A QUANTITATIVE APPROACH TO THE HISTORY OF

MUSIC BINDER'S VOLUMES (1820–1900)

Brian K. Anderson, M.M.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2022

APPROVED:

Peter Mondelli, Major Professor
Frank Heidlberger, Committee Member
Mark McKnight, Committee Member
Benjamin Brand, Chair of the Division of
    Music History, Theory, and
    Ethnomusicology
Jaymee Haefner, Director of Graduate Studies
    in the College of Music
John Richmond, Dean of the College of Music
Victor Prybutok, Dean of the Toulouse
    Graduate School

Anderson, Brian K. *A Quantitative Approach to the History of Music Binder's Volumes (1820–1900)*. Doctor of Philosophy (Music), December 2022, 213 pp., 11 tables, 41 figures, 2 appendices, works cited, 111 titles.

Music binder's volumes, or collections of sheet music typically bound by women in the nineteenth century, constitute an informative and underutilized set of historical artifacts. Each binder's volume can be viewed as a Spotify playlist frozen in time. An individual volume contains more than just the volume's individual pieces; it also holds the marginalia, the choices women made on what to include in a binder, and information on where and how music was produced. This dissertation examines music binder's volumes quantitatively, processing information found in binder's volumes by using the MARC and other cataloguing data to construct a relational database. I engage with broad questions of music publishing and consumption and provide a method to contextualize qualitative results on a larger scale. In doing so, I make two distinct contributions to music research and the digital humanities. First, this project offers a clear path for engaging with music binder's volumes and material history of nineteenth-century America in ways that scholars have rarely engaged in prior to this point. I highlight how data analysis provides new framings for binder's volumes and for sheet music consumption both at the song-level and at larger levels of the data. Second, and more broadly, this project serves as a framework for engaging in digital projects on multifaceted items that can be difficult to catalog and process. I describe in detail the metadata schema, methodological concerns in establishing a database, and data cleaning process in a way that provides transparency for the project and can also serve as a starting point for others considering data projects in the digital humanities.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND RATIONALE

In the nineteenth century, music binder's volumes contained a representative summation of an amateur musician's music library. As such, they have much to tell us about both the individuals and the societies that assembled them. Women in the United States, as in Europe, were trained in music and other ornamental arts.[1] During this time they accumulated sheet music, often as individual pieces of music.[2] As these women reached a marriageable age, they would often take their sheet music and have it bound together, reifying their musical collection, and preserving it as a memento.

This dissertation aims to investigate the contents, material history, and metadata available from music binder's volumes bound in the United States in the nineteenth century. More specifically, I hope to address the following questions based on the information in music binder's volumes: first, when are binder's volumes being compiled, and when is the music published that goes into the volumes; second, where was music published and disseminated in the United States and did this change over time; third, whose works are included in volumes and when is this work included (is it during the composer's life); and fourth, how do the answers to the first three questions help us understand music and social life in nineteenth century America? To do this I take several binder's volumes as individual case studies of the types of music a binder contains and focus on the qualitative aspect of what an in-depth examination of the music and marginalia reveals. After considering these volumes on a small scale, I next apply this concept to a larger

---

[1] Petra Meyer-Frazier, "American Women's Roles in Domestic Music Making as Revealed in Parlor Song Collections: 1820–1870," (PhD diss., University of Colorado, 1999), 89.

[2] Candace Bailey, "Binder's Volumes as Musical Commonplace Books: The Transmission of Cultural Codes in the Antebellum South," *Journal of the Society for American Music* 10, no. 4 (2016): 446–69. In this article, Bailey shows that while binder's volumes are typically thought of as an upper-class luxury, people of more modest means also compiled binder's volumes.

scale via a relational database. Throughout this project I refer to my database, and this project, as the binder's volumes research initiative, or the BVRI for short. Ultimately, this study pursues two broad goals. First, it creates an expandable compendium of materials for researchers on nineteenth-century American print culture. I intend to make this compendium standardized, searchable, and available. Second, I use this data set to develop a nuanced understanding of how nineteenth-century Americans consumed music over time and in different regions.

This chapter serves to set the stage for discussions to come. First, I define what a binder's volume is, giving examples from various repositories. Next, I describe the state of the literature, focusing on both traditional musicological sources and on the ways digital projects launched by other scholars have shaped my own goals and methods. I also look at material history and how scholarship that designates the materials themselves as the driver of the stories being told offers value and insight. I then discuss such materiality in the context of how contemporaneous users of binder's volumes experienced the musical work before exploring digital historiography and its impacts in the long term.

<center>What is a Binder's Volume?</center>

Figure 1 shows a binder's volume from the Library of Congress. It is rather thin comparatively speaking, has approximately 80 pages, and contains 17 works. The binder belonged to Laura E. Cooke and is one of two at the Library of Congress that she once owned. Figure 2 shows another volume, this one assembled by a collector with the presumptive last name of Schieffelin, currently in the collection of the University of North Texas Library.[3] Where Cooke's volume is a modest 80 pages, Schieffelin's volume is nearly four times the size at 310

---

[3] Schieffelin, Untitled binder's volume, https://digital.library.unt.edu/ark:/67531/metadc1706061/. This example highlights the challenges of binder's volumes; provenance is not always found.

pages. Lengths range widely between, and at times beyond, these two exemplars in size and

scope. Some volumes are ornately decorated, and others are entirely plain. The musical contents

of each collection also vary greatly from volume to volume.
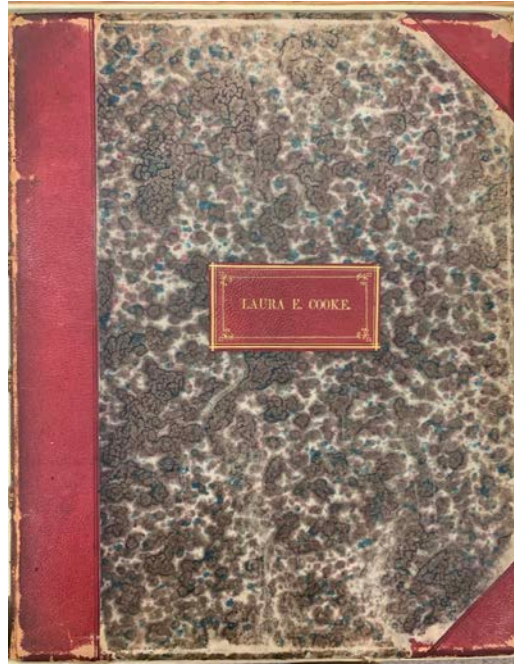
Figure 1: Laura E. Cooke's Binder's Volume, photo taken at the Library of Congress.
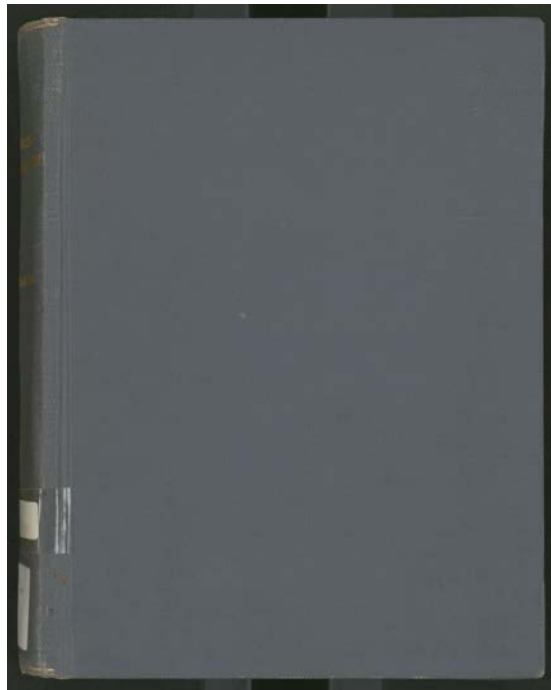
Figure 2: S. L. Schieffelin's volume, housed at the University of North Texas.

Staying with Schieffelin's volume, we can probe the inside to explore some differences between the styles of volumes. Some have a table of contents and compilers who numbered every page. Others just start with the music, as is the case with Schieffelin's volume. One only has to look at a few volumes to see the range of styles. Some include notes, doodles, fingerings on piano works, and other marginalia; others are bare and seemingly unused. This volume is relatively plain in terms of musical markings, as seen in Figure 3.

Figure 3: Music from Schieffelin's binder's volume.

Despite being relatively unadorned with markings, the duet above is a rich source of data when taken in context. Figure 4 shows the title page for the four-hand piano work in Figure 3. On its face, the data might seem mundane, yet it is worth highlighting. The composer listed by the publisher is Henri Herz. Rossini composed the opera *William Tell*, on which this set of variations is based. B. Schott of Paris and London published the work. This is some of the

information that has value for my research, but there are additional things to note. The instrumentation matters, and the seller's stamp–added by the vendor at the time of purchase-- contains useful information. Various works reference performers and dedicatees, some vocal works have lyricists and translators. All of this data is noteworthy: if not as individual data points, then in aggregate.
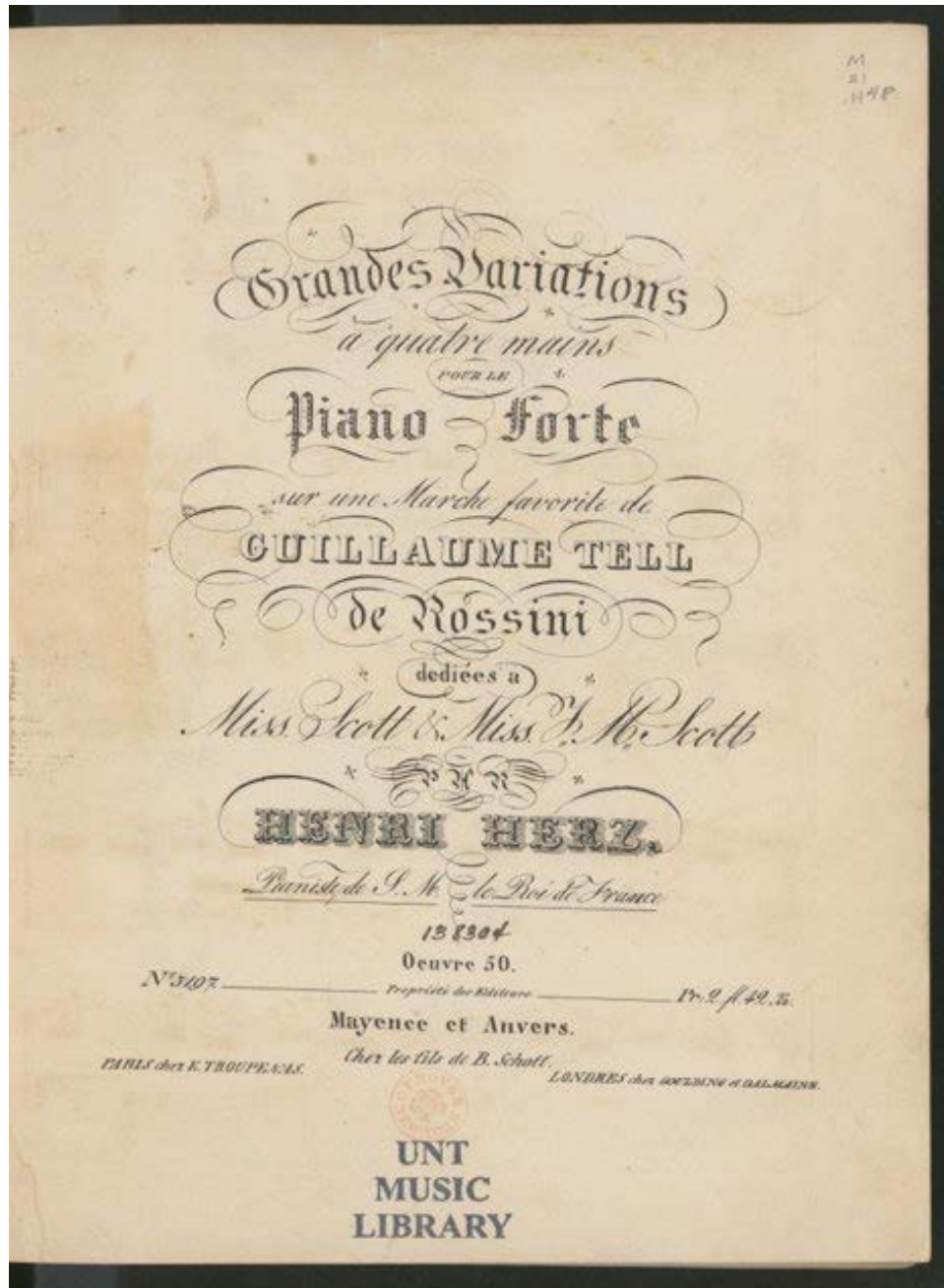
Figure 4: Title page for piano duet shown in Figure 3.

In addition to the printed page, many volumes contain fascinating information about the compiler scattered throughout; flipping through these volumes is engrossing. The stories contained in a volume like this are worth telling, and they offer a window into the lives of people from all classes and backgrounds. My literature review highlights some of the work done telling these stories, and it is vital work to understanding nineteenth-century America. While I hope to tell a handful of stories, I intend to leave that work to others. My work instead seeks to create a structure for understanding the volumes and data contained therein, using that data to contextualize these stories in ways that might not be readily apparent when considering individual volumes.

After their use, and typically after the compiler's passing, binder's volumes have gone to many different places. In some cases, families preserved these as legacy items—especially in the case of wealthy or famous individuals like Emily Dickinson. In other cases, they were destroyed or given to libraries to be preserved. Libraries have struggled with what to do with these artifacts. Some just unbound the volumes and listed the sheet music, while others kept the books. The Library of Congress recently made a listing for binder's volumes (M1.A15), but this is not yet widely adopted. Several libraries have collections of varying size; Middle Tennessee State, UNC, Harvard, the Library of Congress all have larger collections, and places like the University of North Texas have smaller collections of closer to 25 or 50 volumes. The challenges cataloging the volumes combined with the scattered numbers all contribute to the goals of the BVRI, and its necessity.

A Brief History of the Binder's Volume

Collections of sheet music have been popular throughout the eighteenth and nineteenth centuries, and the United States is no exception to the practice. Before printing became quite as

prevalent as it did in the 1800s, sheet music collections tended to have just as much handwritten music as they did published music. As Glenda Goodman documents, collections tended to be very personal and largely (or even entirely) consist of music copied from another source.[4] This tradition underpins much of American music through the revolution, and only begins to change en masse with the rise of cheaply printed sheet music in Europe and America.[5]

The conclusion of the eighteenth century saw the start of a massive revolution in print culture and printmaking. Because the rise of print culture is integral to understanding the rise of binder's volumes, I explore this growth in detail here. One of the first noteworthy printmaking advancements of the time is the invention of lithography in 1796, which enabled small print runs to be produced more quickly and more cheaply after the 1860s, though engraved plates remained part of the process.[6] Over the next century, lithography grew in prominence and techniques evolved that pushed lithography as a process. Chromolithography allowed multiple colors to be applied to one piece of paper, and offset printing (a variant of which is still in use today for magazines) provided a means to keep the paper dry and ensure the process could move more quickly. In addition to, and in conjunction with lithography, plate engraving improved dramatically in the nineteenth century. During this time, engravers completed the shift from copper to pewter plates, and as specialized tools for engraving continued to be refined, plates were the main source of music through most of the nineteenth century. In the latter half of the nineteenth century, plates and offset lithography were used in conjunction, with plates serving as

[4] Goodman, *Cultivated by Hand: Amateur Musicians in the Early American Republic*. (New York: Oxford University Press, 2020): 4.

[5] Ibid.

[6] Deirdre Loughridge, "Making, Collecting, and Reading Music Facsimiles before Photography," *Journal of the Royal Music Association* (May 2016): 34.

the template for the lithographs, which preserved the lifetime of the plates. This is especially useful for pewter plates, given the increased malleability as compared to copper.

In addition to improvements in printing methods, paper itself became much cheaper and easier to produce in large quantities. By 1807, the Fourdriniers machine established the basic mode of paper production to today.[7] By the time the Fourdrinier machine arrived in the US almost three decades later, paper production had rapidly increased, providing yet another incentive towards larger production runs for all types of printed materials, including sheet music. The advancements in papermaking achieved ca. 1800 continue to influence paper production today.

With paper and printing both becoming cheaper, print became increasingly ubiquitous throughout the nineteenth century. Peter Mondelli catalogs the pace and scope of this expansion in Paris throughout the nineteenth century, noting in particular the rapid growth of print culture and the relative rise in literacy.[8] In Paris, print capacity increased dramatically as rotary presses came into increased usage and facilitated larger publications with a smaller crew of workers.[9] As more copies could be produced on cheaper materials with fewer people, prices went down as print runs increased in size, and literacy grew at least partly as a result of these changes. One way this increase in literacy has been documented in France is through Maggiolo's detailed tracking

---

[7] B.J. McMullin, "Machine-Made Paper, Seam Marks, and Bibliographical Analysis," *The Library: The Transactions of a Bibliographic Society* 9, no. 2 (March 2008): 64=65.

[8] Peter Mondelli discusses Maggiolo's study, which he encountered in James Smith Allen, *In the Public Eye: A History of Reading in Modern France, 1800–1940* (Princeton: Princeton University Press, 1991); and in Martyn Lyons, *Readers and Society in Nineteenth- Century France: Workers, Women, Peasants* (New York: Palgrave Macmillan, 2001). Peter Mondelli, "The Sociability of History in French Grand Opera: A Historical Materialist Perspective," *19th-Century Music*, Vol. 37, No. 1 (Summer 2013), pp. 48.

[9] Gilles Feyel, "Les transformations technologiques de la presse," in *La Civilisation du journal: Historire Culturelle et littéraire de la presse française au XIXe siècle*, ed. Dominique Kalifa, Philippe Régnier, Marie-Ève Thérenty, and Alain Vailland (Paris: Nouveau Monde, 2011), 122-123.

of the percentage of men and women who could sign their own name on a marriage certificate.[10]

This rise in mass publishing is mirrored in the United States, and newspapers provide one lens to document this rapid growth. James Crouthamel tracks the rise of newspapers in the United States from 1830 through 1860 in New York City, noting that New York offers a reasonable proxy for American papers in general. In 1830, the successful model for most mercantile papers was to circulate papers that focused on business to a small audience: the circulation for these papers was often between 1,000 and 3,000 copies (near the maximum capacity for printing presses at the time).[11] These mercantile papers were business focused and sold on subscriptions for eight to ten dollars per year. Many newspapers also sold individual copies for around 6 cents, and the "six cent" papers often avoided news in the 1830s, which was slow to arrive via transatlantic travel until the telegraph changed that process.[12]

The time from the 1830s onward saw a massive rise in "penny papers," which were marketed directly to the average consumer for daily consumption. Several factors contributed to this change. First, penny papers demonstrated that there was a market for things outside of business. The *Sun,* founded in 1833, achieved a daily circulation of about four thousand papers within four months—the same circulation as the nation-leading *Courier and Enquirer*.[13] Where the *Courier and Enquirer* focused on news for merchants and businessmen, the *Sun* focused on items that were "readable, sensational, or shocking." This focus reflects a growing class of readers outside the business community, as does the continuing rise of circulations. In 1856, the

---

[10] Maggiolo, via Smith and Lyons. Mondelli p. 4, footnote 7.

[11] James L. Crouthamel, "The Newspaper Revolution in New York: 1830–1860," *New York History: Quarterly Journal of the New York State Historical Association*, Vol. 45, No. 2 (April 1964): 93.

[12] Ibid. The Associated Press formed in 1846 to streamline the process of gathering transatlantic news by introducing a cost sharing system for distributing European news that arrived via telegraph.

