaits: volumes by women make up 16% (7 volumes) of the 50, even though such women-authored volumes make up only 6.5% of the TWDB volumes during these last three decades.[96] These titles include descriptions of France after the Napoleonic Wars by Lady Morgan and Helen Maria Williams, as well as more personal memoirs, including *A Memoir of the Life and Ministry of Ann Freeman* by the eponymous Bible Christian (1826, EBuWY8jmOMAC) and *Letters From The Mountains* by

---

[94] Joseph Dupuis' *Journal of a residence in Ashantee* (1824, hvd.32044011474574) discusses similar topics, but in what is now Ghana. Titles about the French Revolution and Napoleonic Wars are generally not surprising to the pre-1801 models.

[95] 75% of the volumes are more surprising to the post-1801 model. For context, in Underwood's similar example looking at three decades and after 1940, only 28.8% of title predictions were more surprising to the post-1930 model.

[96] Of the 973 travel titles in the five 1801–1830 models, 68 are by women (7%). The seven titles are *Letters from the mountains* by Anne MacVicar Grant (1806, nyp.33433069350431); *A narrative of the events which have taken place in France* by Helen Maria Williams (1815, nyp.33433069337081); *A History of the Isle of Man* by H. A. Bullock (1816, hvd.hxjvgw); *Narrative of the demolition of the monastery of Port Royal* by Mary Anne Galton Schimmelpenninck (1816, uc1.$b297562); *France [in 1816]* by Lady Morgan (1817, hvd.hn4dkg); *A Memoir of the Life and Ministry of Ann Freeman* by Ann and Henry Freeman (1826, EBuWY8jmOMAC); and *France in 1829–30* by Lady Morgan (1830, hvd.hwkzer).

Anne MacVicar Grant (1806, nyp.33433069350431), "one of Britain's best-known and most successful commenters on the landscape and culture of the Scottish Highlands" (Perkins 246). These latter are the most "surprising" women-authored volumes to the post-1801 models, with the pre-1801 models predicting *Memoir* 76.6% and *Letters* 58.2% higher than the models of these volumes' actual period.

What within these books makes them so surprisingly "old-fashioned"? By comparing the relative strengths of the models' most influential terms to make a "surprise metric" and applying it to chunks of text within each volume, we can shift to a more familiar style of close reading—but one that is still informed by our models. In *A Memoir of the Life and Ministry of Ann Freeman*, according to Amy Culley, Freeman "constructs herself in opposition to institutional religion and styles herself as a preacher in exile," and her "narrative is structured around moments of conflict that express the difficulty of reconciling spiritual autonomy with family commitments and religious fellowship" (97). Culley focuses on *Memoir* as religious women's life writing rather than travel writing, but the above themes are also evident in our models' differences: pre-1801 models, compared to the post-1801 models, rely more on *preached*, *Christians*, *religious*, *devotion*, *wonder*, *speak*, *speaking*, *and spoke.* But more interestingly, some features that the pre-1801 models are most drawn to—and what the post-1801 models are skeptical of in comparison—are the elements that, at the beginning of this chapter, our models identified as a main feature of travel writing in general: words that directly discuss the mechanics of travel. Consider the following passages, with the pre-1801 model's most influential terms in italics:

> H[enry] *preached* to the people in the market-place. We *walked* to friend Brownlow's in the evening; and next day we *rode* to Monaghan.

> 1st of tenth month.—I *parted* from my husband and came to Dublin, intending soon to go for England if the Lord *permit***,** as it is thought the change may *restore* my feeble *frame*; but I feel no *choice* of either life or death but feel this is now my way to take.
> 3d.—I am *waiting* to have my way made plain for *leaving* Ireland. I cannot take one *step***,** without satisfaction from the Lord.
> 10th.—My husband arrived this morning in Dublin from the country. We *rode* to Kingstown, *accompanied* by a few friends. (100–1)

> …to *inform* thee of my safe *arrival* here. After *leaving* thee at Bristol, I *rode* on comfortably, on the outside of the coach, till about eleven at night: then I had an inside passage all the way to Barnstaple. I was as well as I could expect. I *stopped* at the inn about an hour and the coachman got a conveyance for me to Biddeford; (228)

For models after 1801, some—though not all—of the explicit descriptions of travel become less important. The post-1801 models do still favour *depart* and *departed***,** for example; but in general, paragraphs discussing the action of travel at length are more important to pre-1801 models than post-1801 models. This is also evident in the second volume of Grant's *Letters from the Mountains*, where pre-1801 models point to moments in her correspondence that focus on travelling:

> I am glad you do not go to Edinburgh because then you will *possibly cross* Drimochter.—Short stages, and slow *travelling***,** might *prevent* any risque to your health; the *accommodation* and arrangements may be managed much *easier* than last summer. (2: 52–3).

> I found my relations there very well, and cordially kind, and was sorry I could *spend* but a day with them. - - - - - I had an agreeable day's journey from Dunkeld, or Blair rather, with C. who *proves* an excellent *travelling companion*—always cheerful, and full of *observation***,** and easily silenced when I wish to *indulge* my meditations.—We took many long considerate *walks* for I *dreaded* Paddy's being overloaded. (2: 158)[97]

This shift of the models' valuation of particular travel descriptions is not just because of these titles' personal nature; several titles, including the third volume of Grant's *Letters from the Mountains*, declare themselves as epistolary or a memoir and still surprise the early models.[98] Once again, our definitions and models of travel writing—whether created by a bibliographer, a cataloguer, or a computer—are nuanced by intersections and comparisons with the long eighteenth century's diverse and sometimes contrasting genres, subgenres, modes, and trends. Close readings of dozens of titles within such categories may indicate a "turn to the Romantic" (Parks) or professed desires for novelty, or any number of other changes in style or content, but at the scale of our models' hundreds upon hundreds of books—or thousands upon thousands of counted features—travel writing does not actually change dramatically between 1701–1830.

Conclusion

This exploration of models of travel writing—our view from the metaphorical top of the

---

[97] Grant organized the volumes by correspondent (Perkins 248). The difference in prediction between the pre-1801 and post-1801 models are 10.3% (Volume 1, nyp.33433069350423); 58.2% (Volume 2, nyp.33433069350431); and -16.0% (Volume 3, nyp.33433069350449).

[98] For example, William Stewart Rose's *Letters from the north of Italy* (1819, mdp.39015065298849), *Mementoes, historical and classical, of a tour through part of France, Switzerland, and Italy, in the years 1821 and 1822* (1824, nyp.33433082469721), and Thomas Crofton Croker's *Legends of the lakes: or, sayings and Doings at Killarney* (1829, nyp.33433061821561).

mountain—looks at large-scale perspectives of the entire Travel Writing Database, subgenres of travel and location both imagined and real, authorship both claimed and real, and changes spanning decades and over a century. The high accuracy of these models, often around 90%, suggests that despite significant cultural and literary shifts and a seemingly disparate collection of titles, travel writing published in Great Britain does cohere across the long eighteenth century. Rather than travel writing as a borderless genre, these findings support arguments that travel writing has recognizable conventions, such as those that Charles Batten suggests "govern a travel writer's actions and descriptions" (4) and Barbara Korte's "particular strategies—including specific artistic principles and designs" that authors rely on when participating in a "particular form of writing" (15; 2–3). For example, imaginary voyages—the most common excluding factor in definitions of travel writing—do have different conventions than real travel writing, as their lower rates of mutual recognition demonstrate. Such divisions are fuzzy rather than distinct, however, especially the more that a fictional journey emulates nonfiction. Instead, style is more important: for example, models show that travel-poetry is more similar to nontravel poetry rather than travel-prose. In history, models of travel-history texts prioritize features distinct from other travel writing models, even those based on true journey, suggesting that history is not a potential "parent" genre as suggested by some authors. This redrawing of the map of travel writing also applies to geographical locations, where models suggest that descriptions of different regions cohere around specific features rather than those prominent for travel writing more generally.

These results also intersect with authorial voice, especially when that voice is publicly gendered. This study thus offers caution when relying solely on gender as a "salient variable" when analyzing these texts, instead suggesting that other genre features are more influential (Foster and Mills 3). The models' categorization of authors such as Aubin draws our gaze to

texts outside of traditional definitions of travel writing, suggesting that expanding definitions of travel writing to include fiction is critical for understanding both women's travel texts and the genre more broadly. Although scholars often describe women authors as breaking new ground, the examples above also show that their texts align with common travel writing conventions, even if those patterns are out-of-date. Finally, across time frames and file sources, models of travel writing suggest that the genre remains recognizable throughout the long eighteenth century, despite the chronological divisions imposed by researchers and archives (both physical and digital). These representations of travel writing, in the form of computational models, both quantify differences and contribute nuance at the scale of both genre and individual titles, drawing attention to how authors of this period and scholars with our hindsight alike blur or create boundaries, sometimes at the level of a single title, volume, or paragraph.[99]

This chapter is also a test of data and methods, experimenting with and building on approaches developed by Underwood, his collaborators both individual and institutional, and the larger digital humanities community. Underwood's original work relies on more modern genres, a single file source, and a few bibliographies and LCSH tags. My application of perspectival modelling in both methodology and code demonstrates that we can extend this approach to older genres, despite their permeable borders, multiple definitions, assorted file sources, and relatively dirty OCR. In typical digital humanities fashion, the data, models, and interpreted results confirm some expectations about the genre, but offer many more branching paths for future research in fields as varied as postcolonialism, aesthetics, and women's writing. These opportunities are in

---

[99] A further exploration could look more closely at differences between "Observations" and "Reflections," especially as connected with subjective description. For an overview, see Batten, chapter 3: "Descriptive Conventions in Eighteenth-Century Nonfiction Travel Literature."

part due, however, to the TWDB's rich metadata and the documentation of its creation, drawing from Underwood's approach to sampling and comparison as well as bibliography and book history. Even, or especially, when data and metadata are endless, fractured, or obscured, a curated dataset more akin to Katherine Bode's "scholarly edition of a literary system" allows for an expansion of questions that shift beyond, but are complemented by, close reading, bibliography, and the canon. Humans decide the labels, algorithms identify and count the features, and humans interpret the results; a model "sucks intelligence in ev'ry clime, / And spreads the honey of his deep research / At his return, a rich repast for me" (Cowper 111–3).

Conclusion: Future Itineraries

> That I should make some reflections, or write down some observations, in the course of a long journey, is not strange; that I should present them before the Public is I hope not too daring.
>
> —Hester Thrale Piozzi, the Preface to *Observations and Reflections made in the course of a journey through France, Italy, and Germany* (1789, 5)

Hester Thrale Piozzi's *Observations and Reflections made in the course of a Journey through France, Italy, and Germany* (1789) is a stereotypical travel narrative in many ways. It describes a journey that actually happened, when Piozzi—criticized by her friends and family for marrying the Catholic Italian Gabriel Piozzi—departed on a three-year honeymoon.[1] Their route is a conventional one: she travels through France, Italy, and Germany, all familiar locations to readers after decades of the Grand Tour and "extremely numerous" travel accounts (*Critical Review* 5: 294). She was a woman travel author—a rarity—but she was already famous for her friendship with Samuel Johnson and infamous for her scandalous marriage to her daughters' music teacher.[2] She writes her *Observations and Reflections* in part because her estranged friends, including Johnson and Frances Burney, "rather chose to amuse themselves with

---

[1] Johnson responded to Piozzi's marriage announcement harshly: "If I interpret your letter right, You are ignominiously married, if it is yet undone, let us once talk together. If You have abandoned your children and your religion, God forgive your wickedness; if you have forfeited your Fame, and your country, may your folly do no further mischief" (July 2, 1784; IV: 338).

[2] Piozzi wrote *Anecdotes of the late Samuel Johnson* (1786) and edited *Letters to and from the Late Samuel Johnson* (1788).

conjectures, than to flatter me with tender inquiries during my absence" (vi).[3] Piozzi claims that this lack of correspondence is why she must abandon the epistolary form: "I have not thrown my thoughts into the form of private letters; because a work of which truth is the best recommendation, should not above all others begin with a lie" (vi).[4] While travel writing, including *Observations*, is always about a journey in some sense, that journey—and its eventual "voyage into narration"—is often accompanied by such justifications, apologies, and critiques (Bourguet qtd. Keighren et al. 2). This careful positioning applies to travel writing in general, with Shef Rogers suggesting that "travel writing was probably the most self-consciously print-informed genre of the period" (784). Reading Piozzi's defensive *Observations and Reflections* in relation to other authors and traditions is thus necessary, but also a risk: scholars are prone to reading Piozzi as merely a reflection of the famous Doctor Johnson.[5] Even this dissertation relies on the great lexicographer to frame its approach to travel writing.[6]

Models, however, know nothing about Piozzi and Johnson's relationship. Similarly, they do not read through the lens of epistolarity, empire, or the Enlightenment. Instead, models offer generalizations which we can then interpret and apply alongside these theories, "synthesiz[ing]

---

[3] Johnson died on December 13, 1784. Piozzi received the news in Milan two weeks later (*Thraliana* 2: 284n4). On January 2, 1785, Piozzi writes "Oh poor Dr Johnson!!!" and it is not until January 25 that she continues: "I have recovered myself sufficiently to think what will be the Consequence to me of Johnson's Death, but must wait the Event as all Thoughts on the future in this World are vain" (2: 284–5).

[4] Women's travel writing before this was traditionally in letters, leading Marianne D'Ezio to suggest that with *Observations*, "a specifically woman's travel narrative was born and began to take the conscious form of cultured prose" (168).

[5] Mirella Agorni opens her chapter on Piozzi, titled "Hester Piozzi: A Dilettante or a Remarkable Woman Writer?", with the observation that "Piozzi is usually remembered today merely for being Samuel Johnson's landlady at Streatham Park" (111). In "Hester Piozzi, Italy, and the Johnsonian Ether," Dussinger suggests that "no one was more influenced by Johnson's opinions than Piozzi," pointing to at least 15 references to Johnson in *Observations* (48). In contrast, Angela Wachowich suggests that *Observations* "was the first of Piozzi's works to escape Johnson's long shadow."

[6] For example, see the epigraphs of chapters 1 and 2 and the conclusion of chapter 1. See "Travel, Trade, and the Expansion of Empire" for a collection of Johnson's definitions that are related to travel writing.

humanistic and computational approaches" (Long and So 236) much in the way that travel texts draw on different styles and subgenres.[7] These data, codebases, and models can (and should) be shared, tested, and challenged, a goal for the TWDB since its inception.[8] This position is not just practical, but epistemological and ethical.[9] In Andrew Piper's advocacy for modelling in literary and cultural studies, he describes how "You can enter into my model and I yours. A model allows us to be more vulnerable with our ideas, making knowledge cumulative but also conglomerative. It emphasizes collectivity, departing from our field's historical focus on singularities" ("Think Small" 657). Our models can look beyond the Johnsons of the eighteenth century to consider not just the Piozzis, but also thousands of other unstudied accounts, narratives, novels, journals, guides, tours, and trips.

Coded Voyages addressed the question of "what is travel writing?" by assembling fourteen bibliographies, combining them into a single dataset, and then comparing models based on definition, subject, and authorship. Chapter 1 confirmed expectations that one's choice of bibliographical "guide" to the genre often limits the perspective based on the truth of the journey, the author's gender, or periodization. This focus on vantage points continued in chapter 2 where metadata and textual data created the final Travel Writing Database. Chapter 3 examined the affordances and constraints of the TWDB definitions, metadata, and data. Finally, chapter 4

---

[7] Although generalization may have a negative connotation, Andrew Piper argues that it "is an essential, one might even say existential, scholarly practice that until now has remained all but invisible in critical debates in the humanities. The failure to generalize well puts at risk nothing less than our credibility as scholars and cultural commentators" (4). See Piper's *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data* for his full discussion and examples. See also Richard Jean So's "All Models Are Wrong" and Hoyt Long and So's "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning."

[8] Limitations on publishing the dataset are still being established. At a minimum, the *Coded Voyages* code and file ids will be published in an open-source format.

[9] See "Appendix A: Data" and "Appendix B: Methods" in Underwood's *Distant Horizons*. See also Katherine Bode's critiques of inaccessible data in "The Equivalence of 'Close' And 'Distant' Reading; Or, toward a New Object for Data-Rich Literary History," especially 80–5.

applied Underwood's perspectival modelling, where combining the perspective of different predictive models revealed travel writing's "multiple crossings" (Borm 26).[10] Significantly, models could easily identify travel writing as a large mixed bag and as distinct subgenres. Models also revealed the conceptual distance between imaginary voyages and real voyages, and they collapsed geographical distance to demonstrate similarities between locations thousands of miles apart.[11] Notably, models supported the scholarly position that women's voices were a "salient variable" in travel writing, but suggested that such "feminine" authors participated in conventions rather than departing from them (Foster and Mills 3). As the first computational project to collect, model, and analyze travel writing texts of the eighteenth century, *Coded Voyages* thus offers new insights into this extremely popular eighteenth-century genre, as well as the potential of computational analysis in literary studies.

This dissertation's exploration of the Travel Writing Database, however, is just a single path. Many other routes remain open to different theoretical or methodological itineraries, such as studies focusing on postcolonialism, world literature, or female authors. The Travel Writing Database also enables and invites different computational approaches. *Coded Voyages*' chapter 4 relied primarily upon perspectival and predictive modelling, but as digital humanists know, new methodologies appear quickly. A decade ago, scholars were exploring topic models.[12] Now, academic research, pedagogy, and the broader world are reckoning with easier access to machine learning, large language models, and generative artificial intelligence. In documenting their

---

[10] Jan Borm suggests that travel writing is "a useful heading under which to consider and to compare the multiple crossings from one form of writing into another and, given the case, from one genre into another" (26).
[11] Piozzi, paraphrasing *Letters from an English Traveller* (1780), makes a similar claim: "if a hundred men of parts travelled over Italy, and each made a separate book of what he saw and observed…no two should be alike; yet all new, all resembling the original, and all admirable of their kind" (2: 384).
[12] See Rob Churchill and Lisa Singh on "The Evolution of Topic Modeling," which points to Matthew L. Jockers and David Mimno's 2013 "Significant themes in 19th-century literature" as an example in literary studies.

approaches to these technologies, digital humanists are not so different from travel writers. Whether for themselves, future generations, or funding bodies, scholars are expected to keep *notanda* of their progression, and "such descriptions of project building have become a genre in themselves" (Ceserani).[13] These accounts are popular for their novelty and practicality, but their prominence in documentation, at conferences, and in academic publications also suggests that humanities scholars remain largely self-taught in computational approaches. Formal training in creating and critiquing these methods, ranging from coding to project management, is scarce in curricula, despite suggestions of including statistics in English programs or counting a coding language for a PhD language requirement.[14] "Observations" and "reflections" thus remain critical to the field, even when passing over seemingly well-trodden ground.

Travel writing has also persisted as an influential and profitable genre. Observations from the turn of the nineteenth century, such as "One of the most distinguishing features in the literary history of our age and country, is the passion of the public for voyages and travels," could apply equally to the twenty-first century, especially if one includes personal blogs and Instagram (qtd. in Rogers 781). Traditionally published literature can also reap huge profits. *Eat, Pray, Love: One Woman's Search for Everything across Italy, India and Indonesia* (2006) by Elizabeth Gilbert spent 199 weeks on the *New York Times* Bestseller List, earned Penguin $15 million dollars (US), and was made into a film starring Julia Roberts with box office sales of over $180

---

[13] For descriptions of eighteenth-century journals and other writing technologies, see Malcolm Andrews (73–6) and Innes M. Keighren et al. (42–8). Piozzi's notes on her travels were collected in two notebooks, "Italian Journey 1784" and "German Journey 1786" (Agorni 113). See Giovanna Ceserani's extended description of the Grand Tour Explorer and *A World Made by Travel*, as well as Piper's discussion of "Building a Team, Building a Model" in *Can We Be Wrong?* (17–27).

[14] See Underwood (*DH* 165), Andrew Goldstone's "Teaching Quantitative Methods: What Makes It Hard (in Literary Studies)," and Zoe LeBlanc et al.'s "From Precedents to Collective Action: Realities and Recommendations for Digital Dissertations in History."

million (US).[15] Despite the 200 years between Piozzi and Gilbert, defining travel writing has not gotten easier: as Pamela Thoma points out, *Eat, Pray, Love* has "been variously described as 'foodie romance,' 'confessional memoir,' 'culinary adventure,' 'gastronomic travelogue,' or 'priv-lit'" (109). And as the subgenres above suggest, gender still influences perspectives on travel writing. *Eat, Pray, Love*'s author and narrator, like Piozzi, is also a professional writer, "a role that neatly encapsulates tropes of female agency, commercial success, and creativity" (Thoma 109)—a description that applies to female authors of both now and then.[16] Expectations from hundreds of years ago still influence the ways that modern travellers and writers move through, and describe, the world.

Coded Voyages began with epigraphs describing the many "Heads" of travellers and "faces" of travel writing. Whether "Idle Travellers, Inquisitive Travellers, Lying Travellers, Proud Travellers, Vain Travellers, Splenetic Travellers[, or]…Travellers of Necessity," it is impossible for us, so distant from their time, to meet these figures on their own terms. By comparing them side-by-side, however, we can gain a sense of their silhouettes, their relationships, and their long shadows over both travel writing, but also literature of the eighteenth century more broadly.

---

[15] As Ruth Williams notes, this does not count the various merchandise, such as yoga mats, tea, and so on (1). Williams also suggests that "one of the key products *EPL* marketing encourages women to consume is tourism" (4). In addition to Williams, see "Eat, Pray, Spend: Priv-lit and the New, Enlightened American dream" by Joshunda Sanders and Diana Barnes-Brown. Gilbert's publisher funded her travels.

[16] Similar to Piozzi's estranged friendships encouraging her extended travels, part of Gilbert's inspiration was the ending of her marriage and another romantic relationship. In the "Love" section of *Eat, Pray, Love*, set in Bali, Gilbert falls in love with and marries "Felipe" (José Nunes). See also Dussinger's description of Piozzi's *Observations*: "it does have finally a coherence as the autobiography of a middle-aged, affluent, middle-class, learned, and fair-minded Protestant woman confronting the sublime mystery of a many-layered pagan, Gothic, and early modern Italian culture" (53).

Appendices

Appendix 1.1: TWDB Metadata Codes

Table 1.1.1: TWDB Metadata Codes

Tag	Context	Description
andrews	Bibliographical source	Andrews, Malcolm. *The Search for the Picturesque: Landscape Aesthetics and Tourism in Britain, 1760–1800*: "Select Bibliography." Stanford UP, 1989.
bdanth	Bibliographical source	Bohls, Elizabeth A, and Ian Duncan. *Travel Writing 1700–1830: An Anthology*. Oxford UP, 2005.
brynmawr	Bibliographical source	"European Travel Accounts Printed Before 1850." Bryn Mawr College Library Special Collections. 2018. https://wayback.archive-it.org/230/20210310150611/http://bascom.brynmawr.edu/library/speccoll/guides/travel/index.html
btw_europe	Bibliographical source	Colbert, Benjamin. "Bibliography of British Travel Writing, 1780 - 1840: The European Tour, 1814 - 1818 (excluding Britain and Ireland)." *Romantic Textualities*, 13: 5–43.
btw_w	Bibliographical source	Colbert, Benjamin. "Women's Travel Writing, 1780–1840: A Bio-Bibliographical Database." *British Travel Writing*. Designer Movable Type Ltd. https://www.british-travel-writing.org/. Accessed Jan. 29, 2019.
cox	Bibliographical source	Cox, Edward Godfrey. *A Reference Guide to the Literature of Travel, Including Voyages, Geographical Descriptions, Adventures, Shipwrecks and Expeditions*. The University of Washington, 1935–1949. https://catalog.hathitrust.org/Record/000883216.
gove	Bibliographical source	Gove, Philip Babcock. *The Imaginary Voyage in Prose Fiction: A History of Its Criticism and a Guide for Its Study, with an Annotated Check List of 215 Imaginary Voyages from 1700 to 1800*. Columbia UP, 1941. HathiTrust, http://hdl.handle.net/2027/mdp.39015062655330.
irishmcveagh	Bibliographical source	McVeagh, John. *Irish Travel Writing: A Bibliography*. Wolfhound Press, 1996.
leask	Bibliographical source	Leask, Nigel. *Curiosity and the Aesthetics of Travel Writing, 1770–1840*: "From an Antique Land." Oxford UP, 2002.
murray	Bibliographical source	Innes M. Keighren, Charles W. J. Withers, and Bill Bell. *Travels into Print: Exploration, Writing, and Publishing with John Murray, 1773–1859*: "Appendix: Books of Non-European Travel and Exploration Published by John Murray Between 1773 and 1859: by Date of First Imprint, with Notes on Edition History Before 1901." U of Chicago, 2015.
ncco_c19trav	Bibliographical	"19th Century Travel Literature, from the Bryn Mawr

Tag	Context	Description
	source	College Library." *Nineteenth Century Collections Online*, Gale. Accessed Feb. 19, 2019.
ncco_travelnarr	Bibliographical source	"Travel Narratives and Personal Reminiscences." *Nineteenth Century Collections Online*, Gale. Accessed Feb. 19, 2019.
robinson_w	Bibliographical source	Robinson, Jane. *Wayward Women: A Guide to Women Travellers*. Oxford UP, 1991.
tee	Bibliographical source	*Travels, Explorations and Empires, 1770–1835*, edited by Tim Fulford and Peter J. Kitson, 2 sets of 4 vols. Pickering & Chatto, 2001–2.
alltravel	Content	All volumes in the TWDB with a travel tag from either/both a bibliographical source or a LCSH.
random	Content	Non-travel writing, randomly selected from HT and ECCO. The only filtering was for OCR quality (same process and standards as the TWDB files) and deduplication, as well as limiting periodicals.
lcsh_source	Bibliographical source	Title was added to the TWDB via an LCSH search for travel tags (travel, voyage, antiquit, discover, guidebook, or explor) on ECCO and HT.
lcsh_traveltag	Content	LCSH metadata contains at least one of the travel tags (travel, voyage, antiquit, discover, guidebook, or explor).
lcsh_title_poetry	Content	The LCSH contains "poetry" (poetry, English poetry, narrative poetry, etc). Alternatively, I added it by examining the title text.
nontravel_poetry	Content	The LCSH has no travel tags, but does have "poetry".
lcsh_imaginary	Content	The LCSH contains "voyages, imaginary".
fictitious	Content	Cox lists the title in his "fictitious voyages and travels" section (vol 2, 470–87).
Lcsh_fiction	Content	The LCSH contains "fiction".
lcsh_history	Content	The LCSH contains "history".
lcsh_antiquities	Content	The LCSH contains antiquities".
gender_au	Content	The author claims to be a woman.
gender_non	Content	No TWDB gender tags.
htrandomSecondTest	Content	Non-travel writing, gathered from HT, 1770–1830. Deduplicated from previous data.
htrandomfileSource	File Source	Non-travel writing, gathered from HT.
eccorandomfileSource	File Source	Non-travel writing, gathered from ECCO.
loc_gb	Content	The LCSH contains "description and travel" and location strings related to Great Britain (but no other location categories).
loc_ireland	Content	The LCSH contains "description and travel" and location strings related to Ireland (but no other location categories).
loc_latinam	Content	The LCSH contains "description and travel" and location strings related to Latin America (but no other location categories).