[13] Crouthamel, 93-94.

four leading penny papers were the *Herald* (57,840), the *Sun* (50,000), the *Times* (42,000), and the *Tribune* (29,000 daily, with a weekly edition reaching 163,000 copies).[14] For comparison, the census places New York City's population between 515,547 (1850) and 813,669 (1860).[15]

The growing list of readers outside of the business class is one factor driving newspaper distribution; another is the improvement of technology and reduction of material costs in printing. Paper fell dramatically in price between 1828 and 1840, dropping from 20–25 cents per pound to 8–14 cents per pound.[16] At the same time, printing presses were consistently able to produce more copies per hour. In the 1830s, most mercantile presses used Napier single-cylinder press, which produced 2,000 sheets per hour. In 1846, Hoe and Company's (the makers of the Napier press) "lightning press" was installed in Philadelphia to print the *Public Ledger*, and the four-cylinder press could print 8,000 sheets per hour, which each cylinder printing about 2,000 sheets. The number of impression cylinders continued to rise, reaching 10 cylinders (and 20,000 sheets). For newspapers, this commitment to more copies also meant a need to upgrade to keep up with the *Times* and other newspapers pushing mass distribution; this meant that publishers and newspapers sought to use the presses once purchased. A new press with more cylinders also meant a need for more space for the press, which at times also necessitated a new building. Such capital investments paired with a desire to use the new investment to offset the initial costs of the purchase.

The growth of New York City newspapers between 1830 and 1860 thus reflect two

---

[14] Crouthamel, 103.

[15] The US Census Bureau lists the ten largest cities and their population on their fast facts website for each decade. "United States Census Bureau 1850 Fast Facts," *United States Census Bureau*, https://www.census.gov/history/www/through_the_decades/fast_facts/1850_fast_facts.html and "United States Census Bureau 1860 Fast Facts," *United States Census Bureau*, https://www.census.gov/history/www/through_the_decades/fast_facts/1860_fast_facts.html

[16] Crouthamel, 105.

trends: the growth of printing in America, and the increasing desire of a public to consume printed materials. The growth of both in the consumption of newspapers did not stop there; books, music, and all manner of printed materials rose with the advent of mass production and a growing consumer base. This printing revolution is a core part of the nineteenth-century American story, just as the internet and digital consumption has fueled our twenty-first century world today.

The impact on sheet music was similarly dramatic; music went from being a luxury to something middle-income families and even some lower-income families could afford. While some binder's volumes collections come from wealthy young women, recent inquiry into the subject has shown how working-class children also had their own binder's volumes.[17] This expansion of print media—musical and otherwise—marks a shift from the American musical traditions highlighted by Glenda Goodman into the musical scene of the mid-nineteenth century. Music was more affordable, so families of all incomes could purchase more of it. Each individual musical item was still worth having, and worth preserving, but the increased access combined with this value contributed to making musical preservation in binder's volumes a common nineteenth-century American practice.

As Crouthamel suggests in his analysis of the newspaper revolution, the end of the twentieth century was marked by another turn in materials, which in turn fueled another shift in consumption. Paper itself changed multiple times in the nineteenth century, with pulp paper replacing rag paper in the mid-nineteenth century and with the rise of sulfite (pulp) paper by the end of the nineteenth century.[18] *Fin di siècle* America had a particularly robust paper industry,

---

[17] Bailey, "Binder's Volumes as Musical Commonplace Books."

[18] Judith A. McGaw, *Most Wonderful Machine: Mechanization and Social Change in Berkshire Paper Making, 1801–1885* (Princeton: Princeton University Press, 1987).

which fueled yet another increase in production. Where the mid-nineteenth century saw the acme of scrapbooking, and collecting individual items for binding was common throughout the country, the further increase in paper production led printed materials to become more ephemeral, effectively making printed material less valuable, and less worth keeping. Combined with the rise of the curated musical anthology, print music consumption changed and binder's volumes largely fell out of favor by the early twentieth century.

It is worth taking a moment to discuss what music is included in binder's volumes, and what limitations and opportunities are associated with examining these collections of music. The volumes themselves are undoubtedly gendered, and contain a certain type of music popular in America at the time. Charles Hamm discusses three broad categories in his description of American music: Germanic music, the modern school (composers like Chopin who were associated with Romanticism), and the music of democratic sociability.[19] This final category includes music from popular operas and arrangements of famous arias. In short, this is the music also denoted as parlor music, and the category that makes up most binder's volumes. Binder's volumes, at least based on current sampling, do not include much in the way of Germanic music or music of the modern school; their compilers focus primarily on parlor music.

Thus, an examination of binder's volumes does not provide a wide view of the music in America. Instead, the perspective offered by binder's volumes might be likened to looking in the side window of a house. Most scholars have not peered in this window or looked with much seriousness at the subject; some thought the gendered perspective less valuable, while others did not see particular value in the collections of music as collections and focused on other areas of American music. The view offered by binder's volumes is a different perspective on music, but

---

[19] Charles Hamm, *Music in the New World* (New York: W. W. Norton and Company, 1983).

this is also the reason that binder's volumes offer value and insight into American music. First, the volumes highlight a musical practice too often relegated to the sidelines—parlor music is one of the most commonly practiced traditions of American music, and scholars are increasingly showing that this is not just the elite who participated in creating these volumes. Second, the sheer fact that this is a different perspective helps us see the broader picture of American music more clearly. Indeed, the more vantage points from which the academic community can approach a multifaceted subject like Antebellum American music, the clearer our musical history becomes. This dissertation uses binder's volumes, combined with a few new digital approaches, to provide insight and highlight another angle—another perspective—to interrogate American music of the nineteenth century, and does so by situating itself at the intersection of musicology, library science, and the digital humanities.

Musicological Literature Review

This project combines methodologies drawn from historical musicology with those typically associated with the digital humanities. As such, this literature review covers several different areas to establish how a quantitative approach to binder's volumes relates to scholarly projects in the digital humanities and musicology. This section of the literature review focuses on discussing the historical context of binder's volumes and musicological approaches taken to understand binder's volumes. The following section situates the digital components of this project within relevant digital humanities projects and methods.

Antebellum American Music and Its Publishers

Numerous scholars have researched music publishing and consumption in the US, though the approach of applying metadata to describe the material history of music publishing and consumption in nineteenth-century America is original. My project intersects with four broad

groups of secondary sources: writings on publishers and the objects produced, writings on

consumers and objects as they are consumed, writings on advertisers and the mediation between

production and consumption, and recent approaches to digital humanities projects.

Publishers

A great deal of research has been done on American music publishing, though little of it

takes a broad chronological or geographic approach. Several scholars have done detailed

research into music publishing, but such studies frequently focus on one specific geographical

area.[20] Some studies go beyond production to consider other cultural forces, but these also tend

to focus on a specific place.[21] These studies are informative and foundational to the work I plan

to do; my examination of material culture across more than one region relies on a detailed study

of regional production.

Broader studies are rare.[22] One notable exception is Karen Stafford's work; she has

turned to binder's volumes to study publishers and the distribution of sheet music using one

particular set of binder's volumes: the Library of Congress collection. Her dissertation came out

in the very late stages of this project, and I integrate her electronic records into my database,

which will be a significant aid in developing future regional comparisons between various parts

---

[20] Examples of this include Ernst C. Krohn, *Music Publishing in St. Louis*, Ed. J. Bunker Clark (Warren, MI: Harmonie Park Press, 1988), Christa Pehl, "Men, Women, and Their Manuscripts: Musical Culture in Pennsylvania in the 1790s" (PhD diss., Princeton University, 2017), and Frank W. Hoogerwerf, *Confederate Sheet-Music Imprints*, 21 Vol. (Brooklyn, NY: 1984).

[21] Nancy Regan Ping, "Music in Antebellum Wilmington and the Lower Cape Fear of North Carolina" (PhD diss., University of Colorado, 1979) and Mark Slobin, James W. Kimball, et al. *Emily's Songbook: Music in 1850s Alban* (Middleton, WI: A-R Editions, 2011) are two studies that focus on one area. Candace Bailey also focuses on the South, but her work tends to deal much more with the compilers of the binder's volumes and is discussed there.

[22] One resource for broad information on publishing is in the 1870 catalog of musical works. This is not a resource for analysis but helps provide a list of things published in the United States. Board of Music Trade of the United States of America, *Complete Catalogue of Sheet Music and Musical Works, 1870*, edited and new introduction by Dena J. Epstein (New York: Da Capo Press, 1973).

of the country. Stafford's method is most similar to mine and adding multiple regions with this focus will be helpful for future projects.

Consumers

With respect to consumers and consumerism, my project is informed by both research on music consumption at the time and on consumerism more broadly in nineteenth-century America. Edgar W. Martin and Charles Grier Sellers have both written about the United States' economy before the Civil War, and their perspectives help establish the climate for sheet music consumption.[23] Other scholars, including James Thompson and William Harrison Barnard, have written about other American literature and consumerism.[24] One scholar who has looked at consumerism in music is Richard Crawford.[25] His examination is broad in scope; the sections that cover antebellum and postbellum America are foundational in my own look at music consumption through the lens of binder's volumes.

While each of these authors provides useful perspectives on understanding consumerism in nineteenth-century America, none use binder's volumes as a lens for understanding consumer practices. This is an area where my research presents a different, and useful perspective on consumer behavior and trends. Chapters 4 and 5 provide context on some of these possibilities, with Chapter 4 focusing on large-scale observations, and Chapter 5 giving case studies for how the database can answer specific questions on specific topics.

---

[23] Charles Grier Sellers, The Market Revolution: Jacksonian America, 1815–1846 (New York: Oxford University Press, 1994) and Edgar W. Martin, The Standard of Living in 1860: American Consumption Levels on the Eve of the Civil War (Chicago: Univ. of Chicago Press, 1942).

[24] James Thompson, *Models of Value: Eighteenth-Century Political Economy and the Novel* (Durham: Duke University Press, 1996) and William Harrison Barnard, "Significant Objects: American Literature and Consumer Culture, 1850–1900" (PhD diss., University of Michigan, 2001).

[25] Richard Crawford, The American Musical Landscape: The Business of Musicianship from Billings to Gershwin (Berkeley: University of California Press, 2000).

Binder's Volume Literature Review

Candace Bailey has been working on binder's volumes for several years. Her most recent book, *Unbinding Gentility*, focuses on the binder's volumes of three women from Charleston, using their music collections as a lens through which the lives of women of the past can be studied in detail.[26] While Bailey has worked with hundreds of binder's volumes, her approach has consistently been to look at these on a very personal level, seeking the story of the woman behind the volume.[27] Bailey, Maristella Feustle, and I are currently working together on taking this research and creating a collaborative database project that will involve cataloging thousands of binder's volumes in the nation, but the project is in its infancy. This dissertation is one step in the infancy of this larger project.

While Glenda Goodman does not write directly about binder's volumes, her research on handwritten volumes of bound sheet music was integral to this project, and to identifying its early chronological bounds. My goal in working with binder's volumes is to deal with the printed materials contained therein, and to interrogate the space between the printed work and the bound printed work to see what information can be gleaned from materials in these formats, and how the bound work provides context to printed sheet music. Goodman's work deals with handwritten volumes, and so while our work does not directly intersect, her ideas and methods were useful in providing context to the American musical scene before 1800 and to examining a related trend in musical consumption and use.

---

[26] Candace Bailey, *Charleston Belles Abroad: The Music Collections of Harriet Lowndes, Henrietta Aiken, and Louisa Rebecca McCord* (Columbia, SC: University of South Carolina Press, 2019). Bailey's other works, noted in the bibliography, are also relevant both for their work with binder's volumes and for the contextualization of women in the South. The most recent book is *Unbinding Gentility: Women Making Music in the Nineteenth-Century South* (Champaign, IL: University of Illinois Press, 2021).

[27] I had a chance to converse with Candace Bailey at the Society for American Music conference in 2019. She estimated then that she had easily looked at more than 1,000 volumes

Petra Meyers-Frazier has also worked extensively with binder's volumes in her research. She articulates the goals of her dissertation as thus: "This study, then, examines parlor songs for what they say about those who bought, used, displayed, and bound them."[28] The questions she addresses focus on the parlor song and the role of women in society, using binder's volumes as primary sources to contextualize the women who assembled them. In many ways my research serves as a counterpart to Meyers-Frazier's approach. We use the same materials, but where she examined an extensive amount of music binder's volumes with her eye towards parlor songs and gendered expectations within society, I intend to take the same materials and look towards publishing, and consumption. We are interested in the same type of primary materials, but we seek to answer different questions with those materials.

Digital Humanities Literature Review

The Digital Humanities

In the last decade, the digital humanities have expanded their scope and have received greater acceptance within academia. Projects ranging from computational linguistics to the analysis of tags on social media demonstrate the prevalence of database-driven research and the various ways it can quantitatively validate our qualitative understanding of issues in the humanities.[29] As Matthew Gold and Lauren Klein argue, from 2012 to 2019 the field of the

---

[28] Meyer-Frazier, "American Women's Roles in Domestic Music Making as Revealed in Parlor Song Collections," 20–21.

[29] While computational text analysis has been a part of digital humanities since its inception, a recent example can be found by Ryan L. Boyd, "Psychological Text Analysis in the Digital Humanities," in *Data Analytics in the Digital Humanities*, ed. Shalin Hai-Jew (eBook: Schirmer, 2017): 161–190. For examples of tagging and social networking, see Shalin Hai-Jew, "Parsing Related Tags Networks from Flickr® to Explore Crowd-Sourced Keyword Associations," in *Data Analytics in the Digital Humanities*, ed. Shalin Hai-Jew (eBook: Schirmer, 2017): 191–214.

digital humanities has matured and is ready to take on substantial questions.[30]

The creation of these and other databases has facilitated numerous projects in what might be termed "computational humanities," which includes such projects as statistical analyses of Palestrina's masses.[31] My research resides at the intersection of database creation and what was once termed computational humanities. Projects in the digital humanities naturally require an interdisciplinary approach, and I look forward to using my background in mathematics to enhance my musicological research.

Music Database Review

Musicologists have become active lately in the digital humanities, and several projects are worth highlighting. Recent projects include work on several large databases such as the SALAMI project and CANTUS, and these two projects piqued my interest in the digital humanities.[32] A few years after encountering these databases, I encountered the English Broadside Ballad Archive (EBBA) at a session of the American Musicological Society in 2017, and this system is much more germane to the project.[33] EBBA is a large database of broadside ballads, which are themselves complex objects. Broadside ballads often contained many

---

[30] Matthew K. Gold and Lauren F. Klein, "Introduction" in *Debates in the Digital Humanities 2019,* ed. Matthew K. Gold and Lauren F. Klein (Minneapolis: University of Minnesota Press, 2019), xi.

[31] Claire Arthurs gave one such presentation at the University of North Texas on February 20, 2019, where she used a statistical analysis of Palestrina's masses to test part-writing rules articulated by Zarlino.

[32] Michelle Urberg, "Past and Futures of Digital Humanities in Musicology: Moving Towards a Bigger Tent," *Music Reference Services Quarterly* 20, no. 3–4 (2017): 134–150. Urberg discusses several recent musicological projects in the digital humanities, highlighting their potential impact and relevance. SALAMI is a collaborative, computational musicology project out of McGill; information about SALAMI can be found at https://ddmal.music.mcgill.ca/research/SALAMI/. CANTUS is a database of chant, with searchable incipits.

[33] In this session, I participated with several other people in groups to workshop how to catalog complex objects like broadside ballads and binder's volumes according to the FRBR model, whose impact on my project I consider in detail in Chapter 2. The fruits of this conversation appear later in this chapter and in Chapter 2, where I discuss how the FRBR impacts my metadata schema. This AMS discussion, and the ensuing research into the setup and tables constituting EBBA, mark the start of this dissertation project.

different woodcuts that might be used on different broadsides, and in different places. Thus, EBBA contains indices of the wood cuts that could be reused in different ballads as well as the larger broadsides. In this way, both the larger broadside and the smaller objects that constitute it are searchable, and this dual functionality was a central aim for my project. Binder's volumes are individual works bound in larger volumes, but with even more potential for individuality; pieces tend to contain marginalia and volumes possess covers, binding, orderings of pieces, and other factors that should be accounted for. Thus, EBBA became a useful model for keeping track of information, but it is by no means the only one.

The Hofmeister project is one of the larger database projects that relates to nineteenth century publications. This project combined the resources of the Royal Holloway Music Department with the Center for Computing in the Humanities to engage in a massive digitization of more than 330,000 records drawn from Leipzig music publisher Friedrich Hofmeister's catalogs of printed music issued between 1829 and 1900.[34] The project is very thorough, especially given the last update posted was September 22, 2008, and the database very effectively helps users find a particular document. The project regularized names and detailed XML markup help ensure that data is available for ensuring that the data about a particular place, name, or entry stays attached to the original data and that presentation becomes a later concern.[35]

One limitation of this database is the challenge of exporting large data sets to users. As stated on the help page,

> While Hofmeister XIX is an indispensable resource for the bibliography of nineteenth-century printed music, it is essential to remember that it is a series of publisher's catalogues placed online, not a purpose-designed musicological database based on

---

[34] Hofmeister XIX <https://hofmeister.rhul.ac.uk/2008/content/about/project.html>. Here I cite the about discussion of the project, as published in October of 2022.

[35] Ibid.

independent research. To use Hofmeister XIX effectively, it is necessary to be aware of the limitations that result from this, as well as the inherent complexity of the material.[36]

This limitation is important for all projects, and it highlights what cannot be done reliably from the database. While very useful in focusing research to a particular source, the Hofmeister catalog does not allow for a reliable search for Friedrich Burgmüller's published works in the catalog. Even with the process of regularizing names, problems abound when contemplating whether Fr. Burgmüller is Friedrich or Francois, and it requires individualized attention to parse these nuances, and it is not always possible. Other issues abound with abbreviations like "idem." complicating how several lines in a catalog from one composer might not always credit her or him.[37] Thus, while the Hofmeister catalog is an indispensable resource, it is not geared towards quantitative research.

Several other databases have impacted my vision for the project. I was fortunate to hear presentations on the William Blake Archives and the Seward Archive at the 2022 Nineteenth Century Studies Association,[38] and the presentations were helpful in contemplating how free text and marginalia might be effectively cataloged. This is one challenge for capturing the most interesting aspects of the page: how each volume becomes unique. Capturing marginalia in a meaningful yet time-efficient way is one of the challenges for binder's volumes and is discussed in Chapter 2 with the metadata schema. With that said, I looked to both of these projects with aspirational goals in mind. In terms of scope, both fall somewhere between the EBBA sized

---

[36] Hofmeister XIX <https://hofmeister.rhul.ac.uk/2008/content/help/help.html>.

[37] Ibid.

[38] Oishani Sengupta , "'Images of Various Species': Remediating Visual Data in the William Blake Archive," Lauren Davis, "Beyond the Archive: The Seward Family Digital Archive as Research, Instruction, and Community Engagement." Both presentations took place at a student panel I was a part of in the Nineteenth Century Studies Association's *Radicalism and Reform* conference March 17–19, 2022, and the discussions that took place as a part of the panel have helped spark ideas for future direction in the project.

project and the current scope of my project. While I cannot emulate these databases yet, I look

forward to advancing research in pursuit of these larger projects.

Data Cleaning: Philosophical Considerations

Chapter 3 deals with the use of data, but a brief word about data is merited here. Data

within this database is taken from existing cataloged materials; indeed, one of the core aims in

this database is to pursue and define the limits which existing cataloging can be integrated into a

larger, research-oriented searchable database. The process for integrating data from other sources

is worth examining further (and Chapter 3 does so), but it is important to understand here that the

data does come from other sources, and that this is a key feature for the BVRI's expansion. I

know that I cannot visit all repositories that I hope to include at some point (a fact thoroughly

reinforced by the recent pandemic), but establishing the baseline for how to integrate information

from disparate places and cataloging modes is a central tenet of the project.

When I use the term searchable database, I mean that the user interface and database

itself are designed with the goal of helping direct users to a particular source or output. This in

turn means I must design my project in a way that anticipates my users' searches. The data is

carefully added, with steps that help ensure the validity of the information being included, and

this rigor helps to ensure that the database is indeed sufficient for musicological research.[39] This

in turn necessitates that I often need to modify or "clean" my data to maintain its consistency and

integrity. In cleaning the data for the database, there are a few principles that undergird my

database: 1.) preserving original documentation as a general rule, 2.) going through data to

correct clear mistakes and convert to my schema as part of the importing process.

---

[39] Chapter 3 details the process for cleaning data and importing it, when data was not directly entered by myself (or in the future a person who has read the process for inputting said data.

In importing data, my primary goal was to preserve data as it is entered by the catalogers. There are several reasons for this position, but the most important is that I want to preserve the original record of cataloging unless clear evidence exists that the item was entered incorrectly. For example, unless I am confident that a piece was incorrectly labeled, I defer to what is written. Examples of incorrect labelling include multiple entries for composers that could be the same person, differing punctuation in the title, and other minor details.

The primary exceptions to this stance are when the original gives clear evidence of different information. The basic process that I used to catch these types of inconsistencies are clustering in OpenRefine (see Chapter 3), where it is possible to look for examples of closely matched text with very slight differences (commas instead of periods). In the case of a discrepancy, I would always refer to the original text before making a change. In many cases, different publishers would make minor differences to title text or list different composers, and I determined that these differences are worth preserving for musicological research purposes.

There are other times where data would be slightly altered: not in substance, but in format. One example is when a date is uncertain. Dates like 1842-43 are entered as 1842.5 to allow for numeric searches, and I adjust the format of data to account for caps. At other times, special characters need to be reinserted into the text after data imports and exports. This is one reason I commit to reference numbers for each composer, song, publisher, and other information; it reduces the number of corrections.

The goals behind all the data cleaning changes are consistency and transparency in storing data, ease of use in manipulating data, and clarity in presenting data. Keeping each composer separate allows for one edit of a composer to impact all instances, and it prevents splits in composer type later as more information comes to light on composers underrepresented in

current research. Changes to keep data types as numbers allows researchers the ability to search data quickly and efficiently, and as the project grows, there are options for presenting dates like 1842.5 as 1842-43 using conditional functions. For now, this is one choice that gives researchers the data that they would seek in manipulating the data. Finally, all the changes, the cleaning, and the research in this section focuses on making sure data reaches its target audience as clearly and usably as possible. The separate tables and powerful search functions of a database allow users to zero in on particular dates, creators, and musical works with ease. The decisions outlined in Chapter 2 for the organization of material, and in Chapter 3 for data cleaning, all were made with these goals as the driving focus.

As the project continues to grow (after the completion of this dissertation), one of my future research goals is to make a functional API that allows users to easily download the tables contained in the database as well as tables from individual searches. This is the type of data that I hope to make available; in some ways the dissertation is the guide for upscaling the project as well as the documentation for how the database is assembled and samples of what can be done with it.

Sorting Through Abstract Works and Material Manifestations

When writing about these musical items and material history, it has been and continues to be important to have an idea both of how people used and viewed the item at the time, as well as of how we might use and view items today. Women at the time would have dealt with the music in binder's volumes in two different ways. First, the music was used as individual pieces of sheet music. This is important for people who used it in its unbound state. It was often used for practice and at times doodled on. Marginalia and other factors become important in understanding how a person (usually a woman in the case of binder's volumes) played from the

sheet music. The second way is as the collected whole, with the sum of all music being a different experience than each individual part. In each volume, the work takes on a significance beyond just the individual song; it becomes a part of a whole, similar to how an album can (and often is) a much greater statement than the individual songs. Catalogers and librarians have wrestled with this issue, and this section explores the challenges with conceptually holding the volume and the works contained therein as individually important. This section explores how catalogers accomplish the goal when working with various materials, and how the methods used by librarians become relevant for binder's volumes.

In constructing the idea of a work, Lydia Goehr's research on the work-concept is important to consider here. Her framework supports the idea of a work states, "(i) that it is an *open* concept with *original* and *derivative* employment; (ii) that it is correlated to the ideals of a practice; (iii) that it is a regulative concept; (iv) that it is projective; and (v) that it is an emergent concept."[40] This work concept arose near the end of the eighteenth century, and shaped the nineteenth century concept of musical works.[41] The work concept is an important part of this project because it is appropriate both to the period in which these pieces were collected and bound, and thus a relevant framework for modern researchers. In unpacking the idea of a work, Goer's broad definition is not the same as what generated the FRBR model outlined below, but both Goehr's framework of a work concept and the one that arose from library practice share several key points. Perhaps most essential to the current discussion is the broad, open-ended nature of a work and its various realizations, and I attempt to unpack and analyze this relationship below.

---

[40] Lydia Goehr's frameworks of a work. Emphasis present in original. Lydia Goehr, *Imaginary Museum of Musical Works: An Essay on the Philosophy of Music* (New York: Oxford University Press, 1992): 89-90.

[41] Ibid., 2–3.

In discussing the sheet music and binder's volumes, it is worth considering the work concept, and what constitutes a work. Functional Requirements for Bibliographic Records (FRBR) provide a framework for engaging with the question of creative endeavors, and how to view each part of the process. The overview is the work, expression, manifestation, item (WEMI) model, which moves from work as an abstract idea to an item as a single, tangible instance of that work. FRBR defines **the work** as a distinct intellectual or artistic creation.[42] This work is then *realized* in various ways. Realizations of a work are deemed **expressions**, which are themselves not yet objects. Expressions are then *embodied* in **manifestations**, which are physical templates for the creation of individual **items.** Each item is thus an *exemplar* of a manifestation.

In the case of a book, this concept is often relatively straightforward. I have on my bookshelf a copy of Tom Clancy's 1991 American bestselling novel *The Sum of All Fears*. That individual book is an item. The book is a paperback, published by Berkley Books. It is a reprint; many subsequent books are advertised on the cover. The physical book that I can hold is the item. It is one example created by the print run of books, which might be thought of as the manifestation level of the work. Manifestations are the promise of an item and as such a print run, or a specific saved copy of the file, is the manifestation of the work. This print run (be it thought of as a plate or file) is one embodiment of the expression, which might be thought of as the English-language text of the story. A Spanish-language audiobook would be a different expression of the overall story that is *The Sum of All Fears*. Thus, the metaphysical idea of Jack Ryan saving us all from a nuclear war is the work, the various ways that the story might be realized (print, audiobook, language) is the expression, the process of physically realizing one

---

[42] Many sources describe FRBR, but the OCLC website provides an effective summary. "OCLC Research Activities and IFLA's Functional Requirements for Bibliographic Records," https://www.oclc.org/research/activities/frbr.html.

expression (print run, digital file, etc.) is the manifestation, and one individual copy is the item.

Generally, this idea is straightforward in a book. Music immediately presents challenges within the FRBR model. Composers, arrangers, and lyricists all share some of the creative credit. Music also creates a separate challenge for understanding what the item is, largely because of the numerous natures of what constitutes music. Sheet music is, for historical purposes, an item; however, sheet music simultaneously serves as the instructions—or the foundation—for numerous different musical performances. Thus, sheet music can function as a manifestation that underpins a separate production of items (should the piece be recorded). Performing sheet music would not fall into the item category by itself, just as reading *The Sum of All Fears* on a street corner does not by itself produce an item unless it is recorded and preserved in some form. This is not meant to devalue socially and culturally music-making that does not produce, but simply to point out that it is not a part of the FRBR model unless it becomes archived into some long-lasting form. This has traditionally meant paper, but in the digital age numerous ways of archiving works exist.

Returning to the past, FRBR terminology offers a useful framework to consider issues of musical consumption in Antebellum America. In the nineteenth century, manifestations of a musical work were generated at a rapid pace. Composers might write an opera, but copyists quickly made piano arrangements of varying degrees of difficulty of the most prominent arias and excerpts. This process, which included publishers occasionally stepping on each others' rights and credit at times being misattributed (both intentionally and accidentally), creates problems in sorting out the exact information for an individual work. Not every piece is the same even if it has the same title and melody. In cataloging works, I make sure that my database preserves each version created, which does create more copies of a title but also recognizes that

26

different titles from different print runs are different manifestations/items and should be understood as such. I explain how exactly FRBR has shaped the organization of my database in Chapter 2.

The discussion thus far has centered on individual pieces and how performers might interact with those pieces, but the volumes themselves are a separate entity. Each volume is a creative effort, typically undertaken by the person who owned the music being bound. Some took care and diligently made tables of content, included pagination, or otherwise put clear thought into the ordering of the music. Others did not produce binder's volumes that indicate such care or precision. In either case, the ordering of the sheet music and reification of those pieces into an order produce a separate creation, a separate item reliant on its constituent parts but at the same time more than any one song. For contemporaneous users, this bound volume presented a different use case for enjoyment. Music was now in order, and order often impacts our own biases in how we engage with art. Just as albums are ordered to create an experience, so too did binder's volumes have the potential to influence their users.

Thus, any attempt to catalog binder's volumes must preserve the order and information contained therein, as well as how each piece has become different from its original manifestation. My goal in cataloging the data is to come as close as possible to documenting the two phases that the music was used, but the final solution in this dissertation and database is slightly different. The reason is because our present perspective does not allow for differentiation between indicators of use like marginalia pre-binding and post-binding. As is discussed further in Chapter 2, I created separate layers in my database for sheet music on the one hand, and binder's volumes on the other, considering unique details like marginalia in the latter. This does not precisely align with past use, but this system does preserve the distinction between music printed

and music bound.

One additional bonus in this system is the mechanism to document accurately who likely took agency in creating the item at hand. With respect to printed music, this means that creators (composers, arrangers, lyricists, translators, etc.) are cataloged with the manifestation of the work, which captures their input. Publishers are also referenced here, and the manifestation catalog becomes a record of the music as first printed. Thus, the catalog of the manifestations of music contained in binder's volumes (discussed in Chapter 2) houses all information about the creation of the work before it reaches the consumer.

In total, I have sought to develop a system that provides a framework to recognize both what each item is, and who made creative contributions to each item. The ultimate goal of this framework is to document in an easily accessible way who contributed to the final binder's volume, and in what way. To say that another way, this framework focuses on what the material is, how it changes, and who makes changes to the material itself throughout its history. Understanding the sheet music and binder's volumes—the material in the history—in this way will help to understand the social implications of music and musicking at the time.

<center>Engaging with the Musical Materials</center>

Indeed, understanding how people experience musical materials is a central factor, as well as a primary delimiter, of the types of musical items that this dissertation sets out to engage with. Sheet music in its bound form was popular in a particular window of time, and while I mentioned those bounds from a material sense as what underpins this change earlier in the chapter, I also want to highlight what those bounds mean for the consumer, who experiences them. To that end, this section identifies the materials that I engage with, and the materials I exclude. While the dissertation (like most dissertations) includes temporal boundaries in its title,

these temporal delimiters are the product of the samples incorporated thus far. As the BVRI

expands, and the bounds of outliers change (see Chapter 4 for a discussion of outliers in dates),

the time range of the BVRI will also shift and possibly expand. But everything hinges on the

materials. To that end, the below sections discuss musical materials and how I determined the

bounds of the project. The first section defines the early bounds of the project, which deals with

rising print culture and the decline of handwritten volumes. The later bound is driven by

twentieth-century advances in music and culture.

Handwritten Volumes and the Absence of Standardization before 1820

Until the print revolution dramatically reduced the cost of printed materials, most people

who possessed music transmitted much of it through handwritten copies. Amateur musicians

kept manuscript books, which provide a lens into the types of musicmaking of the time.[43] The

manuscript book was also popular in Europe, and studies of the time, particularly Glenda

Goodman's work, offer insight into music and culture with each book. A manuscript book by its

nature is personal, and individual. The person who copies a book actively puts their labor into

procuring the original to copy, and into writing each note. Standardization is not as important for

these volumes; it cannot be with the personalized volume predicated on individual copyists. This

makes each volume unique and gives each character.

For all of the reasons that these sources give a uniquely individualized look into the past,

they also become problematic for my current work. In copying a work, attribution is important

for some, but often unnecessary. Where publishers often are required by contract to acknowledge

the labor of others, home copyists face no obligation, and attributions are incomplete, inaccurate,

---

[43] Goodman, 2.

or at times nonexistent. This is not the fault of the copyist; I often take notes which are incomplete to others, and my desire to look back historically should obviously not imply that people at the time "missed something." Nevertheless, the individuality here negates the manifestation level of musical works entirely, as each hand copy becomes removed from the past in a way difficult, if not impossible, to trace en masse.

Thus, the individualized experience of copying music, while important, falls outside the scope of my research, though I start soon after. The transition from copying individual works to buying them is the starting place for my inquiries. The date is not firm, and is informed by materials, not by a line I place. Nevertheless, the terminus post quem, as determined by the latest published work bound into a volume, tends to place my focus as beginning after 1830 based on the current materials.

Twentieth Century Considerations

While the beginning of binder's volumes is relatively easy to trace in its time, the end of the practice is somewhat harder. Numerous factors influence the tapering of the practice, which seems to happen between 1875 and 1900. I make no attempt to assign causality or proportionality to any of these factors, but cooccurrence and the possibility of causality merit each one being mentioned here. The assignation of cause is a project for future research projects, and not necessarily my own.

The first item to note is a general decline in scrapbooking, especially in the way that was associated with bound volumes in the dowry. This trend coincides with the end of the Civil War Reconstruction and the rise of several categories of women's groups. Taken together, there appears to be a notable move from women staying in the house to spaces where women were

much more frequently out and about.[44]

Along with the end of scrapbooking comes the rise of anthologies. While sheet music was often purchased one piece at a time before 1875, the last third of the nineteenth century saw a marked increase in printed anthologies that contained multiple pieces by multiple authors, some of which might be 10-20 works long. Such anthologies could almost function as a replacement for binder's volumes themselves; they were certainly not as personal, but the time saved to get many of one's favorite pieces is an important factor to consider.[45]

At the same time, some of the journals that had actively encouraged scrapbooking and binder's volumes faded as print culture and social culture changed. The *Godey's Lady's Book* was one such prominent periodical, beginning in 1830, changing hands in the 1870s, and closing entirely in the 1890s. In some ways the rise and fall of the *Godey's Lady's Book* mirrors the popularity of binder's volumes and provided for their contents. Figure 5 shows one of the first pages of the July 1855 Godey's Lady's Book, and such pieces were easily collectible and could be readily removed for piano playing, and eventually for binding. "Shells of Ocean" appears multiple times in the BVRI. Thus, these lady's books—of which *Godey's* was just one— provided content for binder's volumes and, given their symbiotic relationship, it makes sense that a fall of popularity one would be at least in part reflected by the other.

---

[44] Meyer-Frazier, p. 40–41.

[45] On a personal level, I purchased music by looking for books of music that had a piece I wanted, then played through the rest of the book to see if I found any other piano pieces I liked. Because books and individual copies were so close in price, I tried to avoid just a single copy of a work when possible. Several factors contribute to the rise of anthologies, notably the improvements in paper developments and print continually becoming cheaper and faster. Other scholars have also engaged with this issue, notably Deirdre Loughridge in "Making, Collecting, and Reading Music Facsimiles before Photography" and James Davies in "Julia's Gift: The Social Life of Scores, ca. 1830." *Journal of the Royal Musical Association* 131, no. 2 (2006): 287-309. Peter Mondelli also discusses these matters in chapter 2 of his forthcoming book and at the end of his chapter "Parisian Opera between Commons and Commodity, ca. 1830," *Consuming Music: Individuals, Institutions, Communities, 1730–1830* (Rochester: University of Rochester Press, 2017).

Figure 5: First page of the July 1855 issue of the Godey's Lady's Book, from the University of Vermont. https://www.uvm.edu/~hag/godey/glb185507/index.html#

Lastly, significant changes in music technology coincide with the end of binder's

volumes. The anthology certainly impacted print music, but the twentieth century also saw the

beginning of a shift in how the average person experienced music. The nineteenth century musical experience in which audiences experienced works live, performed them at home, and read about them in newspapers, is markedly different from the twentieth century rise of hearing music from recordings and player pianos. This shift, taken with the other aforementioned factors, helps frame my research.

Pragmatic Bounds for the Project

To be clear, the project is not bound by any factor other than whether a binder's volume exists, and when its various components were created and assembled. Everything mentioned before this point influences the creation of binder's volumes, but if the binder's volume does exist, I hope to someday record it into the project. The early bounds become the creation of the volumes, and the late bounds largely deal with tapering. For data purposes, I anticipate most volumes falling between 1830 and 1875 as my first and third quartiles as the data progresses, but time will tell as volumes continue to be added.

Lastly, I see several possible uses for this project beyond those considered in the final chapters of this dissertation. I present three here, but this is not meant to be restrictive. First, for performers, this is a way to find repertoire, and specifically repertoire that has marginalia intact. This offers performers a way to look at nineteenth-century music, with fingerings intact, and to examine multiple contemporaneous copies of the sheet music.[46] Second, for performers and researchers, this presents a catalog of where binder's volumes and specific works exist. This will make it easier to find a copy of a specific work, and to compare various editions and works. The third case is for data-driven researchers, and it involves looking at the metadata in aggregate.

---

[46] For an example of how a detailed analysis of musical annotation and marginalia can influence one's understanding of the music itself, see Fabio Morabito, "Rehearsing the Social: Beethoven's Late Quartets in Paris, 1825–1829," *The Journal of Musicology*, Vol. 37, Issue 3, pp. 349–382.

Chapters 4 and 5 explore some of these cases, but as the size of the project increases, I intend to explore more potential use cases and to present these over time.

Meditation on Digital Historiography

One central theme throughout this dissertation is the possibilities that open up when one takes traditional materials and uses them in nontraditional ways. In terms of the music itself, I certainly value the notes on the page and the sounds created by the music—subsequent research projects will undoubtedly focus more on these aspects of the music—but this dissertation aims to critically examine the data and metadata that typically falls by the wayside, and to pursue ways the music scholarship can expand and benefit from an increased inclusion of small data in aggregate.

Such a discussion of the materials and hidden data is particularly necessary in the COVID and post-COVID age of research. The last thirty years has seen many research resources move from paper to more digital forms. Databases like JSTOR have made access to all types of secondary literature extremely rapid, and sources themselves have moved online. As more and more archives and other repositories of primary sources continue to move online, it is incumbent on all of us to interrogate the ways we present online information. While online resources, including the database produced in conjunction with this project, exist as a way of cataloging and increasing access to a print source, we are also moving into a space where the online source itself should seek to move beyond simply representing the volume and into new ways of allowing users to interact with primary sources.

This dissertation thus seeks to demonstrate the new possibilities afforded by changing how we interact with archival materials. This mindset shift is a central goal of my project; it developed in conjunction with my teaching during the pandemic. Early in 2020, everyone was

forced from the comfort of in-person teaching to a new world of online pedagogy. I noticed

many people (including myself) working to replicate in-person strategies on various virtual

platforms. By mid-to-late 2020, I had shifted my online teaching philosophy to embrace a

different approach: what can I now do online that I could not do in a classroom before? This shift

led me to thinking of the online world as an opportunity. Virtual pedagogy is inherently neither

better nor worse than in-person pedagogy; it is different. Any attempt to make one modality a

lesser version of the other is doomed to inadequacy. Instead, I came to view online pedagogy,

and eventually online research, as a separate category that functions best without the constraints

of traditional methods.

One way to illustrate the difference in approaches is to compare the *Journal of Music

Theory* (JMT) with *Music Theory Online* (MTO). Both are important periodicals and serve as

pillars of the theory community. JMT is the oldest theory journal published in the United States

and has provided the forum for high-level theory papers in American music theory. MTO, a

much newer journal, exists only for the online space and capitalizes on the opportunities.

Footnotes and sources are embedded in the article itself in addition to being available at the

bottom of the page, and musical examples are also available.[47] In addition to being able to see

musical examples, MTO often incorporates the ability to listen to examples—a nuance simply

unavailable to JMT and other print outlets.

This type of innovation is one example of what happens when the digital interaction is

prioritized as the means of how scholars interact with objects; that is, when there is an explicit

change of behavior to recognize that materials are being dealt with differently, and that the

---

[47] One of my favorite examples is this article on backbeat and polyrhythms. Zachary Cairns, "Switching the
Backbeat: The Quick Flip and Polymetric Pogo in 1980s-era Rock Music," *Music Theory Online* Vol. 28, No. 1
(March 2022). https://mtosmt.org/issues/mto.22.28.1/mto.22.28.1.cairns.html. This is a very different, and in my
opinion much less convincing article in plain PDF format.

materials themselves are now different. This project began as a material history of sheet music, and its scope has become defined by the binder's volume as a means of defining both how the materials were consumed and used. Binder's volumes themselves are complicated, as they both are an object to study and contain individual objects of study. This dichotomy has forced me to consider objects, how individuals interact with them, and the impact organizing and presenting materials has on their use, consumption, and on the object themselves.