Tag	Context	Description
loc_northam[1]	Content	The LCSH contains "description and travel" and location strings related to North America, except Mexico (but no other location categories).
loc_africa	Content	The LCSH contains "description and travel" and location strings related to Africa (but no other location categories).
loc_arctic[2]	Content	The LCSH contains "description and travel" and location strings related to the Arctic (but no other location categories).
loc_carib	Content	The LCSH contains "description and travel" and location strings related to the Caribbean (but no other location categories).
loc_easia	Content	The LCSH contains "description and travel" and location strings related to east Asia (but no other location categories).
loc_seasia	Content	The LCSH contains "description and travel" and location strings related to southeast Asia (but no other location categories).
loc_meast	Content	The LCSH contains "description and travel" and location strings related to the Middle East (but no other location categories).
loc_oceania	Content	The LCSH contains "description and travel" and location strings related to Oceania (but no other location categories).
loc_russia[3]	Content	The LCSH contains "description and travel" and location strings related to Russia (but no other location categories).
loc_europe	Content	The LCSH contains "description and travel" and location strings related to Europe (but no other location categories).

[1] Search terms did not include "America" on its own, since that conflicted with other terms such as "South America."
[2] Arctic was not modeled because it only had 3 volumes.
[3] Russia was not modeled because alone, it only had 8 volumes.

Appendix 1.2: Description of Sources

1.2.1: Bibliographies

Cox, Edward Godfrey. *A Reference Guide to the Literature of Travel, Including Voyages, Geographical Descriptions, Adventures, Shipwrecks and Expeditions*. University of Washington, 1935–1949. https://catalog.hathitrust.org/Record/000883216.

The largest bibliography on travel writing published in Britain is Edward Godfrey Cox's three-volume *A Reference Guide to the Literature of Travel: Including Voyages, Geographical Descriptions, Adventures, Shipwrecks and Expeditions* (the third volume is subtitled, instead, *Including Tours, Descriptions, Towns, Histories and Antiquities, Surveys, Ancient and Present State, Gardening etc.*). Published in three volumes in 1935, 1938, and 1949, the bibliography lists "in chronological order, from the earliest date ascertainable down to and including the year 1800, all the books on foreign travels, voyages, and descriptions printed in Great Britain, together with translations from foreign tongues and Continental renderings of English works" (1: v). Cox also lists works first published post-1800, as long as the journey occurred before 1800. Although he does not include essays in Royal Society tracts or voyages in the early modern collections (or their later reprints) by Richard Hakluyt or Samuel Purchas, he does record entries in other collections, including Awnsham and John Churchill's *A Collection of Voyages and Travels* (1704; Cox cites the 1732 edition), John Harris's *Navigantium atque Itinerantium*

Bibliotheca (1705; Cox cites the 1744 edition), and Thomas Astley's *A New General Collection of Voyages and Travels* (1745–7).[4]

While the three volumes hold thousands of titles, over three hundred titles appear multiple times across volumes and are cross-referenced with their other entries.[5] Cox notes the impossibility of this categorical task, writing that "The geographical sections under which works are listed are not and cannot be sharply dividing. It will be evident sometimes that a given title could just as well have been placed elsewhere" (1: v). He does not, however, describe his methodology, and he clearly did not read or research every text; for example, he includes the novel *Millenium Hall* under "Descriptions" rather than "Fictitious Voyages and Travels." He knows other titles better, however, with many entries noting later editions, translations into or from other languages, and other details. He does not distinguish between works with no author on the title page or with other taglines or pseudonyms.[6]

The thirteen sections of Volume 1 focus on "The Old World," dividing Europe and the "East" into variations on "Continental," "West," "Near," and "Far." Siberia gets its own category. Cox also uses Volume 1 to include general sections on collections, circumnavigations, and general travels and descriptions, many of which appear cross-referenced in later categories. As with Volume 1, Volume 2, "The New World," uses geographical categories for many of the section headers but includes more specific locales. Areas like "The Northwest Passage," "The Northeast Passage," "Arctic Regions," and "The South Seas" all get their own section, in

---

[4] *A Collection of Voyages* is commonly known as the Oxford Collection or, after the editor, Osborne. *A New General Collection* is sometimes called the Harleian collection.
[5] For example, James Cook's *A voyage to the Pacific Ocean, for making discoveries in the northern hemisphere* (1784) is listed in "Circumnavigations," "North Pacific," and "Arctic Regions."
[6] For an overview of anonymity and publishing in the eighteenth century, see Gillian Paku's "Anonymity in the Eighteenth Century" and James Raven's "The Anonymous Novel in Britain and Ireland, 1750–1830."

addition to continents such as "North America" and "Australia." The second volume also has sections based on subgenre, including directions for travellers, military expeditions, naval expeditions, adventures, disasters, shipwrecks, and fictitious voyages and travels.

Great Britain is the focus of Volume 3, including "Tours by Natives," as well as several categories that may not immediately appear to be "literature of travel," including "Natural History," "Canals, Rivers, Fen Drainage," "London," and three sections on different kinds of antiquities. Because of these ambiguous subtitles, Cox adds headnotes to each section of Volume 3, which "attempt to tell somewhat of the lay of the land and to account for the choice of sections and items listed" (3: ix). At 638 pages of bibliography (over 700 if including his list of general reference works, other bibliographies, and an index) and 22 sections, this third volume is the largest.[7] In his preface, he notes that Volume 3's "ever-expanding bulk shut out the discovery of Ireland" (an uncomfortable phrase, considering the relationship of England and Ireland); the planned Volume 4 on Ireland exists only in manuscript.[8] Instead, John McVeagh's *Irish Travel Writing: A Bibliography* (1996), also included among the TWDB sources, picks up where Cox leaves off.

Cox's bibliography remains the largest bibliography on travel writing published in Britain, but still not definitive. Although Cox's definition of travel writing is the broadest of all the TWDB's bibliographers, his expectation that "Many titles must of necessity have escaped my net…But I can well believe that what is missing will be found to have little renown," combined with the limitations of working before mass digitization efforts and metadata sharing practices is,

---

[7] Volume 1's 401 pages consisted of 13 sections, and Volume 2 has 556 pages and 21 sections.
[8] The papers of Edward Godfrey Cox are held in Special Collections, University of Washington Libraries, Accession No. 0476–001.

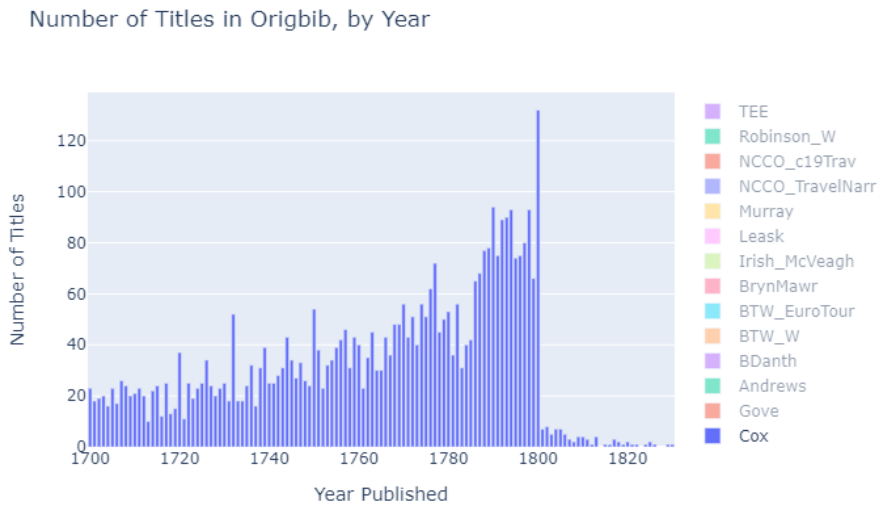
on its own, cause for caution (1: v). Indeed, though Cox's work is exceptional considering his constraints, as William H. Sherman notes, "Cox's listing is neither comprehensive nor entirely accurate: his coverage of "Fictitious Travels" (particularly in plays and poems) is especially incomplete, and some of the authors, titles, and dates of publication have been corrected by more recent reference materials" (21). Benjamin Colbert calls Cox's work a "fair representation of travel-related books published in English since the dawn of printing, as well as many helpful textual annotations, yet his volume on Europe stops at 1800, his regional categories are often too inclusive for specialised use, nearly all his entries lack imprint details, and there are many errors and omissions" ("European" 7). Still, at over 4,000 unique entries, his work remains a monumental starting point. For this project, I omitted entries primarily listed in Cox's sections on maps, atlases, and navigational charts. I also removed entries referring to letters and essays published in the *Gentleman's Magazine*, as well as entries referring to excerpted texts in the many collections.[9]

Table 1.2.1: Cox Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
cox	3846	4072	0.944499

[9] Although Cox does not include individual entries for collections by Purchas and Hakluyt, he does include individual entries for excerpts in Callander, Churchill, Astley, and Harris.

Figure 1.2.1: Cox Titles in TWDB-origbib



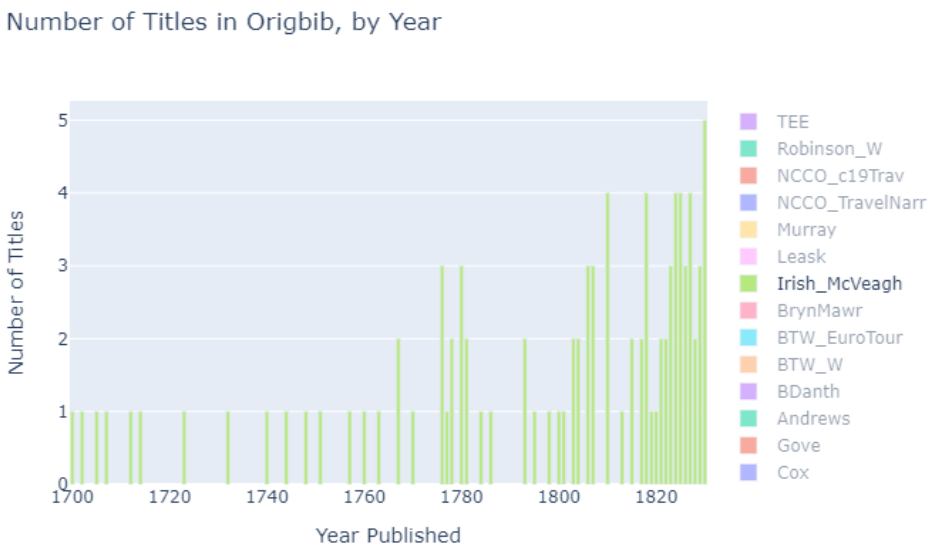
McVeagh, John. *Irish Travel Writing: A Bibliography*. Wolfhound Press, 1996.

This bibliography is the first to focus on travel writing in Ireland, and fills an important gap left by Cox's missing fourth volume. John McVeagh uses a large time frame, from approximately 1200 to the bibliography's publication in 1996. For McVeagh, considering the "vexed question of categories…an inclusive policy on the matter…seemed the most sensible option" (14–5). Therefore, rather than a narrow definition, he includes texts on subjects as varied as "sailing, lepidoptery, famine relief, evangelism, or some other primary concern…as they involve the description of travel through Ireland" (15). Unlike Cox's bibliography, McVeagh's contains guidebooks.

Table 1.2.2: McVeagh Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
irishmcveagh	81	94	0.861702

Figure 1.2.2: Cox Titles in TWDB-origbib



Colbert, Benjamin. *Database of British Travel Writing, 1780–1840*.

A significant forthcoming resource is the *Database of British Travel Writing, 1780–1840* (BTW), a project led by Benjamin Colbert. The project aims to provide an online bibliographical database of travel writing published in Britain and Ireland between 1780 and 1840. Colbert defines travel writing in "terms of genre (narratives, guidebooks, illustrated letterpress plate books, topographical descriptions, and collections); witness (accounts derived from actual tours); and place of publication (Britain and Ireland)" ("About"). BTW also excludes fictional travel writing, such as novels and poetry, though it includes storybooks for children that are based on real travels. The project also does not include many practical paratexts of travel—such as city and commercial directories, seats and buildings, and travellers' phrasebooks—and excludes some subgenres such as shipwreck literature ("unless containing important travel content"), histories, geographical works derived from travellers' accounts second hand, works about travel, and

descriptions of seats, buildings, or churches. In total, BTW includes information for 5,000 travel books published by 3,000 authors, a final count higher than that of the TWDB-origbib even though the TWDB-origbib spans 130 years compared to BTW's 60. According to the online excerpt *Women's Travel Writing, 1780–1840*, BTW draws on over 25 different sources and databases for its texts. While many of the TWDB and BTW sources overlap, BTW does not list Gove, Andrews, the Bryn Mawr Library, or Keighren, Withers, and Bell's Murray bibliography.

Although the entire database is still in development and unavailable (as of April 2024) to the public, two excerpts are published.[10] Because of their differences, I have recorded each as a separate source in my bibliography.

"Bibliography of British Travel Writing, 1780 - 1840: The European Tour, 1814 - 1818 (Excluding Britain and Ireland)." ***Romantic Textualities: Literature and Print Culture, 1780–1840*****, vol. 13, 2004, pp. 5–43.**

This checklist focuses on European (or continental) tours (therefore excluding Britain and Ireland), published in Great Britain between 1814 and 1818. Colbert published this checklist in 2004 and he has since refined the criteria for BTW: for example, in BTW–Europe, Colbert includes English-language books published abroad, but intended for British tourists, as well as the "Histories" and "shipwreck literature" eschewed in his current project description. In the article accompanying the checklist, Colbert notes how even this small frame of time can offer "intriguing statistics when placed in the context of total travel book production" (9). He also draws attention to regional coverage and women writers in the list.

---

[10]As BTW was the subject of funding bids, Colbert declined to share the larger dataset in early 2017. A larger release, public or paywalled, has not yet appeared (Colbert *Database*).

Full Description of btw_europe criteria

1. First-hand accounts of actual travels or residences abroad, including translations, new editions of older material, campaign journals, and shipwreck, castaway, and captivity narratives.

2. Travellers' aids, including itineraries, guidebooks, 'companions', road manuals, regional descriptions, and atlases (when accompanied by letterpress).

3. Collections, anthologies, digests, abridgements, and histories of travel and exploration, including travel compendiums designed for children, and biographies of travellers.

4. View books (collections of engravings and etchings accompanied by letterpress).

5. 'Virtual' tour books accompanying panoramas and exhibitions. (9)

***British Travel Writing: Women's Travel Writing, 1780–1840: A Bio-Bibliographical Database*. 2018, https://btw.wlv.ac.uk**

The second BTW excerpt, the online *Women's Travel Writing, 1780–1840* (BTW–Women's excerpt) contains all the known books of travel published in Britain and Ireland by women between 1780 and 1840. In a 2016 description of the excerpt and some preliminary findings, Colbert indicates that of the BTW's roughly 5,000 travel books by 3,000 authors, "only 204 can be identified as produced by 146 women," meaning that "women accounted for only around 5% of travel books published in Britain and Ireland during this period (a 20 to 1 ratio of men to women)" ("Reflections" 156). More recent iterations of the database have more titles, perhaps because he has expanded the project to include women translators ("About").[11] He includes texts

[11] For example, in June 2019, the BTW–Women's excerpt listed 220 titles.

where women are principal authors or coauthors, as well as editors, illustrators, and translators, "so long as contributions to the book are not peripheral" ("About").

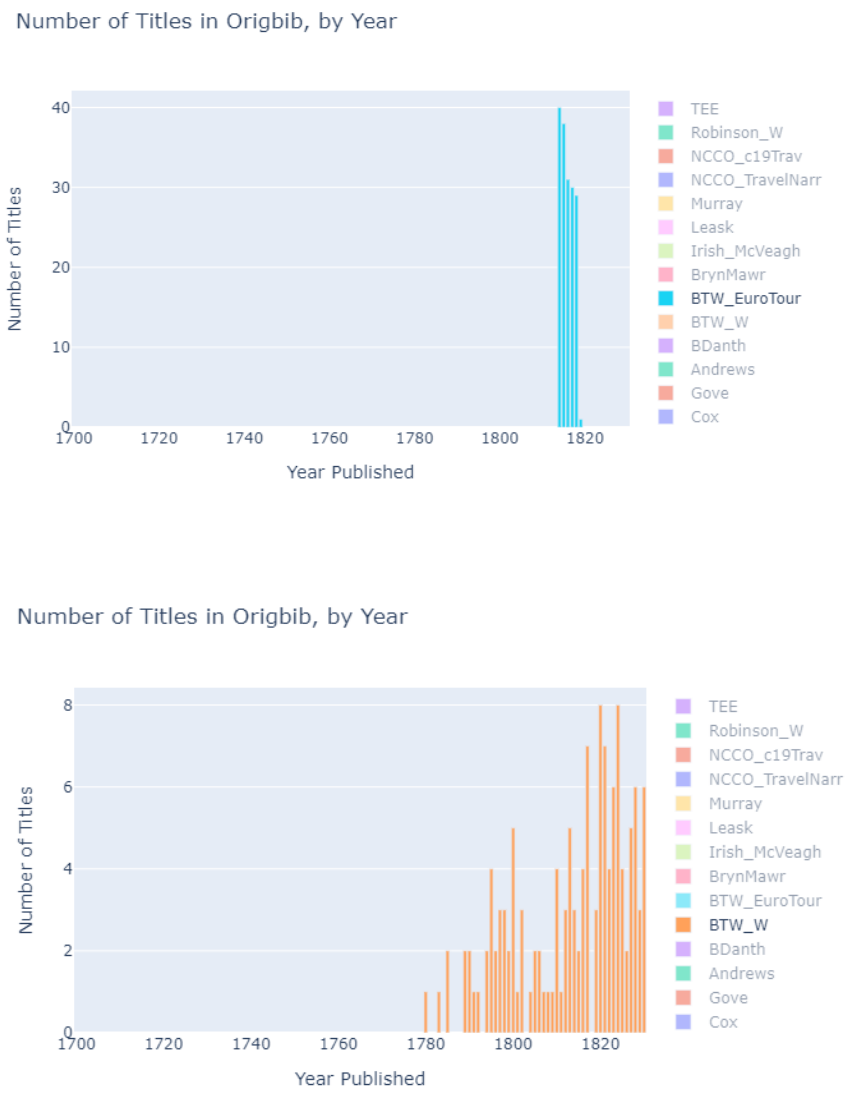
The *Women's Travel Writing* database is not currently downloadable en masse; rather, users can browse by author, illustrator, publisher, or printer. Creator fields include author, contributor, editor, and translator, though the database (at least, in its public iteration) does not include fine-tuned variables such as location of authors' names.[12] Nonetheless, Colbert and his team often inspected physical editions and created biographies of all the figures in their database; as Peiser notes in her review of the Women excerpt, "several of the authors featured in the database do not have an entry in [the *Oxford Dictionary of National Bibliography* or *The Orlando Project*]" ("Review"). *Women's Travel Writing* also lists previous sources that list each text, contemporary reviews, and links to online digital sources.[13] These links were especially useful when adding digital editions to the TWDB.

Table 1.2.3: BTW Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
btw_w	74	134	0.552239
btw_europe	152	169[14]	0.899408

[12] See chapter 2 for a discussion of encoding gender.
[13] For Colbert's two publications that draw on his statistics from *Women's Travel Writing,* see "British Women's Travel Writing, 1780–1840: Bibliographical Reflections" (2016) and "'Our observations should not be disunited': Collaborative Women's Travel Writing, 1780–1840" (2016).
[14] The lone title in 1819 is Uno von Troil's *Iceland; or the Journal of a Residence in that Island*. NCCO and Bryn Mawr (which I collected before BTW) listed the publication date as 1819 rather than btw_europe's 1818.

Figure 1.2.3: BTW Titles in TWDB-origbib



Robinson, Jane. *Wayward Women: A Guide to Women Travellers*. Oxford UP, 1991.

Jane Robinson's select bibliography lists over four hundred women travellers, from Margery Kempe into the twentieth century, recording the first editions of only their "*first-hand* travel accounts *in book form*…not [including] articles, fictional or historical works, and so on" (ix). Like Colbert in the BTW, she pays similar attention to the truth of works, though scholarly

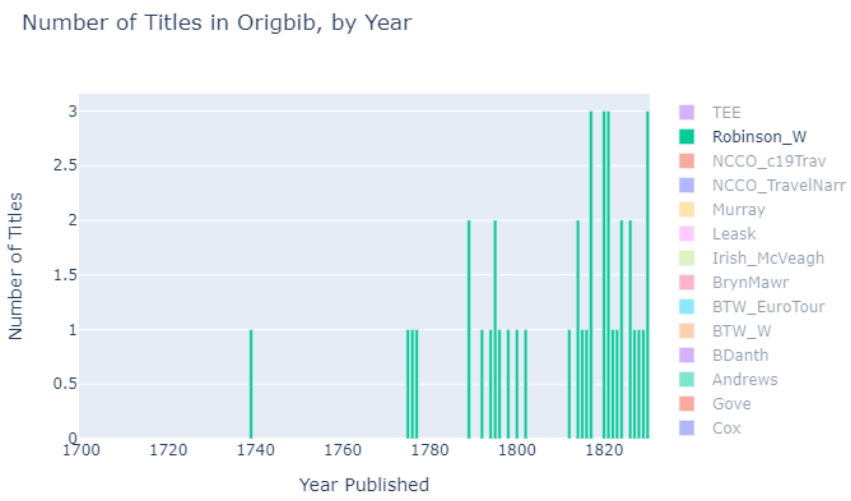
positions on some of the listed titles have changed.[15] As Robinson notes, her selections "all us[e] English as a first language, [are] mostly of British extraction, and [are] always travelling beyond the frontiers of their native land" (x). Rather than time period, the organization focuses on the "sort of traveller": "pioneer (as in the opening chapter 'Untrodden Peaks and Unfrequented Valleys'), evangelist ('Quite Safe Here with Jesus'), emigrant ('Life in the Bush'), and so on" (ix). Most of the texts that she records are cross-referenced with others in the corpus, particularly in the BTW–Women's excerpt (which lists Robinson as a source).

Table 1.2.4: Robinson Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
robinson_w	5	40	0.125

[15] For example, she includes Anna Jameson's *A Lady's Diary* (republished as *The Diary of an Ennuyée*) in 1826, which the BTW–Women's excerpt describes as "a fictional account of the travels and early death of its romantic heroine and supposed author, actually based on Jameson's own experiences on the continent" ("Anna Brownell Jameson")

Figure 1.2.4: Robinson Titles in TWDB-origbib



Gove, Philip Babcock. *The Imaginary Voyage in Prose Fiction: A History of Its Criticism and a Guide for Its Study, with an Annotated Check List of 215 Imaginary Voyages from 1700 to 1800*. Columbia, 1941.

Philip Babcock Gove is one of the few critics to focus on the imaginary voyage in eighteenth-century travel writing, filling a gap in travel criticism which tends to focus on "real" travel. His analysis of the history of the term of imaginary voyage and its related scholarly criticism is invaluable, particularly for his mixture of English, French, and German literature. The TWDB-origbib included texts in other languages upon their first translation to English.[16]

Gove identifies the imaginary voyage as "an organic, shifting division of fiction, recognizable, but indefinable as a static, fixed, and exclusive genre" (viii). Definitions that might work for an early part of the century do not work for the later, indicating the changing nature of

[16] Of the 215 texts that Gove lists, 67 were originally published in English, 65 in French, 59 German, 10 Dutch, with Danish, Swedish, Italian, Latin, and Japanese all at 5 or less.

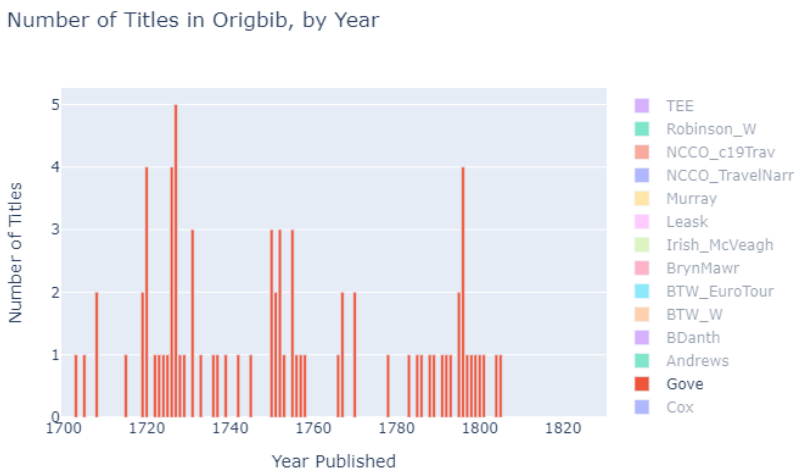
the genre over the decades. His description of the checklist spans several pages, describing in detail his requirements. In general, Gove focuses on the terms of the voyage itself: the voyage cannot be merely summarized in a paragraph, as it is in Psalmanazar's *Description of Formosa*, but must be a significant portion of the plot. Some qualifications are, as Gove admits, arbitrary: in addition to "the terrestrial voyage" being "preferably oceanic, not coastal or thalassic," he also discards any travel primarily in the Mediterranean Sea, as well as any narratives limited to historical battles and expeditions or that use travel just to move characters from one familiar port to another. Gove also eliminates voyages in dreams, by means of familiar, cabalistic, or Rosicrucian, or to the land of the dead.[17] Despite these limitations, nearly two-thirds of the titles from his checklist were not listed in any of the other sources, demonstrating the significance of his work.

Table 1.2.5: Gove Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
gove	49	77	0.636364

[17] Gove notes that the dreams qualification is "one of the most difficult to apply, he and recommends the following test: "after removing the prefatory remarks, the as-I-fell-asleep first paragraph, and the then-I-awoke final paragraph, one should try to decide whether the narrative is free enough from the dream atmosphere to justify calling it an imaginary voyage" (176n277).

Figure 1.2.5: Gove Titles in TWDB-origbib



Keighren, Innes M., Charles W. J. Withers, and Bill Bell. *Travels into Print: Exploration, Writing, and Publishing with John Murray, 1773–1859*: "Appendix: Books of Non-European Travel and Exploration Published by John Murray Between 1773 and 1859: by Date of First Imprint, with Notes on Edition History Before 1901." U of Chicago P, 2015.

Innes M. Keighren, Charles W.J. Withers, and Bill Bell's work is a comprehensive list of Murray's 239 books of non-European travel published between 1773 and 1859. They include works that focus on Europe, as long as the texts include some travel beyond the continent. Their bibliography "does not include works of formal geographical instruction, such as geographical gazetteers, but it does include works of topographical description . . . [as well as] practical guides to scientific travel" (22). They also omit narrative poems such as Lord Byron's *Childe Harold's Pilgrimage*. In addition to title, author, and imprint information, they also include the book format, the price, and the ESTC or NSTC number.