Thus, we return to the issue of sources, their organization, and presentation. As mentioned, sources were moving online as a mode of access as early as the 1990s, and the pandemic radically accelerated the movement of items online. As we emerge from a world defined by the ebbs and flows of infection and hospitalization rates, this is exactly the right time to look at how sources are moved online, and what advantages come from the process. Too often the conversation is limited to what is lost, and I hope with this dissertation to show in detail how this move might be accomplished in a research oriented and rigorous way that creates a database useful for musicological inquiry. Along the way, I provide the rationale for my decisions and give some takeaways.

Chapter 2 moves from the rationale and literature review for the database into the metadata schema. First, I provide an overview of what a database is, and how it helps research. Next, I give a detailed description of the data and metadata included in the project. I also include the metadata schema itself for each table in the database (also Appendix A). Chapter 3 focuses on how I clean data and what choices I make to generate my database. Appendix B is gives standard operating procedures for systematically ensuring that data is included in the document in a reasonable way. Chapter 4 contains two case studies showing a sample of what research possibilities can be accomplished with this database. Chapter 5 offers concluding remarks,

highlighting future research goals and steps to take with this project as well as suggestions for researchers attempting projects like this endeavor.

CHAPTER 2

ORGANIZING AND CONSTRUCTING THE BVRI

The purpose of this project is to assemble the information contained in music binder's

volumes in an approachable and usable way for researchers. To this end, I have created a

relational database that organizes and manages relevant data concerning binder's volumes,

makes the data searchable, and facilitates adding new information into the database with ease.

This chapter addresses two primary questions: first, why a relational database; and second, how

should such a database be structured to make it usable?

What is a Relational Database?

Given that this dissertation is directed towards the humanities, the best place to start is

with the mechanics of what a relational database is, and what it offers beyond a typical

spreadsheet or table. A table, at its core, is information organized into rows and columns, which

facilitates easy viewing. Examples include websites that use tables to lock images and icons in

place relative to other graphics and text, and word documents or monographs that use tables as

inserts to present data. Spreadsheets are tables that offer additional mechanisms such as searches,

formulas, graphs, and other functions that enhance what users can do to both manipulate the data

and display it to others. For example, a spreadsheet can calculate the average of several grades

in a class, weight different types of grades based on how the instructor desires tests and daily

work to contribute proportionally to a final grade, and display both a percentage and an

associated letter grade. Most people use spreadsheets and tables throughout their day, whether it

be to organize lists, calculate grades, view websites, or any number of tasks that involve data

organization and manipulation.

Spreadsheets provide ways to connect tables but do so most effectively when the link is

thought of in advance and designed for a specific purpose. The most used spreadsheet programs are Microsoft Excel, Google Sheets, and Apple Numbers, and all of these programs both arrange data in tables and provide functionality for editing cells in the table and for presenting the data in graphs and charts. This can work well if values are being totaled or calculations are being conducted on one table and the result is displayed in another, as in a monthly budget. Figures 6 and 7 contain images of the two sheets contained in Google's monthly budget prototype, which is a typical example of how sheets, or tables, should be linked to be effective. Figure 6 is the lower half of the summary page; the upper half contains graphs to display visually the data contained in the section shown here. At the very top of Figure 6, the formula bar contains the instructions for connecting to the transactions sheet seen in Figure 7. The instructions in the formula bar read as follows: If cell B28 is blank, then cell E28 (the highlighted cell) should be left blank. If cell B28 is not blank—in this case B28 is Food—then cell E28 should contain the sum of all cells in the transactions page that meet the following criteria. For any cell in column E that contains the word Food (cell B28), then the corresponding row in column C should be included in the sum. Stated more plainly, any amount marked as food should in the transactions page is included in the food total. Despite the laborious nature of restating this formula in English, it is worth doing once to see exactly how spreadsheets map data from one table to another. This can be done many times and, for people fluent in writing and understanding formulas, this process does not take a great deal of time. It does, however, require a specific plan for each link between tables, and it requires the intent to be completely thought through before a link is made, if it is to be efficient and effective. Also, as is the case in this example, these links work extremely well for numbers or data. These links become much more laborious when the data type is free text fields without any constraints on the data type within the field.

Figure 6: Example of income and expenses in Google Spreadsheet, sheet 1.

Figure 7: Example of income and expenses in Google Spreadsheet, sheet 2.

Databases operate on the same principal as spreadsheets, but they also provide more robust mechanisms to connect multiple tables. The differences here can be challenging to understand, so it is worth taking a moment to emphasize this point. Instead of linking tables through specific instructions for each cell, a database uses one specific column in a table to link to another column. This can even happen through chains of tables and creates more complex

relationships. Due to the freedom afforded by this expanded range of connections, databases can also generate tables in a single search and export them, which creates significantly more opportunities for searching and reporting data, especially when the outcomes are not yet determined. One way to think of the potential of databases is that it turns a two-dimensional spreadsheet into something with three dimensions. Databases, like spreadsheets, incorporate tables of two-dimensional data, but databases are much more capable of linking data between tables. While spreadsheets can do so to a limited degree (see figures above), databases exist with this functionality as axiomatic to their purpose. Databases are also much more capable of expanding to incorporate massive amounts of data provided they are formed well at the outset. For all of these reasons, I have used a database to contain and model my data.

Before delving directly into databases, many readers have used databases at some point. In terms of building them, multiple database management systems exist. SQL-based platforms like mySQL and Microsoft SQL Server, and other programs like Oracle Database, Microsoft Access, and FileMaker. Each program offers slightly different terminology and features, and those lie outside the scope of this dissertation. Instead, this section is a more rudimentary guide to what databases are and what they can do.

Some database terminology is useful going forward, given the data-heavy nature of this project. Tables are the primary mechanism for organizing data. A database contains several tables, each of which is designed to hold information and organize it. Each table has rows and columns. In a database, the rows in the tables are called records. A record is a set of data about a particular object. If I had a table organizing address labels for my Christmas card list, a record would be all the information related to a person or household receiving a single card. It would include the name, address, city, state, and zip code. Each one of these separate types of data

would be a column in the database, which is a field. Address, city, state, and zip code are all

fields in this example. Even a straightforward example like this demonstrates the ways in which

a database needs to be crafted carefully with its goals in mind. Adding names, for example, can

be challenging and necessitates planning and choices early in the design process. It is possible to

leave a person's whole name as a field. In my Christmas card list, I often write names as I would

like them to appear on the envelope: "Rob and Pam Anderson," "The Jones Family," or simply

"Jebediah Smith." In more complex lists, it often makes sense to break out names into the

separate fields of first name and last name. This is especially useful if one might want to search

for a particular last name later in the process, or if alphabetizing might be useful.

Databases become far more powerful tools when tables are connected. Keys connect one

table to another. The basic way to connect tables is to choose one field that is unique to every

record and designate that as the primary key. A primary key is the field other tables use to access

a table, and it has several distinguishing factors. First and foremost, a primary key must be

unique. Sometimes a field naturally fits this description. ISBNs can be primary keys when

dealing with books, just as social security numbers can do so with people. The need for a unique

point of reference, however, often requires the use of a separate field for that specific purpose. In

my Christmas card example, there is no field that I can guarantee is unique. The most common

solution to this problem is to make a numerical field that serves as the primary key. This way

every record has a unique number, and no number can be repeated. A second rule for primary

keys is that they cannot be null, or blank. This would prevent a record from being referenced by

another table and can create cascading errors as the database looks for information it cannot find.

Once a table has a primary key, another table can use that key to call upon all the data in

that table. The parent table seeking to draw upon another table does so through a secondary key.

When a field is designated as the secondary key, it is linked to the primary key of another table, the parent table. This allows the second, child table to draw information from the parent table through a single number. For example, a shipping company could use the list of customers from a mailing list to assign packages a delivery location simply by calling up the primary key. Many of us have seen this feature work when filling out online forms. Some forms ask for a zip code, and then it auto fills the city, state, and country from that single zip code. This is all done through databases drawing on other tables to get data. Secondary keys do not face the same restrictions as primary keys: they can repeat, contain null values, and can be deleted in the child table. None of these is true for a primary key in the parent table.

Having explained some terminology pertinent to relational databases and their structure, we can now consider when it makes sense to link two tables together. First, the data in the tables should exist in a one-to-many relationship. Two sets can be related in three different ways: one-to-one, one-to-many, and many-to-many. When two sets of data exist in a one-to-one relationship, that data generally should be put in the same table. For example, it does not make sense in a mailing list to have people in one table and addresses in another table. People generally have one address, and with limited exceptions, an address has one family of residents. Thus, people and addresses have a one-to-one relationship and should exist in the same table. While it is possible to relate tables that have a one-to-one relationship, this usually creates an unnecessary extra step.

One-to-many relationships are generally where databases work well. In this case, tables exist in a parent-child relationship and one table feeds into another. The example of shipping packages works well here. A person may order multiple packages, but each package is only shipped to one person. This is the basic idea behind a one-to-many relationship. In situations like

this, data from one table can help populate a second table and databases function quite well. Databases can have numerous one-to-many relationships.

Many-to-many relationships are complicated in a database and should be used sparingly. If they are necessary, the best way to accomplish the goal is through a join table. Otherwise, the many-to-many relationship creates circular reference problems where a table attempts to act as a parent to another table, providing it information, while also receiving information from the same table, which is the fed back into the table and the cycle continues indefinitely. A join table solves this problem. One of the most common examples of a many-to-many relationship is between students and classes in a high school setting. Students often have eight classes, and each class— biology, math, physics, history, etc.—might have hundreds of students depending on the size of the school. A join table is a table where both tables in the many-to-many relationship feed information, creating a space where that information can come together. Join tables can also provide extra context to the relationship between the parent tables. In the example of students and classes, a join table might note that Student 123 (Ezekiel) is taking Class 26 (biology), and that this juncture occurs during fifth period. Other tables can also contribute information (Teacher 13, Ms. Monadjemi). Because the join table uses every table leading to it as a secondary key, both students and classes can be repeated. The join table (schedule) can list all eight classes Ezekiel is taking as well as the 590 students taking biology at Ezekiel's very large school.

In general, the method for building databases should use as many one-to-many tables as needed and should be mindful of the number of join tables. Often one join table is ideal for displaying the culmination of the databases data, but problems can arise if more than one join table exists in a database. The rule for tables only feeding in one direction extends through any

tables to which the original table is a parent. Multiple join tables have the potential to create circular reference errors that follow from attempting a many-to-many relationship without a join table, and they place restrictions on the construction of future tables due to the potential for these errors. One step that can help prevent such errors is a diagram mapping the table structure within the database. Figures 8 and 9 contain the table relationship for my database, and I describe the structure in detail in the next section.

When it comes to generating reports through databases, the find/query function is powerful. Databases have different names, but the find function in FileMaker mirrors the query function in Microsoft Access users; every database has some process like this. FileMaker saves finds for repeated use, which is particularly useful as the database grows in size. Finds can be simple, or they can be complex and parse multiple fields using operators (like AND, OR, etc.).

Why a Database in This Case

Thousands of music binder's volumes are known to exist throughout the United States; those identified in Chapter 1 are just a small sample. If each binder's volume contains between 20 and 80 works, hundreds of thousands of copies of songs and pieces are contained in these volumes. A spreadsheet is useful for smaller studies, but any project that envisions scaling up past 1,000 binder's volumes needs a way to keep information separate and to query select pieces of information. A relational database provides this functionality. As discussed, such one-to-many relationships are ideally suited to database queries.

In addition to the number of works, music binder's volumes pose a particularly challenging set of questions to catalogists and researchers alike. The problem is that one extant object from a specific time—a song, for example—has been incorporated into another catalogable object—the volume itself. To truly document the contents of binder's volumes, it is

45

important to make note of both the identifying information for each work (publisher, arranger, composer, date), but also for each volume (compiler, order of works, date bound, marginalia).[48] Each of these questions merits its own set of information if both parts of the issue are to be handled equally. To be sure, there are valid reasons to focus on either the volumes or the works, but my aim has always been to consider both parts together.

Relational databases provide a method to keep all this information organized accurately and usably. They specialize in organizing vast quantities of information about sheet music and about books more generally. Moreover, databases allow for tables of known information—such as composers or publishers—to populate other tables through keys, which makes constructing queries about the publishing, locations, and composers straightforward. Finally, it is possible to draw connections between the musical pieces as objects on the one hand, and the binder's volumes as separate but equally important objects on the other, without forcing one or the other to be the primary object in this process. The ability to preserve this distinction allows for consideration of specific musical works and consideration of the binder's volumes writ large, without imposing a choice on the analytical process before analysis can happen.

<div align="center">Formatting Conventions Used in the Dissertation</div>

Certain formatting conventions have been created for this dissertation, to enhance comprehension and readability. Names of specific data tables used in the Binder's Volume Research Initiative relational database (BVRI) appear in green small capital letters, e.g., CREATORS TABLE. The data tables showing fields, descriptions, and data types are formated in green. Lastly, names of fields used in the data tables appear in italics, e.g., *Last Name*.

---

[48] This section intersects with a discussion of FRBR in the Musical Works Section.

As mentioned above, the Binder's Volume Research Initiative relational database (BVRI) contains several tables, and various tables have different functions. The easiest way to view these relationships is by looking at the following figures.

Figure 8 shows the database layout. Read from left to right, information contained in one table, or set of tables, populate the next table. For example, the PUBLISHERS TABLE feeds into the MUSICAL WORKS TABLE , which then populates into the BV [BINDER'S VOLUME] CONTENTS TABLE.
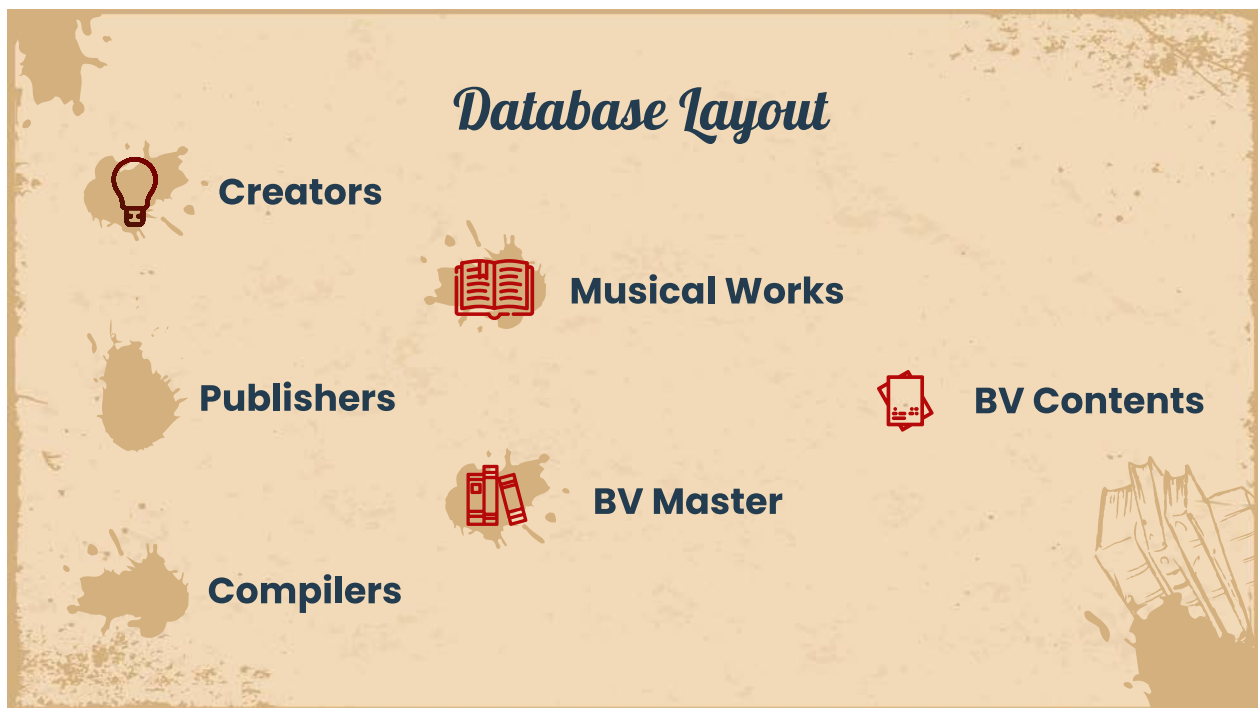
Figure 8: Database layout, list of tables.

Figure 9 shows all the different tables in the BVRI and how they relate. Arrows indicate parent/child relationships between tables. For example, the CREATORS TABLE is a parent of the MUSICAL WORKS TABLE, which means that creators can be associated with a specific work. If a table does not have any arrows pointing into it, then it does not serve as a child for any parent. This is true of CREATORS, PUBLISHERS, and COMPILERS. If a table does not have any arrows going out of it, then it is not a parent of any table. Tables that have two arrows coming into it are

join tables; they facilitate a many-to-many relationship between information from two or more parent tables. The two join tables are the MUSICAL WORKS TABLE and the BV CONTENTS TABLE. The MUSICAL WORKS TABLE provides a place for multiple publishers and creators to be assigned to a manifestation of an individual work.
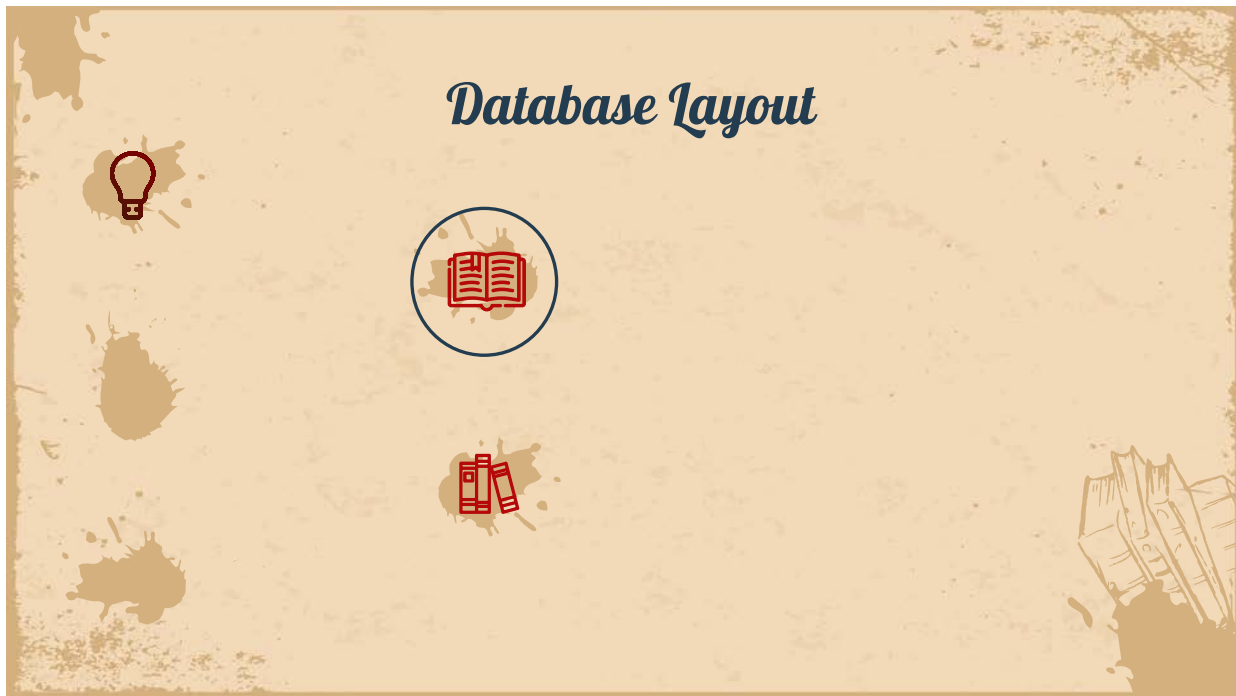
Figure 9: Database tables. Arrows indicate parent/child relationships, join tables are circled.

Following a note about location terminology used in BVRI, the subsequent sections break down what information is contained in various tables and how they relate. The data tables are found at the end of each discussion; however, Appendix A presents the data tables together in the order discussed in this chapter, for readers who may want to see the fields of each table alongside their reading.

A Brief Note about Locations

One challenge with such a database is the constantly changing geopolitical landscape. Over the course of the nineteenth century, Europe underwent several revolutions, and American

territory changed not only in terms of states being added to the Union, but also which country laid claim to parts of what is now the United States. To simplify the task of determining is the nation in which a city is located, I define countries and regions as they exist as classified by the U.N. on January 1, 2020. This means that cities such as Berlin, Leipzig, and Bonn are considered German, that Austria and Hungary are considered by their current, smaller borders, and that a hypothetical binder's volume from what is now Hays, Kansas is considered Kansan even if compiled in the early to mid 1800s, before Kansas was a state.

Creators Table

This table is focused on individuals involved in the creation of a work. One problem that arises in accounting for the labor of multifaceted musicians is that some individuals may be listed on different musical works as composers, arrangers, or performers. To solve this, I created a table of people involved in contributing creative labor to the music and some important information about each of these contributors. Other tables draw upon this one as a reference to describe the role of the creator at that time. The primary key (in my case, *Creator Counter*) functions as a reference number, simply due to the probability of having composers with the same name (fathers and sons) and to allow for different spellings.

In addition to the name and reference number of a creator, this table also contains important information about the creators to help with future searches. This is where alternate spellings of a creator's name are cataloged, to help with identifying works as they are added to the database. Also, the table contains several dates. The most obvious dates are the birth and death dates of a composer, but the table also has a field that describes when a composer is known to be active, *Flourished*, which is a range of dates, while birth and death dates are actual dates. The *Flourished* field helps track whether a composer or arranger was active at a certain time,

which helps disentangle ambiguity in determining if a specific person was involved in creating a work.[49]

Ferdinand Beyer (1803–1863) exemplifies the complicated role creators play in various fields of the MUSICAL WORKS TABLE. He was known in the United States both as a composer and as an arranger.[50] In Tables 1 and 2, the Place column shows where Breyer is associated as a creator (composer or arranger). The Composer columns shows individuals for whom Breyer was the arranger; the total number of times Beyer arranged for each composer is shown in the last row. Table 2 shows the same data with places further broken down by publisher. Not only does he appear over 50 times as the primary composer listed within the UNC volumes, but he is also the most commonly cited arranger of musical works. Breyer is listed in the database both as a primary and secondary composer, and his name is used in several titles as a means of advertising the composition.[51] Arrangements with such titles as "Beyer's Bouquet de Melodies" highlight not just the original works by Donizetti and others, but foreground Beyer's name as a reason to purchase the collection. Beyer's representation on over 70 pieces within the binder's volumes highlights his importance, and this system captures his multifaceted contributions to musical works bound in the United States.

---

[49] Two things that I contemplated putting in this table are associated publishers and known locations, but ultimately, I decided against this. In both cases, tracking these relationships is an important aspect of the BVRI's work, and these relationships need to be evaluated on an ongoing basis. Embedding locations and/or publishers within the CREATORS TABLE might be advantageous in the short term, but it creates challenges in the future, when more repositories are added to the BVRI. This illustrates the importance of keeping tables separate when relationships between the two categories are one-to-many; this is the strength of the relational database.

[50] Isaac Nagao, "Beyer, Ferdinand," Oxford Music Online, 2001; Accessed 20 Jan. 2022. https://www-oxfordmusiconline-com.libproxy.library.unt.edu/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000051838

[51] For this dissertation, a primary composer is one who is listed first on the work. A secondary composer is listed second or is the known composer of a related work (such as the opera variations are derived from.

Table 1: Places Beyer was published in the BVRI, as of January 2022.

Place	Composer for Whom Beyer Arranged							Grand Total
	Auber	Bellini	Breyer	Offenbach	Reissiger	Rouge de Lille	blank	
Baltimore, MD			1					1
Boston, MA			5	1				6
Brussels, Belgium			2					2
London, England			1					1
Mainz, Germany	1	2	14		1		1	19
New Orleans, LA						1		1
New York, NY			12				1	13
Philadelphia, PA			12				2	14
(blank)			1					1
Bonn, Germany		1						1
Grand Total	**1**	**3**	**48**	**1**	**1**	**1**	**4**	**59**

Table 2: Places and publishers associated with Breyer, as of June 1, 2022, in the BVRI.

Place and Publisher(s)	Composer for Whom Beyer Arranged							Grand Total
	Auber	Bellini	Breyer	Offenbach	Reissiger	Rouge de Lille	blank	
Baltimore, MD			**1**					**1**
G. Willig Jr.			1					1
Boston, MA			**5**	**1**				**6**
J.E. Gould			1					1
Oliver Ditson			3					3
Oliver Ditson & Co.			1	1				2

Place and Publisher(s)	Composer for Whom Beyer Arranged							
	Auber	Bellini	Breyer	Offenbach	Reissiger	Rouge de Lille	blank	Grand Total
Brussels, Belgium			**2**					**2**
B. Schott			1					1
Schott			1					1
London, England			**1**					**1**
Schott			1					1
Mainz, Germany	**1**	**2**	**14**		**1**		**1**	**19**
B. Schott	1		6		1		1	9
B. Schott's Sons		1	3					4
Les fils de B. Schott		1	5					6
New Orleans, LA						**1**		**1**
A.E. Blackmar & Bro.						1		1
New York, NY			**12**				**1**	**13**
Firth & Hall							1	1
William Hall & Son			12					12
Philadelphia, PA			**12**				**2**	**14**
Fiot			1				1	2
J.E. Gould			9					9
Lee & Walker			2				1	3
(blank)			**1**					**1**
(blank)			1					1
Bonn, Germany		**1**						**1**
N. Simrock		1						1
Grand Total	**1**	**3**	**48**	**1**	**1**	**1**	**4**	**59**

During the database creation process, I recognized the need for and thus created a *Company* field. This field captures the fact that some creators—especially performers, but also some composers—are listed as troupes or companies. The *Company* field captures this and allows for a clear indication that the entity listed should not be viewed as a person, but rather as a collective or group of people.

The last five fields are automatically entered, and they primarily exist for the purpose of organizing entries as the database expands. A major purpose of this dissertation is codifying the processes used in creating the BVRI's alpha version so that they can be expanded and updated. The *Creator Counter* is the reference number I use as the index for other tables. FileMaker's primary key exists in each table, but serves an auto archiving tool, not a reference between tables.

CREATORS TABLE		
Field Title	**Description**	**Data Type**
Last name	Standardized Spelling of Creator's Last Name	Text
First Name	Standardized Spelling of Creator's First Name	Text
Alternate Spellings, Last Name	List of alternate spellings of a creator's last name, found in the BVRI	Text
Alternate Spellings, First Name	List of alternate spellings of a creator's first name, found in the BVRI	Text
Birth Date	Four-digit year creator was born, if known	Number
Flourished	Years when a composer was known to be active	Number
Death Date	Four-digit year creator died.	Number
*Creator Counter	Automatically generated number, serves as Primary Key	Integer
Company name	Name of company (if creator is a company). This primarily applies to performers.	Text
Original name	Name originally in cell (pre-edits). This is useful if a name is disputed or if data somehow becomes corrupted.	Text
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp

CREATORS TABLE		
Field Title	**Description**	**Data Type**
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

Publishers Table

The PUBLISHERS TABLE is designed to catalog any establishment that prints or sells music. While I contemplated keeping the two separate, several publishers also were involved in selling music. This table contains all people and institutions who might accomplish either task, and the join tables that draw upon this information delineate the roles of the establishment in each case.

Cataloging publishers presents several challenges, due to the consistently changing nature of the establishments involved. Several publishers change addresses with some frequency, and this creates challenges when it comes to describing whether a different address or a different name merits the creation of multiple publishers. Questions include:

1. To what degree are publishers distinct entities if they publish from the same address?

2. If one publisher lists two to four different addresses in the same ten-year span, should these be considered the same establishment in a spreadsheet?

3. If both the name and address are different, does this make the two different publishers, even if obvious relationships exist?

4. Does it make a difference if the business has changed hands from father to son?

These questions are some of the issues that have arisen in the course of compiling this volume. Perhaps the easiest way is to illustrate how this has been reconciled with the case of the Firth, Hall, Pond, and Ditson families.

John Firth came to the United States from England in 1810. Firth and William Hall served together in the War of 1812 and went to work for Edward Riley in New York. Firth set

out on his own as a printer in 1815, and Hall joined him in 1820. Firth and Hall worked together and were joined by Sylvanus Pond. In 1833, the company Firth, Hall, and Pond was formed, and this trio remained together until 1847. At that point William Hall and his son, John formed the company William Hall & Son. The company was dissolved in 1866.

John Firth, his son Thaddeus, and William A. Pond formed the company Firth, Pond, & Co. in 1847, after William Hall left to publish on his own. The company remained together until 1862, at which point it split again. William A. Pond formed the company William A. Pond & Co., which he ran with his son. The company ended in 1917.[52] John Firth continued his company with Thaddeus and his grandson, Edward, for another five years after the 1862 corporate split, and then the Firth publishing company was sold to Oliver Ditson.

This is the history of one publishing company and its affiliates. Table 3 lists relevant members of each family by generation, as well as their birth and death dates. Numerous works are published by Firth, Hall, Pond, or some combination of the three. The same company changed addresses multiple times, and often before or after a change in ownership, the prior address remained occupied, at least for a time. My solution to how to treat this network of publishers is as follows.

Table 3: Firth, Pond, Hall, and Ditson family members.

Firth Family	Pond Family	Hall Family	Ditson Family
John Firth (1789–1864)	Sylvanus B. Pond (1792–1871)	William Hall (1796–1874)	Oliver Ditson (1811–1888)
Thaddeus Firth (1821–1913)	William A. Pond (1824–1885)	John Hall (1823–?)	C.H. Ditson (Charles Healy) (1845–1929)
Edward Firth (1836–?)	William A. Pond, Jr. (1845–1884)		

---

[52] Christina Wilkinson, "The Musical Legacy of Maspeth Firth's Family," *The Juniper Berry,* March 6, 2020, https://junipercivic.com/juniper-berry/article/the-musical-legacy-of-maspeths-firth-family.

A primary goal of the database is to preserve granularity as effectively as possible. To this end, I catalog every instance of an address as a separate entry in the PUBLISHERS TABLE, and every instance of a different name is treated similarly. Thus, the BVRI contains 19 entries in the PUBLISHERS TABLE related to Firth, Hall, and Pond publishing firms. With this said, the BVRI also contains a generic category for publishers. The generic category of Firth, Hall, and Pond contains all three publishers and their respective companies. Oliver Ditson and his publishing company, which purchased Firth, is treated as the breaking point for this family of publishers. Firth, Hall, and Pond are listed as a related family for Ditson, and the converse is true, but they are viewed as distinct in the database.

In general, I have decided to treat each variance of a publishing company as worthy of its own unique row in the table. To keep the publishers from becoming mixed up, I use a reference number as my primary key. Each time a publisher changes its name or address, it receives a different reference number and be considered a distinct publisher in the table. The process of tracking changes requires a certain degree of granularity, and the only way that the BVRI can provide this granularity is if each change becomes its own entry in the table. Such detail can be useful for dating works based on where they are published and by whom.[53]

In addition to the publisher's name, the PUBLISHERS TABLE has two ways to aid in dating musical works. First, the table keeps track of both the earliest and the latest plate number associated with each entry. Just as plate numbers can be associated with years, plate numbers can also be associated with addresses. Additionally, the table tracks the earliest date known for the entry and the last date known for the entry.

---

[53] Anik Devriès and François Lesure, *Dictionnaire des éditeurs de musique française*, Vol. 2 (Geneva: Minkoff, 1979) provides a mechanism for dating nineteenth-century French publications by publishers' addresses and plate numbers.

Table 4: Presence of Firth, Hall, Pond, and Ditson across various cities from UNC Musical Works Table (*N* = 1168).

Pub ID#*	Publisher Name (as stated)	Location Unstated		Boston, MA		New York, NY		Publisher Total	
		n	% of *N*	*n*	% of *N*	*n*	% of *N*	*n*	% of *N*
79	C.H. Ditson & Co.					7	0.60	**7**	**0.60**
122	Ditson & Co.	1	0.09	1	0.09			**2**	**0.17**
167	Firth, Hall & Co.					2	0.17	**2**	**0.17**
168	Firth, Hall & Pond					39	3.34	**39**	**3.34**
169	Firth, Pond & Co.					169	14.47	**169**	**14.47**
170	Firth, Son & Co.					4	0.34	**4**	**0.34**
171	Firth & Hall	1	0.09			188	16.10	**189**	**16.18**
403	O. Ditson & Co.	1	0.09	1	0.09			**2**	**0.17**
409	Oliver Ditson	2	0.17	228	19.52			**230**	**19.69**
410	Oliver Ditson & Co.			245	20.98	1	0.09	**246**	**21.06**
412	Oliver Ditson & Company			1	0.09			**1**	**0.09**
413	Oliver Ditson Company			4	0.34	1	0.09	**5**	**0.43**
422	Parker & Ditson			11	0.94			**11**	**0.94**
495	W.A. Pond & Co.					4	0.34	**4**	**0.34**
523	William A. Pond & Co.					6	0.51	**6**	**0.51**
526	William Hall					1	0.09	**1**	**0.09**
527	William Hall & Son					171	14.64	**171**	**14.64**
535	Wm. A. Pond & Co.					62	5.31	**62**	**5.31**
542	Wm. Hall & Son					17	1.46	**17**	**1.46**
	Totals	**5**	**0.43**	**491**	**42.04**	**672**	**57.53**	**1168**	**100.00**

*Pub ID# = Last 3 digits of the publisher number.

This data was not entered in the first draft; instead, I used a report from the MUSICAL WORKS TABLE to work through the table a second time. Together, this provides context in dating works by various publishers.

Table 4 provides some detail on the Firth/Hall/Pond/Ditson family of publishers. There are 1,168 works that cite one of the four as the primary publisher (24.3% of total works from UNC), and the list shows what percentage of those 1,168 works are published under each name.. From this list, I want to highlight a few things that the database adds in terms of context. First, it is much easier to get an idea of the consumption of each publisher, and as the BVRI grows, consumption by area and over time. In this case, 35% of works list the Firth family as involved with publishing the work: Firth, Hall & Co., Firth, Hall & Pond, Firth, Pond & Co., Firth, Son & Co., and Firth & Hall would all qualify Firth as being involved. Similarly, 36% list Hall, 24% list Pond, and 43% list Ditson.

The BVRI can also offer insights into the dates for publications. Tables 5 to 8 and Figure 10 show the first and last dates for each publisher name, and this helps provide a functional range for how publishers are listing themselves on the title page at different points. William Hall is never listed working with Firth or Pond after 1847, but we do have instances of Firth and Pond publishing together as late as 1862. In this case, we know the Firth/Pond company disbanded at this time, but information about other, lesser-known publishers can be intuited this way as well. Data in this format is also much easier to visualize. Based on the UNC data, this is what publishing looks like for the Firth family of publishers, which makes it easier to visualize peaks and valleys in publishing activity. Again, as the BVRI grows in size and scope, this data can be nuanced for place and for time. This data can also be used to quantify and visualize the publishers being pursued in various areas, as well as tracking the popularity of their works.

Figure 10 presents the Firth family publishing by year in the UNC data. Each color highlights a

different publisher, and the graph presents bursts of publishing by year

Table 5: Firth family of publishers: First and last date for each.

Company: Firth	Publisher No.	First Yr Pub	Final Yr Pub
Firth & Hall	171	1823	1847
Firth, Hall & Co.	167	1831	1834
Firth, Pond & Co.	169	1837	1862
Firth, Hall & Pond	168	1840	1847
Firth, Son & Co.	170	1863	1865

Table 6: Hall family of publishers: First and last date for each.

Company: Hall	Publisher No.	First Yr Pub	Final Yr Pub
Firth & Hall	171	1823	1847
Firth, Hall & Co.	167	1831	1834
William Hall & Son	527	1838	1866
Firth, Hall & Pond	168	1840	1847
Wm. Hall & Son	542	1848	1867
William Hall	526	1856	1856

Table 7: Ditson family of publishers: First and last date for each company.

Company: Ditson	Publisher No.	First Yr Pub	Final Yr Pub
Parker & Ditson	422	1833	1842
Oliver Ditson	409	1835	1894
Oliver Ditson & Co.	410	1841	1886
C.H. Ditson & Co.	79	1869	1887
Ditson & Co.	122	1869	1869
O. Ditson & Co.	403	1860	1860
Oliver Ditson Company	413	1889	1913

Table 8: Pond family of publishers: First and last date for each.

Company: Pond	Publisher No.	First Yr Pub	Final Yr Pub
Firth, Hall & Pond	168	1840	1847
Firth, Pond & Co.	169	1837	1862
William A. Pond & Co.	523	1853	1869
Wm. A. Pond & Co.	535	1850	1892
W.A. Pond & Co.	495	1872	1880
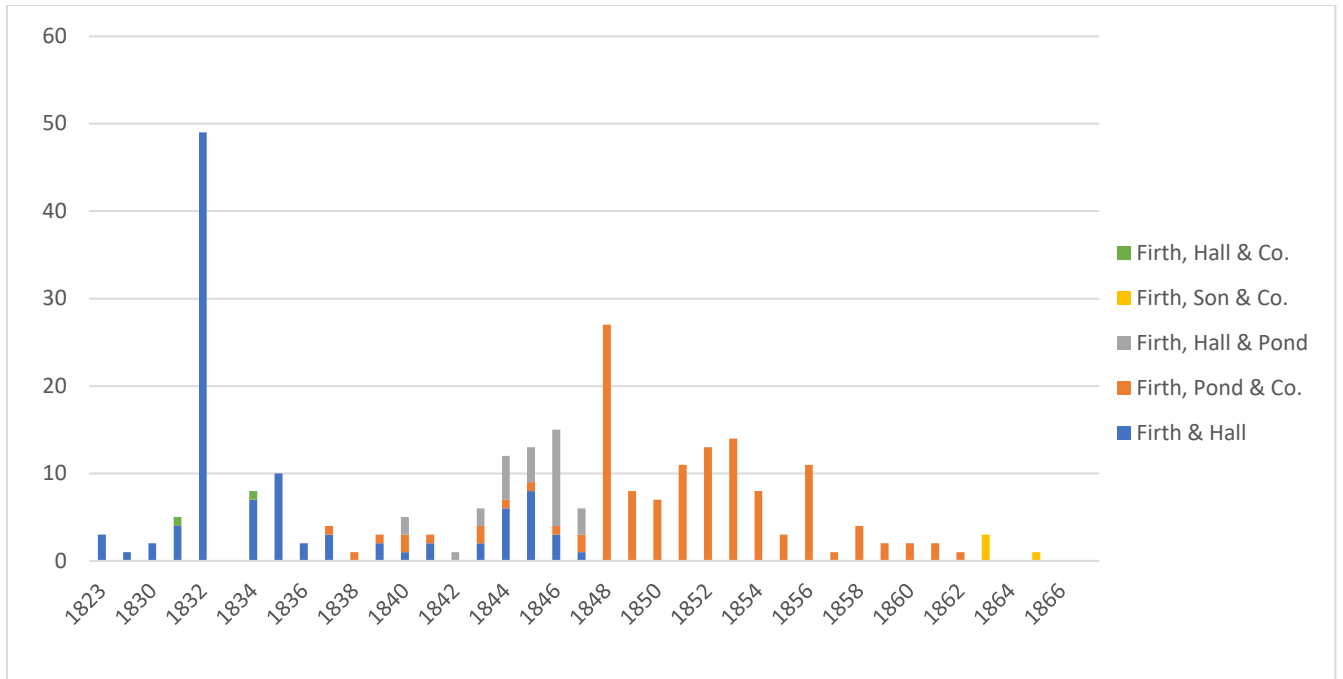
Figure 10: Firth family publishers showing activity of each publisher by year.

One additional factor that I did not anticipate but that came up during the creation of the BVRI is that publishers might list multiple cities on a single print run. This led to the creation of *Cities/States/Countries Fields 2* and *3*. These six additional fields more fully capture the reality that publishers listed multiple places on a single print to save money with the printing.

Cities and the known first and last plates are entries that take some time to maintain. Cities requires periodic reviews of the PUBLISHERS TABLE to ensure that multiple entries are not created by accident, either through manual entry or through large data imports from repositories. The cleaning process takes time, and while I have worked to ensure good data enters the project, the process is ongoing. Similarly, filling the first and last publishers' plates takes some review. Eventually, this will be done with a calculation, but the BVRI is not yet at the scope where such a calculation produces meaningful results. This is part of the upscaling process that I look forward to engaging with in future research.

One final note: the last five fields listed are the exact same as creators, and they

accomplish the same purpose. Every table has these entries to keep track of data as it is generated

and modified.