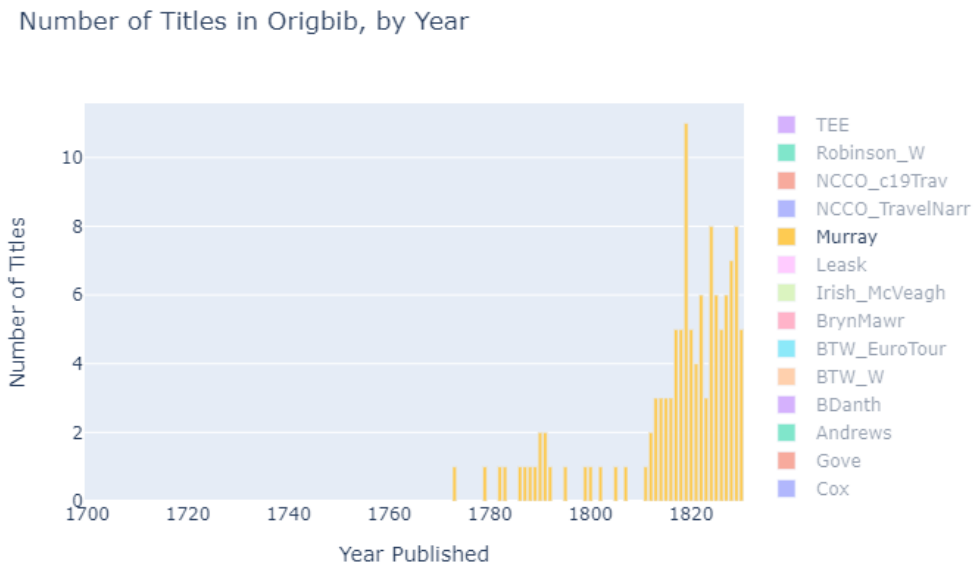
This source provides another access point to titles that are not focused exclusively on Europe, as well as a corpus that spans the century divide. It also offers quick access to texts published by a single printer—and one important to travel writing in general—without requiring

my own intervention. Humphrey Carpenter suggests that Murray's publication of travel books, rather than literary publications, is the publishing house's "greatest contribution to the advancement of knowledge and of human understanding of the world" (qtd. in Keighren et al. 6). Murray House's involvement in the world of travel writing also extended to the political and critical. From 1813, it was the official publisher to the Admiralty and to the Board of Longitude, and therefore published "most British narratives of Arctic and African exploration" (7). From 1809, the house of Murray also published the *Quarterly Review*, wielding a second level of influence on the travel writing industry. As the authors of this source assert, "the house of John Murray was at the center of the production of travel narratives" (32).

Table 1.2.6: Murray Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
murray	63	118	0.533898

Figure 1.2.6: Murray Titles in TWDB-origbib



1.2.2: Anthologies

Bohls, Elizabeth A., and Ian Duncan, editors. *Travel Writing, 1700–1830: An Anthology*. Oxford UP, 2005.

Elizabeth A. Bohls and Ian Duncan's collection is one of the most accessible contemporary anthologies; Carl Thompson calls it "an absolute godsend" for instructors teaching eighteenth-century travel writing ("Review" 304). The anthology contains only non-fiction prose travel writing mostly produced between 1700 and 1830, and it includes unpublished sources such as Captain Cook's original journals, as well as documents by women. The editors acknowledge their caution in choosing texts: they did not favour writing that is "romantic or ironic or otherwise highly 'literary' in the present-day sense of the term" or that "self-consciously foreground[s] the writer's personality or persona" (xvii). Instead, they seek "to do justice to the range and diversity of material that was written, published, purchased, and read concerning travel…and the variety of writers who produced it" (xvii). A key part of this goal is their inclusion of excerpts from autobiographies of enslaved persons and other non-leisure travellers, "partly to encourage readers to reconsider and broaden their understanding of what constitutes travel" (xvii).

At least one reviewer criticizes this decision: Kate Durie characterizes the inclusion of "involuntary travellers" such as Olaudah Equiano and Mary Prince, or those travelling for service or work, as the "most striking—and debatable" element of the anthology. She warns that "If the label 'travel writing' is stretched too far, it becomes almost meaningless," and states that while texts by such authors "may reflect something of the context of travel (such as the conditions on board ship), [they] rarely have an eye to place, to custom, to strangeness and difference" (927). Thompson, on the other hand, praises the anthology's inclusion of "'landmark'

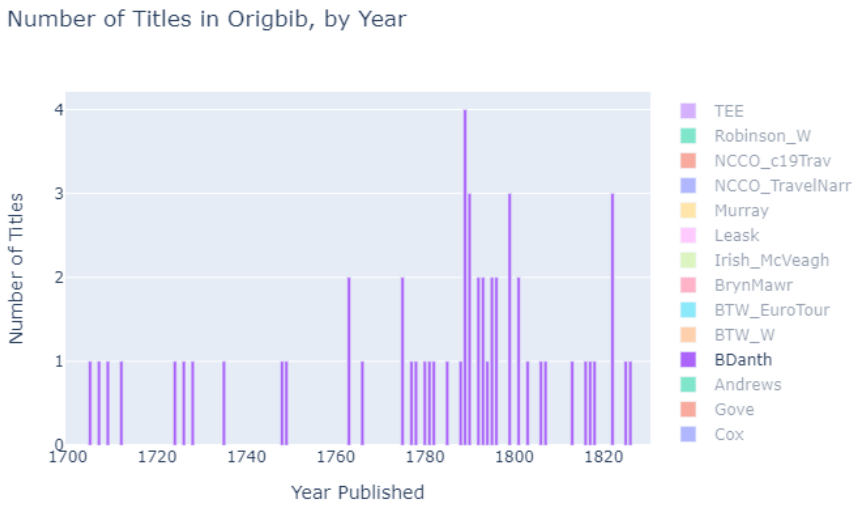
texts of the form" and the "contrasting perspectives" of Equiano, Prince, and sailor John Nicol, as well as the editors' choice of excerpts that draw attention to "unlettered individuals" and "numerous unfortunates who in many cases might be more accurately categorized as 'displaced persons'" (303). He does note, however, that the editors do not include sections on the Arctic, South America, or India.[18] As these reviews reflect, *Travel Writing 1700–1830: An Anthology* is an accessible entry point (in content, critical context, and price) for leisure readers and, more importantly, for the classroom. The choice of titles and content in comparison to the rest of my corpus gives insight to the perspective this anthology presents to students. Like Bohls and Duncan, the TWDB also tends toward "stretched" definitions of travel writing, making this anthology especially useful.

Table 1.2.7: Murray Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
bdanth	8	55	0.145455

[18] Though South America does not get a section in the volume, Thompson notes that "the Caribbean section does include John Gabriel Stedman's account of his experiences in Suriname" (303).

Figure 1.2.7: Murray Titles in TWDB-origbib



Fulford, Tim and Peter J. Kitson (editors). *Travels, Explorations and Empires, 1770–1835*. 8 vols. Pickering & Chatto: 2001–2.

This anthology spans 8 volumes and 93 facsimile excerpts from 88 different published works published between 1772 and 1857. The editors' goal is to present "exploration" narratives that had "a direct influence on imperial policy": texts by ambassadors, envoys, explorers on government missions, or other figures that influenced policy, where the authors could "see, with imperial eyes, lands and cultures that Europeans had rarely seen" (xxvii). They omit writing on Britain and continental Europe, and some locations, such as Africa, are still underrepresented (Carretta 258). The authors are largely British, but also American and European, as their texts— often published in London, sometimes in translation— "shaped British perceptions and policies" (xxvii). Women and Indigenous people are also missing from the corpus, except for the English Mary Ann Parker (*A Voyage Round the World, in the Gorgon Man of War*, 1795): although the editors tried to "reflect the different perspectives" of these groups, the "central concentration is

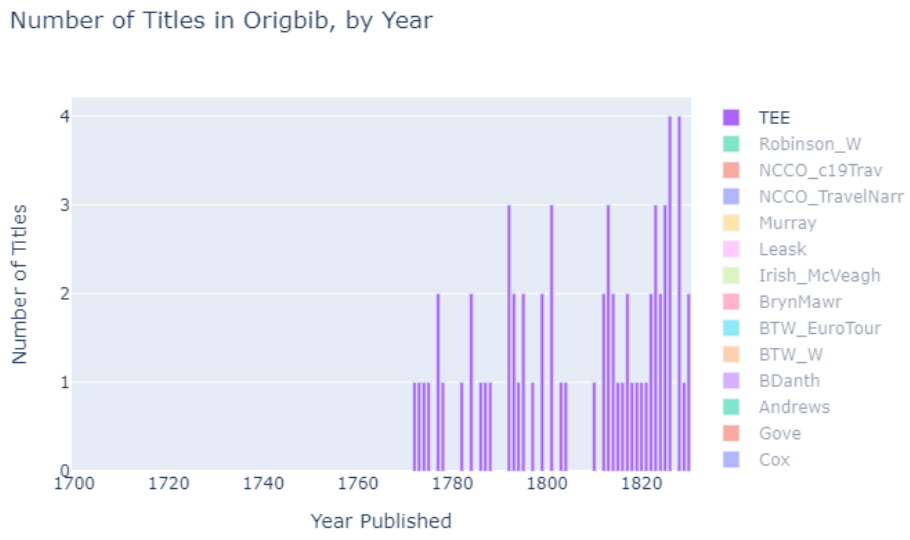
on white men, since it was mostly they who, in the service of science, commerce and their nations, explored areas largely unfamiliar to the West" (xxvii). Early in the introduction, the editors also emphasize the inclusion of texts that influenced literary authors, scientific development, or the construction of Britain's empire.

The nature of the collection—eight volumes, reproducing facsimile pages, in hardback, and costing hundreds of dollars—places this anthology in the realm of academic libraries or the dedicated scholar rather than an undergraduate classroom or casual reader. This source is most useful, then, as a shortcut to texts that influenced British policy.

Table 1.2.8: TEE Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
tee	13	66	0.19697

Figure 1.2.8: TEE Titles in TWDB-origbib



1.2.3: Archives

***Nineteenth Century Collections Online*: Mapping the World: Maps and Travel Literature. Gale: 2014.**

Nineteenth Century Collections Online (NCCO) has had a long influence on digital research and draws directly from particular archives. Although the availability of free eighteenth- and nineteenth-century facsimile resources via HathiTrust and Google Books may be reducing the dominance of Gale products, NCCO—in its now defunct original interface or the newer Gale Primary Sources interface—is a familiar resource (with more modular, topical organization than HathiTrust or Google Books) for educators and researchers looking for works as teaching tools or an introduction to a topic. If scholars rely primarily on these collections, how is their exposure to texts different from browsing travel writing elsewhere?

In NCCO's set of modules, "Mapping the World: Maps and Travel Literature" is the most explicitly focused on travel.[19] The module's description does not include an advisory board or curator; however, on one promotional item for the module, a note lists Dr Jordana Dym, Associate Professor (now Professor) of History at Skidmore College as curator ("Mapping the World: Maps and Travel Literature" fact sheet).The majority of Mapping the World's collections (and therefore, its description on NCCO) consists of maps of various kinds, but a few modules include extensive textual resources. The module draws attention to these specific collections in the final paragraph of the description: "Bryn Mawr's extensive collection of European Travel accounts provides a sweeping glance of the travel narrative genre. In addition to the Bryn Mawr Collection, selected travel narratives have been included from the collections of the American

[19] Other modules, such as "Science, Technology, and Medicine: 1780–1925," also include texts that other TWDB sources would consider travel writing.

Antiquarian Society and the British Library" ("Maps and Travel Literature"). Instead of combining these two modules into a singular NCCO source in the TWDB data, I separate them to highlight their significant differences.

NCCO: 19th Century Travel Literature, from the Bryn Mawr College Library

This collection is derived from the Bryn Mawr College Library and includes titles ranging from the late eighteenth to the early twentieth century.[20] According to NCCO's online description, the archive was chosen not only because of its size, but because "Among the collection's strengths are the travel accounts written by Europeans travelling throughout the world, beginning with works issued in the late fifteenth century" ("19th Century Travel Literature").[21] NCCO's description of the collection is brief, including a few topical highlights of the collection; instead, they link directly to a (now defunct) Bryn Mawr guide, which describes how the titles "cover a striking variety of topics" and "all fields of interest": "texts that will be of interest to scholars of nearly every discipline." The Bryn Mawr librarians include only "writings [that]…are considered 'factual' (in contrast with works of evident fiction)," despite their immediate acknowledgement of the "tension between truth and fiction" in the genre ("European Travel Accounts Printed Before 1850: Description").[22] Because NCCO is drawing from this archive, then, we might assume the "19th Century Travel Literature" collection has a similar shape.

---

[20] This description is different from the accompanying "Collection Facts," which lists the date range as 1800–1899.
[21] NCCO also points to the significance of Bryn Mawr's history as one of the first women's colleges in their overview of the collection, but it is unclear how exactly the collections (either in general, or travel specifically) "reflect" this unique history, other than including "accounts by women travelers" in a list of "highlights."
[22] This Guide was also linked to by the main Bryn Mawr library page, which called the Guide "a partial bibliography." Updates to the Bryn Mawr page have since removed these descriptions.

However, when comparing these two sources, several discrepancies in both raw numbers and titles arise.[23] These differences highlight one simultaneous dis/advantage of digital projects: they are relatively easy to update, but those changes may not be recorded, explained, or used to update other resources in turn. NCCO's description of "19th Century Travel Literature" does not explain whether their editorial decisions tended toward total or selective inclusion of the Bryn Mawr "catalogue" (or even what exactly that catalogue was, considering that it may not be the linked Guide) or when the digitization took place.[24] The catalogue is now only accessible in an archived version via the Wayback Machine.[25] Without contacting Bryn Mawr librarians or Gale representatives directly, the user cannot know whether some titles are missing because they are new acquisitions or excluded because of an unknown editorial mechanism.

Table 1.2.9: NCCO-Brynmawr Statistics in TWDB-origbib

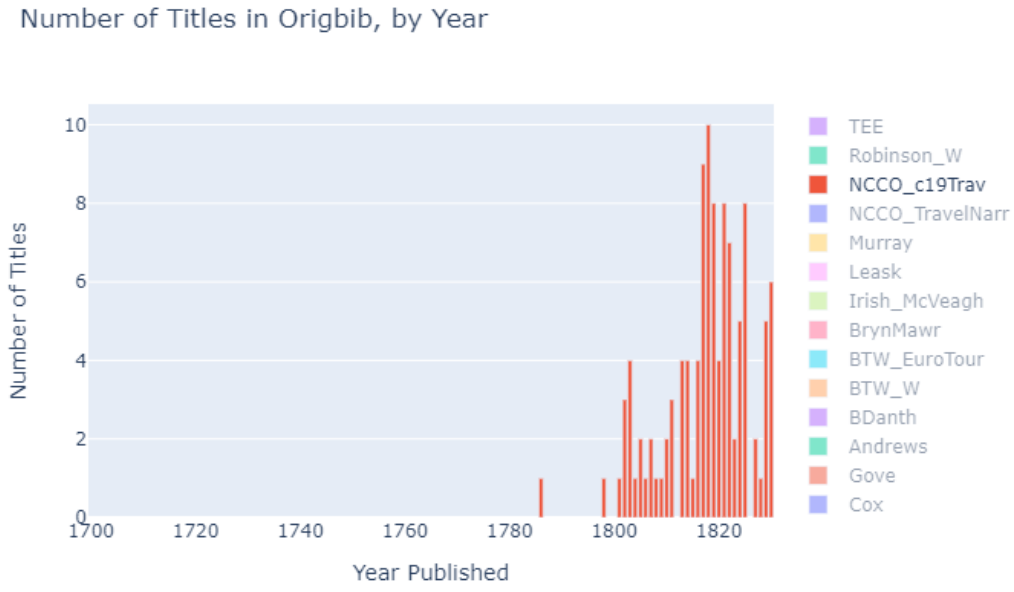
TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
brynmawr	43	200	0.215
ncco_c19trav	36	111	0.324324

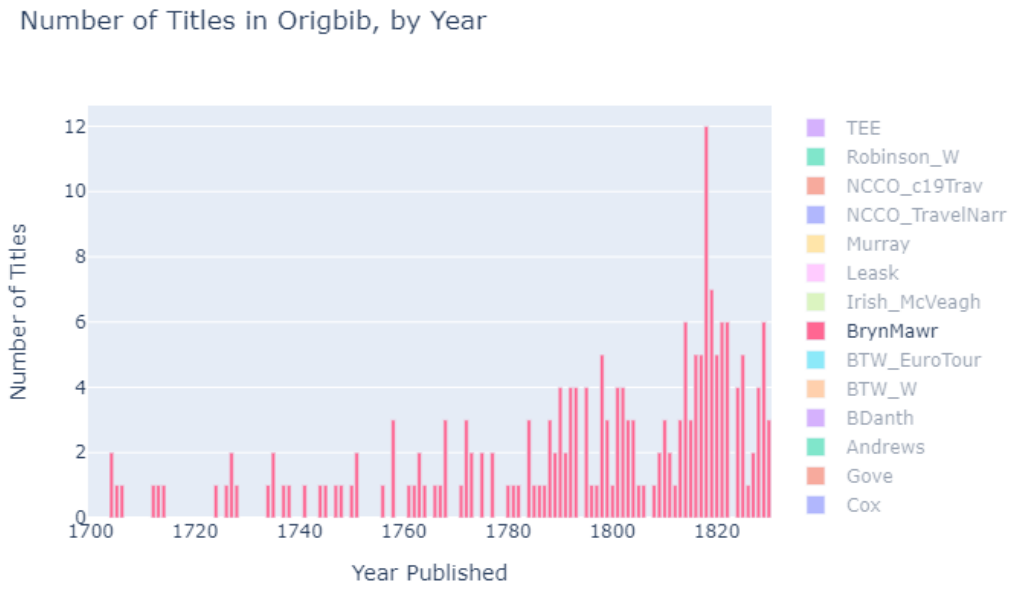
[23] If we take as our start date 1776 (the earliest work included in NCCO_c19Trav is the 1776 *Mémoires Concernant l'Histoire, les Sciences, les Arts, les Mœurs, les Usages, &c. des Chinois*; a 1733 edition of *Lettres Édifiantes et Curieuses: Écrites des Missions Étrangères* is misdated in the NCCO metadata), 85 titles are listed in Bryn Mawr, but not in NCCO_c19Trav. Bryn Mawr lists an additional 47 titles from 1704 to 1775. And this discrepancy goes in the other direction as well: 50 titles are listed in NCCO_c19Trav but not in the Bryn Mawr Guide. I searched for a handful of these titles in the main Bryn Mawr Library catalogue, and the Library does hold them; they are just not listed in the Guide linked to by NCCO.

[24] Depending on the digitization date, NCCO would not include recent Bryn Mawr acquisitions. Similarly, the Bryn Mawr Guide webpage was last updated January 24, 2018, but the update's content is unknown.

[25] Bryn Mawr Library is experimenting with its own display of the page, though it is broken as of May 18, 2024. See https://digitalcollections-staging.tricolib.brynmawr.edu/node/315.

Figure 1.2.9: NCCO-Brynmawr Titles in TWDB-origbib





Travel Narratives and Personal Reminiscences, from the British Library

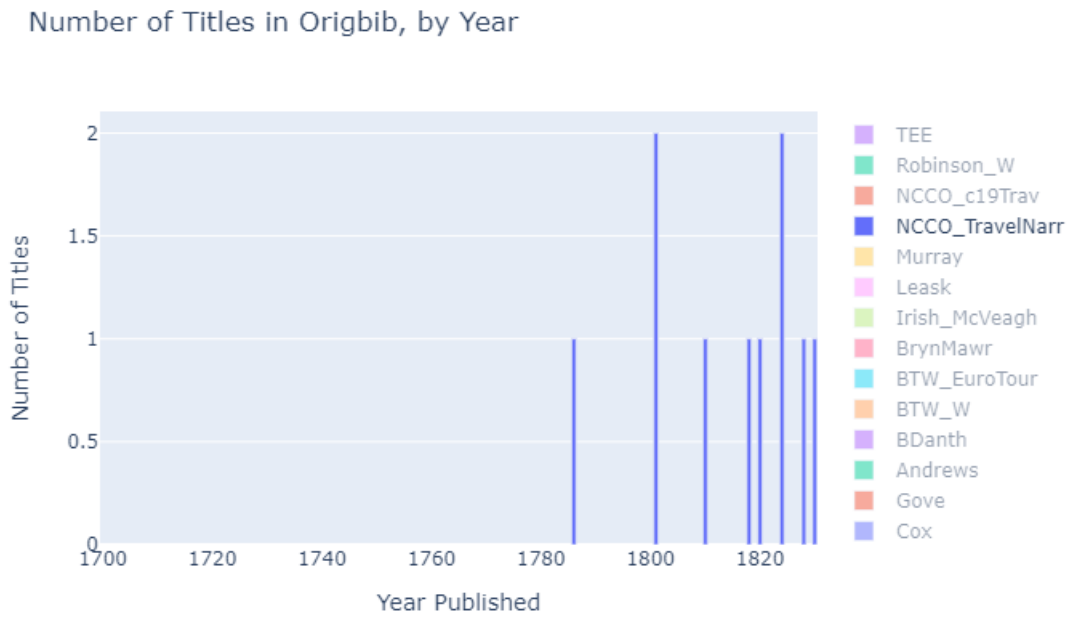
This second NCCO collection, sourced from the British Library, also focuses on the nineteenth century and reaches back to the eighteenth century and forward into the twentieth.[26] The collection's stated purpose is to "give researchers a perspective on English and French travelers and the places they traveled to" ("Travel Narratives"). The description includes no other information about editorial decisions, and it does not link to any resources from the British Library. Overall, the collection includes 288 volumes and 160 English-language titles.[27] Although the British Library certainly holds more than 160 such works, the reasons for focusing on these titles is unknown beyond the above stated purpose. Only ten titles, however, fit within the scope of the TWDB, which limits use of the collection.[28]

Table 1.2.10: NCCO-Travel Narrative Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
ncco_travelnarr	5	10	0.5

[26] The earliest work is *Baron Munchausen's Narrative of His Marvellous Travels and Campaigns in Russia* (1786); the collection also includes a 1496 edition of Mandeville's *Itinerarius.*
[27] The description does not list the total number of titles.
[28] For example, predictive models' accuracy of this subset range from 70% to 90%. Ted Underwood notes that when "modeling a homogenous group of works, accuracy ordinarily increases as one gathers more data" (*DH* 55).

Figure 1.2.10: NCCO-Travel Narrative Titles in TWDB-origbib



1.2.4 Critical Works

Andrews, Malcolm. *The Search for the Picturesque: Landscape Aesthetics and Tourism in Britain, 1760–1800*: "Select Bibliography." Stanford UP, 1989.

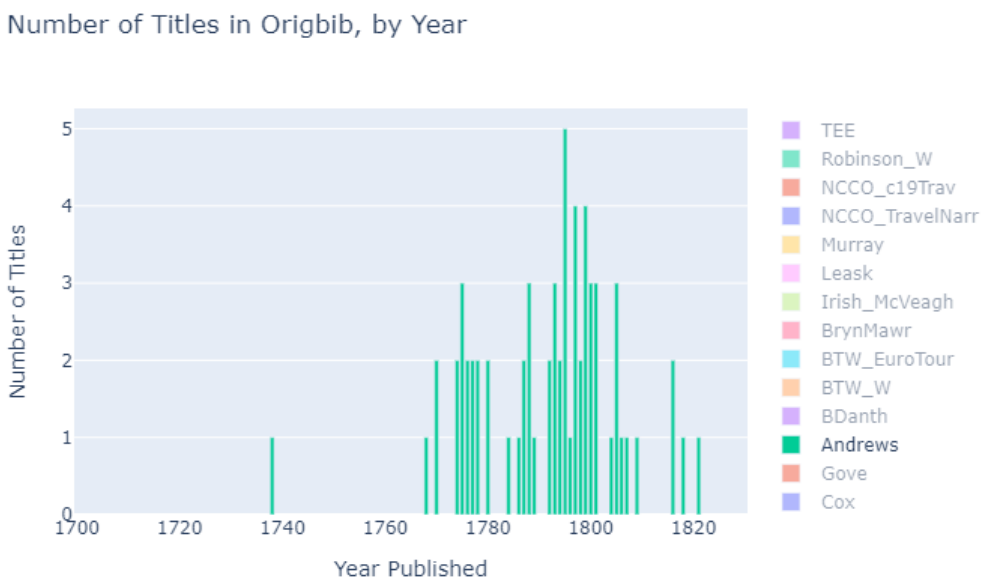
In Malcolm Andrews' work on the picturesque, he includes a select bibliography of texts consulted for his work on the primary picturesque tours in Britain: the Wye valley, North Wales, the Lake district, and the Highlands. This bibliography offers the dual opportunities of regional comparisons and texts considered key to the development of the picturesque. Reviews at the time of its publication varied; Stephanie A. Ross appreciated both the theoretical analysis of the picturesque and the descriptions of the tours, noting that "my only disappointment with this book is that he did not tell us in much more detail to what extent the picturesque tours he describes can still be traced today" (250). Ronald Paulson, however, was skeptical of *Search*, seeing it as a "felicitous, popular survey of materials that are available in more original and speculative

formulations elsewhere," criticizing how Andrews relies mostly on work published more than a decade earlier, rather than more contemporary scholarship (873). His final, damning judgment, that "the virtues of the book tilt in the direction of the coffee table," seems to have been ignored by subsequent scholars, however; *Search* has hundreds of citations, placing it firmly among the most seminal of works on the picturesque tour.[29]

Table 1.2.11: Andrews Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
andrews	13	65	0.2

Figure 1.2.11: Andrews Titles in TWDB-origbib



[29] In June 2019, Google Scholar listed 712 citations, and in April 2024, over 1000 citations. In a future iteration of the TWDB, Peter Bicknell's 1990 *The Picturesque Scenery of the Lake District: 1752–1855: A Bibliographical Study* would offer a useful comparative corpus.

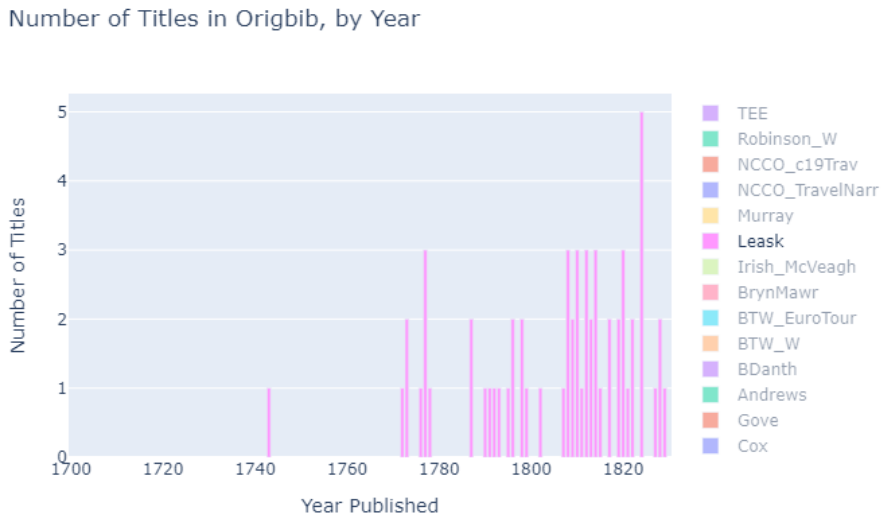
Leask, Nigel. *Curiosity and the Aesthetics of Travel-Writing, 1770–1840*: "From an Antique Land." Oxford UP, 2002.

Nigel Leask's criticism is an important intervention in discussions of the subjectivity of travel writing. Leask focuses on travel outside of the usual boundaries of Europe (and, in particular, Italy). Instead, he pays close attention to travel writing about Ethiopia, Egypt, India, and Mexico, "countries situated within the 'torrid zones' of Africa, Asia, and America, which, despite their cartographic and cultural distance from one other, shared the fate of being considered 'antique lands' by Europeans" (1). In keeping with the focus on antiquity, he positions his work as being "principally concerned with aesthetic and archaeological (in contemporary terminology, antiquarian) discourses of travel" (1). While examining these texts—rather than the more popular critical trend of science of natural history books—he challenges many of the prevailing critical assumptions about the style and genre of travel writing. In particular, he takes issue with the proposed "disjuncture between scientific and literary travel writing" in the decades around 1800, demonstrating how the division happened "in the decades *after* 1790–1820 (decades which saw the retrospective construction of 'romantic ideology')" rather than…[being] essentially constitutive of the genre in the period itself" (7). He criticizes the purported split between "subjective" and "objective" travel writing, asserting that "To generalize travel writing in the romantic period as merely 'subjective' is to ignore not only the majority of travelogues produced during the period but also the testimony of contemporary commentators" (8). In examining his claims in the context of my corpus, I take up the challenge from Bohls and Duncan, who note that though his argument and analysis are limited to particular locations, his book "deserves to be tested against a wider range of writings" (xxv).

Table 1.2.12: Leask Statistics in TWDB-origbib

TWDB Tag	Number of Titles Only in This Source	Total Titles Listed in This Source	Ratio of Unique Titles to Total Titles
leask	24	60	0.4

Figure 1.2.12: Leask Titles in TWDB-origbib



Appendix 1.3: Extended Methodology of TWDB-origbib

1.3.1: Notes on Cox

Cox makes up a significant portion of the TWDB, both in bibliographical entries and matched digital files. For users who want to rely on these entries or Cox more specifically, I include additional details of the process of transcribing and condensing Cox's bibliography below. I only recorded titles, authors, and publication dates, ignoring metadata such as place of publication, edition information, format, and volumes. My shorthand ignores the opening "the" or "an". When searching for matching texts, a title of 5–7 notable words (i.e., non-stop words), in combination with author and publication date, was most important. During the transcription phase, I also entered a short version of repeated words and then expanded them later using a find

and replace function, such that “voy” and “desc” would become “voyage” and “description.”[30] I also recorded cross-references, though some of these may have been missed due to human error (by myself, Cox, or his transcriber). In total, 348 titles appeared in two sections, 47 in three, and 12 in four; most were only in one section.

Table 1.3.1: TWDB-origbib excerpt, showing cross-references

pubdate	author	title	xref	cox_section
1784	cook, james	voy to the Pacific Ocean	xref	arctic regions
1784	cook, james	voy to the Pacific Ocean	xref	north pacific
1784	cook, james	voy to the Pacific Ocean	xref	circumnavigations

Table 1.3.2: TWDB-origbib excerpt with consolidated cross-references

pubdate	author	title	cox	cox_section_1	cox_section_2	cox_section_3
1784	cook, james	voyage to the Pacific Ocean	x	circumnavigations	north pacific	arctic regions

Finally, I removed all the extracts from periodicals and collections (by relying on the TWDB-origbib notes column), any later editions (often found through deduplication, as Cox did not consistently list editions), and non-English titles. Assuming the language of a text based on the short title is sometimes an error, however. Some titles, like George Martine’s *Reliquiae Divi Andreae; or, The State of the See of St. Andrews,* use Latin for the title and expand the subtitle in English. Others are entirely in Latin. Of course, even with a title entirely in English, a significant

[30] Other such words included ant/iquities, hist/ory, disc/overy, acc/ount, journ/ey. Generally, these nouns were in the form seen above; any occasional “journeys” would be handled by our search algorithm.

portion of the text proper may be non-English quotations of Italian poetry or scientific language in Latin. While leaving the Latin-titled texts (because of the prominence of antiquities in Cox), I removed all other foreign-language texts that I could easily identify, leaving it to later methods (such as metadata on language of a text and algorithms recognizing English words) to filter remaining titles.[31]

1.3.2 Additional notes on transcription and deduplication

Rather than creating scripts to match titles, I used a mixture of manual (sorting through title and author) and digital methods. Microsoft Excel's Fuzzy Search Add-In helped to identify potential matches.[32] For combining, searching, and manipulating data, I relied on the pandas Python toolkit.[33] Instead of ID numbers, the TWDB-origbib uses a combination of a short title + author formula ("authentic narrative of the shipwreck + bradley, eliza") to connect the TWDB-origbib and the final TWDB. This stable and clear identifier was also very practical in deduplication efforts.

---

[31] Cox lists the majority of these non-English titles in the section "Tours by Foreigners" (3: 67–124).

[32] Excel is notorious for changing data encoding, especially of characters with diacritics, numbers with leading zeros, and so on. Fuzzy Search's use of Levenshtein distance and token transformations accommodated typographic errors and title variations. Considering that Fuzzy Search is free with Microsoft Excel and has a number of customizable parameters, it may be useful for others who need a tool for quick and reliable matching, but do not have the resources for customized code. Though not noted in documentation, Fuzzy Search will ignore text beyond a certain character limit. While I did not test to find this exact limit, I experimented with titles shortened to the first 25–50 characters. For my project, I set Fuzzy Search's Similarity Threshold to .60, though matches became dramatically less useful approaching .65.

[33] When using Python, I used a mixture of the command line and Jupyter notebooks. For visualizations, I used the Plotly graphing library for Python.

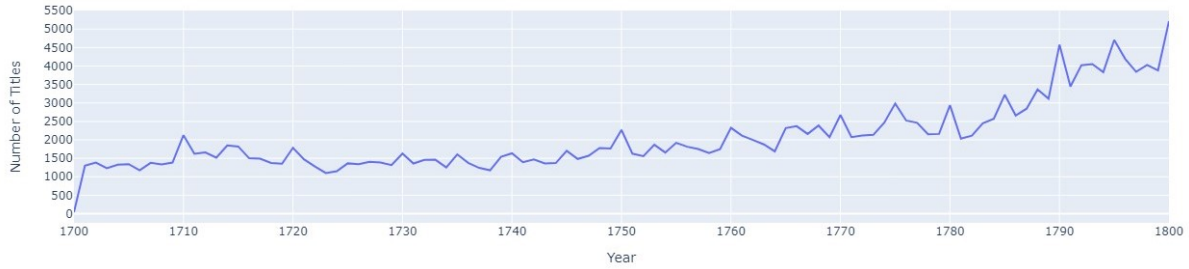
Appendix 1.4: Cross-References in TWDB-origbib

1.4.1: Table of Single and Multiple Sources

Source	Number of Titles Listed Only in This Source	Total Number of Titles	Ratio of Unique Titles to Shared Titles
cox	3846	4072	0.944499
btw_europe	152	169	0.899408
irishmcveagh	81	94	0.861702
gove	49	77	0.636364
btw_w	74	134	0.552239
murray	63	118	0.533898
ncco_travelnarr	5	10	0.500000
leask	24	60	0.400000
ncco_c19trav	36	111	0.324324
brynmawr	43	200	0.215000
andrews	13	65	0.200000
tee	13	66	0.196970
bdanth	8	55	0.145455
robinson_w	5	40	0.125000

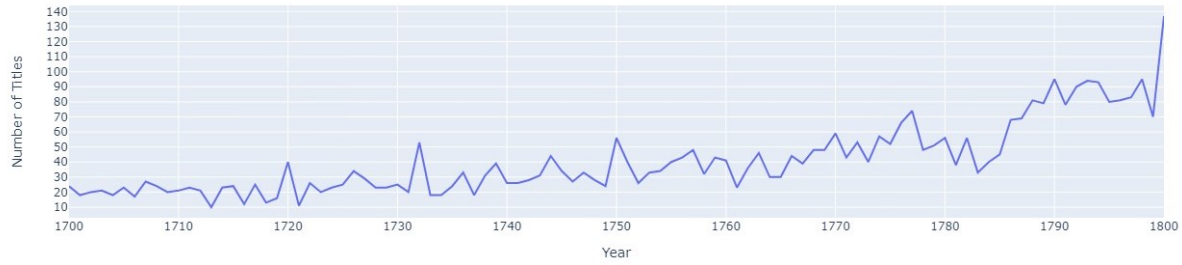
Appendix 2.1: Publication Numbers in ECCO

Figure 2.1.1: ECCO Publication Numbers, by Year[34]



[34] See also Cassidy Holahan (816–8) and Mikko Tolonen et al. (“Anatomy” 101–2).

Figure 2.1.2: TWDB Publication Numbers, by Year



Appendix 2.2: LCSH on ECCO and HT

2.2.1 LCSH on ECCO

While most of my analysis focuses on travel writing in connection with specific TWDB titles, understanding the overall composition of ECCO's subject headings, including travel titles not included in the TWDB, provides context for both the random samples and the travel samples. Note that while the TWDB includes only titles published in Great Britain, the below numbers include titles published anywhere in the world and multiple editions and reprints.[35] Unless otherwise noted, I relied on the ECCO hard drive data.[36] These differences may be related to the number of volumes on the ECCO hard drives compared to the ECCO interface or to how my code handles terms compared to the Gale algorithms.[37]

---

[35] Although ECCO is supposed to only include titles between 1700 and 1800, there are titles outside this time frame. I used the TWDB limitation of 1700–1830.

[36] For example, *The new Bath guide; or, useful pocket companion* (documentID 905900100, ESTC ID T091324), returned with the `guidebook` keyword, is on the hard drives in the General Reference module, but is not on the ECCO or ECCO-GPS interface (search conducted June 25, 2020).

[37] Overall, the ECCO hard drives at McGill hold 207,614 volumes (gathered by scraping for XML files in each module folder). The ECCO interface returns 207,627 volumes total. ECCO-GPS returns 207,477 volumes. These numbers vary even more if including parameters such as publication date 1700–1830. One potential explanation is a different handling of publication dates, especially those with a date range. Even on the ECCO interface, there are discrepancies: there is an approximate 5,000 volume difference comparing the overall number of volumes and the 1700–1800 volumes that can not be accounted for by examining volumes outside of that time frame or by selecting the "Include documents with no known publication date" option (searches conducted June 26, 2020).

The most common travel term on ECCO, `travel`, corresponds with one of the most popular subdivisions, `description and travel`.[38] This LCSH subdivision is also the eleventh most common subdivision in ECCO. While this prominence is not direct evidence of travel writing's reputation of being a popular genre, it does demonstrate the recognition of such texts by cataloguers. The other travel terms draw attention to patterns in other modules. For example, the `voyage` keyword in the "Literature and Language" module is boosted by the LCSH `voyages, imaginary`, important to analysis of the subgenre in chapter 4. Over half the volumes in the "Fine Arts" module are added via `antiquit`. Overall, the addition of ECCO Part II was also a boon for travel writing; while ECCO Part I holds 4,072 of volumes with a travel term, ECCO Part II increased the total count of travel volumes by 967, or nearly 17%.[39]

Table 2.2.1: Keyword Results in ECCO (Online vs Hard Drives)

Keyword	Number of Volumes on ECCO (online, English)	Number of Volumes on ECCO (hard drive)
travel	3485	3516
voyage	1058	1065
discover	282	284
explor	223	225
antiquit	580	583
guidebook	285	286
Total	4745	4782

[38] When I discuss the travel terms in this section, I always assume a wildcard character at the end of the term. For example, `travel` includes `travellers`, `travels`, and so on.

[39] The only cleaning of this data was to modify the publication date to four characters and processing of LCSH into main headings and subdivisions. I did not correct any spelling errors.

Table 2.2.2: Counts of Volumes Holding Each LCSH Keyword in ECCO Modules[40]

Module	travel	discover	explor	voyage	guidebook	antiquit
History and Geography	2374	204	171	610	206	361
Social Sciences	622	46	30	136	63	33
Literature and Language	270	17	17	230	4	46
Religion and Philosophy	85	5	1	51	2	38
Medicine, Science and Technology	64	9	2	14	4	4
General Reference	49	2	2	10	3	13
Fine Arts	37	1	1	13	3	86
Law	15	0	1	1	1	2
Total Volumes with Keyword	3516	284	225	1065	286	583

Even if condensing all volumes' metadata into one representative title, the frequency ratios in the travel corpus stay within 2% of each other and the order of the top terms stay relatively stable, indicating that multivolume works, even the large collections or periodicals, do not overly skew the results.

Overall, table 2.2.3 shows that most common main headings in ECCO are related to locations (Such as Great Britain, France, the United States, and Ireland), religion, or genres. Within travel writing more specifically, table 2.2.5 shows the influence of LCSH structure, where locations are again very popular (along with variations on voyages), but subdivisions have more variety (table 2.2.6).

[40] Because keywords can occur multiple times with a single volume, these numbers do not reflect the total number of volumes per module.

Table 2.2.3: Most Common Main Headings in ECCO (1700–1800)

Main Heading	Number of Occurrences of Main Heading	Frequency
great britain	19900	0.060450
bible	14450	0.043895
english poetry	8588	0.026088
church of england	6248	0.018980
sermons, english	4603	0.013983
english fiction	4254	0.012922
france	3836	0.011653
english drama	3556	0.010802
united states	3518	0.010687
ireland	3053	0.009274
songs, english	2611	0.007931
christian life	2016	0.006124

Table 2.2.4: Most Common Subdivisions in ECCO (1700–1800)

Subdivision	Number of Occurrences of Subdivisions	Frequency
early works to 1800	135919	0.317965
18th century	29388	0.068749
great britain	21811	0.051024
history	18756	0.043877
sermons	15847	0.037072
england	11527	0.026966
politics and government	10083	0.023588
n.t.	6027	0.014099
o.t.	5541	0.012962
periodicals	5467	0.012789
description and travel	3957	0.009257
ireland	3617	0.008462

Table 2.2.5: Most Common Main Headings in ECCO Travel Writing[41]

Main Heading	Number of Occurrences of Main Heading	Main Heading Frequency Within Travel Volumes
great britain	825	0.078542
voyages and travels	624	0.059406
europe	358	0.034082
england	237	0.022563
france	234	0.022277
voyages around the world	220	0.020944
voyages, imaginary	220	0.020944
united states	214	0.020373
america	210	0.019992
london (england)	167	0.015899
italy	164	0.015613
scotland	150	0.014280

Table 2.2.6: Most Common Subdivisions in ECCO Travel Writing (1700–1800)

Subdivision	Number of Occurrences of Subdivisions	Frequency
description and travel	3802	0.287616
early works to 1800	3608	0.272940
history	738	0.055829
antiquities	526	0.039791
guidebooks	315	0.023829
18th century	276	0.020879
discovery and exploration	238	0.018004
periodicals	219	0.016567
social life and customs	202	0.015281
great britain	174	0.013163
fiction	153	0.011574
england	132	0.009986

[41] For these tables, note that I did not deduplicate main headings or subdivisions (for example a single volume may hold: `Great Britain---Antiquities` and `Great Britain---Description and Travel`). Considering title rather than volume data results in a similar top ten with some changes in the bottom six.

2.2.2: ECCO and HT Travel Keyword Comparisons

The proportion of the TWDB's travel keywords are relatively the same in ECCO and HT, especially for the most common keywords of `travel`, `voyage`, and `antiquit`. This ratio is important for studies that may want to rely on particular keywords or data sources.

Table 2.2.7: Raw Counts and Ratios of Keyword Results in ECCO and HT Searches[42]

Travel Keyword	ECCO Titles (hard drive, English)	ECCO Ratio of Travel Titles with Keyword	HT (online, English)	HT Ratio of Travel Titles with Keyword
travel	2585	0.731	3457	0.759
voyage	694	0.196	849	0.186
discover	166	0.047	373	0.082
explor	140	0.04	369	0.081
antiquit	412	0.117	516	0.113
guidebook	273	0.077	239	0.052
Total Titles	3534		4557	

Appendix 2.3: Place of Publication on ECCO

As with ECCO titles overall, travel writing on ECCO is predominantly published in London, the center of the British empire.[43] ECCO's publication place metadata shows that nearly 70% of the travel writing volumes were printed in London, with Dublin (9.5%), Edinburgh (3.4%), Oxford (2%), and Philadelphia (1.2%) rounding out the top five.[44] For digitization, the British Library provides 3,486 volumes, or 73%, with the Bodleian contributing another 314 (6.5%), and the remaining 23 institutions providing the remaining 982. To further contextualize the data that the models draw from, all the citation metadata for these items (though not necessarily their LCSH)

[42] HT search conducted July 16, 2020.
[43] For geographical patterns in ECCO overall, see Mikko Tolonen et al. ("Anatomy" 106–9) and Cassidy Holahan (816–7).
[44] If French-language titles are included, Paris is fifth on this list.

is sourced from the ESTC. Over half of the volumes (58.6%) are single volume titles; only four titles go up to 20 volumes, and of those, some are missing earlier volumes. Additionally, of the 4,782 volumes, 1,168, or nearly a quarter, have some sort of note in the edition field.[45] However, consider how there are 42 volumes with "Travels into several remote nations" in the title (the original *Gulliver's Travels* was published in two volumes), but only six with metadata in the edition field.[46] Future research could examine or model differences between works published in different locations, or in different formats.

Appendix 2.4: Notes on the Data Model

The final TWDB dataset is a spreadsheet consisting of a row for each volume; columns of metadata from the TWDB-origbib (if applicable) indicating bibliographical sources and data sources; columns of metadata harvested from the data source (i.e., from ECCO, HT, NCCO, or Google); columns of metadata about claimed author gender; and columns of select data exported for modelling purposes. Below, I describe in further detail several of the practical and editorial decisions that created the TWDB dataset.

2.4.1: First Editions and Place of Publication

Some projects favour having only first editions, and only one copy of each first edition, while others include multiple editions, reprints, and copies. Depending on the researcher's questions, these decisions may have varying levels of impact; for example, Ted Underwood et al. found that some patterns "follow nearly the same diachronic arc…whether we emphasize prominent books,

---

[45] Matthew Day points to unpublished research by Shef Rogers, which indicates that "between 10% and 15% of all travel accounts first published in the eighteenth century were reissued, most often to assert a new edition, even though little if anything had changed in the text" (2).

[46] A few of these are keys, such as *Gulliver decypher'd: or remarks on a late book, intitled, Travels into several remote nations* (Gale ID CW0112271495).

balance authorial gender, remove duplicate volumes, or select texts completely at random" ("NovelTM" 5). The TWDB holds only one title (with multiple volumes) per work, with no duplications, prioritizing first editions. If no first edition was available, the TWDB includes the next earliest available edition published within five years of the TWDB-origbib's claim of publication date. Imposing a strict match to the TWDB-origbib would have resulted in a smaller corpus: not only did publishers sometimes print inaccurate dates on their title pages for marketing purposes, but there are also errors in metadata.[47] Limiting to first editions also reduces the inclusion of rare titles, such as those by women, where later editions are all that are available digitally. Finally, I do not obfuscate my data differences; the metadata in the TWDB allows the removal of first or later editions based on a variety of different features. Because of the TWDB's foundation in bibliography and interest in change within the genre, I use the first edition (or its proximate descendants) to narrow the field to one copy that, at least in publication history (though not necessarily reading reception or reference), is chronological.[48] The TWDB also focuses on publication in Great Britain. However, if a title was originally published in a foreign location, the TWDB includes the first edition published in Britain. This is especially relevant for the Cox bibliography, which did not always list the place of publication.

2.4.2 : Matching the TWDB-origbib to HathiTrust, ECCO, NCCO, and Google

Of the two main digital sources for my data, the TWDB first draws files from HT, for reasons both practical and ethical. The HT Solr Proxy API allowed immediate querying of HT metadata,

---

[47] This may be on the part of bibliographers (ranging from typographical errors to working on different imprints), my transcription, or institutional metadata.

[48] Change in the genre can also be evident in reprints, revisions, and abridgements, such as in Samuel Richardson's editorial changes to Daniel Defoe's *Tour Thro' the Whole Isle of Great Britain* (originally published 1724–1727, with Richardson's 5 editions between 1738 and 1762. As T.C. Duncan Eaves and Ben D. Kimpel note, "as it is revised [by Richardson], the *Tour* becomes less and less like a travel book" (73).

and unlike ECCO, the metadata and some textual features (via the Extracted Features dataset) were publicly accessible, allowing future scholars to more easily use the TWDB and test its findings.[49] Stéfan Sinclair and I created a script that considered tokenized titles, authors, and publishing dates from the TWDB-origbib, identified the most similar HT titles, and gave each match a score based on its similarity.[50] Some works, especially those with similar title tokens (*History and Antiquities of* [locale] or *Tour in France*) and anonymous authors, had multiple options for matches, or the suggested match was inaccurate. I reviewed each title match manually based on the accompanying metadata, though I did not check for details such as whether the match was a reissue published in the same year. After matching 1,220 of the TWDB-origbib entries to files on HT, a similar script was applied to an XML file of ECCO metadata provided by the eMOP project.[51] The filenames on the ECCO hard drives are not listed on the online interface; nor are the "Gale Document Numbers" of the online interface listed in the hard drive XML. Instead, any connection between these two interfaces relies on the ESTC number.[52]

Both methods run into the challenge of multivolume works: because of the structure of the TWDB-origbib data (one row per title) and the matching method, only one volume was matched with each title in the TWDB. For HT, the "record identification number" refers to a bibliographic entity such as the typical catalog record, while the "volume ID" refers to "a discrete object that was digitized and cataloged as one unit," or a particular physical copy of that

---

[49] HathiTrust has retired the Solr Proxy API.
[50] We normalized the text of the original entry and the ECCO metadata: lowering all capitalization and removing short stop words (such as "and" or "the"), non-alphabetical characters, and multiple spaces. Other features for future consideration include place of publication and edition information.
[51] My thanks to Laura Mandell, Matthew Christy, and Bryan Tarpley at eMOP for providing the eMOP ECCO metadata.
[52] Because I relied primary on the hard drive XML, the TWDB uses these file IDs for ECCO-sourced files.

text.[53] Therefore, I deduplicated based first on record id, then selected only volumes held by one institution (prioritizing Google-OCR texts), and collected the relevant volume ids from that cross-section of the metadata.[54] On ECCO, the most accurate way to gather all volumes is to consider the ESTC number and gather all associated files.[55] Finally, I ensured every multivolume carried the TWDB-origbib metadata of the parent volume using a script to group and fill related rows and columns.

After HT and ECCO, I integrated the two smallest data sources of the TWDB: NCCO and Google Books. Because the TWDB relies on NCCO modules as both a bibliographical source and a data source, the process involved collecting the Gale Document Number from the online NCCO interface. Then, from the hard drive XML, I gathered the assetID which corresponded with the Gale Document Number, and the PSMID and filename, which were the same. Finally, through these multiple cross-references, I was able to identify which XML files corresponded to the bibliographical metadata.[56] For Google Books, after identifying a matching volume, I downloaded the connected file from the Google Books interface and renamed it

---

[53] For example, *A journey from London to Genoa* by Joseph Baretti has the record id of 000812990, with the volume ids of gri.ark:/13960/t6k10qg1h, nyp.33433082474853, and mdp.39015062995116 indicating copies of volume 1 contributed by the Getty Research Institute, the New York Public Library, and the University of Michigan, respectively. See https://catalog.hathitrust.org/Record/000812990/ as an example. In my initial application of the SolR proxy, these fields were named htrc_ht_id and htrc_idvol.

[54] Note that some institutions may have multiple copies of one volume, requiring another level of deduplication. Regarding Google OCR, Peter Organiasciak, formerly of HathiTrust, recommended Google over other digitizers for early titles because it handled the long f/s better. On the "About the Collection" page in the HT Research Centre documentation, it suggests prioritizing these documents as well: "A proxy for OCR quality that researchers have found helpful is to select volumes that were most recently digitized (as OCR processes have improved with time) by Google (as their OCR technique is considered to be high quality)."

[55] Although ECCO has a field for "totalVolumes," the listed number did not always correspond to the number of digitized volumes. While the filenames may seem to indicate all volumes in a set (1705500101, 1705500102, and so on), this naming structure is inconsistent (see, for example, John Moore's *A journal during a residence in France*, with the filenames of 0610900501 and 0611000102). Unfortunately, this means cross-referencing the filename with the ESTCID field in each XML file, or, in my case, relying on the eMOP XML's basic metadata, which was all in one file.

[56] The XML structure of ECCO and NCCO differs in tags used and other structural issues that requires adjustments.

according to the "volume id" in the address bar.[57] As the above efforts show, even data from the same company can require different modes of access and documentation, and relying on different datasets requires iterative and diverse processes of finding, comparing, and integrating metadata.

2.4.3: Harvesting LCSH: Notes on Technical Steps

LCSH Structure

Even with a list of the technical rules and the subject headings in hand, the application of subject headings is a subjective process during which cataloguers perform subject analysis and determine the "aboutness" of a work.[58] In general, a cataloguer will "Assign to the work being cataloged one or more subject headings that best summarize the overall contents of the work and provide access to its most important topics," following the practice of "Assign[ing] headings only for topics that comprise at least 20% of the work" ("Assigning and Constructing Subject Headings H 180"). LC also recommends assigning no subject headings to works of a "very general or amorphous nature, for example, a general periodical or a collection of essays with no discernable theme," as well as "texts of sacred works or to individual works of *belles lettres* with no identifiable theme or specific form" ("Assigning"). While up to ten headings may be assigned, LC suggests "a maximum of six is appropriate." Multiple headings should also be assigned "in order of predominance" ("Order of Subject Headings H 80").[59]

Structurally speaking, LCSH are hierarchical, consisting of at least a main heading and, optionally, followed by subdivisions. The categories of subdivisions are topical, form,

---

[57] For example, Elizabeth Strutt's *A Spinster's Tour in France* (1828) at https://books.google.co.uk/books?id=Vu_iksWgQloC has the ID of Vu_iksWgQloC. Cursory tests show these ids to be stable across different domains, including books.google.com, .ca, or .co.uk.
[58] For examples, see "How Do We Determine Aboutness?" from the Library of Congress.
[59] Because my harvest for HT iterates by tag through the metadata, rather than by tag order, I cannot rely on which "came first" in my data.

chronological, and geographical, and each have their own rules.[60] As of 1974, cataloguers may also apply free-floating subdivisions ("Free-Floating Subdivisions H 1095"). [61] These free-floating subdivisions significantly impact the TWDB, as `antiquities`, `description and travel`, `discovery and exploration`, and `guidebooks` are all free-floating subdivisions, unlike `voyages and travels`. These LC rules present only a small fraction of the organization's complex documentation, but they establish the policies that influence my practices below.

1. **LCSH Implementation**
   Display:
   Slave-trade—Africa—Early works to 1800
   or
   Slave-trade
   —Africa
   ——Early works to 1800

   Machine readable:

   650$a Slave-trade**$z**Africa**$v**Early works to 1800[62]

2. **ECCO Implementation**
   Display (online, original interface):
   Slave-trade--Africa--Early works to 1800

---

[60] See "Library of Congress Subject Headings: Online Training" for examples, including an 18-minute video exclusively on applying the `history` subdivision (https://www.loc.gov/catworkshop/lcsh/index.html).

[61] A free-floating subdivisions is a "form or topical subdivision that may be used under designated subjects without the usage being established editorially, and, as a consequence, without an authority record being created for each main heading/subdivision combination that might be needed." This includes the caveat that "Free-floating status does not allow the assignment of a subdivision under any topic of interest without regard for appropriateness. All free-floating subdivisions may be assigned only in accordance with subject cataloging rules, policies, and practices. Most subdivisions are usable only under limited categories of headings in specifically defined situations." Many subdivisions now authorized as free-floating were printed in LCSH before 1974, so they may still appear under individual headings belonging to a category. See "Free-Floating Subdivisions H 1095" and "Library of Congress Subject Headings: Online Training" for specific examples.

[62] Each subfield code below (ie, a, z, x, v) indicates a different type of subfield, such as topical, geographical, form, or chronological. For more details on field 650 (topical), see "650 - Subject Added Entry-Topical Term (R)."

Machine readable (XML):

```
<locSubjectHead type="topicalTerm">
<locSubject subField="a">Slave-trade</locSubject>
<locSubject subField="z">Africa</locSubject>
<locSubject subField="v">Early works to 1800</locSubject>
</locSubjectHead>
```

3. **HathiTrust Implementation**

Display (online interface):
Slave-trade > Africa > Early Works to 1800.