PUBLISHERS TABLE		
Field Title	**Description**	**Data Type**
Generic Publisher Name	Publisher name for general set of publishers. This name is broad and repeats across publishers to denote a family (Firth, Pond, etc.).	Text
Specific Publisher Name	Specific name of publisher. Given unique addresses, this name does not repeat.	Text
Address	Address for publisher	Text
City	First city listed for publisher, corresponds to address if given	Text
State	State or territory where city resides as of January 1, 2020 (if relevant or given), particularly in U.S.	Text
Country	Country (January 1, 2020 boundaries) where publisher resides	Text
City 2	Second city listed with publisher	Text
State 2	Second state/territory listed with publisher	Text
Country 2	Second country listed with publisher	Text
City 3	Third city listed with publisher	Text
State 3	Third state/territory listed with publisher	Text
Country 3	Third country listed with publisher	Text
Start Date	Earliest known date publisher at this address: 4-digit year.	Number
End Date	Last known date publisher at this address: 4-digit year.	Number
Plate number start	First plate number for publisher at this address, if known.	Text
Plate number end	Final plate number for publisher at this address, if known.	Text
Related Publishers	Generic name of related publishers.	Text
Related Publishers 2	Generic name of related publishers (if needed twice)	Text
*Publisher Counter	Automatically generated number, serves as Primary Key	Integer
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

Compilers

This is a simpler table designed to keep track of the people known to have contributed a binder's volume to one of the repositories that the BVRI catalogs. This table contains such information as birth date, death date, known locations (with the degree of specificity possible), and professions (if applicable). This table contains any people who may have compiled a binder's volume or who is known to be related to the compiler—husbands, mothers, daughters, cousins, etc. The relationships are not stated here; those are outlined in the VOLUMES TABLE.

COMPILERS TABLE		
Field Title	**Description**	**Data Type**
Last name	Standardized Spelling of Compiler's Last Name	Text
First Name	Standardized Spelling of Compiler's First Name	Text
Middle Name	Standardized Spelling of Compiler's Middle Name or Middle Initial	Text
Last Name Alternate	Piped List of Alternate Spellings of compiler's Last Name	Text
First Name Alternate	Piped List of Alternate Spellings of compiler's First Name	Text
Birth Date	Four-digit year compiler was born, if known	Number
Death Date	Four-digit year compiler died, if known.	Number
Address	Address for compiler, if known	Text
City	First city listed for compiler, corresponds to address if given	Text
State	State or territory where city resides as of January 1, 2020 (if relevant or given), particularly in U.S.	Text
Country	Country (January 1, 2020 boundaries) where compiler resides	Text
City 2	Second city listed with compiler	Text
State 2	Second state/territory listed with compiler	Text
Country 2	Second Country listed with compiler	Text
*Counter	Automatically generated number, serves as Primary Key	Integer
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text

COMPILERS TABLE		
Field Title	Description	Data Type
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

Musical Works

The MUSICAL WORKS TABLE—the most substantive component of this database—

constitutes a list of every musical work in a volume the BVRI contains. It draws upon the

CREATORS and PUBLISHERS TABLES to catalog people involved in composing, arranging, and

publishing the work, and houses additional information about each individual work, such as

instrumentation, plate numbers (if available), and other notes applicable to all copies of this

particular piece. Each work has been assigned a reference number, and unless the two can be

shown that they are likely the same work (publisher, composer, title), two entries remain distinct.

While this information is discussed in more detail below, it is worth stating briefly that

the date here is the date of the manifestation's creation, not of the item's use or consumption.

That date of use, if ascertainable, resides in the BV CONTENTS TABLE. Similarly, the location

contained here is a reference to the location where the manifestation was created; the location of

the binder's volume is contained in the BV MASTER TABLE. In the case that an individual piece of

sheet music is trackable to a place independent of the volume itself, that is kept in the BV

CONTENTS TABLE. Before going further, I must explain what I specifically mean with the terms

manifestation and item, and functional requirements for bibliographic records (FRBR)[54] provides

a useful framework to discuss these issues.

---

[54] At the 2018 AMS national meeting, I had the opportunity to workshop the idea of what it might mean to classify
music binder's volumes with a small group who, like me, had just attended a presentation on FRBR and databases.

FRBR, and What is a Work?

Perhaps one of the most significant challenges in creating a database of binder's volumes is discerning what the actual object is. This is a question that librarians, and especially music librarians, must consider carefully when classifying objects and labeling them in a database. The issue is especially vexing for binder's volumes. Compilers take one musical object, group it with several others, write on them and personalize the original object, then bind it with a set of other musical works to create a new object.

When contemplating these issues, librarians often use a concept model such as the functional requirements for bibliographic records (FRBR) to discern how works might be classified and recorded in a database. FRBR is divided into three groups, each of which relates to an artistic creation (literary, musical, or other) in a different way. The first group concerns manifestations of an artistic creation and how those are classified. The basic model is WEMI: work, expression, manifestation, and item. The second group deals with custodianship/ownership of an artistic creation. Group 3 concerns subjects of an artistic creation: concepts, objects, events, and places. The portion of FRBR that is relevant for this project is Group 1: how to classify and consider an artistic creation.

To understand the WEMI model, it is easiest to begin at the work level, which is the most abstract, and work towards an individual item. A work is the idea of a creation, perhaps thought of in the metaphysical or noumenal sense.[55] One such example would be the concept of

---

is discussion and thought exercise were illuminating for this project in many ways, which merit explanation. In late 2018, I was in the very early stages of contemplating what it might mean to work with binder's volumes. I had only looked at my first volume in the previous 45 days, and I had the idea that I wanted to understand these artifacts at a large scale if possible, and this meant classifying them. Up to that point, I had not really thought about what it meant to classify a musical work in a library database, and FRBR is designed for this very topic.

[55] Goehr, 89-90 and Bekiari, Chryssoula, Martin Doerr, Patrick Le Bœuf, and Pat Riva, editors. "Definition of FRBR$_{OO}$: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism." *International*

Beethoven's Fifth Symphony.[56] When someone mentions the piece in conversation, they usually mean the work in the abstract sense.[57]

Using the WEMI model in FRBR, the next level below the work is the expression of the work, which is a realization of the work.[58] Expressions represent the way that the work can be represented, which might include in score form, as a recording, as a performance, or other possibilities. For text, the expression might be an English or Spanish translation. Expressions, in this case, are not items or physical representations of a work. It is not the score for Beethoven's Fifth Symphony, or even an engraving. An expression is the idea that one way to encounter Beethoven's Fifth Symphony is to read the sheet music as a score. Similarly, an expression of the movie *How to Train Your Dragon* is the idea of watching the movie in English or Spanish, not the actual copy in English or Spanish. In this way, expressions serve as a link between the work and the manifestation of the work.

The manifestation of a work is "the physical embodiment of an expression of a work. As an entity, manifestation represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form."[59] Put more simply, a manifestation is a physical embodiment, of which many copies could or do exist. A single performance is a manifestation whether or not it is recorded. Leonard Bernstein conducting the London Philharmonic in a performance of Beethoven's Fifth Symphony is a manifestation of the

---

Federation of Library Associations and Institutions. Netherlands: IFLA, 2016. https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo_v_2.4.pdf. It is striking how closely these two concepts come close to aligning, and from quite different perspectives.

[56] Bekiari, 17.

[57] FRBR is a model for understanding the work from the present point of view. It does not attempt to claim values for people at the time.

[58] Bekiari, 18.

[59] Ibid.

symphony. A printing of Beethoven's Fifth Symphony is a manifestation of the symphony. The expression was the sound of the music generally or the notes generally written: one draft.

Items are individual, concrete objects. They are a single example of the manifestation.[60] Staying with Beethoven's Fifth Symphony, an individual score that can be checked out of the library is an item, as is a single copy of the recording. These distinctions matter when it comes to binder's volumes. The persistent question that has at times vexed librarians is "what is the item?" The individual musical works in a binder's volume are what traditionally count as items for librarians, but the volumes themselves are separately items. Any attempt to track these volumes and their contents must reckon with both, and the BVRI does so.

Musical Works vs. BV Contents

The MUSICAL WORKS TABLE plays a vital role in this process; it is an account of a manifestation of a printing. The fields in this section pertain to the creation of a manifestation, not an item. Contrariwise, the individual details of each individual item—things added to a manifestation or produced from the manifestation—are found in the BV CONTENTS TABLE. Determining which aspects of a work belong to the musical work inherently and which traits are added to a work can be especially challenging. The MUSICAL WORKS TABLE catalogs what is specific to the work that does not change. The title, the language of lyrics, and opus are all examples of things that do not change. Some factors that clearly change from item to item and are thus contained in the BV CONTENTS TABLE include order and page number in the volume itself, marginalia, and handwritten dedications. Inscriptions and dedications can be quite difficult to parse, and for the most part I have cross listed them between both tables. In a few instances,

---

[60] **Error! Hyperlink reference not valid.**Bekiari, 58.

such as when the dedication is clearly handwritten and personal to the compiler, I have removed the information from the MUSICAL WORKS TABLE, but most of the time this is difficult to tell, especially in bulk imports from metadata.

Two factors have proved particularly challenging to locate: price and seller information. Price is frequently printed on the music, which would seem logical to keep in this table; however, prices are also written on the work with pencil or pen. Because of this, I place price—whether printed on the piece or not—in the BV CONTENTS TABLE. Seller information is challenging largely because it can become tangled with publisher information. I list the first publisher here with MUSICAL WORKS, and document additional sellers and publishers in the BV CONTENTS TABLE.

It is worth taking a moment to discuss how the tables that feed into the MUSICAL WORKS TABLE are used, as this is the first example under consideration of how several tables have been joined to construct a larger table. Information from the CREATORS and PUBLISHERS TABLES factor into the musical work, and the relationships possible are extensive. Thus, I take each table in turn.

Creators in the Musical Works Table

The MUSICAL WORKS TABLE is the primary place where relationships between CREATORS and other data tables are established. The MUSICAL WORKS TABLE draws upon the CREATORS TABLE for several categories, and as this is a dissertation focused on music, composers are the natural place to start. The category of primary composer is designated for the person listed as the composer of a work. This is the person whom the publisher credits as the composer of the music, and I do not take steps to question or correct the publisher's accreditation outside of accounting for misspellings. For example, Ferdinand Beyer is listed as the composer on a work entitled

"Repertoire des jeunes pianists petites fantaisies composees par Ferd Beyer… No. 27, Rigoletto." Obviously, Beyer did not write Rigoletto, and a logical way for the publisher to list this would be with Verdi as the composer and Beyer as an arranger. With that said, if the publisher credits Beyer with composition, I do not challenge this statement. This is to eliminate questions about the degree to which a composition is an arrangement as opposed to a recomposition. When Bellini is credited by the publisher as the composer and an arranger is listed, that arranger is listed in the CREATORS TABLE and credited in the arranger's column. When Bellini is listed as an "influence" or is in some other way given secondary credit, he is listed as a secondary composer. The primary purpose of this category is to cite who the publisher designates as the composer, or whom the publisher gives primacy to as a composer of the work.

Many of the musical works that have made their way into binder's volumes are arrangements, and the *Arrangers* field draw upon the CREATORS TABLE to note these arrangers. The people cited in the *Arrangers* field are often listed as composers in other rows, and several useful queries arise from the database. First, the overlap between arrangers and composers requires queries to be well formed; that is, queries must consider whether they are interested in a creator's contributions to any part of a musical work, to works where they are listed as the arranger specifically, or whether to pursue works where they are likely the arranger, however the work is listed. This final category is best pursued by searching for a creator in the *Arranger* field or a creator in the *Composer* field if and only if a secondary composer is listed. Such a search requires a bit of proofreading, but the BVRI is designed for such queries (as seen above with the Beyer example). Second, it is useful to see how some people are designated by publishers, and if this marketing changes over time as the fame of an arranger/composer increases.

The category of *Secondary Composer* is one that is designated for acknowledgement

outside of the composer that the publisher acknowledges as primary. This category could be used for several purposes: (1) this could list the composer of a larger work (opera, song cycle) from which a song or aria is adapted or excerpted; (2) this could be the original composer if the person listed might be more properly thought of as an arranger, especially in the case of a piano work; (3) this person could be someone who actually collaborated on the music itself; and/or (4) the relationship of the primary composer to the secondary composer could lie outside of the scenarios listed above. The *Secondary Composer* populates from the CREATORS TABLE and contains a field that identifies how the secondary composer relates to the musical work in question. The relationship is exactly as identified above: Larger Work, Actual Composer, Collaborator, or Other. In the case of the Larger Work, a third field contains the name of the larger/related work.

The field of *Tertiary Composer* functions very much like the *Secondary Composer* field but is likely to be used much less frequently. It requires that a third person have one of the aforementioned relationships to the work. In general, this database strives to order composers as follows. First, the primary composer is the one given primacy on the page. Next, the secondary composer is the actual composer of the work. If the primary composer is believed to be the actual composer, the secondary composer a person who collaborated on the composition. If no one fits that description, the secondary composer is the composer of the Larger Work. If there is no larger/related work and other aforementioned criteria have no occupant, the secondary composer is a person who has some other relationship to the composition. The tertiary composer is the next person in line should the secondary composer already be filled. If a person is an arranger or lyricist, they are listed in the appropriate field, not as a secondary or tertiary composer.

When originally designing the metadata schema, three composers seemed like a sufficient

number. It was not; some compositions had as many as six composers listed, and I preserve each composer in their own field. Should more than six composers be listed in a work as the BVRI expands, I will expand the number of composers included. Designations about roles for each composer are primarily limited to the first and second composers, and occasionally the third composer listed on a work.

Performers are another field that factor into the MUSICAL WORKS TABLE. Many pieces in the 1800s were advertised by the person who performed the work, not just the person who composed it. This is not surprising and remains a common practice today. Because performers and composers overlap, both are kept in the CREATORS TABLE, even though performers are not typically thought of as participating in the creation of a musical work. This allows the BVRI to store all the data on people in one place and display it in multiple places when called upon by the MUSICAL WORKS TABLE or the BV CONTENTS TABLE. In building the BVRI, some works listed as many as five performers. Because the performers are an area I want to investigate further, I parsed performers into several fields. The *Performers* fields indicate each of the performers (referencing its own *Creator Counter*), the method of performance (sung, played, sung with great applause), and how the original performed by claim was worded for ease of retracing.

The final fields that use CREATORS as foreign keys are *Primary Lyricist* and *Secondary Lyricists*. Many composers wrote both the words and lyrics to their musical works. Instead of keeping track of the two separately, the BVRI draws upon the CREATORS TABLE for lyricists as well. Some lyricists listed in the CREATORS TABLE have never composed, which poses no problem to the database. Secondary lyricists are used in the case of two people contributing lyrics, or in the case of one person contributing lyrics and another translating. As with composers, whoever is credited by the publisher as lyricist, or listed first by the publisher, is

considered primary lyricist in the database. It is the BVRI's practice that the translator be listed as the secondary lyricist should a primary lyricist exist in the absence of clear accreditation from a publisher.

Several fields are related to the CREATORS TABLE but do not draw upon the table directly. A few examples have been mentioned above, and the list includes opus number of the work, relationships between secondary composers and the work, whether a lyricist is serving as a translator, the language of the work (which relates to the lyricist/translator), and the instrumentation chosen by the arranger. Each of these can be queried in the BVRI, and the associations are easily seen through the relational database.

Publishers in the Musical Works Table

Publishers, like creators, fill in large sections of each row in the MUSICAL WORKS TABLE. Often, multiple publishers are listed on a single musical work. I currently have three fields for publishers listed, and that number may expand in the future as required by the works themselves. The publishers are listed in the order of primacy as indicated on the work (top to bottom, left to right), though no inference should be drawn past the first listed publisher. Sellers, though drawn from the PUBLISHERS TABLE, do not factor into the MUSICAL WORKS TABLE; instead, they are accounted for in the BV CONTENTS TABLE.

As with CREATORS, several fields in the MUSICAL WORKS TABLE closely relate to the PUBLISHERS TABLE that do not directly draw on it as a foreign key. This is to say that some fields in the MUSICAL WORKS TABLE, such as *Plate Number, Number of Pages, Date, And Margin of Error,* become much more meaningful when the publisher's information can be pulled from the PUBLISHER'S TABLE and placed alongside the musical work in context. The plate numbers are useful to track. Several sources give plate number ranges and dates, which are very useful in

tracking down the precise date of a piece when a lack of information or other factors make the task much more difficult.

To understand the dating system used in the BVRI, a bit of context in field types is required. Databases have several field types, most of which I have avoided discussing here as the type of field chosen for a field is often self-evident. The default is to use the smallest possible field type whenever possible. Thus, when a statement is true or false, yes or no, or some other binary, a Boolean-type field is appropriate. If a Boolean is not possible, a number-type field is preferred. This can include integer fields or number fields more broadly, but in either case, no alphabetical or other characters besides numbers and decimals can be used. number fields should be used whenever numbers are the exclusive entry in the field. Text fields contain all alphanumeric data that falls outside of these ranges.

The reason for seeking the most precise field possible is that it facilitates more precise searches and queries in the BVRI and allows for more extensive data processing and manipulation. Numbers offer a wide variety of options as a field type but require precision in the type of field used. The case of dates is an instructive example. Dates, if input into the BVRI as numbers, can be used for sorting and numerous types of statistical analyses. It is useful to know several indicators, including the range of dates at a site, the mean and standard deviation of a set of binders' volumes, the median and quartiles for a set, and such specific questions as the mean and median age for a binder's volume containing a particular musical work. Such specific data analysis is only possible with number fields. This is a problem for many publications, because the dates are given as ranges or as approximations. Sometimes this approximation is by decade, and other times it is simply a range that is given. It is important to capture the degree of precision with which dates are evaluated; this presents an essential caveat to any inferences drawn from

the BVRI; however, this cannot be done at the expense of making dates into text.

The solution I use is to make input the date given as a number field, but to also input a margin of error for each date as a separate number field. When a precise date is known, such as 1843, I put 1843 into the *Date* field and 0 into the *Margin of Error* field. When a date is given as the range [184-], 1844.5 is the date and 4.5 is the margin of error. This allows for a work to be published in any of the years 1840–1849. Any other range of dates are treated the same way, with the average of the range being listed as the date, and the margin of error calibrated to incorporate the range. At times, other factors allow for greater specificity in dating. Plate numbers are one example where reasonable inferences can be made into the year a document is published. In these cases, the date and range is adjusted accordingly.[61]

The final thing to note on the dating musical works is that this field is interested in the earliest date of a work published. For example, Oliver Ditson acquired Firth's plates in 1867 and began to publish with the acquired plates. A work that contains Ditson's and Firth's publishing information and Firth's plate numbers is cataloged here under the earlier date. The *terminus post quem* of when the document might have been purchased, inferred here as after 1867 and validated in the data thus far, would be included in the BV CONTENTS TABLE. The MUSICAL WORKS TABLE is concerned with the creation of the standard form of the document, which I take to be the plate itself. The BV CONTENTS TABLE is concerned with individual documents, which here should be understood as the copy sold to a person.

Other Information in the Musical Works Table

In addition to drawing on the aforementioned tables, this table contains information on a

---

[61] The calendar only deals in years, not in months or days. Also, slight dating changes by country, such as holdouts not using the Julian Calendar, are not corrected in the data entry.

common title of a musical work. This is probably best thought of as listing the work itself in the WEMI model. While the table is designed to track manifestations, documenting the work, and drawing these relationships is an important part of the project, and it is important that this information does not become buried. Thus, the common work title documents a work or expression-level title, and the specific work title tracks the title on the printed page. The reference number is the unique identifier that separates to instances of the same title, should other factors differ. Once again, the five auto-generated fields at the end are useful for general upkeep in the database, especially as I expand this project into a more broadly usable format.