Machine readable (JSON):

```
<datafield tag=\"650\" ind1=\" \" ind2=\"0\">
     <subfield code=\"a\">Slave-trade</subfield>
     <subfield code=\"z\">Africa</subfield>
     <subfield code=\"v\">Early Works to 1800.</subfield>
</datafield>
```

4. **Travel Writing Database Implementation**

Machine readable and display in CSV:

```
Slave-trade---Africa---Early works to 1800_Africa---
Description and travel
```

LCSH Identification and Processing

After identifying thousands of potential titles via the LCSH in the HT "topic_heading" and ECCO "<locSubjectHead>" searches, I removed duplicates, later editions, and titles already in my corpus. I then cross-referenced these LCSH titles with the TWDB-origbib, as dozens of these titles had not scored high enough to show up as matches in the earlier algorithmic matching.[63] Relying on LCSH can offer multiple files and metadata for analysis, but because of the TWDB's goals, including them required a significant amount of labour.

[63] Based on my manual reviews of the data, this was most often the case where some metadata fields were much longer, such as "buchanan, john" compared to "Buchanan, John Lanne, fl. 1780–1816."

2.4.4: Gender Encoding

I rely on several descriptive categories to create richer metadata around gender, especially as informed by scholars such as Laura Mandell, Megan Peiser, and Katherine Bode, focusing on the roles that women, real or claimed, held in creation of these works, as well as those roles' public nature. Note that the categories below may hold more than one element. For example, in *Excursions in Madeira and Porto Santo…by the Late T. Edward Bowdich; to Which Is Added, by Mrs. Bowdich, I. A Narrative of the Continuance of the Voyage to its Completion* (NCCOF0257-C00000-B0186500), Sarah Bowdich both edits her husband's work and adds her own writing.

gender

claimed: the text claims that a woman is the author, editor, or translator.

confirmed: scholarly work confirms that a woman is the author, editor, or translator.

speculation: while not recorded in the text, contemporaries or scholars have speculated that a woman is the author, editor, or translator.[64]

gender roles (individual columns or twdb_genderRoles)

au - a woman is claimed to be the primary author of a title.

ed - a woman is claimed to be the editor of a title.

translator - a woman is claimed to be the translator of a title.

pt - a woman is credited with part of the title, such as in a diversion from the main text.[65]

gender_claim

---

[64] This category only applies to *Memoirs of the Court of Lilliput,* which Alexander Pope claims that Eliza Haywood wrote, though David Brewer notes that "Few twentieth-century Haywood scholars have accepted Pope's attribution or even taken it particularly seriously" (217).

[65] See William Rufus Chetwood's *The voyages and adventures of Captain Robert Boyle…Intermix'd with the story of Mrs. Villars, an English lady…* (0388100100), as a collaborator, or as authoring their own second part or addendum to the work, such as Bowdich's contributions to her husband's *Excursions in Madeira and Porto Santo*, advertised on the title page.

title: on the title page, either in the title or listed as the author

peritext: in materials such as a preface or dedication

research: evidence from epitextual materials such as reviews or correspondence

text: the narrator uses gendered references in the text, such as to her husband

2.4.5: Processing Plain Text and OCR Quality

Files from different sources require different steps to get to a similar state. In general, I convert each file into a plain text file that includes prefatory materials, body text, and end matter, while removing running headers. Then, I use Ted Underwood's OCR normalizer to normalize the text of each work ("OCRnormalizer.py" in tedunderwood/datamunging).[66] As Underwood describes,

> OCRnormalizer corrects and normalizes OCR versions of English books published after 1700. It addresses the notorious "long S" problem, rejoins words broken across a linebreak, standardizes word division, and normalizes spelling to modern British practice. The name is "normalizer" rather than "corrector" because its goal is explicitly not to reproduce the original page image but to produce a standardized corpus that permits meaningful comparisons across time and across the Atlantic Ocean (e.g. "to day" and "to-day" turn into "today," "honor" turns into "honour", "fame" turns into "same" in 18c contexts where we can infer that it was originally "same").[67]

As a final step to prepare for predictive modelling, I use another of Underwood's scripts ("tokenizetexts.py") to tokenize each text and count its features, including several such as word

---

[66] Ryan Heuser also uses this script for his dissertation project on eighteenth-century texts and word embeddings (*Abstraction: A Literary History*).
[67] See also Underwood's collection of "new_normalizers" in "tedunderwood/DataMunging."

length or probability of capitalization.[68]

Steps for Each File Source

HathiTrust

Original file state: ZIP files holding each page as a separate plain text file. The text file includes page numbers and running headers, as well as all prefatory materials and end matter. I accessed these files via a direct download[69] or, later, through the HT Data Capsule.[70]

1. Run Underwood's OCR normalizer Python script on each zip file, which concatenates the pages into a single plain text file with <pb> indicating page breaks ("OCRnormalizer.py").[71]
2. Remove test pages for the Canadian Institute for Historical Microreproductions (common for a subset of files from University of Alberta).[72]
3. Run a modified version of Underwood's Python script to remove running headers (ie, repeated text and page numbers) from each page ("HeaderFinder.py").[73]
4. Extract features using "tokenizetexts.py"

---

[68] Unless otherwise noted, I relied on Underwood's code, available at "tedunderwood/horizon: Data and Code to Support Distant Horizons" for any steps after the initial plaintext. These steps include normalizing the text, counting features, and modelling.

[69] This request required an affiliated institutional sponsor, a description of the proposed research, and an agreement other terms and conditions, such as deleting any volumes that are later found to be under copyright. See "Data Availability and APIs: Datasets" for more information. Only certain institutions (including McGill) have a signed Google Distribution Agreement, providing access to Google-digitized items in HT.

[70] See full documentation at "HTRC Data Capsule Environment."

[71] I did not modify any of the OCR normalizer rules.

[72] For example, see *A chronological history of voyages into the Arctic Regions* (1–4) (https://catalog.hathitrust.org/Record/100263028).

[73] The script treats running headers "as lines, near the top of a page, that are repeated within the space of two pages, in either direction." For full documentation, see "HeaderFinder.py" in "tedunderwood/DataMunging." My only modification was adding ".splitlines()" to ensure splitting at the sentence level rather than page to character.

ECCO

Original file state: XML files with text indicated by <wd> tag.

1. Extract text using lxml library, focusing on the <wd> tag. This extraction includes prefatory material, end matter,
2. Run Underwood's OCR normalizer Python script on each file ("OCRnormalizer.py").
3. Extract features using "tokenizetexts.py"

NCCO

Original file state: XML files with text indicated by <wd> tag.

1. Extract text using lxml library, focusing on the <wd> tag. This extraction includes prefatory material, end matter, etc.
2. Run Underwood's OCR normalizer Python script on each file ("OCRnormalizer.py").
3. Extract features using "tokenizetexts.py"

Google

Original file state: EPUB files downloaded individually from the Google Books website. The files that I examined did not include running headers or page numbers, but at least one title includes signatures (symbols aiding in page collation).

1. Convert to plain text using the Calibre software. Remove the Google copyright information at the beginning or end of each file.[74]
2. Run Underwood's OCR normalizer Python script on each file ("OCRnormalizer.py").
3. Extract features using "tokenizetexts.py"

---

[74] Otherwise, predictive models may identify titles as belonging together based on the frequency of terms like "Google."

OCR Quality

OCR quality is "the most well-known topic in discussions of ECCO" (Gregg *Old Books* 63), and it has long been a concern for scholars working on digitized historical documents on any platform, impacting research techniques ranging from search results to advanced text mining.[75] Most analysis of eighteenth century focuses on ECCO, and ECCO and NCCO (via Gale Primary Sources) are the only digital sources to include "OCR Confidence" scores for each page and text.[76] The Linguistic DNA project decided that "there are too many problems within the OCR dataset to use" ECCO, and relied on the ECCO-TCP dataset. On the other hand, Mikko Tolonen et al.'s more recent work concludes that "We believe that investigating language by the use of ECCO is possible, given that careful work is put into taking different aspects into consideration and the research questions are matched with what is possible to do with such a biased and largely inaccurate corpus" ("Corpus Linguistics" 32). To enable the "robust approaches" that Tolonen et al. suggest, the TWDB records the percentage of recognized tokens from "OCRnormalizer.py," allowing for future applications to choose appropriate accuracy levels. Furthermore, for my work in chapter 4, I use travel and nontravel volumes that have over 85% of their tokens in the normalizer's dictionary.[77] In an online discussion, Underwood said rates of 90% did not cause

---

[75] In particular, see Mark J. Hill and Simon Hengchen's "Quantifying the impact of dirty OCR on historical text analysis: Eighteenth Century Collections Online as a case study"; François Dominic Laramée's "How to Extract Good Knowledge from Bad Data: An Experiment with Eighteenth Century French Texts"; Ryan Cordell's "'Q i-Jtb the Raven': Taking Dirty OCR Seriously"; David A. Smith and Cordell's "A Research Agenda for Historical and Multilingual Optical Character Recognition"; Paddy Bullard's "Digital Humanities and Electronic Resources in the Long Eighteenth Century"; and Matthew Christy et al.'s "Mass Digitization of Early Modern Texts With Optical Character Recognition." For another approach to enriching ECCO's data, see Sherif Abuelwafa et al. in "Detecting Footnotes in 32 million pages of ECCO."

[76] GPS' linked information for "Learn how this text was created" describes how "OCR Confidence represents the OCR engine's confidence in the accuracy of the conversion from image to text. Many aspects affect the quality of the conversion including: Condition of the original document itself[,] Type of text (handwritten vs printed)[,] Year the original document was created[,] Equipment used to scan the original document[, and] Maturity of the OCR algorithm used at the time of creation."

[77] Tests relying on files with over 90% recognized tokens did not have significant changes in accuracy.

issues in his work, and Ryan Heuser found in cursory experiments that OCR affects longer words more (Tweets Jul 6, 2017).[78] For my work, OCR errors do have an impact, especially if comparing across different file sources or, as I suspect, in my models of the early eighteenth century.[79]

Appendix 3.1: Bibliographical and Data Source Trends

Diachronically, the 60% average of found TWDB-origbib is relatively stable, ranging from closer to 50% in the 1700s, 1730s, 1740s, and 1810s, while the 1800s and 1820s are around 70% (figure 3.1.1). These numbers differ more when considering bibliographical sources, where the TWDB typically holds matches for at least 75% of their files. For Benjamin Colbert's European Tour, John McVeagh's Irish travels, and Edward Godfrey Cox's *Reference*, however, the percent of found titles are at 40%, 53%, and 58%, respectively (figure 3.1.2). If the calculations disregard any entries solely contributed by Cox, the overall found rate climbs from 59% to 72%. Some of the works cited in these sources are probably rarer than those in anthologies. The more thorough a bibliography, the more likely that it will list rarer titles, or titles from smaller archives that may not have been digitized.

---

[78] In that discussion, Underwood also points to Shlomo Argamon et al.'s work with supervised classification that found "even a relatively high level of errors in the OCRed documents does not substantially affect stylistic classification accuracy."

[79] See appendix 4.1 for comparisons of ECCO and HT files in the 1790s, and chapter four, where I model pace of change ("The Changing Faces of Travel"). Tolonen et al. note that "median accuracy [of OCR] improves with time, particularly from 1700 to 1750" ("Corpus Linguistics" 29).

Figure 3.1.1: Ratio of Found to Remaining Titles in TWDB-origbib

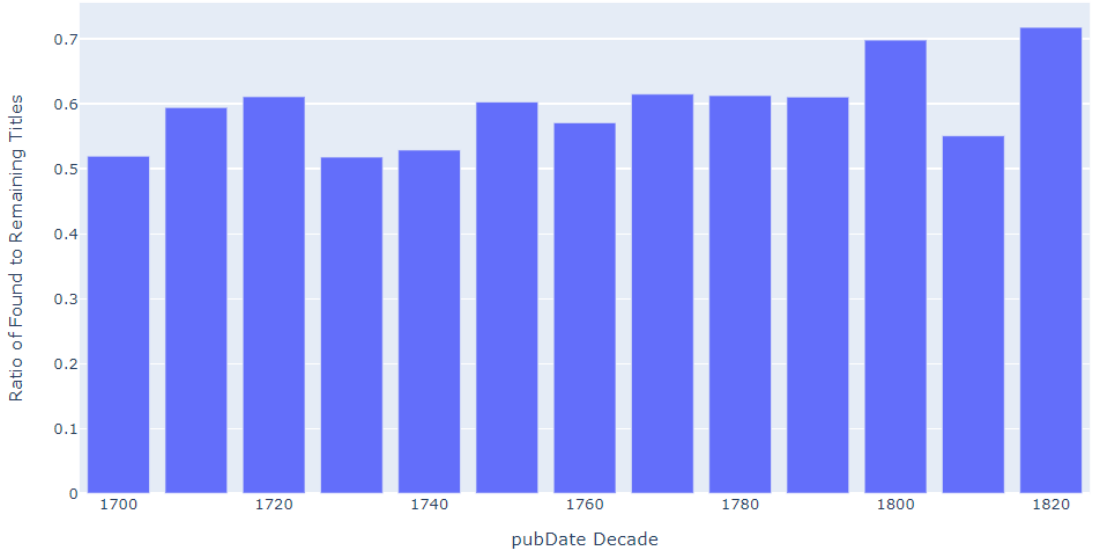
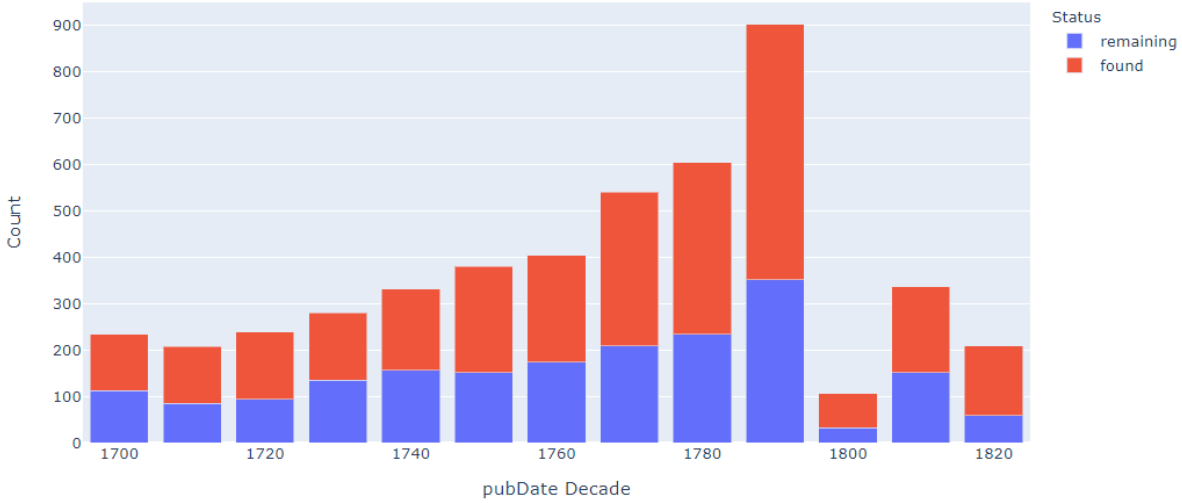


Table 3.1.2: Number of Found TWDB-origbib Titles[80]

Bibliographical Source	Number of Found Entries	Number of Remaining Entries	Ratio of Found Files
btw_europe	67	102	0.40
irishmcveagh	50	44	0.53
cox	2374	1698	0.58
murray	89	29	0.75
leask	47	13	0.78
gove	63	14	0.82
andrews	54	11	0.83
tee	56	10	0.85
btw_w	115	19	0.86
brynmawr	177	23	0.88
robinson_w	36	4	0.90
bdanth	50	5	0.91
ncco_c19trav	110	1	0.99
ncco_travelnarr	10	0	1.00

[80] Because of cross-references, the number of found entries do not add up to the titles in TWDB-origbib.

Figure 3.1.3: Raw Count of Found to Remaining Titles in TWDB-origbib



3.1.1: Sources: Place of Publication and Holding Institutions

Two other elements connect to the material creation and continued existence of the physical and digital books in the TWDB.[81] First, more than 80% of the volumes were published in London (table 3.1.3). While another 5% were published in Edinburgh and 1.3% in Oxford, the remaining volumes are dispersed across 149 geographic entries.[82] The libraries that hold the physical copies of the TWDB books are also concentrated in London: the British Library alone provides over 41.2% of the volumes in the TWDB, all through ECCO (table 3.1.4). The next six institutions with the most volumes are from HT, and reflect its North American connections, with the University of Michigan, the New York Public Library, and the universities of California,

[81] See also Mikko Tolonen et al. in "Corpus Linguistics and Eighteenth Century Collections Online (ECCO)" regarding bias in ECCO around place, language and dating of publications.
[82] Eleven of these twdb_pubPlaces are combined lists of cities, such as London; Edinburgh. A total of 49 volumes do not have entries for the pubPlace of publication, though some may be inferred from the imprint data.

Alberta, Harvard, and Princeton providing from 10% to 2.8% of the total volumes in the corpus. The remainder of the volumes come from another 43 institutions located in the United States, the United Kingdom, Ireland, and Spain, with 22 of them contributing less than 10 volumes.[83] Just as with other research relying on these collections, then, the TWDB is a reflection of what has survived and been digitized in primarily academic institutions; smaller archives may hold more titles that were privately printed about local features, but may not be able to afford digitization.

Table 3.1.3: Place of Publication of TWDB Volumes

Place of Publication	Raw Count	Percent
London	4509	82.61
Edinburgh	276	5.06
Oxford	71	1.30
Bath	51	0.93
York	37	0.68
Glasgow	33	0.60
Newcastle upon Tyne	24	0.44
Cambridge	20	0.37
Bristol	19	0.35
Norwich	19	0.35

Table 3.1.4: Top Ten Institutions Contributing to the TWDB

Institution	Total Volume Counts	HT Results	ECCO Results	NCCO Results	Google Books Results	Overall Percent
British Library	2267	0	2257	9	1	41.17
University of Michigan	556	556	0	0	0	10.10
New York Public Library	503	503	0	0	0	9.13
University of California	410	410	0	0	0	7.45
University of Alberta	309	309	0	0	0	5.61
Harvard University	211	208	0	0	3	3.83
Princeton University	158	158	0	0	0	2.87
Bodleian Library (Oxford)	154	0	154	0	0	2.80

[83] *Venice under the Yoke of France and Austria*, gathered from Google, does not list a source library for its two volumes (IDs 6ooKAAAAIAAJ, G4sKAAAAIAAJ; search conducted July 9, 2020).

Institution	Total Volume Counts	HT Results	ECCO Results	NCCO Results	Google Books Results	Overall Percent
Bryn Mawr College Library	147	0	0	147	0	2.67
Harvard University Houghton Library	93	0	93	0	0	1.69

Appendix 3.2: LCSH Keywords in the Overall TWDB

Table 3.2.1: Titles with No LCSH, by TWDB fileSource

TWDB fileSource	Titles with no LCSH	Total Titles	Percent of no LCSH - TWDB	Total Titles - origbib only	Percent of no LCSH - origbib only
ecco	9	2361	0.38	1773	0.51
htrc	46	1766	2.60	925	4.97
google	19	19	100	19	100.00
ncco	13	113	11.50	113	11.50

Table 3.2.2: Number of Titles Returned by Each Travel Keyword

Travel Keyword	Titles with Keyword (TWDB)	Percent of Titles with Keyword (TWDB)	Titles with Keyword (origbib)	Percent of Titles with Keyword (origbib)	Titles with Keyword (LCSH Harvest)	Percent of Titles with Keyword (LCSH Harvest)
travel	2108	49.50	1013	35.80	1095	76.63
voyage	336	7.89	156	5.51	180	12.60
antiquit	327	7.68	112	3.96	215	15.05
discover	145	3.40	66	2.33	79	5.53
guidebook	124	2.91	62	2.19	62	4.34
explor	123	2.89	56	1.98	67	4.69

Table 3.2.3: Titles with Travel Keywords Cross-referenced with Travel

Travel Keyword	Titles with Keyword Except Travel (TWDB)	Titles with Keyword and Travel (TWDB)	Percent of Titles with Keyword and Travel (TWDB)
travel	0	2108	100.00
voyage	107	229	68.15
antiquit	238	89	27.22
discover	112	33	22.76
guidebook	102	22	17.74
explor	96	27	21.95

Appendix 3.3: Subjects and Trends: Travel Keywords

Table 3.3.1: Most Common Complex LCSH (Main Heading and First Subdivision Only)

Main Heading	Subdivision 1	Raw Count
great britain	description and travel	125
voyages and travels		117
europe	description and travel	100
england	description and travel	98
italy	description and travel	85
united states	description and travel	83
france	description and travel	82
voyages and travels	early works to 1800	62
great britain	history	58
voyages around the world		55
london (england)	description and travel	55
united states	history	54
scotland	description and travel	53
great britain	colonies	49
middle east	description and travel	47
canada	description and travel	46
great britain	antiquities	44
india	description and travel	40
america	discovery and exploration	36
wales	description and travel	36

Table 3.3.2: Count and Frequency of All LCSH Elements (Main Headings and Subdivisions)

LCSH Element	Raw Count	Frequency
description and travel	2544	0.131725
early works to 1800	2020	0.104593
history	828	0.042873
great britain	798	0.041319
england	494	0.025579
antiquities	310	0.016051
voyages and travels	205	0.010615
united states	205	0.010615
france	196	0.010149
scotland	185	0.009579

Table 3.3.3: Most Frequent LCSH Elements by Decade

Decade	1	2	3	4	5
1700	early works to 1800	description and travel	great britain	history	18th century
1711	early works to 1800	description and travel	great britain	history	england
1721	early works to 1800	description and travel	great britain	history	england
1731	early works to 1800	description and travel	great britain	england	history
1741	early works to 1800	description and travel	great britain	history	england
1751	early works to 1800	description and travel	history	great britain	england
1761	early works to 1800	description and travel	history	great britain	england
1771	early works to 1800	description and travel	history	great britain	england
1781	early works to 1800	description and travel	great britain	history	england
1791	early works to 1800	description and travel	history	great britain	england
1801	description and travel	great britain	history	france	england
1811	description and travel	history	antiquities	france	great britain
1821	description and travel	history	antiquities	italy	united states

Table 3.3.4: Top Five Most Frequent LCSH Main Headings by Decade

Decade	1	2	3	4	5
1700	great britain	london (england)	spain	voyages and travels	indians of north america
1710	great britain	roads	voyages and travels	england	scotland
1720	great britain	voyages, imaginary	voyages and travels	europe	voyages around the world
1730	great britain	voyages and travels	london (england)	georgia	jews

Decade	1	2	3	4	5
1740	great britain	voyages and travels	france	england	geography
1750	great britain	united states	voyages and travels	canada	natural history
1760	great britain	agriculture	united states	france	italy
1770	great britain	united states	england	agriculture	voyages and travels
1780	great britain	voyages and travels	england	united states	london (england)
1790	great britain	voyages and travels	france	england	united states
1800	great britain	france	england	united states	voyages and travels
1810	france	united states	great britain	voyages and travels	greece
1820	united states	italy	canada	europe	voyages and travels

Appendix 3.4: Subjects and Trends: No Travel Keywords

Table 3.4.1: Count and Frequency of Main Headings (No Titles with Travel Keywords)

Main Heading	Raw Count	Frequency
great britain	173	0.062613
agriculture	79	0.028592
united states	60	0.021716
france	52	0.018820
mineral waters	42	0.015201
london (england)	40	0.014477
india	36	0.013029
gardening	33	0.011944
geography	23	0.008324
botany	20	0.007239

Table 3.4.2: Count and Frequency of Subdivisions (No Titles with Travel Keywords)

Subdivision	Raw Count	Frequency
early works to 1800	687	0.213819
history	456	0.141923
england	241	0.075008
great britain	171	0.053221
18th century	72	0.022409

politics and government	64	0.019919
commerce	61	0.018985
london	53	0.016495
colonies	49	0.015251
social life and customs	41	0.012761

Appendix 3.5: Counting Gendered Titles

Table 3.5.1: Unmatched Gendered Titles in TWDB-origbib

Publication Date	TWDB-origbib Title	TWDB-origbib Author	Bibliographical Source
1727	fruitless enquiry being a collection of several entertaining histories and occurrences, which fell under the observation of a lady in her search after happiness	haywood, eliza	gove
1755	The life and suprizing adventures of Friga Reveep	Reveep, Friga	gove
1772	manchester directory for the year 1772	raffald, elizabeth	cox
1788	letters of dr. johnson	piozzi, hester	cox
1792	Voyages to the Madeira, and Leeward Caribbean Islands	Riddell, Maria	cox, robinson_w, btw_w
1796	dunnotar castle by the rev james walker	carnegie	cox
1797	views in kent a series of 17 coloured	noel, amelia	cox
1806	An Excursion from London to Dover	Gardiner, Jane	btw_w
1809	city scenes, or a peep into london	taylor, ann	brynmawr, ncco_c19trav
1810	Views in Orkney	Gower, Elizabeth Leveson	btw_w
1812	The Lowestoft Guide	anonymous	btw_w
1813	The Young Northern Traveller	Hofland, Barbara	btw_w
1813	The Beauties of Leamington Priors and Its Environs	Medley, Sarah	btw_w
1815	Battle of Waterloo	Eaton, Charlotte Anne	robinson_w
1817	A Short Journal of a Tour, Made through Part of France, Switzerland, and the Banks of the Rhine, to Spa, Antwerp, Ghent, &c	Southwell, Mary Elizabeth, Baroness de Clifford	btw_w, btw_europe
1820	An Appendix to the Descriptions of Paris	Domeier, Esther Lucie	btw_w

Publication Date	TWDB-origbib Title	TWDB-origbib Author	Bibliographical Source
1821	Notes Relating to the Manners and Customs of the Crim Tatars	Holderness, Mary	robinson_w, btw_w
1821	Journal of a Tour in France	anonymous	btw_w
1821	authentic narrative of the shipwreck and sufferings of mrs eliza bradley	bradley, eliza	robinson_w
1821	A Ten Years' Residence in France	West, Charlotte	btw_w
1824	Tour in Holland in the Year MDCCCXIX	Murray, Anne Elizabeth Cholmley	btw_w
1824	Journal of Excursions through the Most Interesting Parts of England, Wales & Scotland	Selwyn, Elizabeth	btw_w
1827	Recollections of Egypt	Minutoli, Wolfradine-Auguste-Luise von	btw_w
1827	Souvenirs of a Tour on the Continent	anonymous	btw_w
1828	Narrative of a Three Years' Residence in Italy	Martin, Selina	btw_w
1829	Six Views of the Most Important Towns, and Mining Districts, upon the Table Land of Mexico	Ward, Emily Elizabeth	btw_w
1829	Narrative of a Journey from Calcutta to Europe	Lushington, Sarah	btw_w, murray
1830	o'donoghue, prince of killarney, a poem in seven cantos	bourke, hanna maria	irishmcveagh
1830	Continuation of Journals in the Years 1824, 25, 27, 28, and 29	Selwyn, Elizabeth	btw_w

Table 3.5.2: Gendered TWDB Titles of the 1720s

Title	Author	Publication Date	LCSH
The noble slaves: or, the lives and adventures of two	Aubin, Penelope	1722	[emilia]---biography---early works to 1800_teresa---biography---early works to 1800_slaves---biography---early works to 1800
The life of Charlotta Du Pont, an English lady; taken	Aubin, Penelope	1723	[charlotta du pont,]---biography---18th century---early works to 1800_virtue---early works to 1800
A stage-Coach journey to Exeter. Describing the humours on the	Manley, Delariviere, d. 1724	1725	voyages and travels---early works to 1800

Title	Author	Publication Date	LCSH
The dumb projector: being a surprizing account of a trip	Haywood, Eliza Fowler	1725	holland---description and travel---early works to 1800
The voyages and adventures of Captain Robert Boyle, in several	Chetwood, W. R. (William Rufus)	1726	voyages, imaginary---early works to 1800_pennsylvania---description and travel---early works to 1800
Philidore and Placentia: or, l'amour trop delicat. Part II. By	Haywood, Eliza Fowler	1727	english fiction---18th century
Memoirs of the court of Lilliput. / Written by Captain	NaN	1727	voyages, imaginary---fiction_imaginary places---fiction_imaginary societies---fiction_gulliver, lemuel (fictitious character)---fiction
The adventures of Abdalla, son of Hanif, sent by the	Bignon, Jean Paul, 1662–1743.	1729	abdalla, son of hanif_utopias

Appendix 3.6: Gendered Roles and Locations

Table 3.6.1: Ratio of Claim Locations

Claim Location	Raw Count	Ratio
title	132	0.737430
research	21	0.117318
peritext	19	0.106145
peritext, research	4	0.022346
text, research	1	0.005587
text	1	0.005587
research, text	1	0.005587

Table 3.6.2: Gendered Roles in the TWDB

Role	Raw Count	Ratio
au	139	0.776536
translator	17	0.094972
pt	9	0.050279
ed	8	0.044693
ed, pt	4	0.022346
translator, pt	1	0.005587
translator, au	1	0.005587

Table 3.6.3: Location of Claim of Gendered Author Role

Location of Claim	Raw Count from Search: au	Ratio from Search: au
title	112	0.800000
peritext	12	0.085714
research	12	0.085714
peritext, research	2	0.014286
text, research	1	0.007143
text	1	0.007143

Table 3.6.4: Location of Claim of Gendered Translator Role

Location of Claim	Raw Count from Search: translator	Ratio from Search: translator
title	12	0.631579
research	4	0.210526
peritext	2	0.105263
peritext, research	1	0.052632

Appendix 3.7: LCSH in the Gendered Corpus

Table 3.7.1: Main Heading Frequencies in Gendered TWDB Corpus

Main Heading	Raw Count	Frequency
france	25	0.081433
great britain	10	0.032573
europe	9	0.029316
italy	9	0.029316
united states	6	0.019544
voyages and travels	6	0.019544
crimea (ukraine)	4	0.013029
voyages, imaginary	4	0.013029
switzerland	4	0.013029
ireland	3	0.009772
london (england)	3	0.009772
indians of north america	3	0.009772
turkey	3	0.009772
women	3	0.009772
soviet union	3	0.009772
paris (france)	3	0.009772
scotland	3	0.009772
english letters	3	0.009772
natural history	3	0.009772
english fiction	3	0.009772

Table 3.7.2: Main Heading and Subdivision Frequencies in Gendered TWDB Corpus

Subdivision	Raw Count	Frequency
description and travel	109	0.161721
early works to 1800	55	0.081602
france	27	0.040059
history	20	0.029674
social life and customs	15	0.022255
antiquities	13	0.019288
fiction	13	0.019288
18th century	12	0.017804
great britain	12	0.017804
italy	11	0.016320

Table 3.7.3: Titles in the TWDB holding `wom`

Title	Author	Publication Date	LCSH	TWDB Gender Role
An excellent ballad, call'd The Blind beggar of Bednal Green.	NaN	1720	lifestyle---early works to 1800_courtship---early works to 1800_man-woman relationships---early works to 1800_travellers---great britain---early works to 1800_broadsides---great britain---early works to 1800_young women---conduct of life---early works to 1800	NaN
The spy at Oxford and Cambridge. Containing, many remarkable transactions,	Perspective, John _ Critical Wou'd-be	1744	women---conduct of life---early works to 1800_political satire, english---early works to 1800_women---moral and ethical aspects---early works to 1800_travelers' writings---great britain---early works to 1800	NaN
Letters written by a Peruvian princess. Translated from the French.	Grafigny, Mme de (Françoise d'Issembourg d'Happoncourt) (1695–1758)	1748	peruvians---france---fiction---early works to 1800_women---france---fiction---early works to 1800_france---social life and customs---fiction---18th century	au

Title	Author	Publication Date	LCSH	TWDB Gender Role
The loyal lovers garland. In four parts.	NaN	1760	love---early works to 1800_disguise---early works to 1800_merchants---early works to 1800_ballads, english---18th century_lifestyle---early works to 1800_courtship---early works to 1800_women sailors---early works to 1800_parent and child---early works to 1800_shipwrecks---poetry---early works to 1800_fathers and daughters---early works to 1800_broadsides---great britain---early works to 1800_voyages and travels---ocean travel---early works to 1800	NaN
Facts. The female spy; or Mrs. Tonkin's account of her	Tonkin, Mary	1783	women spies---great britain---early works to 1800_france---history---18th century_great britain---history---1760–1789	au
A trip to Paris in July and August, 1792.	Twiss, Richard, 1747–1821.	1793	women publishers---great britain---18th century_france---history---1789–1793---early works to 1800_paris (france)---description and travel---early works to 1800	NaN
An account of two charity schools for the education of	Cappe, Catharine	1800	charity-schools---england---york_poor---education---england---york_women---education---england---york_friendly societies---england---york_children---employment---england---york	au
The traveller in Asia: or, A visit to the most	Wakefield, Priscilla, 1751–1832	1817	voyages and travels---women_china---description and travel_india---description and travel	au
Narrative of the operations and recent discoveries within the pyramids,	Belzoni, Giovanni Battista, 1778–1823.	1820	women---egypt_nubia---antiquities_egypt---antiquities	pt
Letters from the Caucasus and Georgia: To Which Are Added,	Freygang, Frederika Kudrëiìavskaëiìa von	1823	women travelers---anecdotes_voyages and travels---women authors_caucasus---description and travel_iran---description and travel_iran---history---qajar dynasty, 1794–1925_travelers' writings---women authors	translator

Appendix 4.1: Models and "Accuracy"

For this chapter, "accuracy" describes whether a model trained on a category of writing can correctly identify a volume as fitting into that category. It considers both false positives (nontravel volumes that the trained model wrongfully identifies as travel) and false negatives (travel volumes that the model incorrectly identifies as nontravel). For example, one model of the entire TWDB with 300 travel volumes and 300 nontravel volumes has the following results:

> True positives: 256
> True negatives: 261
> False positives: 39
> False negatives: 44
> bestaccuracy: 0.86

A low accuracy close to 50% means that the model struggles to find features that distinguish the labeled category (travel writing) from the random category (nontravel writing). If there is a strong difference, however, then the accuracy percentage (i.e., how many volumes were correctly identified by the model) will be higher. As Underwood notes, models are good at discerning patterns: when he creates a "ghastly stew" of Gothic, detective, and science fiction, "even this sprawling category can be distinguished from a random background, on average, 78% of the time" (*DH* 47). The important next step, to see the differences between models, is to

> train a model on the contrast between detective fiction and a randomly selected background and then ask the same model to distinguish works of science fiction from the same background. As we might expect, this model fails utterly: it's right less than half of the time…Although these two genres have a few things in common (theories and laboratories, for instance), their common elements seem not to be the features that most saliently define them. (47)

Underwood demonstrates the approach with his chosen genres, as well as on two other aspects (character gender and reviews/prestige). My approach differs from Underwood's not just in that I

am focusing primarily on travel writing, but also in my data, which is from multiple sources that see a large shift in both printing and digitization practices and technology. The 130 years after 1700 saw printing become more consistent and higher quality, as well as the disappearance of the long-s, especially following the year 1800. [84] That time period is also exactly when ECCO's database ends and I shift to relying on HathiTrust, which often uses the more modern Google OCR technology.[85] So, if we examine whether travel writing becomes more consistent or distinctive in the latter part of the dataset, we need to know if the model is just interpreting features that are better digitized or OCRed. Of course, there is no guaranteed way to know without doing comparisons with reviewed and corrected OCR, but we can compare ECCO and HT in the 1790s, the closest period of overlap possible, to see if there are any big swings in accuracy.

Models are good at discerning patterns, regardless of whether the features are a word with semantic meaning. If I select a group of random ECCO files compare it to the rest of my random sample, then the accuracy is still around 50%, which is what we would expect for a random sample; the same happens for files from HathiTrust. However, when I compare such "randomly selected random volumes" from ECCO to a dataset made of random HathiTrust files (all from the decade 1790), the results get more interesting: the model is over 90% accurate at picking out which files are from ECCO, and which are from HT. Some features in these texts are making these texts, originating from the same time period, different enough that the model can distinguish whether they came from ECCO or HathiTrust. This obviously suggests significant challenges for anyone looking to simply combine these two datasets.

---

[84] See Paul W. Nash on "The Abandoning of the Long S in Britain in 1800."
[85] I only use files that were OCRed by Google (based on the advice of Peter Organisciak during his tenure at HT).

With the regular caveats about overstating the significance of individual features, many of the tokens that are most indicative of ECCO files are OCR or tokenizing errors: features such as om, re, ia, and even single letters such as z and t. Despite my early attempts to clean up the OCR, the model is a very discerning judge. However, travel models end up stronger than the OCR quality: an ECCO travel model can still recognize HT travel volumes against ECCO non-travel volumes 85% on average, and HT on ECCO, 82%, rates that are higher than the mutual recognition of other categories of the TWDB. If we do not separate the non-travel corpus by file source, models of travel writing in the 1790s that use only ECCO-sourced TWDB volumes can recognize HT-sourced TWDB volumes above 90% on average. Overall, then, the patterns of travel writing seem to be stronger than OCR, but for the pace of change models, I include options to compare the models based on file source.

Finally, for overall averages above and for the rest of this chapter, the numbers are based on three modelling runs. Averages of overall model comparisons are based on nine comparisons (each of the 3 Cox models compared to each of the 3 Gove models, for example). For volume predictions, if the volume was used in the training set for the model, the prediction rate comes from that model; if it is not in the model, then the prediction rate comes from applying the model to the text. Unless otherwise noted, these predictions are also based on three separate modelling runs, meaning that for larger models that do not always use the same training set, some averages may consist of predictions from both the model in training and in application.

Appendix 4.2: Mutual Recognition

Table 4.2.1: Mutual Recognition of Bibliographical Sources

category	andrews	bdanth	brynmawr	btw_europe	btw_w	cox	gove	irishmcveagh	leask	murray	ncco_c19trav	ncco_travelnarr	robinson_w	tee
andrews		88.3	88.0	88.2	85.7	75.5	66.1	90.4	84.5	85.5	87.4	77.8	85.2	88.7
bdanth	88.3		90.5	85.4	87.1	82.0	82.0	78.0	88.1	86.6	84.8	83.1	89.0	88.5
brynmawr	88.0	90.5		89.0	91.1	83.9	78.5	82.1	92.1	90.2	92.8	79.8	89.6	90.3
btw_europe	88.2	85.4	89.0		89.8	75.1	66.3	84.3	87.2	86.3	90.8	79.6	89.4	87.5
btw_w	85.7	87.1	91.1	89.8		75.2	74.3	82.5	85.7	85.8	88.4	77.1	89.8	85.6
cox	75.5	82.0	83.9	75.1	75.2		70.8	75.3	79.3	78.3	77.6	76.1	73.9	76.7
gove	66.1	82.0	78.5	66.3	74.3	70.8		58.7	74.7	77.6	72.2	73.7	71.8	72.7
irishmcveagh	90.4	78.0	82.1	84.3	82.5	75.3	58.7		77.7	75.7	83.4	72.2	82.9	77.4
leask	84.5	88.1	92.1	87.2	85.7	79.3	74.7	77.7		91.3	88.0	79.4	85.1	92.7
murray	85.5	86.6	90.2	86.3	85.8	78.3	77.6	75.7	91.3		88.0	78.8	85.0	93.1
ncco_c19trav	87.4	84.8	92.8	90.8	88.4	77.6	72.2	83.4	88.0	88.0		79.5	88.9	89.2
ncco_travelnarr	77.8	83.1	79.8	79.6	77.1	76.1	73.7	72.2	79.4	78.8	79.5		76.3	80.8
robinson_w	85.2	89.0	89.6	89.4	89.8	73.9	71.8	82.9	85.1	85.0	88.9	76.3		86.4
tee	88.7	88.5	90.3	87.5	85.6	76.7	72.7	77.4	92.7	93.1	89.2	80.8	86.4	

Table 4.2.2: McVeagh's Irish Travels and Location Mutual Recognition

Location Model	**Mutual Recognition with McVeagh's Irish Travels**
loc_oceania	71.6
loc_africa	73.7
loc_carib	74.2
loc_seasia	75.6
loc_latinam	75.9
loc_meast	76.0
loc_easia	77.2
loc_northam	80.8
loc_europe	86.6
loc_ireland	91.9
loc_gb	92.6

Table 4.2.3: Voyages, Imaginary: Select Titles

docid	author	title	date	fictitious	gove	lcsh_imaginary	lcsh_fiction	imagfiction	nontravel_fiction	alltravel	lcsh_traveltag	btw_w	robinson_w
dul1.ark+=13960=t4mk7258x	Scott, Sarah	A description of Millenium Hall, and the count...	1762	95.2	86.0	99.2	85.6	96.3	96.9	36.4	28.2	29.2	23.2
0509400601	Smollett, Tobias George	The expedition of Humphry Clinker. By the auth...	1771	50.6	67.5	80.4	71.2	83.8	76.1	41.0	40.7	19.9	43.0
0509400602	Smollett, Tobias George	The expedition of Humphry Clinker. By the auth...	1771	65.4	75.3	86.1	79.1	88.2	83.8	50.8	49.6	22.0	36.8
0509400603	Smollett, Tobias George	The expedition of Humphry Clinker. By the auth...	1771	71.4	77.1	93.8	78.5	89.8	84.6	67.2	62.8	34.7	52.0
1269300201	Sterne, Laurence	A sentimental journey through France and Italy...	1768	65.9	75.4	94.9	74.7	72.2	86.3	43.8	38.1	25.7	59.5
1269300202	Sterne, Laurence	A sentimental journey through France and Italy...	1768	68.8	81.9	96.9	76.7	78.9	92.0	55.9	43.6	26.4	71.2
0647100501	Swift, Jonathan	Travels into several remote nations of the wor...	1726	96.6	87.0	100.0	86.6	97.6	77.0	86.8	83.9	29.9	79.7
0647100502	Swift, Jonathan	Travels into several remote nations of the wor...	1726	97.3	84.7	100.0	85.9	98.3	79.8	84.2	77.3	25.0	55.0
mdp.39015078568592	[unlisted-twdb]	Travels into several remote nations of the wor...	1727	98.0	92.4	100.0	90.7	99.6	97.1	94.3	91.2	46.8	79.7

Table 4.2.4: Poetry Model Mutual Recognition

Model	Mutual Recognition with lcsh_title_poetry
gove	60.4
leask	60.5
bdanth	63.3
cox	64.2
robinson_w	65.8
tee	66.9
alltravel	68.0
lcsh_traveltag	69.5
brynmawr	69.7
murray	70.4
btw_europe	70.8
btw_w	71.2
irishmcveagh	72.4
ncco_travelnarr	73.1
ncco_c19trav	73.6
andrews	74.3
nontravel_poetry	84.5

Table 4.2.5: History Model Accuracy

Model	Average Accuracy	Total Volumes
nontravel_history	78.7	300
lcsh_history	88.0	300
lcsh_antiquities	90.7	279

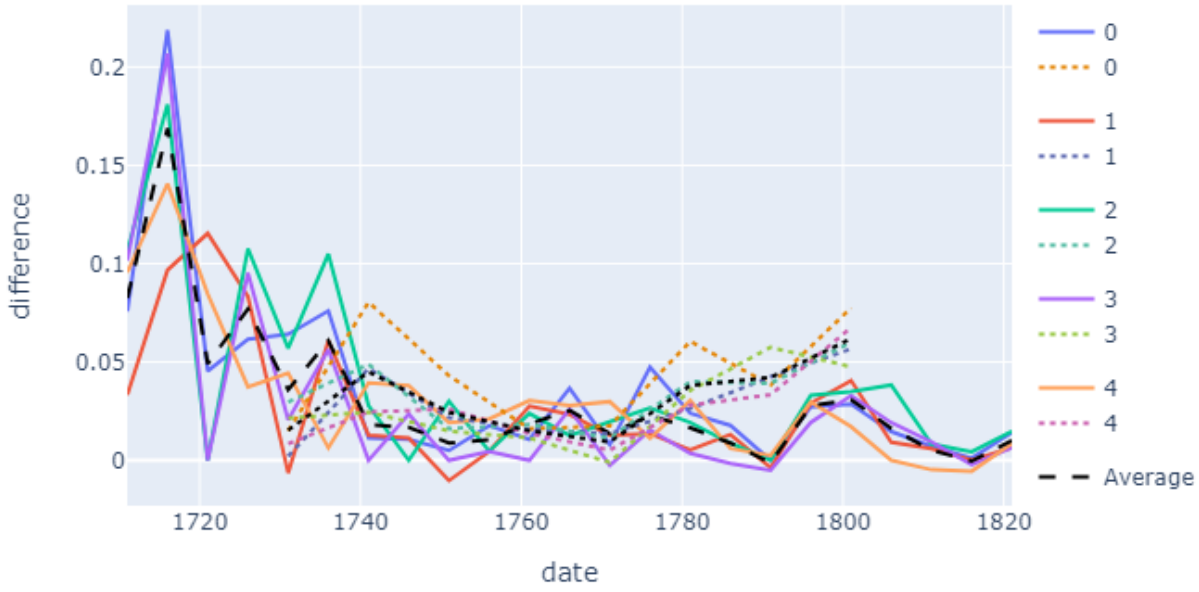
Table 4.2.6: History Model Mutual Recognition

Model	Mutual Recognition with lcsh_history
gove	67.4
robinson_w	69.7
btw_europe	70.9
andrews	71.0
irishmcveagh	71.9
ncco_travelnarr	73.1

Model	Mutual Recognition with lcsh_history
btw_w	74.1
tee	74.1
murray	76.6
ncco_c19trav	78.3
leask	78.6
bdanth	78.7
brynmawr	84.8
cox	84.8
alltravel	86.1

Appendix 4.3: Pace of Change

Figure 4.3.1: Pace of Change in LCSH `Description and Travel` Model



Bibliography

Primary Sources

Berchtold, Leopold. *An Essay to Direct and Extend the Inquiries of Patriotic Travellers; with Further Observations on the Means of Preserving the Life, Health, & Property of the Unexperienced ... By Count Leopold Berchtold, ...*, vol. 1, printed for the author, and sold by Mr. Robinson, Mr. Debrett, Mr. Payne, Mr. Jeffery & Mr. Faulder, 1789. *ECCO*, https://link-gale-com.proxy3.library.mcgill.ca/apps/doc/CW0105108831/ECCO?u=crepuq_mcgill&sid=gale_marc&xid=87b77da5&pg=46.

Boswell, James. *James Boswell: The Journal of His German and Swiss Travels, 1764*. Edited by Marlies K. Danziger, Edinburgh UP, 2008. https://www.oxfordscholarlyeditions.com/display/10.1093/actrade/9780748618064.book.1/actrade-9780748618064-book-1.

Cowper, William. *The Poems of William Cowper, Vol. 2: 1782–1785*. Edited by John D. Baird and Charles Ryskamp, Oxford UP. https://www.oxfordscholarlyeditions.com/display/10.1093/actrade/9780198123392.book.1/actrade-9780198123392-book-1.

Freeman, Ann, and Henry Freeman. *A Memoir of the Life and Ministry of Ann Freeman, Written by Herself, and an Account of Her Death, by H. Freeman*. Harvey and Darton, 1826. *Google Books*, https://books.google.ca/books?id=EBuWY8jmOMAC.

Grant, Anne MacVicar. *Letters from the Mountains: Being the Real Correspondence of a Lady, between the Years 1773 and 1803.* Printed for Longman, Hurst, Rees, and Orme, 1806. *HathiTrust*, https://catalog.hathitrust.org/Record/008646088.

Johnson, Samuel. *The Letters of Samuel Johnson, Volume II: 1773–1776*. Edited by Bruce Redford, Hyde Edition, Princeton UP, 2014. *Project MUSE*, https://muse.jhu.edu/pub/267/monograph/book/34292.

---. *The Letters of Samuel Johnson, Volume IV: 1782–1784*. Edited by Bruce Redford, Hyde Edition, Princeton UP, 2017. *Project MUSE*, https://muse.jhu.edu/pub/267/monograph/book/61213.

---. *The Yale Edition of the Works of Samuel Johnson*. Edited by John M. Bullitt et al., vol. 2, Yale UP, 1958.

Piozzi, Hester Lynch. *Observations and Reflections Made in the Course of a Journey through France, Italy, and Germany*. Printed for A. Strahan; and T. Cadell in the Strand, 1789.

Cambridge Library Collection, vol. 1, Cambridge UP, 2011, https://doi.org/10.1017/CBO9780511732003.

Piozzi, Hester Thrale. *Thraliana: The Diary of Mrs. Hester Lynch Thrale (Later Mrs. Piozzi) 1784–1809*. Edited by Katharine C. Balderston, vol. 2, Oxford UP, 2015. https://www.oxfordscholarlyeditions.com/display/10.1093/actrade/9780198739029.book.1/actrade-9780198739029-book-1.

Pope, Alexander. *The Dunciad in Four Books*. Edited by Valerie Rumbold, 2nd ed., Routledge, 2014, https://doi.org/10.4324/9781315833897.

Sterne, Laurence. *A Sentimental Journey and Other Writings*. Edited by Tim Parnell and Ian Jack, Oxford UP, 2003.

West, Thomas. *A Guide to the Lakes: Dedicated to the Lovers of Landscape Studies, and to All Who Have Visited, or Intend to Visit the Lakes in Cumberland, Westmorland, and Lancashire. By the Author of The Antiquities of Furness*. Printed for Richardson and Urquhart, under the Royal Exchange, and W. Pennington, Kendal, 1778. *Gale*, https://link-gale-com.proxy3.library.mcgill.ca/apps/doc/CW0102246963/ECCO?u=crepuq_mcgill&sid=gale_marc&xid=405b5f1f&pg=4.

Secondary Sources

Abuelwafa, Sherif, et al. “Detecting Footnotes in 32 Million Pages of ECCO.” *Journal of Cultural Analytics*, 2018, https://doi.org/10.22148/16.029.

Adams, Percy G. *Travel Literature and the Evolution of the Novel*. UP of Kentucky, 1983.

---. *Travelers and Travel Liars, 1660–1800*. U of California P, 1962.

---. “Travel Literature of the Seventeenth and Eighteenth Centuries: A Review of Recent Approaches.” *Texas Studies in Literature and Language*, vol. 20, no. 3, 1978, pp. 488–515.

Agorni, Mirella. *Translating Italy for the Eighteenth Century: British Women, Translation and Travel Writing (1739–1797)*. Routledge, 2014.

Alacovska, Ana. “Genre Anxiety: Women Travel Writers’ Experience of Work.” *The Sociological Review*, vol. 63, no. S1, 2015, pp. 128–43, https://doi.org/10.1111/1467–954X.12246.

Algee-Hewitt, Mark, et al. “Canon/Archive. Large-Scale Dynamics in the Literary Field.” *Pamphlets of the Stanford Literary Lab*, vol. 11, Stanford Literary Lab, 2016.

Anderson, John Parker. *The Book of British Topography. A Classified Catalogue of the Topographical Works in the Library of the British Museum Relating to Great Britain and Ireland*. Theatrum Orbis Terrarum, 1966.

Andrews, Malcolm. *The Search for the Picturesque: Landscape Aesthetics and Tourism in Britain, 1760–1800*. Stanford UP, 1989.

Baird, Ileana, editor. *Data Visualization in Enlightenment Literature and Culture*. Palgrave Macmillan, 2021, https://doi.org/10.1007/978–3-030–54913–8.

Batten, Charles L., Jr. *Pleasurable Instruction: Form and Convention in Eighteenth-Century Travel Literature*. U of California P, 1978.

Beasley, Jerry C. *Novels of the 1740s*. U of Georgia Press, 1982.

Belgum, Kirsten, et al. “Mapping Travel Writing: A Digital Humanities Project to Visualise Change in Nineteenth-Century Published Travel Texts.” *Studies in Travel Writing*, vol. 22, no. 3, July 2018, pp. 306–24, https://doi.org/10.1080/13645145.2019.1575765.

Benedict, Barbara M. *Making the Modern Reader: Cultural Mediation in Early Modern Literary Anthologies*. Princeton UP, 2018, https://www.degruyter.com/isbn/9780691193977.

Biber, Douglas. “A Typology of English Texts.” *Linguistics*, vol. 27, no. 1, Jan. 1989, pp. 3–44, https://doi.org/10.1515/ling.1989.27.1.3.

Board, C. “Maps as Models.” *Physical and Information Models in Geography (Routledge Revivals)*, Routledge, 1967.

Bode, Katherine. “The Equivalence of ‘Close’ And ‘Distant’ Reading; Or, toward a New Object for Data-Rich Literary History.” *Modern Language Quarterly*, vol. 78, no. 1, Mar. 2017, pp. 77–106, https://doi.org/10.1215/00267929–3699787.

---. “Why You Can’t Model Away Bias.” *Modern Language Quarterly*, vol. 81, no. 1, Mar. 2020, pp. 95–124, https://doi.org/10.1215/00267929–7933102.

---. *A World of Fiction: Digital Collections and the Future of Literary History*. U of Michigan Press, 2018.

Bohls, Elizabeth A. “Age of Peregrination: Travel Writing and the Eighteenth-Century Novel.” *A Companion to the Eighteenth-Century English Novel and Culture*, John Wiley & Sons, 2005, pp. 97–116, https://doi.org/10.1002/9780470996232.ch6.

---. *Women Travel Writers and the Language of Aesthetics, 1716–1818*. Cambridge UP, 1995.

Bohls, Elizabeth A., and Ian Duncan, editors. *Travel Writing, 1700–1830: An Anthology*. Oxford UP, 2005. http://site.ebrary.com/id/10177896.

Bolufer Peruga, Mónica, et al. *Gender and Cultural Mediation in the Long Eighteenth Century: Women across Borders*. Palgrave Macmillan, 2024, https://doi.org/10.1007/978–3-031–46939–8.

Booth, Alison. “Distant Horizons: Digital Evidence and Literary Change by Ted Underwood (Review).” *Victorian Studies*, vol. 62, no. 3, 2020, pp. 547–50.

Borm, Jan. “Defining Travel: On the Travel Book, Travel Writing and Terminology.” *Perspectives on Travel Writing*, edited by Glenn Hooper and Tim Youngs, Routledge, 2004.

Brewer, David. “‘Haywood,’ Secret History, and the Politics of Attribution.” *The Passionate Fictions of Eliza Haywood: Essays on Her Life and Work*, edited by Jeremy W. Webster et al., UP of Kentucky, 2001, pp. 217–39.

Bryn Mawr College Library Special Collection. “European Travel Accounts Printed Before 1850.” *Bryn Mawr College Library Special Collections*, 2018, https://wayback.archive-it.org/230/20210310150611/http://bascom.brynmawr.edu/library/speccoll/guides/travel/index.html.