MUSICAL WORKS TABLE		
Field Title	**Description**	**Data Type**
*Musical Works Counter	Automatically generated number, serves as Primary Key	Integer
Specific Work Title	Statement of work title as printed.	Text
Common Work Title	Commonly known title of work.	Text
Caption Title	Caption listed on work	Text
Other Titles	Titles listed as other on the work; Second Title	Text
Title 3	Third Title listed, originally part of main title	Text
Title 4	Fourth Title listed, originally part of main title	Text
Title 5	Fifth Title listed, originally part of main title	Text
Title 6	Sixth Title listed, originally part of main title	Text
Title 7	Seventh Title listed, originally part of main title	Text
**Composer	First Composer listed on sheet music (SK, Creators)	Integer
Opus	Opus and work number of work	Text
**Composer: 2nd	Second Composer listed on work or composer of larger work (SK, Creators)	Integer
Relationship to work (2nd)	Statement of how 2nd composer is related to work	Text
Larger Work: Related Work	Name of work that is a related to specific work at hand.	Text
**Composer: 3rd	Third Composer listed on work or composer of larger work (SK, Creators)	Integer

MUSICAL WORKS TABLE		
Field Title	**Description**	**Data Type**
Relationship to work: 3rd	Statement of how 3rd composer is related to work	Text
**Composer: 4th	Third Composer listed on work or composer of larger work (SK, Creators)	Integer
**Composer: 5th	Third Composer listed on work or composer of larger work (SK, Creators)	Integer
**Composer: 6th	Third Composer listed on work or composer of larger work (SK, Creators)	Integer
**Lyricist: 1st	First Lyricist listed (SK, Creators)	Integer
Translator: 1st	Is the first lyricist a translator?	Boolean
**Lyricist: 2nd	Second Lyricist listed (SK, Creators)	Integer
Translator: 2nd	Is the second lyricist a translator?	Boolean
Language	Language or Languages listed in performance (Listed in English)	Text
**Lithographer	Lithographer listed (SK, Creators)	Integer
Lithograph Design	Short description of lithograph	Text
**Lithographer 2	Second Lithographer listed (SK, Creators)	Integer
**Lithographer 3	Lithographer listed (SK, Creators)	Integer
**Performer 1	Performer advertised with piece (SK, Creators)	Integer
**Performer 2	Second Performer advertised with piece (SK, Creators)	Integer
**Performer 3	Third Performer advertised with piece (SK, Creators)	Integer
**Performer 4	Fourth Performer advertised with piece (SK, Creators)	Integer
**Performer 5	Fifth Performer advertised with piece (SK, Creators)	Integer
Performer Method	Description of how the performer executed the performance	Text
Performer Full	Full text of the performer description	Text
Performer By Long	Full Text String that captures performers contributing to the work (for later disambiguation if needed)	Text
Performer By Short	Shortened Text String that captures performers contributing to the work (for later disambiguation if needed)	Text
Performer Also By Short	Text String that captures performers contributing to the work (for later disambiguation if needed)	Text
**Arranged by	Arranger of this particular edition, if specifically listed as arranger (SK, Creators)	Integer
Instrumentation	Description of voice parts and overall instrumentation	Text

MUSICAL WORKS TABLE		
Field Title	**Description**	**Data Type**
**Publisher: 1st	First publisher listed on work (SK, Publishers)	Integer
Plate numbers Publishers Numbering	Plate numbers (listed as range)	Text
Date Published	Date published, or average of range of dates published	Number
Date Margin of Error	Number of years that the date might be off by	Number
Number of Pages	Number of pages, including title page	Number
**Publisher :2nd	Second publisher listed on work (SK, Publishers)	Integer
**Publisher: 3rd	Third publisher listed on work (SK, Publishers)	Integer
Location Listed	Location listed for the work publisher	Text
First Line of Text	First line of text	Text
First Line of Chorus	First line of chorus	Text
Edition	Edition of Work	Text
Topic	Topics listed as associated with work	Text
Inscription	Printed Inscription on Work	Text
Full Title	Full Title as Listed	Text
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

BV [Binder's Volume] Master Table

This table contains everything that the BVRI knows about the binder's volumes themselves. In addition to the compiler and date compiled (if known), it also has information about whether this is part of a known set, the color, size, and other information about the actual volumes. Here relevant marginalia are also documented. The date of the earliest and latest work in the volume is not included in this table; that information is more easily ascertained by query.

One particular note that came up during the creation of the database is the question of how to designate the institution. The original draft of BVRI used a call number, but it became apparent that not all volumes are listed as separate call numbers. To rectify this, a broader category of institutional designation now controls for how any institution might denote a particular volume.

To create this table, the database draws upon the COMPILERS TABLE. The locations are used to pinpoint as precisely as possible the volumes, so as to facilitate searches that link a location with a publisher, as well as facilitating the potential for geolocation and other types of graphs. The COMPILERS TABLE is a natural table to feed into this one, as the BV MASTER TABLE serves to provide a level of granularity to the compilers, just as the BV CONTENTS TABLE fleshes out the detail contained here in the BV MASTER TABLE.

BV [BINDER'S VOLUME] MASTER TABLE		
Field Title	**Description**	**Data Type**
Institution	Current Location of BV	Text
Call Number	Call number at Institution	Text
Institution Designation	Number or way institution denotes particular volume	Text
Series Collection	Name of Series or Collection the volume belongs to	Text
**Compiler	Compiler of BV (SK, Compilers)	Text
Date Bound	Date volume was bound, or average of date range that volume was bound (if known).	Number
Date Margin of Error	Margin of error for date bound	Number
**Location Bound	Location Volume was bound, if known (SK, Locations)	Integer
Color	Main color of the outside of the binder |colors of bindings	Text
Physical Description	Description of the outside of the volume	Text
Notes	Relevant information from the BV (sections delimited by vertical pipe)	Text
*Counter	Automatically generated number, serves as Primary Key	Integer
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text

BV [BINDER'S VOLUME] MASTER TABLE		
Field Title	**Description**	**Data Type**
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

BV [Binder's Volume] Contents Table

This is the table that serves as the end goal of the BVRI. All of the information and tables described up to this point have been created to best display the information contained here. The goal for this table is to contain everything relevant at the item level of FRBR (functional requirements for bibliographic records); that is, one particular instance of a work in one particular volume. In addition to data unique to this table, the BV CONTENTS TABLE is populated by information from all of the other tables, either directly or indirectly.

The MUSICAL WORKS TABLE provides information on the manifestation of the musical work, of which the BV CONTENTS TABLE documents one item. The *Musical Works Counter* reference number calls forward all of the information related to a manifestation, or of an expression of a work (documented in the common title). Some data are listed in both the MUSICAL WORKS TABLE and the BV CONTENTS TABLE, but the data should be different, as each documents a separate set of information.

The CREATORS TABLE is simplest here. Because the MUSICAL WORKS TABLE contains all of the relevant information on creators, none is needed here. It can be accessed and displayed here through relationships, but that data is not stored or input at this level of the database.

The PUBLISHERS TABLE is used here and in the MUSICAL WORKS TABLE, but in different ways. The publishers referenced in the MUSICAL WORKS TABLE are the ones that created

manifestation, or print run, of the object. Those listed here are the people who sold the object, if known to be distinct than the creator. Sellers stamps may or may not play a significant role in the database as it grows, but where the seller's information is available, it is cataloged.

Marginalia and other alterations to the work are noted and kept in this table. These are the types of changes that make one item unique from another, and that information is distinctly worth preserving in this process. If the cataloger left specific notes on an item, that information is also documented in this table. Essentially, everything that makes the item unique from the manifestation set is important here and documented here.

Just as the MUSICAL WORKS TABLE populates a large section of this table, so too does the BV MASTER TABLE. This makes information on the volume readily obtainable in the BV CONTENTS TABLE. In addition, this table contains some information as to how the individual work relates to the larger binder's volume. The order of the works in the volume is especially notable: the database can be queried for which works tend to be bound toward the front, but at the outset this is not a primary focus of the project.

Dates and *Locations* are two fields that have complicated relationships to both the MUSICAL WORKS TABLE and the BV MASTER TABLE. The MUSICAL WORKS TABLE is the most likely to have a date and location, and the majority of manifestations should have both, even if the date is imprecise. The BV MASTER TABLE likewise has, if nothing else, a *terminus ante quem* based on the last published work, and some dates are much more precise on the binding of a volume. Locations may be more difficult, but I hope to have several as I research provenance of several binder's volumes. The BV CONTENTS TABLE has no need to input dates present in these other two tables unless there is reason to believe that the date for a specific item in the volume contributes information distinct from both other tables that populate it. That is, unless the date for acquiring

an item in the BV can be shown to be distinct from both the manifestation's creation and the binding of the volume, it need not be included in this table. The same is true of locations.

There are several fields generated that are useful for referencing the original item. Some repositories have their items archived and available online, such as North Carolina. The reference URL can get a user back to the original work, provided the URL is still active and unchanged. The OCLC and Contentdm numbers provide additional means of locating sources, especially as the digital landscape is constantly changing.

Some fields are initially cross listed. *Inscriptions, Dedications,* and *Prices* are all fields also listed with the MUSICAL WORKS TABLE. A later stage of this project will disentangle the fields, but for now I decided that having the information available twice was better than losing or improperly assigning the information. Should this decision be untenable in the future, I have backups that can easily facilitate a rewrite of the data. Once again, the final five fields listed are for noting revision history with an eye toward project expansion.

BV [BINDER'S VOLUME] CONTENTS TABLE		
Field Title	**Description**	**Data Type**
**Work Information	Description of manifestation of musical work (SK, Musical Works)	Integer
**BV Master	Binder's Volume containing the work (SK, BV Master)	Integer
**Seller	Person who sold work to compiler, if known (SK, Publishers)	Integer
Location Address	Address for seller, if known	Text
City	First city listed for seller, corresponds to address if given	Text
State	State or territory where seller resides as of January 1, 2020 (if relevant or given), particularly in U.S.	Text
Country	Country (January 1, 2020 boundaries) where seller resides	Text
Order in Binder	Number in binder	Number
Notes	Relevant information from the work (sections delimited by vertical pipe)	Text
*Counter	Automatically generated number, serves as Primary Key	Integer

BV [Binder's Volume] Contents Table		
Field Title	**Description**	**Data Type**
Content dm number	An in-house repository code that I track when applicable	Number
Previous owner	Information on the previous owner, if available	Text
Marginalia	Notes on marginalia contained within a piece.	Text
Dedication	The dedication and/or dedicatee of the work	Text
Inscription	Inscription on the work	Text
Price	Price of the work, if given.	Text
Defects	Defects noted on the work.	Text
Reference URL	URL that links to where the work is hosted, if given.	Text
Notes 2	Additional notes on the work.	Text
OCLC Number	WorldCat Reference number.	Text
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

Conclusion

This is the structure of the BVRI, and the rationale for what data are located in the

metadata schema. The next chapter details the process for getting data into the schema outlined

here, and Chapters 4 and 5 show how all of this data functions in a music historiographical way.

Chapter 4 provides a broad analysis of some questions I engage with and demonstrates how

statistical analysis is useful for answering these large questions. Chapter 5 then turns to the

details, and how the BVRI can provide context to two case studies: (1) the consumption of a

musical work "Home Sweet Home," and (2) an analysis of a single binder's volume.

CHAPTER 3

DATA CLEANING METHODS AND CHALLENGES

This chapter is an account of how I cleaned and organized the data for the database. While it does not contain every step, particular emphasis is placed on the processes used in cleaning UNC's machine-readable cataloging (MARC) records. I hope to accomplish two things in the chapter. First, by documenting my methods, this chapter provides clarity and transparency to what would otherwise be a murky process. For people with extensive coding or data science experience, this chapter might be tedious. For humanists interested in the digital humanities but without any foundation in coding or cleaning data, this chapter should serve as an approachable account of how and why I modified a curated set of records. This leads me to the second use for this chapter: ideas for others. I made mistakes along the way, not all of which are listed here. With that caveat, I also aim to highlight the major pitfalls and inefficiencies in my process for others who might wish to recreate the process more quickly.

Most of the data for the BVRI comes from existing records. Strategies for importing and cleaning data are therefore essential for the viability of the project going forward. The term MARC record consists of two parts: machine-readable means that a computer can read the indicators and decipher the information, and a cataloging record indicates that the record contains bibliographic information on an item in the library.[62] At this point, the MARC system is over 50 years old, and MARC 21 (the current standard) has been in use since 1987. To be clear, MARC is not the only system from which the BVRI can incorporate data, but it is certainly the most common one. As this dissertation was being written, Andrea Cawelti has worked on a

---

[62] "What Is a Marc Record, and Why Is It Important?" *Library of Congress* information booklet, 2009, https://www.loc.gov/marc/umb/um01to06.html.

system for how to standardize binder's volumes' cataloging practices, and her work is forthcoming. Its publication should standardize the way that this project incorporates data from binder's volumes in the future, but data need not be in MARC format to be incorporated; it simply must be consistent.

The question remains: why import any data? Why not just catalog the data volume by volume? This question is essential to the methodology, and in general taking data from catalogs made by others is not the ideal way to incorporate data for a project like this. However, in the short term, it is the most realistic way to expand the reach of the BVRI, and to ensure that documents become centralized and public as quickly as possible. This is a tradeoff: speed for personally assuring that each entry is valid. However, the remainder of this chapter is aimed at describing how I manage that tradeoff, and ultimately why I believe the trade is worthwhile. My methods assure, moreover, that any modifications to imported data were not only brought to my attention as part of the process, but also finalized and incorporated into the project deliberately.

This does mean that my dissertation relies on, and builds upon, previous expertise. There is inherent risk here, but the opportunity cost of significant time and limited scope seems to be much greater than the risk. I also want to note that when in doubt, I return to the source. For example, clustering is discussed in the chapter below as a process where large numbers of titles are grouped by similar titles. This similarity could be based on characters, on phonemes, or other clustering methods, and for each one I made a point to check the source if I was going to change the work of another, or if I suspected something inaccurate. This is one reason that I have confidence in the outcomes: data is dealt with in aggregate, but whenever needed I return to the source as the basis for making editorial decisions on cataloging.

Throughout the chapter I refer to this process of grouping and verifying data for the

database as *data cleaning*. When I use the phrase data cleaning, I mean the process by which I go through, check data for accuracy, and convert it from its existing format (MARC records, but also other formats) into the format that I use for the BVRI. Along the way, there are several ways in which I confirm that the data does correlate with past expectations, that I am not entering in a composer multiple times if that can be avoided, and how my information does maintain integrity when it comes from another source. A large part of the data cleaning process is lost periods and commas, but it also includes clustering and other methods for checking past names against new ones.

<div align="center">Observing Patterns in the Data</div>

The first step in cleaning a data set is to spend some time looking for patterns in how it is organized within a field. The example in Figure 11 includes several composers listed in the UNC MARC data. The list represents a sampling of how this field in fact contains information for several different subfields and how those implicit fields are separated in the data.

Hine, James, composer.
Abel, J. L.
Donizetti, Gaetano, 1797-1848.
Shrival, Robert G., b. 1815 or 16.
Rodwell, G. Herbert (George Herbert), 1800-1852.
Braham, John, 1774-1856. ; Nathan, Isaac, 1792-1864.

Figure 11: List of six different composer fields taken from UNC data.

At first, the data seems straightforward, but there are several things to note while sorting and cleaning. First, several types of punctuation delimit the various categories in the example. There are periods, commas, hyphens, parentheses, and semicolons. Each one serves a particular function, but some are used in multiple places. Semicolons reliably delimit multiple composers in a cell, as seen in bottom row (Cell 6) in Figure 11. Parentheses tend to denote either full first

names beyond the printed name or a nickname that might be more recognizable for the creator in question, and they do not seem to be present in other places. Hyphens serve as the space for date ranges, but this can be problematic. Sometimes date ranges are for the entire life of a composer, but in some rare instances date ranges indicate the time that the composer was actively working (flourished, or fl.).

Periods are the most problematic of the punctuation above due to the variety of ways in which they are used. Periods are present in many places. As seen above, many composers are given with an initial. This might be the first initial (G. Herbert), the second (Robert G.), both initials (J. L., and the space is intermittently present), or none at all. There is also a single period at the end of every listing. This could be helpful, but there is only a single period at the end of J.L., and this period serves the function of both abbreviating the L and terminating the entry. Because dates tend to be present on many composers, the period is unreliable to keep attached to a composer's first or last name. It also cannot be removed wholesale, as J. L. would become J. L and the entries would be inconsistent.

Periods also make their way into dates. When a composer's full life is not known, the information is frequently given with a born (b.), died (d.), or the aforementioned flourished (fl.). This can leave some information tagged onto a different section of the composer's dates that needs to be removed before proceeding.

Commas are also challenging for similar reasons. They do separate subfields (last names, first names, dates), but they also are included in some composers' actual names and titles. A common example is "James, composer," but others include "Cleobury, F., Mrs.." Titles like Baron or Esquire are also frequently (but not always) separated by comma. This also makes delimiting by commas challenging while the dates and other sections are attached.

In summary, the inconsistencies in data imported from MARC records could, if left unmodified, easily undermine the stated purpose of the database to provide a means for quantitative historical analysis. Non-standard punctuation in particular poses a challenge, as it separates records that ought to be grouped together. Such divisions of records would make meaningful queries more difficult, as the user would need to account for possible variations in cataloging. Fortunately, a combination of patterns in the data and tools for cleaning it can generate a more consistent set of records.

Part 1: The Tools

To overcome these obstacles when cleaning data, I used three programs: OpenRefine, Microsoft Excel, and FileMaker Pro. Most of the work took place in OpenRefine, but I also moved files between Microsoft Excel and OpenRefine, and between Microsoft Excel and FileMaker. Microsoft Excel is more efficient when viewing larger swaths of the data set and making individual edits. OpenRefine has incredible search and cluster functionality, and it can allow searches of specific characters and strings. It also is useful for editing data sets on a large scale, making it particularly useful on this trial data set with just over 5,000 records before extracting individual subsets of information. Ultimately, the cleaned data resides in FileMaker, where the related tables allow for flexibility in viewing and analyzing data sets. This section discusses important functions that I used within each program to clean data, so that the processes described later in the chapter make sense.

OpenRefine Common Functions and Actions

Per the OpenRefine website, "OpenRefine (previously Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and

extending it with web services and external data."[63] Metadata coming from MARC records is often inconsistent. OpenRefine is a powerful tool for parsing this data, allowing it to be organized more quickly than a record-by-record editorial approach. Because this program is not familiar among musicologists, it is worth taking some time to explain some of the most important functions how they work. Readers already familiar with OpenRefine and GREL can skip this section.
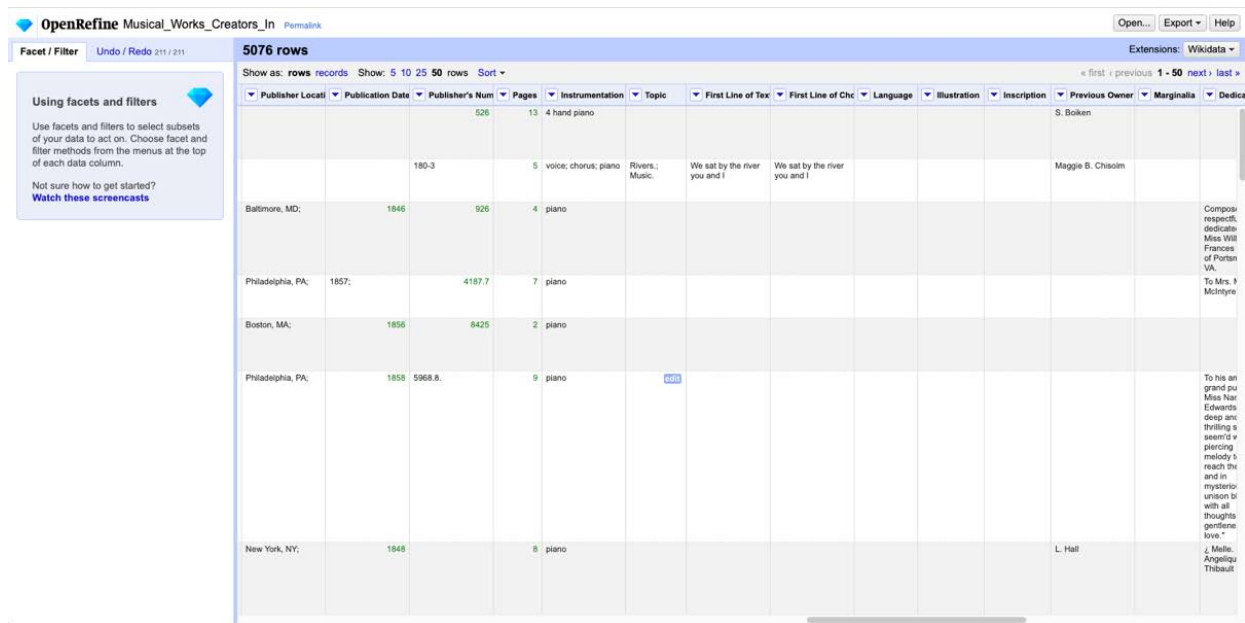
Figure 12: OpenRefine Layout.