Bullard, Paddy. “Digital Humanities and Electronic Resources in the Long Eighteenth Century.” *Literature Compass*, vol. 10, no. 10, 2013, pp. 748–60, https://doi.org/10.1111/lic3.12085.

Burrows, Simon, et al. *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies*. Liverpool UP, 2020.

Buzard, James. “The Grand Tour and after (1660–1840).” *The Cambridge Companion to Travel Writing*, edited by Peter Hulme and Tim Youngs, Cambridge UP, 2002, pp. 37–52. *Cambridge UP*, https://doi.org/10.1017/CCOL052178140X.003.

Campbell, Mary B. “Travel Writing and Its Theory.” *The Cambridge Companion to Travel Writing*, edited by Peter Hulme and Tim Youngs, Cambridge UP, pp. 261–78.

---. *The Witness and the Other World: Exotic European Travel Writing, 400–1600*. Cornell UP, 1991.

Carretta, Vincent. "Strangers in Strange Lands." *Eighteenth-Century Studies*, vol. 36, no. 2, Dec. 2003, pp. 255–58, https://doi.org/10.1353/ecs.2003.0009.

---. *Unchained Voices: An Anthology of Black Authors in the English-Speaking World of the Eighteenth Century*. Expanded ed, UP of Kentucky, 2004, http://site.ebrary.com/id/10674268.

Caughie, Pamela L., et al. "Storm Clouds on the Horizon: Feminist Ontologies and the Problem of Gender." *Feminist Modernist Studies*, vol. 1, no. 3, Sept. 2018, pp. 230–42, https://doi.org/10.1080/24692921.2018.1505819.

Cayley, Seth. "Digitization for the Masses: Taking Users Beyond Simple Searching in Nineteenth-Century Collections Online." *Journal of Victorian Culture*, vol. 22, no. 2, Apr. 2017, pp. 248–55, https://doi.org/10.1080/13555502.2017.1301180.

Ceserani, Giovanna, director. "The Grand Tour Project." *The Grand Tour Project*, https://grandtour.stanford.edu/. Accessed 16 May 2024.

---. *A World Made by Travel: The Digital Grand Tour*. Stanford UP, 2024, https://aworldmadebytravel.org.

Chandler, James. "The Languages of Sentiment." *Textual Practice*, vol. 22, no. 1, Mar. 2008, pp. 21–39, https://doi.org/10.1080/09502360701841886.

Chard, Chloe. *Pleasure and Guilt on the Grand Tour: Travel Writing and Imaginative Geography, 1600–1830*. St. Martin's Press, 1999.

Child, Elizabeth. "'To Sing the Town': Women, Place, and Print Culture in Eighteenth-Century Bath." *Studies in Eighteenth-Century Culture*, vol. 28, no. 1, 1999, pp. 155–72.

Choi, Yoojung. *Women's Mobility, Travel, and Literary Representations in the Long Eighteenth Century*. 2020. Texas A&M U.

Christy, Matthew, et al. "Mass Digitization of Early Modern Texts With Optical Character Recognition." *Journal on Computing and Cultural Heritage*, vol. 11, no. 1, Dec. 2017, p. 6:1–6:25, https://doi.org/10.1145/3075645.

Churchill, Rob, and Lisa Singh. "The Evolution of Topic Modeling." *ACM Computing Surveys*, vol. 54, no. 10s, Nov. 2022, pp. 215:1–215:35, https://doi.org/10.1145/3507900.

Clark, S. H. *Travel Writing and Empire: Postcolonial Theory in Transit*. St. Martin's Press, 1999.

Cohen, Margaret. *The Sentimental Education of the Novel*. Princeton UP, 1999, https://doi.org/10.1515/9780691188249.

Colbert, Benjamin. "9 May 2018: Summary Report for Women Travel Writers, 1780–1840." *Women Travel Writers, 1780–1840*, 5 Sept. 2018, https://btw.wlv.ac.uk/news/1009.

---. "About Women's Travel Writing 1780–1840." *British Travel Writing: Women's Travel Writing, 1780–1840: A Bio-Bibliographical Database*, https://btw.wlv.ac.uk/about. Accessed 17 May 2024.

---. "Arrivals and Departures." *British Travel Writing*, 28 Sept. 2017, https://btw.wlv.ac.uk/news/1007.

---. "Bibliography of British Travel Writing, 1780 - 1840: The European Tour, 1814 - 1818 (Excluding Britain and Ireland)." *Romantic Textualities: Literature and Print Culture, 1780–1840*, vol. 13, 2004, pp. 5–43.

---. *British Travel Writing: Women's Travel Writing, 1780–1840: A Bio-Bibliographical Database*. 2018, https://btw.wlv.ac.uk/.

---. "British Women's Travel Writing, 1780–1840: Bibliographical Reflections." *Women's Writing*, vol. 24, no. 2, July 2016, pp. 1–19, https://doi.org/10.1080/09699082.2016.1207913.

---. *Database of British Travel Writing*. 16 Jan. 2017, Received by Catherine Nygren.

---. "'Our observations should not be disunited' Collaborative Women's Travel Writing, 1780–1840." *Viatica*, vol. 3, Mar. 2016, https://doi.org/10.52497/viatica553.

---. *Shelley's Eye: Travel Writing and Aesthetic Vision*. Ashgate, 2005.

Conant, Martha Pike. *The Oriental Tale in England in the Eighteenth Century*. Routledge, 1908, https://doi.org/10.4324/9780429030710.

Coolidge, W. A. B. *Swiss Travel and Swiss Guide-Books*. Longmans, Green and Co., 1889.

Cordell, Ryan. "'Q i-Jtb the Raven': Taking Dirty OCR Seriously." *Book History*, vol. 20, no. 1, Oct. 2017, pp. 188–225. *Project MUSE*, https://doi.org/10.1353/bh.2017.0006.

Corvey Project. https://extra.shu.ac.uk/corvey/twfset.htm. Accessed 16 May 2024.

Cox, Edward Godfrey. *A Reference Guide to the Literature of Travel, Including Voyages, Geographical Descriptions, Adventures, Shipwrecks and Expeditions*. U of Washington, 1935–1949. *HathiTrust*, https://catalog.hathitrust.org/Record/000883216.

Da, Nan Z. "The Computational Case against Computational Literary Studies." *Critical Inquiry*, vol. 45, no. 3, Mar. 2019, pp. 601–39, https://doi.org/10.1086/702594.

Das, Nandini, and Tim Youngs, editors. "Travel Writing in a Global Context." *The Cambridge History of Travel Writing*, Cambridge UP, 2019, pp. 141–298.

Dasu, Tamraparni, and Theodore Johnson. *Exploratory Data Mining and Data Cleaning*. John Wiley & Son, 2003, https://doi.org/10.1002/0471448354.

"Data Management Plans for NEH Office of Digital Humanities Proposals and Awards." *The National Endowment for the Humanities*, 2019, https://www.neh.gov/sites/default/files/inline-files/Data%20Management%20Plans%2C%202019.pdf.

David Mimno [@dmimno]. "@quadrismegistus Average Accuracy Is Less Important than Distribution. Differences Due to Font, Engine, Scan, Printing Will Swamp Useful Patterns." *Twitter*, 6 July 2017, https://twitter.com/dmimno/status/883056486024400896.

Day, Matthew. "Travelling in New Formes: Reissued and Reprinted Travel Literature in the Long Eighteenth Century." *Mémoires Du Livre*, vol. 4, no. 2, June 2013, https://doi.org/10.7202/1016741ar.

Dekker, George. *The Fictions of Romantic Tourism: Radcliffe, Scott, and Mary Shelley*. Stanford UP, 2005.

DeLucia, JoEllen. "Local Poetry in the Midlands: Francis Mundy's Needwood Forest and Anna Seward's Lichfield Poems." *Representing Place in British Literature and Culture, 1660–1830*, Routledge, 2013, pp. 155–71.

---. "Travel Writing and Mediation in the *Lady's Magazine*: Charting 'the Meridian of Female Reading.'" *13 Travel Writing and Mediation in the Lady's Magazine: Charting 'the Meridian of Female Reading,'* Edinburgh UP, 2022, pp. 205–16, https://doi.org/10.1515/9781474419666–018.

Derrida, Jacques. "The Law of Genre." *Critical Inquiry*, translated by Avital Ronell, vol. 7, no. 1, 1980, pp. 55–81.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. MIT Press, 2020.

Dombrowski, Quinn. *Towards a Taxonomy of Failure*. https://www.quinndombrowski.com/blog/2019/01/30/towards-taxonomy-failure/. On the Benefits of Failure, U of Alberta.

Doody, Margaret Anne. *Frances Burney: The Life in the Works*, Rutgers UP, 1988, https://doi.org/10.36019/9780813568485.

Drabinski, Emily. "Teaching the Radical Catalog." *Radical Cataloging: Essays at the Front*, edited by K.R. Roberto, McFarland & Company, 2008, pp. 198–205.

Drucker, Johanna. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly*, vol. 5, no. 1, Mar. 2011.

Duffy, Eamon P. "Searching HathiTrust: Old Concepts in a New Context." *Partnership: The Canadian Journal of Library and Information Practice and Research*, vol. 8, no. 1, 1, June 2013, https://doi.org/10.21083/partnership.v8i1.2503.

Durie, Kate. "Travel Writing 1700–1830: An Anthology. Oxford World's Classics." *Tourism Management*, vol. 28, no. 3, 2007, pp. 927–28, https://doi.org/10.1016/j.tourman.2006.05.018.

Dussinger, John A. "Hester Piozzi, Italy, and the Johnsonian Aether." *South Central Review*, vol. 9, no. 4, 1992, pp. 46–58, https://doi.org/10.2307/3189480.

Earhart, Amy E. "Can Information Be Unfettered? Race and the New Digital Humanities Canon." *Debates in the Digital Humanities 2012*, U of Minnesota P, 2012. https://minnesota.universitypressscholarship.com/view/10.5749/minnesota/9780816677948.001.0001/upso-9780816677948-chapter-30.

Eaves, T. C. Duncan and Ben D. Kimpel. *Samuel Richardson: A Biography*. Clarendon Press, 1971.

Eggert, Paul. *Securing the Past: Conservation in Art, Architecture and Literature*. Cambridge UP, 2009. http://catdir.loc.gov/catdir/enhancements/fy0909/2008035216-t.html.

Evalyn, Lawrence. *Database Representations of English Literature, 1789–99*. 2022. U of Toronto.

Ezell, Margaret J. M. "'By a Lady': The Mask of the Feminine in Restoration, Early Eighteenth-Century Print Culture." *The Faces of Anonymity: Anonymous and Pseudonymous Publication from the Sixteenth to the Twentieth Century*, edited by Robert J. Griffin, Palgrave Macmillan US, 2003, pp. 63–79, https://doi.org/10.1007/978–1-137–11109–8_4.

---. “Editing Early Modern Women’s Manuscripts: Theory, Electronic Editions, and the Accidental Copy-Text.” *Literature Compass*, vol. 7, no. 2, Feb. 2010, pp. 102–09, https://doi.org/10.1111/j.1741–4113.2009.00682.x.

“Frequently Asked Questions.” *Historical Texts*, https://historicaltexts.jisc.ac.uk/faq. Accessed 19 May 2024.

Fielding, Henry. *The Journal of a Voyage to Lisbon, Shamela, and Occasional Writings*. Edited by Martin C. Battestin, Oxford UP, 2008, http://oxfordscholarlyeditions.com/view/10.1093/actrade/9780199266753.book.1/actrade-9780199266753-book-1.

Fletcher, Loraine. “Introduction.” *Emmeline: The Orphan of the Castle*, by Charlotte Smith, edited by Loraine Fletcher, Broadview Press, 2003, pp. 9–36.

Foster, Shirley, and Sara Mills (editors). *An Anthology of Women’s Travel Writings*. Palgrave, 2002.

Frame, William, and Laura Walker. *James Cook: The Voyages*. McGill-Queen’s UP, 2018. https://canadacommons.ca/artifacts/1878398/james-cook/2627793/.

François, Pieter. “If It’s 1815, This Must Be Belgium: The Origins of the Modern Travel Guide.” *Book History*, vol. 15, no. 1, 2012, pp. 71–92.

Friedman, Emily C. “Afterword: Novel Knowledge, or Cleansing Dirty Data: Toward Open-Source Histories of the Novel.” *Data Visualization in Enlightenment Literature and Culture*, edited by Ileana Baird, Palgrave MacMillan, 2021, pp. 351–70, https://doi.org/10.1007/978–3-030–54913–8_10.

Fulford, Tim, and Peter J. Kitson, editors. *Travels, Explorations, and Empires: Writings from the Era of Imperial Expansion, 1770–1835.* Pickering & Chatto, 2001.

Gale, Nineteenth Century Collections Online. “19th Century Travel Literature, from the Bryn Mawr College Library.” *Nineteenth Century Collections Online*, Gale, A Cengage Company, 2019, https://link.gale.com/apps/collection/6SWS/NCCO?u=crepuq_mcgill&sid=bookmark-NCCO.

---. *Advisory Board Members*. Gale, A Cengage Company. *Zotero*, https://www.gale.com/binaries/content/assets/gale-us-en/primary-sources/nineteenth-century-collections-online/gps_ncco_advisory-board.pdf. Accessed 4 Apr. 2024.

---. *Mapping the World: Maps and Travel Literature Factsheet*. Gale, A Cengage Company, 2014, https://www.gale.com/binaries/content/assets/gale-us-en/primary-sources/nineteenth-century-collections-online/gps_ncco_mappingtheworld_academic_factsheet.pdf.

---. "Maps and Travel Literature." *Nineteenth Century Collections Online*, https://link.gale.com/apps/archive/6RWA/NCCO?u=crepuq_mcgill&sid=bookmark-NCCO. Accessed 18 May 2024.

---. "NCCO Advisory Board." *Nineteenth Century Collections Online*, 10 Mar. 2018, https://web.archive.org/web/20180310110108/http:/gale.cengage.co.uk/product-highlights/history/nineteenth-century-collections-online/ncco-advisory-board.aspx.

---. "Travel Narratives and Personal Reminiscences." *Nineteenth Century Collections Online*, Gale, A Cengage Company, https://link.gale.com/apps/collection/6SWO/NCCO?u=crepuq_mcgill&sid=bookmark-NCCO. Accessed 19 Feb. 2019.

Gale Staff. "Optical Character Recognition (OCR) Improvement: Enabling Deeper Historical Research through Innovation." *Gale Blog: Library & Educator News*, 30 Mar. 2023, https://blog.gale.com/optical-character-recognition-ocr-improvement-enabling-deeper-historical-research-through-innovation/.

Garrett, Jeffrey. "Subject Headings in Full-Text Environments: The ECCO Experiment." *College & Research Libraries*, vol. 68, no. 1, 2007, pp. 69–81, https://doi.org/10.5860/crl.68.1.69.

Gerhard, Kristin H., et al. "An Empirical Examination of Subject Headings for Women's Studies Core Materials." *College & Research Libraries*, vol. 59, no. 2, Mar. 1998, pp. 129–37, https://doi.org/10.5860/crl.59.2.129.

Geriguis, Lora Edmister. "Transplanting the Duchess: Margaret Cavendish and the 'Chronic Dilemmas' of Literary Anthology Construction." *English Studies*, vol. 98, no. 8, Nov. 2017, pp. 897–916, https://doi.org/10.1080/0013838X.2017.1367528.

Gleadhill, Emma. *Taking Travel Home: The Souvenir Culture of British Women Tourists, 1750–1830*. Manchester UP, 2022, https://www.jstor.org/stable/10.2307/j.ctv2j04snq.

Goldstone, Andrew. "Of Literary Standards and Logistic Regression: A Reproduction." *Andrew Goldstone*, 4 Jan. 2016, https://andrewgoldstone.com/blog/2016/01/04/standards/.

---. "Teaching Quantitative Methods: What Makes It So Hard (In Literary Studies)." *Debates in the Digital Humanities 2019*, edited by Matthew K. Gold and Lauren F. Klein, U of Minnesota P, 2019, pp. 209–23.

Goodwin, Jonathan. “Darko Suvin’s Genres of Victorian SF Revisited.” *Jonathan Goodwin*, 17 Oct. 2016, https://www.jgoodwin.net/blog/more-suvin/.

---. “Machine-Classified Microgenres.” *The Microgenre: A Quick Look at Small Culture*, Bloomsbury Academic, 2019, https://doi.org/10.5040/9781501345845.

Government of Canada, Innovation. *Tri-Agency Research Data Management Policy*. Innovation, Science and Economic Development Canada, 14 Mar. 2021, https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management/tri-agency-research-data-management-policy.

Graham, Elyse. “Joyce and the Graveyard of Digital Empires.” *Debates in the Digital Humanities 2019*, edited by Matthew K. Gold and Lauren F. Klein, U of Minnesota P, 2019, pp. 350–68, https://doi.org/10.5749/j.ctvg251hk.32.

Graham, Ruth. “Juvenile Travellers: Priscilla Wakefield’s Excursions in Empire.” *The Journal of Imperial and Commonwealth History*, vol. 38, no. 3, Sept. 2010, pp. 373–93, https://doi.org/10.1080/03086534.2010.503394.

Graham, Shawn. *Failing Gloriously and Other Essays*. Digital Press at the U of North Dakota, 2019, https://digitalpressatund.files.wordpress.com/2019/12/failing_gloriously_final.pdf.

Graham, Shawn, Ian Milligan, et al. *Exploring Big Historical Data: The Historian’s Macroscope*. 2nd ed., World Scientific, 2022.

Gray, Erik. “Doctor Syntax and the Erotics of the Picturesque.” *Literary Imagination*, vol. 24, no. 3, Nov. 2022, pp. 232–50, https://doi.org/10.1093/litimag/imac046.

Gregg, Stephen H. “1748: ‘Fiction’ in the Database.” *Manicule*, 24 Mar. 2016, https://shgregg.hcommons.org/2016/03/24/1748-fiction-in-the-database/.

---. *Old Books and Digital Publishing: Eighteenth-Century Collections Online*. Cambridge UP, 2021, https://doi.org/10.1017/9781108767415.

Gregory, Ian, director. *Geospatial Innovation in the Digital Humanities: A Deep Map of the English Lake District*. https://wp.lancs.ac.uk/lakesdeepmap/. Accessed 16 May 2024.

---. *Mapping the Lakes: A Literary GIS*. https://www.lancaster.ac.uk/mappingthelakes/. Accessed 16 May 2024.

Grenby, M. O. “Tarry-at-Home Antiquarians: Children’s ‘Tour Books’, 1740–1850.” *Pasts at Play*, edited by Rachel Bryant Davies and Barbara Gribling, Manchester UP, 2021, https://doi.org/10.7765/9781526128904.00018.

Gross, Tina, et al. “Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching.” *Cataloging & Classification Quarterly*, vol. 53, no. 1, Jan. 2015, pp. 1–39, https://doi.org/10.1080/01639374.2014.917447.

Guillory, John. *Cultural Capital: The Problem of Literary Canon Formation*. U of Chicago P, 1993.

Hagglund, Betty. “Gendering the Scottish Guidebook: Sarah Murray’s Companion and Useful Guide.” *Women’s Writing*, vol. 17, no. 1, May 2010, pp. 129–46, https://doi.org/10.1080/09699080903533320.

---. *Tourists and Travellers: Women’s Non-Fictional Writing about Scotland, 1770–1830*. Channel View Publications, 2009.

Hagglund, Elizabeth. “Reviews of Travel in the Monthly Review, 1749–1758: An Introductory Survey.” *Studies in Travel Writing*, vol. 2, no. 1, Mar. 1998, pp. 1–45, https://doi.org/10.1080/13645145.1998.9634871.

Hata, Gohei. “Robert Traill: The First Irish Critic of William Whiston’s Translation of Josephus.” *Making History*, edited by Zuleika Rodgers, vol. 110, Brill, 2007, pp. 415–35, https://doi.org/10.1163/ej.9789004150089.i-471.143.

HathiTrust Digital Library. “Requesting and Using Research Datasets.” *Hathi Trust Digital Library*, https://www.hathitrust.org/member-libraries/resources-for-librarians/data-resources/research-datasets/. Accessed 18 May 2024.

---. “Welcome to HathiTrust.” *HathiTrust Digital Library*, https://www.hathitrust.org/about/. Accessed 22 Apr. 2024.

HathiTrust Research Center. “HTRC Data Capsule Environment.” *HathiTrust Research Center Documentation*, 7 Aug. 2023, https://htrc.atlassian.net/wiki/spaces/COM/pages/43286886/HTRC+Data+Capsule+Environment.

---. “Troubleshooting and FAQs - Documentation - HTRC Docs.” *HathiTrust Research Center*, https://wiki.htrc.illinois.edu/display/COM/Troubleshooting+and+FAQs. Accessed 24 June 2020.

Herther, Nancy, et al. “Text and Data Mining Contracts: The Issues and Needs.” *Charleston Library Conference*, Oct. 2016, https://docs.lib.purdue.edu/charleston/2015/plenarysessions/8.

Heuser, Ryan. *Abstraction: A Literary History*. 2019. Stanford U.

---. *Abstraction: A Literary History*. 2020, https://ryanheuser.org/assets/talks/kingscollege2020/.

--- [@quadrismegistus]. "@goldstoneandrew Here's a Micro-Experiment on How Texts of Different OCR Qualities, vs. Texts Randomly Perturbed by Nth Letter, Affects Word Frequency. Https://T.Co/nesTiWV5Lp." *Twitter*, 6 July 2017, https://twitter.com/quadrismegistus/status/883020074545577984.

Higgins, David. "'These Circuits, That Have Been Made around the Globe': William Cowper's Glocal Vision." *Romantic Englishness: Local, National, and Global Selves, 1780–1850*, edited by David Higgins, Palgrave Macmillan UK, 2014, pp. 17–44, https://doi.org/10.1057/9781137411631_2.

Hill, Bridget. "Priscilla Wakefield as a Writer of Children's Educational Books." *Women's Writing*, vol. 4, no. 1, Mar. 1997, pp. 3–15, https://doi.org/10.1080/09699089700200002.

Hill, Mark J., and Simon Hengchen. "Quantifying the Impact of Dirty OCR on Historical Text Analysis: Eighteenth Century Collections Online as a Case Study." *Digital Scholarship in the Humanities*, vol. 34, no. 4, Dec. 2019, pp. 825–43, https://doi.org/10.1093/llc/fqz024.

Hobart, Elizabeth. "Reflecting Diversity in the Library of Congress Subject Headings." *Social Justice and Activism in Libraries*, McFarland, 2019, pp. 110–17.

Hockey, Susan M. "History of Humanities Computing." *A Companion to Digital Humanities*, edited by Susan Schreibman et al., Blackwell, 2004.

Holahan, Cassidy. "Rummaging in the Dark: ECCO as Opaque Digital Archive." *Eighteenth-Century Studies*, vol. 54, no. 4, 2021, pp. 803–26.

Hooper, Glenn. "Anne Plumptre: An Independent Traveller." *Gender Perspectives in Nineteenth-Century Ireland: Public and Private Spheres*, edited by Kelleher, Margaret and James H. Murphy, Irish Academic Press, 1997, pp. 129–39.

Hulme, Peter. *Colonial Encounters: Europe and the Native Caribbean, 1492–1797*. Methuen, 1986.

Hunter, J. Paul. *Before Novels: The Cultural Contexts of Eighteenth-Century English Fiction*. Norton, 1990.

James, Kevin, and Gavin Hughes. "Exploring Explorers, Travellers, and Tourists: Digital Humanities Approaches in North America and the United Kingdom." *Mondes Du Tourisme*, no. 24, 24, Dec. 2023, https://doi.org/10.4000/tourisme.6376.

Kaasa, Janicke Stensvaag. "Travel and Fiction." *The Cambridge History of Travel Writing*, edited by Nandini Das and Tim Youngs, Cambridge UP, 2019, pp. 474–87. *Cambridge UP*, https://doi.org/10.1017/9781316556740.031.

Kaufman, Paul. *Borrowings from the Bristol Library, 1773–1784, a Unique Record of Reading Vogues.* Bibliographical Society of the U of Virginia, 1960.

Keighren, Innes M., et al. *Travels into Print: Exploration, Writing, and Publishing with John Murray, 1773–1859*. U of Chicago P, 2015. https://press.uchicago.edu/ucp/books/book/chicago/T/bo19722688.html.

Kinsley, Zoë. *Women Writing the Home Tour, 1682–1812*. Routledge, 2016.

Kirschenbaum, Matthew. "What Is Digital Humanities and What's It Doing in English Departments?" *Debates in the Digital Humanities 2012*, U of Minnesota P, 2012. https://minnesota.universitypressscholarship.com/view/10.5749/minnesota/9780816677948.001.0001/upso-9780816677948-chapter-1.

Knowlton, Steven A. "Three Decades Since Prejudices and Antipathies: A Study of Changes in the Library of Congress Subject Headings." *Cataloging & Classification Quarterly*, vol. 40, no. 2, Aug. 2005, pp. 123–45, https://doi.org/10.1300/J104v40n02_08.

Korte, Barbara. *English Travel Writing from Pilgrimages to Postcolonial Explorations*. St. Martin's Press, 2000.

Krueger, Misty, et al. *Transatlantic Women Travelers, 1688–1843*. Bucknell UP, 2021, https://www.degruyter.com/isbn/9781684483006.

Krummel, D. W. "The Dialectics of Enumerative Bibliography: Observations on the Historical Study of the Practices of Citation and Compilation." *The Library Quarterly: Information, Community, Policy*, vol. 58, no. 3, 1988, pp. 238–57.

Labbe, Jacqueline M. "'A Species of Knowledge Both Useful and Ornamental': Priscilla Wakefield's Family Tour through the British Empire." *Romantic Geographies: Discourses of Travel, 1775–1844*, Manchester UP, 2000, pp. 35–50.

Lahti, Leo, et al. "Bibliographic Data Science and the History of the Book (c. 1500–1800)." *Cataloging & Classification Quarterly*, vol. 57, no. 1, Jan. 2019, pp. 5–23, https://doi.org/10.1080/01639374.2018.1543747.

Lamb, Susan. *Bringing Travel Home to England: Tourism, Gender, and Imaginative Literature in the Eighteenth Century*. U of Delaware P, 2009.

Laramée, François Dominic. “How to Extract Good Knowledge from Bad Data: An Experiment with Eighteenth Century French Texts.” *Digital Studies/Le Champ Numérique*, vol. 9, no. 1, Jan. 2019, pp. 1–25, https://doi.org/10.16995/dscn.299.

Lawrence, Karen R. *Penelope Voyages: Women and Travel in the British Literary Tradition*. Cornell UP, 2018, https://doi.org/10.7591/9781501732492.

Leask, Nigel. *Curiosity and the Aesthetics of Travel-Writing, 1770–1840: “From an Antique Land.”* Oxford UP, 2002.

---. “Travel Writing in the Eighteenth Century.” *The Cambridge Introduction to Travel Writing*, edited by Tim Youngs, Cambridge UP, 2013, pp. 38–52. *Cambridge UP*, https://doi.org/10.1017/CBO9780511843150.004.

LeBlanc, Zoe, et al. “From Precedents to Collective Action: Realities and Recommendations for Digital Dissertations in History.” *Debates in the Digital Humanities 2023*, edited by Matthew K. Gold and Lauren F. Klein, U of Minnesota P, 2023, pp. 344–66, http://www.jstor.org.proxy3.library.mcgill.ca/stable/10.5749/j.ctv345pd4p.25.

Lee, David. “Genres, Registers, Text Types, Domains and Styles: Clarifying the Concepts and Nevigating a Path through the BNC Jungle.” *Faculty of Arts - Papers (Archive)*, Jan. 2001, pp. 37–72.

Levy, Michelle, and Mark Perry. “Distantly Reading the Romantic Canon: Quantifying Gender in Current Anthologies.” *Women’s Writing*, vol. 22, no. 2, Apr. 2015, pp. 132–55, https://doi.org/10.1080/09699082.2015.1011836.

Library of Congress. *Assigning and Constructing Subject Headings H 180*. Library of Congress, 2016, https://www.loc.gov/aba/publications/FreeSHM/H0180.pdf.

---. *Description and Travel H 1530*. Library of Congress, June 2013, https://www.loc.gov/aba/publications/FreeSHM/H1530.pdf.

---. “History.” *LC Linked Data Service: Authorities and Vocabularies*, 10 June 1999, https://id.loc.gov/authorities/subjects/sh85061212.html.

---. *History H 1647*. Library of Congress, Feb. 2018, https://www.loc.gov/aba/publications/FreeSHM/H1647.pdf.

---. “How Do We Determine Aboutness?” *Library of Congress Subject Headings: Online Training*, Library of Congress, https://www.loc.gov/catworkshop/lcsh/index.html. Accessed 12 Oct. 2022.

---. “Library of Congress Genre/Form Terms PDF Files.” *Library of Congress*, //www.loc.gov/aba/publications/FreeLCGFT/. Accessed 17 May 2024.

---. “Library of Congress Subject Headings.” *LC Linked Data Service: Authorities and Vocabularies*, 26 Apr. 2011, https://id.loc.gov/authorities/subjects.html.

---. “Library of Congress Subject Headings (LCSH).” *Librarianship Studies & Information Technology*, 23 May 2020, https://www.librarianshipstudies.com/2018/01/library-of-congress-subject-headings-lcsh.html.

---. *Order of Subject Headings H 80*. Library of Congress, 2013, https://www.loc.gov/aba/publications/FreeSHM/H0080.pdf.

---. “Travel in Literature.” *LC Linked Data Service: Authorities and Vocabularies*, 17 July 1995, https://id.loc.gov/authorities/subjects/sh85137162.html.

---. “Travel Writing.” *LC Linked Data Service: Authorities and Vocabularies*, 14 June 2012, https://id.loc.gov/authorities/subjects/sh91000035.html.

---. “Travelers’ Writings, English.” *LC Linked Data Service: Authorities and Vocabularies*, 10 Dec. 1992, https://id.loc.gov/authorities/subjects/sh92005951.html.

---. “Voyages, Imaginary.” *LC Linked Data Service: Authorities and Vocabularies*, 11 Feb. 1986, https://id.loc.gov/authorities/subjects/sh85144441.html.

Lincoln, Matthew D. “Tidy Data for the Humanities.” *Matthew Lincoln, PhD*, 26 May 2020, https://matthewlincoln.net/2020/05/26/tidy-data-for-humanities.html.

Lindsay, Claire. “Travel Writing and Postcolonial Studies.” *The Routledge Companion to Travel Writing*, Routledge, 2015.

Long, Hoyt, and Richard Jean So. “Literary Pattern Recognition: Modernism between Close Reading and Machine Learning.” *Critical Inquiry*, vol. 42, no. 2, Dec. 2015, pp. 235–67, https://doi.org/10.1086/684353.

Mak, Bonnie. “Archaeology of a Digitization.” *Journal of the Association for Information Science and Technology*, vol. 65, no. 8, 2014, pp. 1515–26, https://doi.org/10.1002/asi.23061.

Mandell, Laura. “Canons Die Hard: A Review of the New Romantic Anthologies.” *Romanticism on the Net*, no. 7, 1997, https://doi.org/10.7202/005755ar.

---. “Gender and Cultural Analytics: Finding or Making Stereotypes?” *Debates in the Digital Humanities 2019*, U of Minnesota P, 2019, pp. 3–26.

---. “Gendering Digital Literary History.” *A New Companion to Digital Humanities*, John Wiley & Sons, 2015, pp. 511–23, https://doi.org/10.1002/9781118680605.ch35.

Manning, Christopher D., et al. *An Introduction to Information Retrieval*. Cambridge UP, 2009.

Marshall, P. J., and Glyndwr Williams. *The Great Map of Mankind: Perceptions of New Worlds in the Age of Enlightenment*. Harvard UP, 1982.

Martin, Alison. “Performing Scientific Knowledge Transfer: Anne Plumptre and the Translation of Martin Heinrich Lichtenstein’s *Reisen Im Südlichen Afrika* (1811).” *Journal of Literature and Science*, June 2015, pp. 9–26, https://doi.org/10.12929/jls.08.01.02.

May, Jim. “Some Problems in ECCO (and ESTC).” *The Eighteenth-Century Intelligencer*, Jan. 2009, pp. 20–30.

Mazella, David, et al. “‘All the Modes of Story’: Genre and the Gendering of Authorship in the Year 1771.” *ABO: Interactive Journal for Women in the Arts, 1640–1830*, vol. 12, no. 1, May 2022, https://doi.org/10.5038/2157–7129.12.1.1256.

McBurney, William H. *A Check List of English Prose Fiction, 1700–1739*. Harvard UP, 1960, https://doi.org/10.4159/harvard.9780674429666.

McKendry, Andrew. “The Haphazard Journey of A Mind: Experience and Reflection In Samuel Johnson’s *Journey To The Western Islands Of Scotland*.” *The Age of Johnson*, vol. 20, 2010, pp. 11–34.

McKenzie, D. F. *Bibliography and the Sociology of Texts*. Cambridge UP, 1999, https://doi.org/10.1017/CBO9780511483226.

McVeagh, John. *Irish Travel Writing: A Bibliography*. Wolfhound Press, 1996.

McWilliam, Rohan. “Nineteenth Century Collections Online.” *Reviews in History*. https://reviews.history.ac.uk/review/1359.

Miller, David P. “Out from Under: Form/Genre Access in LCSH.” *Cataloging & Classification Quarterly*, vol. 29, no. 1–2, June 2000, pp. 169–88, https://doi.org/10.1300/J104v29n01_12.

Mills, Sara. *Discourses of Difference: An Analysis of Women’s Travel Writing and Colonialism*. Routledge, 1991.

Moretti, F. “The Slaughterhouse of Literature.” *Modern Language Quarterly*, vol. 61, no. 1, Mar. 2000, pp. 207–28, https://doi.org/10.1215/00267929–61–1-207.

Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for a Literary History*.Verso, 2007.

Muri, Allison, director. "About the Project." *Grub Street Project*, 2005, https://www.grubstreetproject.net/.

Mussell, James. *The Nineteenth-Century Press in the Digital Age*. Palgrave Macmillan UK, 2012, https://doi.org/10.1057/9780230365469.

Nash, Paul W. "The Abandoning of the Long s in Britain in 1800." *Journal of the Printing Historical Society*, vol. New series, no. 3, Summer 2001, pp. 3–19.

National Endowment for the Humanities. "Planning Your Next DHAG 3: Managing and Sustaining the Project Assets." *The National Endowment for the Humanities*, 17 Sept. 2020, https://www.neh.gov/blog/planning-your-next-dhag-3-managing-and-sustaining-project-assets.

Naylor, Bernard. *Accounts of Nineteenth-Century South America: An Annotated Checklist of Works by British and United States Observers*. published for the Institute of Latin American Studies [by] the Athlone P., 1969.

Newbould, Mary-Céline. *Adaptations of Laurence Sterne's Fiction: Sterneana, 1760–1840*. 2013.

"NINES." *Nineteenth-Century Scholarship Online*, https://nines.org/. Accessed 17 May 2024.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York UP, 2018, https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=4834260.

Nussbaum, Felicity. *Torrid Zones: Maternity, Sexuality, and Empire in Eighteenth-Century English Narratives*. Johns Hopkins UP, 1995.

O'Kane, Finola. *Ireland and the Picturesque: Design, Landscape Painting and Tourism 1700–1840*. Yale UP, 2012.

O'Loughlin, Katrina. *Women, Writing, and Travel in the Eighteenth Century*. Cambridge UP, 2018, https://doi.org/10.1017/9781316104934.

Organisciak, Peter. "MARC Coverage - Documentation [Version 1]." *HathiTrust Research Center*, 7 Dec. 2015, https://htrc.atlassian.net/wiki/pages/viewpage.action?pageId=43289274&pageVersion=1.

Orr, Leah. "Genre Labels on the Title Pages of English Fiction, 1660–1800." *Philological Quarterly*, vol. 90, no. 1, Winter 2011, pp. 67–95.

---. *Novel Ventures: Fiction and Print Culture in England, 1690–1730*. 2017.

---. "Property, Money, and Benevolence in The Mysteries of Udolpho." *Essays in Romanticism*, vol. 26, no. 1, Apr. 2019, pp. 71–88, https://doi.org/10.3828/eir.2019.26.1.6.

Owen, C. M. *The Female Crusoe: Hybridity, Trade and the Eighteenth-Century Individual*. Brill, 2010, https://doi.org/10.1163/9789042029651.

Oxford English Dictionary. *Topography, n.* Oxford UP, Mar. 2024, https://doi.org/10.1093/OED/1961999965.

Ozment, Kate. "Rationale for Feminist Bibliography." *Textual Cultures*, vol. 13, no. 1, Apr. 2020, pp. 149–78, https://doi.org/10.14434/textual.v13i1.30076.

Paku, Gillian. *Anonymity in the Eighteenth Century*. Oxford UP, 2015, https://doi.org/10.1093/oxfordhb/9780199935338.013.37.

Parks, George B. "The Turn to the Romantic in the Travel Literature of the Eighteenth Century." *Modern Language Quarterly*, vol. 25, no. 1, Mar. 1964, pp. 22–33, https://doi.org/10.1215/00267929–25–1-22.

Paulson, Ronald. "Review: Malcolm Andrews. The Search for the Picturesque: Landscape Aesthetics and Tourism in Britain, 1760–1800. Stanford: Stanford UP. 1989. Pp. xvi, 269." *The American Historical Review*, vol. 96, no. 3, June 1991, p. 873. *academic.oup.com*, https://doi.org/10.1086/ahr/96.3.873.

Pedley, Mary Sponberg. *The Commerce of Cartography: Making and Marketing Maps in Eighteenth-Century France and England*. U of Chicago P, 2005.

Peiser, Megan. *British Women Novelists and the Review Periodical, 1790–1820*. 2016. U of Missouri.

---. "Review of The Bluestocking Archive, Emory Women Writers Resource Project, and Women's Travel Writing, 1780–1840: A Bio-Bibliographical Database." *ABO: Interactive Journal for Women in the Arts, 1640–1830*, vol. 6, no. 2, Dec. 2016, http://dx.doi.org/10.5038/2157–7129.6.2.6.

---. "Reviews as Database: Reading the Review Periodical in Eighteenth-Century England." *The Papers of the Bibliographical Society of America*, vol. 111, no. 4, Nov. 2017, pp. 491–511, https://doi.org/10.1086/694572.

Perkins, Pam. "Anne Macvicar Grant (1755–1838)." *The Routledge Companion to Romantic Women Writers*, Routledge, 2022.

Petto, Christine M. “To Know the Distance: Wayfinding and Roadmaps of Early Modern England and France.” *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 51, no. 4, 2016, pp. 240–62.

Pickford, Susan. “Writing with ‘Manly Vigour’: Translatorial Agency in Two Early Nineteenth-Century English Translations of François Pouqueville’s *Voyage En Morée, à Constantinople et En Albanie* (1805).” *Travel Narratives in Translation, 1750–1830: Nationalism, Ideology, Gender*, Routledge, 2012, pp. 197–217.

Pine-Coffin, Robert S. *Bibliography of British and American Travel in Italy to 1860*. Olschki, 1981.

Piper, Andrew. *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data*. Cambridge UP, 2020, https://doi.org/10.1017/9781108922036.

---. “Enumerations: Data and Literary Study.” *Enumerations*, U of Chicago P, 2018, https://doi.org/10.7208/9780226568898.

---. “Think Small: On Literary Modeling.” *PMLA*, vol. 132, no. 3, May 2017, pp. 651–58, https://doi.org/10.1632/pmla.2017.132.3.651.

Plumptre, Anne. *Something New: Or, Adventures at Campbell-House*. Edited by Deborah McLeod, Broadview Press, 1996.

Pope, Alexander. *The Dunciad in Four Books*. Edited by Valerie Rumbold, 2nd ed., Routledge, 2014, https://doi.org/10.4324/9781315833897.

Porter, Dennis. *Haunted Journeys: Desire and Transgression in European Travel Writing*. Princeton UP, 1991.

Posner, Miriam. “What’s Next: The Radical, Unrealized Potential of Digital Humanities.” *Debates in the Digital Humanities 2016*, edited by Matthew K. Gold and Lauren F. Klein, U of Minnesota P, 2016, pp. 32–41.

Pratt, Mary Louise. *Imperial Eyes: Travel Writing and Transculturation*. Routledge, 1992.

Pretzler, Maria. *Pausanias: Travel Writing in Ancient Greece*. Duckworth, 2007.

Price, Kenneth M. “Edition, Project, Database, Archive, Thematic Research Collection: What’s in a Name?” *Digital Humanities Quarterly*, vol. 003, no. 3, Sept. 2009.

Proteau, Jasmine. “‘Not Your Ordinary Guidebook’: Gender and the Redefinition of the Nineteenth-Century Guidebook.” *Studies in Travel Writing*, vol. 23, no. 2, Apr. 2019, pp. 119–38, https://doi.org/10.1080/13645145.2019.1691808.

Raban, Jonathan. *For Love & Money: Writing, Reading, Travelling, 1969–1987*. Collins Harvill, 1987.

Raven, James. "The Anonymous Novel in Britain and Ireland, 1750–1830." *The Faces of Anonymity: Anonymous and Pseudonymous Publication from the Sixteenth to the Twentieth Century*, edited by Robert J. Griffin, Palgrave Macmillan US, 2003, pp. 141–66, https://doi.org/10.1007/978–1-137–11109–8_7.

---. *British Fiction, 1750–1770: A Chronological Check-List of Prose Fiction Printed in Britain and Ireland*. U of Delaware P, 1987.

---. "The Publication of Fiction in Britain and Ireland, 1750–70." *Publishing History*, vol. 24, 1988, p. 31.

Rawson, Katie, and Trevor Muñoz. "Against Cleaning." *Debates in the Digital Humanities 2019*, U of Minnesota P, 2019, pp. 279–92, https://muse.jhu.edu/chapter/2293548.

Redford, Bruce. *Venice & the Grand Tour*. Yale UP, 1996.

Robinson, Jane. *Wayward Women: A Guide to Women Travellers*. Oxford UP, 1991.

---. *Voyant Tools*. 2016, https://voyant-tools.org/.

Rockwell, Geoffrey, and Stéfan Sinclair. *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. The MIT Press, 2016, https://doi.org/10.7551/mitpress/9522.001.0001.

Roddan, Hector. "'*Orientalism* Is a Partisan Book': Applying Edward Said's Insights to Early Modern Travel Writing." *History Compass*, vol. 14, no. 4, 2016, pp. 168–88, https://doi.org/10.1111/hic3.12307.

Rodgers, Zuleika. *Making History: Josephus and Historical Method*. Brill, 2006, https://doi.org/10.1163/ej.9789004150089.i-471.

Rogers, Shef. "Enlarging the Prospects of Happiness: Travel Reading and Travel Writing." *The Cambridge History of the Book in Britain: 1695–1830*, edited by SJ Suarez Michael F. and Michael L. Turner, vol. 5, Cambridge UP, 2009, pp. 781–90, https://doi.org/10.1017/CHOL9780521810173.045.

Rosenberg, Daniel. "Ted Underwood, Distant Horizons: Digital Evidence and Literary Change." *Modern Philology*, vol. 117, no. 3, Feb. 2020, pp. 151–58, https://doi.org/10.1086/707111.

Rosenthal, Laura J. *Infamous Commerce: Prostitution in Eighteenth-Century British Literature and Culture*. Cornell UP, 2015.

Ross, Shawna. “Toward a Feminist Modernist Digital Humanities.” *Feminist Modernist Studies*, vol. 1, no. 3, Sept. 2018, pp. 211–29, https://doi.org/10.1080/24692921.2018.1505821.

Ross, Stephanie A. “Review of The Search for the Picturesque: Landscape Aesthetics and Tourism in Britain, 1760–1800.” *The Journal of Aesthetics and Art Criticism*, vol. 48, no. 3, 1990, pp. 248–50, https://doi.org/10.2307/431773.

Rybicki, Jan. “Vive La Différence: Tracing the (Authorial) Gender Signal by Multivariate Analysis of Word Frequencies.” *Digital Scholarship in the Humanities*, vol. 31, no. 4, Dec. 2016, pp. 746–61, https://doi.org/10.1093/llc/fqv023.

Sabor, Peter, director. *Reading With Austen*. Burney Centre, 2020, https://www.readingwithausten.com/.

Sanders, Joshunda, and Diana Barnes-Browne. “Eat, Pray, Spend: Priv-Lit and the New, Enlightened American Dream.” *Bitch Media*, 14 Feb. 2012, https://web.archive.org/web/20120214115832/http://bitchmagazine.org/article/eat-pray-spend.

Santini, Marina. “State-of-the-Art on Automatic Genre Identification.” *Information Technology Research Institute Technical Report Series, ITRI,* U of Brighton, 2004.

Sherman, William H. “Stirrings and Searchings (1500–1720).” *The Cambridge Companion to Travel Writing*, edited by Peter Hulme and Tim Youngs, Cambridge UP, 2002, pp. 17–36, https://doi.org/10.1017/CCOL052178140X.003.

Shuttleton, David. “Mary Chandler’s Description of Bath (1733): The Poetic Topographies of an Augustan Tradeswoman.” *Women’s Writing*, vol. 7, no. 3, Oct. 2000, pp. 447–67, https://doi.org/10.1080/09699080000200117.

Siegel, Kristi. “Women’s Travel and Theory.” *Gender, Genre, and Identity in Women’s Travel Writing*, edited by Kristi Siegel, Peter Lang, 2004, pp. 1–13.

Smith, David and Ryan Cordell. *A Research Agenda for Historical and Multilingual Optical Character Recognition - DRS*. https://repository.library.northeastern.edu/files/neu:f1881m409. Accessed 11 Mar. 2019.

So, Richard Jean. “‘All Models Are Wrong.’” *PMLA*, vol. 132, no. 3, May 2017, pp. 668–73, https://doi.org/10.1632/pmla.2017.132.3.668.

Somers, James. “Torching the Modern-Day Library of Alexandria.” *The Atlantic*, 20 Apr. 2017. *The Atlantic*, https://www.theatlantic.com/technology/archive/2017/04/the-tragedy-of-google-books/523320/.

Sorrenson, Richard. “The State’s Demand for Accurate Astronomical and Navigational Instruments in Eighteenth-Century Britain.” *Consumption Of Culture*, Routledge, 1995.

Spedding, Patrick. “‘The New Machine’: Discovering the Limits of ECCO.” *Eighteenth-Century Studies*, vol. 44, no. 4, 2011, pp. 437–53.

St Clair, William. *That Greece Might Still Be Free: The Philhellenes in the War of Independence*. Open Book Publishers, 2008, https://doi.org/10.11647/OBP.0001.

---. *The Reading Nation in the Romantic Period*. Cambridge UP, 2007.

Stein, Sterling Stuart, et al. “The Effect of OCR Errors on Stylistic Text Classification.” *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, 2006, pp. 701–02, https://doi.org/10.1145/1148170.1148325.

Stewart, David. “Landscape.” *The Oxford Handbook of British Romantic Prose*, edited by Robert Morrison, Oxford UP, 2024, pp. 159–76.

Stone, Alva T. “The LCSH Century: A Brief History of the Library of Congress Subject Headings, and Introduction to the Centennial Essays.” *Cataloging & Classification Quarterly*, vol. 29, no. 1–2, June 2000, pp. 1–15, https://doi.org/10.1300/J104v29n01_01.

Sullivan, Megan. *Query: LCSH Reintegration on ECCO / Gale Primary Sources*. 12 Apr. 2024, Received by Catherine Nygren.

Swales, John M. *Genre Analysis: English in Academic and Research Settings*. Cambridge UP, 1990.

Sweet, Rosemary. *Antiquaries: The Discovery of the Past in Eighteenth-Century Britain*. Hambledon and London, 2004.

Tallmadge, John. “Voyaging and the Literary Imagination.” *Exploration*, 1979, pp. 1–16.

Thoma, Pamela. “What Julia Knew: Domestic Labor in the Recession-Era Chick Flick.” *Gendering the Recession: Media and Culture in an Age of Austerity*, edited by Dianne Negra and Yvonne Tasker, Duke UP, 2014, https://doi.org/10.1215/9780822376538–005.

Thompson, C. W. “The Romantic Literary Travel Book.” *The Routledge Companion to Travel Writing*, Routledge, 2015.

Thompson, Carl. “Review of Travel Writing, 1700–1830: An Anthology.” *NWIG: New West Indian Guide / Nieuwe West-Indische Gids*, vol. 81, no. 3/4, 2007, pp. 302–04.

Tolonen, Mikko, Eetu Mäkelä, Ali Ijaz, et al. “Corpus Linguistics and Eighteenth Century Collections Online (ECCO).” *Research in Corpus Linguistics*, vol. 9, no. 1, 1, Apr. 2021, pp. 19–34, https://doi.org/10.32714/ricl.09.01.03.

Tolonen, Mikko, Eetu Mäkelä, and Leo Lahti. “The Anatomy of Eighteenth Century Collections Online (ECCO).” *Eighteenth-Century Studies*, vol. 56, no. 1, 2022, pp. 95–123.

“Travel, Trade, and the Expansion of Empire: Defining Travel and Travels.” *The Norton Anthology of English Literature: The 18th Century: Topic 4: Texts and Contexts*, 26 Oct. 2017, https://web.archive.org/web/20171026125508/https://www.wwnorton.com/college/english/nael/18century/topic_4/johnson1.htm.

Turner, Katherine. *British Travel Writers in Europe, 1750–1800: Authorship, Gender, and National Identity*. Ashgate, 2001.

Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. U of Chicago P, 2019, https://doi.org/10.7208/chicago/9780226612973.001.0001.

---. “The Historical Significance of Textual Distances.” *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, edited by Beatrice Alex et al., Association for Computational Linguistics, 2018, pp. 60–69, https://aclanthology.org/W18–4507.

---. “The Life Cycles of Genres.” *Journal of Cultural Analytics*, May 2016, http://culturalanalytics.org/2016/05/the-life-cycles-of-genres/.

---. “Machine Learning and Human Perspective.” *PMLA*, vol. 135, no. 1, Jan. 2020, pp. 92–109, https://doi.org/10.1632/pmla.2020.135.1.92.

---. “NovelTM Datasets for English-Language Fiction, 1700–2009.” *Journal of Cultural Analytics*, May 2020, https://doi.org/10.22148/001c.13147.

---. *Tedunderwood/DataMunging*. 2017, https://github.com/tedunderwood/DataMunging/tree/a8c5c93d399be36ebdc666cdc27f659527150f24.

---. *Tedunderwood/Horizon: Data and Code to Support Distant Horizons*. v1.1, Zenodo, 25 Mar. 2018. *DOI.org (Datacite)*, https://doi.org/10.5281/ZENODO.1206317.

---. *Understanding Genre in a Collection of a Million Volumes, Interim Report*. 29 Dec. 2014. *Figshare*, https://doi.org/10.6084/m9.figshare.1281251.v1.

---. [@tedunderwood.me 🦋]. "@quadrismegistus I Haven't Tested the Lower Limit, Because My Problems in the 18c Come from Coverage Instead. But for My Methods I Find 90% Is No Prob." *Twitter*, 6 July 2017, https://twitter.com/Ted_Underwood/status/883048201963659265.

Van Netten Blimke, Linda. *Political Affairs of the Heart Female Travel Writers, the Sentimental Travelogue, and Revolution, 1775–1800*. Bucknell UP, 2022, https://doi.org/10.36019/9781684484096?locatt=mode:legacy.

Van Vuuren, Melissa S., and Angela Courtney. *Global Odyssey: A Bibliography of Travel Literature Before 1940*. Indiana U Center for the Study of Global Change, 2006. https://hdl.handle.net/2022/27076.

Viviès, Jean. *English Travel Narratives in the Eighteenth Century: Exploring Genres*. Translated by Claire Davison, Routledge, 2016.

Wachowich, Angela. "Hester Thrale Piozzi's 'Observations and Reflections Made in the Course of a Journey through France, Italy, and Germany' (1789)." *The Women's Print History Project*, 31 July 2021, https://womensprinthistoryproject.com/blog/post/79.

Watt, Ian P. *The Rise of the Novel; Studies in Defoe, Richardson, and Fielding*. U of California P, 1964.

Weber, Shirley Howard. *Voyages and Travels in Greece, the Near East, and Adjacent Regions, Made Previous to the Year 1801; Being a Part of a Larger Catalogue of Works on Geography, Cartography, Voyages and Travels, in the Gennadius Library in Athens.* American School of Classical Studies at Athens, 1953.

Weidman, Sean G., and James O'Sullivan. "The Limits of Distinctive Words: Re-Evaluating Literature's Gender Marker Debate." *Digital Scholarship in the Humanities*, vol. 33, no. 2, June 2018, pp. 374–90. *academic.oup.com*, https://doi.org/10.1093/llc/fqx017.

Welch, Thomas L., et al. *Travel Accounts and Descriptions of Latin America and the Caribbean, 1800–1920: A Selected Bibliography*. Columbus Memorial Library, 1982.

Wernimont, Jacqueline. "Whence Feminism? Assessing Feminist Interventions in Digital Literary Archives." *Digital Humanities Quarterly*, vol. 007, no. 1, July 2013.

White, Willow. "Tracing a Literary Tradition of Women in Comedy from Aphra Behn to Tina Fey." ACCUTE Conference, Vancouver, Canada, 2019.

Williams, Ruth. "*Eat, Pray, Love*: Producing the Female Neoliberal Spiritual Subject." *The Journal of Popular Culture*, vol. 47, no. 3, June 2014, pp. 613–33, https://doi.org/10.1111/j.1540–5931.2011.00870.x.

Williams, William H. A. *Creating Irish Tourism: The First Century, 1750–1850*. Anthem Press, 2010. *Cambridge UP*, https://doi.org/10.7135/UPO9781843318996.

---. "The Irish Tour, 1800–50." *Travel Writing and Tourism in Britain and Ireland*, edited by Benjamin Colbert, Palgrave Macmillan UK, 2012, pp. 97–113, https://doi.org/10.1057/9780230355064_7.

Wolfson, Susan, et al. "Letters from the Editors of the New Romantic Anthologies." *Romanticism on the Net*, no. 7, 1997, https://doi.org/10.7202/038449ar.

Wright, Angela, and Rosie Whitcombe. *The Mysteries of Ann Radcliffe*. Interview by Barbara Burns, 8 Mar. 2023, https://mhra.org.uk/news/2023/03/08/the-mysteries-of-ann-radcliffe.html.

Youngs, Tim. *Travellers in Africa: British Travelogues, 1850–1900*. St. Martin's Press, 1994.

Zold, Elizabeth. "Expanding the Domestic Sphere: Mothers Who Travelled in the Eighteenth Century." *Studies in Travel Writing*, vol. 19, no. 4, Oct. 2015, pp. 324–39, https://doi.org/10.1080/13645145.2015.1108723.