Data in OpenRefine is presented as a spreadsheet. All columns are visible by scrolling horizontally, but only 10 rows are initially shown (though this can be toggled between 5 and 50). Figure 12 shows an excerpt from the MUSICAL WORKS TABLE. Numbers are formatted in green text, and general text is in black. While it is possible to scroll through rows using the next/last in the upper right-hand corner, this is not as convenient for reviewing and editing data as Microsoft

---

[63] The home page of the OpenRefine website provides context, uses and a user guide for the application. https://openrefine.org.

Excel. OpenRefine does not seek this functionality; instead, the power of OpenRefine is to edit data en masse across a column or row.

Facets and Text Filters

OpenRefine exists for sorting data, and the primary way for sorting and viewing the data is facets. As per the OpenRefine user manual:

> Faceting allows you to look for patterns and trends. Facets are essentially aspects or angles of data variance in a given column. For example, if you had survey data where respondents indicated one of five responses from "Strongly agree" to "Strongly disagree," those five responses make up a text facet, showing how many people selected each option.[64]

Facets are the system OpenRefine uses for grouping data. A text facet groups the data in a particular cell, either by occurrences of a string of text, meaning that each entry of a cell is counted and tallied. The example above, with five entries, would be one such occurrence, but this same principle works for True and False and for other typed entries. Text facets can also make outliers appear more obvious in skimming. A data set where users type "True" and "False" will have caps mistakes and spelling mistakes. Facets will immediately highlight misspellings of True and False, allowing for quick mass corrections. Facets also work for numbers, showing the highlights, range, and all numbers. But facets can be used to find other cells and data types. The dropdown arrow by each column name gives options for implementing text filters and facets. Each of these options provides a way to sort data and look at specific rows. Numeric facets filter by range, text facets list all the text options in a cell, and timeline facets organize the data by dates, or a date range. Other facets can find useful information such as blanks and words in a cell within a particular column. The most common facets used in this project were text facets,

---

[64] "Exploring Facets," *OpenRefine Documentation: User Manual and Technical Reference,* July 2022.

numeric facets, duplicate facets, and faceting by blank. Figure 13 shows the lists of facets and

filters that are available in the OpenRefine drop down menu.
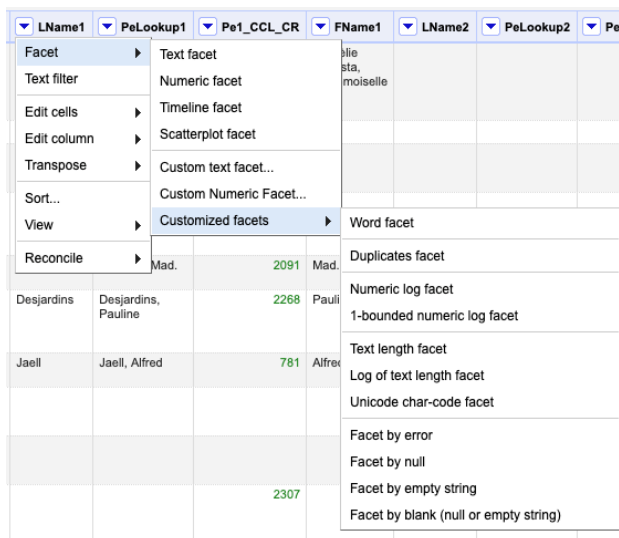
Figure 13: Menu listing types of facets.

Text filters are another way to search data, but where facets are organized as selectable

options from a list, text filters present an active search box to narrow the records shown. Most of

the text filters I used were straightforward searches for strings of characters (the search

disregards capitalization). Because OpenRefine allows for the text currently found in a search to

be edited at once, it is easy to search for a specific word or phrase (or use multiple searches),

then to make edits using some of the functions listed below. For example, several times when

using a formula to fill a column, I had a formula fail to fill every cell, and I wanted to understand

why. I can use the filter or facet feature to search for blank cells, then see what each cell has in

common. Once I find a problem, I can then redo the formula without worrying about changing

any of my existing data (it is already filtered or faceted out of the current workspace), look to see

what worked by using a text filter to see which cells are still blank, refilter, and continue to

pursue problems this way. Learning the best ways to write formulas on a data set of thousands

saved time.

There is a second way to search, and the two boxes in the bottom left of Figure 14 show how to do so. The first box inside the orange oval is for case sensitive filtering, but the second box—"Regular Expressions"—does more. If that box is checked, OpenRefine will parse detailed instructions on where to look for specific characters, to exclude characters, or other instructions. For example, checking the regular expression box and typing in the three characters \.$ as an expression tells OpenRefine to look for a period at the end of a string.[65] This can help when cleaning data that should only have a period when it is the last initial of a person, or when it is an abbreviation (Mr.). Footnote 66 contains a link to more information on regular expressions.
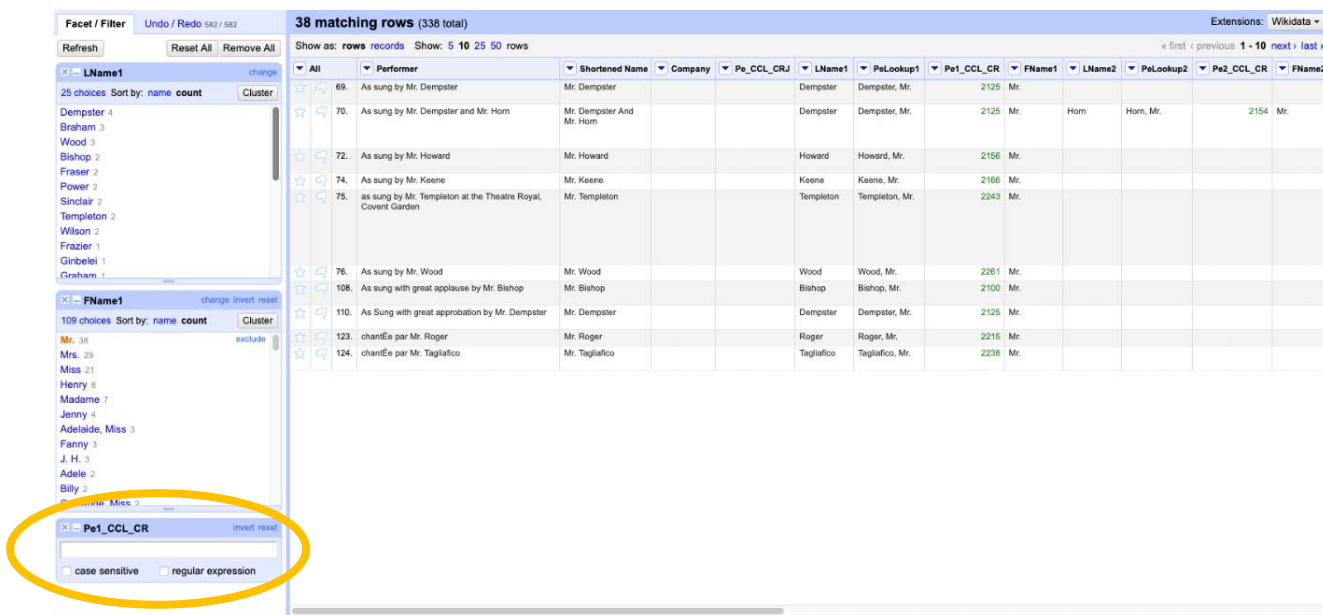
Figure 14: Example showing multiple facets at the same time.

Facets and text filters can work together. Figure 14 shows the facet/filter window on the left side of the screen. The text facet boxes for Last Name 1 (LName1) and First Name 1 (FName1) are open, and the text filter box is open for the column titled Pe1_CCL_CR (the

---

[65] http://regex.bastardsbook.com The description here of regular expressions (regexes) as search and replace on steroids is apt. The more time I spent using OpenRefine, the more I realized that regular expressions could expedite the data sorting more than I did in the initial process. This is a change for the future.

creator counter lookup for performers from the CREATORS TABLE).[66] Using these search

windows, it is possible to view only first names that are Mr. (the current setting), last names that

are Braham (which would reduce the list to 3), and Pe1_CCL_CR values that contain a 1. When

a set of texts is faceted, transformations only act on those rows. Thus, splitting these columns by

," " the action would only happen to the 38 rows, not to all 338 in the OpenRefine file. This is

another example of the power of facets: by restricting the number of entries in the workspace

through facets and filters, mass edit tools can be precisely applied to particular cases.

Undo

OpenRefine tracks each step taken by a user with an extensive degree of detail. In the file

shown (the Performers_Lookup file), OpenRefine has tracked 582 steps. This long memory of

steps is useful when going through data, as it allows for easy retracing of the history of the file's

changes. The list shown in Figure 15 does not include every single action taken in a file; there

were moments when I retraced my steps and went in a new direction; OpenRefine only tracks

one path forward. With that caveat, it is possible to go back through my work and see exactly

which cells were edited, which formulas and transformations were used, and how those

transformations impacted the data. When it became necessary to undo more than 50–100 steps, I

exported the file and archived it, as having an archived version of a file that is not needed is

better than redoing 100 steps. The archived steps can also be extracted as JSON (JavaScript

Object Notation) and applied to other data sets. This is useful if there are several files that need

to be edited in the same manner.

---

[66] In naming columns my system in running lookups was to state the original category (abbreviated), then what I was doing with it in that cell for a lookup, and what I was looking up from. In the case of Pe1_CCL_CR, I was filling in information about the first performer of the string using a lookup based on the *Creator Counter* (the primary key for the CREATORS TABLE), and the CR references that information is coming from the CREATORS TABLE. Thus, Performer 1_Creator Counter Lookup_Creators Table becomes Pe1_CCL_CR.
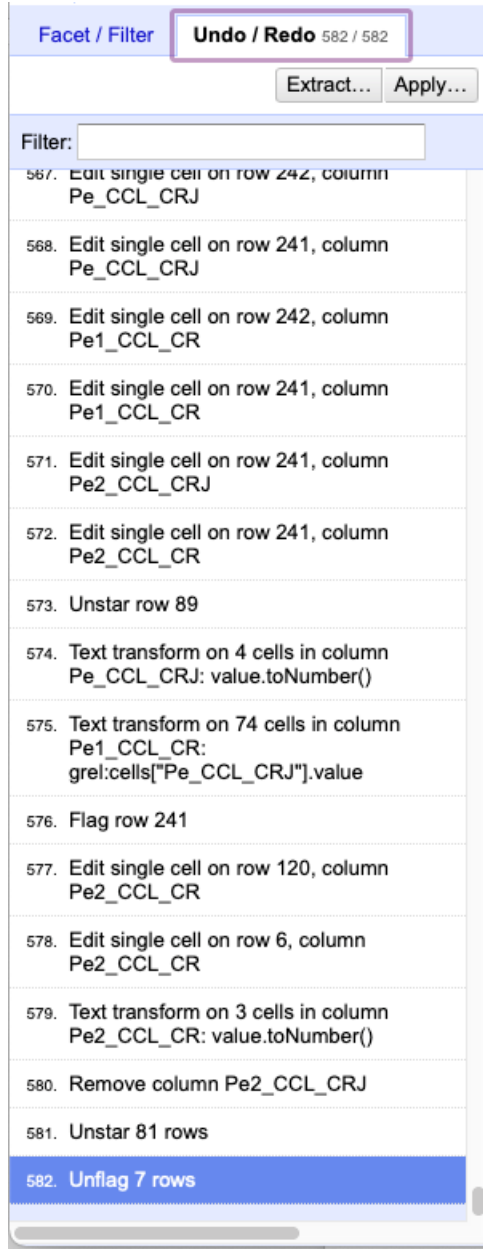
Figure 15: A partial list of steps taken in the Performers_Lookup file.

Common Transformations

     As shown in Figure 16, clicking on the dropdown arrow next to the field (or column) name brings up a list of options. I highlight here a few functions in the "Common transforms" list that proved essential for editing. "Trim Leading and Trailing White Space" directs OpenRefine to look at the beginning and end of every field and delete any spaces it finds. This

saves time when splitting columns, joining columns, or transforming large sets of data. Often, archivists add spaces after punctuation, and this makes the fields consistent. The transformation also helps when splitting a field in two; sometimes separating a field into two fields using a semicolon as the delimiter will produce a space at the beginning or end of some entries. I recommend trimming white space at the start of every file import, and then after every split or joining of columns. It can save time when running lookups and searching for strings of characters. "Collapse consecutive whitespace" looks through the data and, whenever two spaces (or a space and tab) are next to each other, converts them into a single space. This too is useful after editing out text or joining/splitting columns.
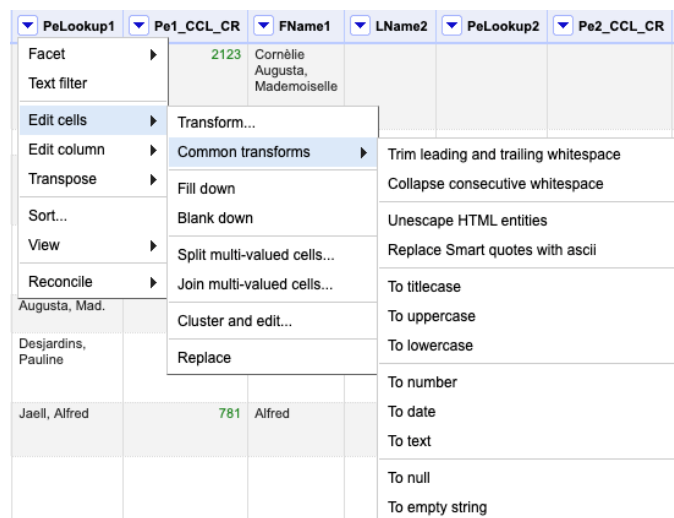
Figure 16: Dropdown menu of common transformations.

Many of the conversions listed as common transformations are helpful for ensuring the data aligns. Converting cells to different cases can be useful when searching for strings of text. Sometimes publishers wrote Op. 12 for opus, others wrote op. 12. Changing the case can reduce the chance that characters are missed. Similarly, converting text to numbers (such as when it is a date) or numbers to text can help align formatting and flag instances where a cell contains data that does not match the rest of the column.

Transformations

As shown in Figure 17, clicking on the "Transform" option brings up a dialogue menu

that allows users to type the text of their desired transformation. I used GREL exclusively, but

Python is also an option here.[67] OpenRefine gives several options, including what to do with an

error and the option to iterate a transformation repeatedly. Some common transformations I used

were:

- value.substring(1) – Remove the first (1) character from the string. This is useful for making a string shorter, especially for removing the first character after doing a search for commas or another character that stuck around from a previous column split.

- value[0,-1] – remove the last character from the string. This was useful in conjunction with regular expressions for finding the last character of a cell that was a period, then trimming the period.

- cells["COLUMN"].value – input the value of the cell in this row from the column entitled "COLUMN." This function is useful for copying column values over from one to another, and can be used in conjunctions with facets to copy specific cells in one column to specific cells in another column.

- "TEXT" + value – Add the string contained in quotations (TEXT) to the cell before the current value in the cell.

- if(logical text, "A," "B") – Execute a logical test on each cell. If true, "A"; if false, "B." "A" and "B" can both be complex sets of instructions. For example, if the criteria in a cell exists in a list, copy the value from the list, and if not, please return the message "value not found."

If unsure what function to use, the help feature contains extensive information on the

variables, functions, other features available using the Transform tool. These same variables and

functions apply when adding a column based on another column.

---

[67] Most people reading this are familiar, at least conceptually, with Python as a programming language, but more information can be found at the website for Python: https://www.python.org. GREL stands for General Refine Expression Language, which is a Java-like language used in OpenRefine. I recommend the OpenRefine user manual for more information on GREL: https://docs.openrefine.org/manual/grel.
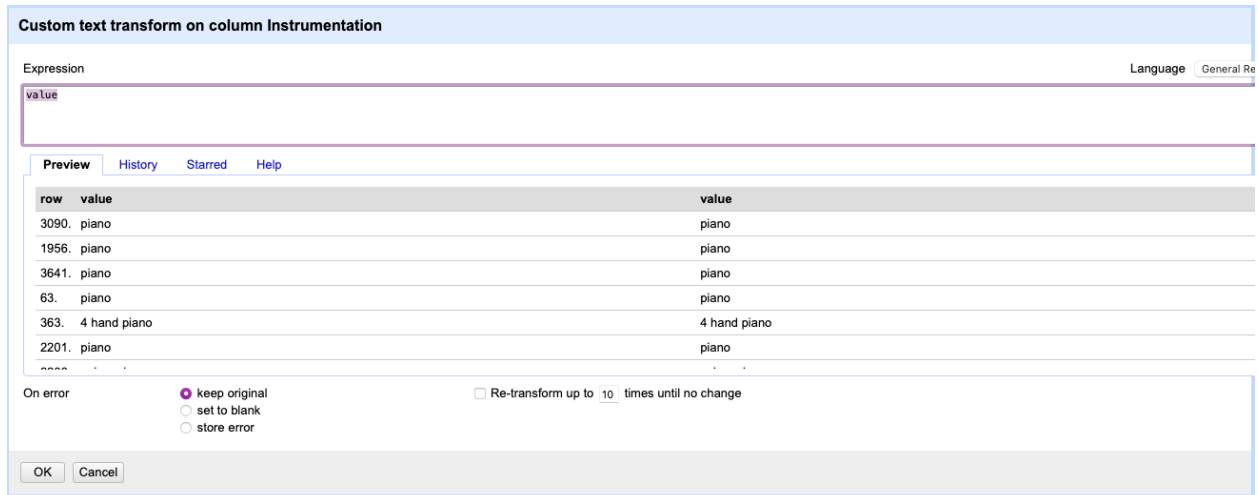
Figure 17: Pop up box for entering transformations on cells in a column.

Cluster and Edit

One of the most powerful things OpenRefine can do is cluster cells. This looks for cells

that are similar and helps reconcile the data in them. This is useful for last names that contain

special characters and can also catch some typos that exist in the data. The dropdown list

contains six different clustering methods, presented here in an order that approximates most

restrictive (requiring a closer match) to least restrictive clusters. Figure 18 employs the more

restrictive fingerprint clustering to first names and finds no matches.[68] Figure 19 uses a much

looser Beider-Morse cluster, and it does suggest some groupings.[69] In this case, the clusters are

not helpful, but they could be in other cases. For example, clusters caught that some works had

the title entered with an exclamation point, while others did not. Looking at each work's original

---

[68] Fingerprint clusters largely function by stripping away punctuation, case, creating each string of text separated by white space as a token whose order does not matter, and comparing the tokens for matches. There is a great deal of complexity in the clustering process of OpenRefine. For details on the precise actions taken when each cluster occurs, please visit: https://docs.openrefine.org/technical-reference/clustering-in-depth.

[69] A Beider-Morse cluster recognizes that words may be spelled differently, but if they sound alike, they could intend to reference the same person or object. Thus, the basic principle is that clustering process matches any two words that could reasonably sound alike: there and their, Schwartz and Chvarts and Шварц (German, French, and Russian spellings of Schwartz). For more information, and for 15 other instances of Schwartz, please see https://stevemorse.org/phonetics/bmpm.htm. OpenRefine

confirmed that these were the same piece. Loose clusters also pick up a less relevant information, including Henri vs. Henry, D.D. and D.B. as initials, and even one case of J.H. and Herr being the same value. Once clustered, OpenRefine allows users to decide if the cluster should be merged, what the new value should be, and opens up the entire row in a new window if additional information would help in making the decision (say, the last names of the Henry/Henri cluster).

Figure 18: Results from clustering first names of composers using the fingerprint clustering method.

Figure 19: Results from clustering first names of composers using the Beider-Morse clustering method.

Column Transformations

Thus far I have focused on what can be done in the "Edit cells" section of the dropdown menu, but "Edit column" tool is also robust, as seen in Figure 20. While some features were not necessary for this project—such as adding columns from URLs or reconciled values—I used every other feature at some point in this process.
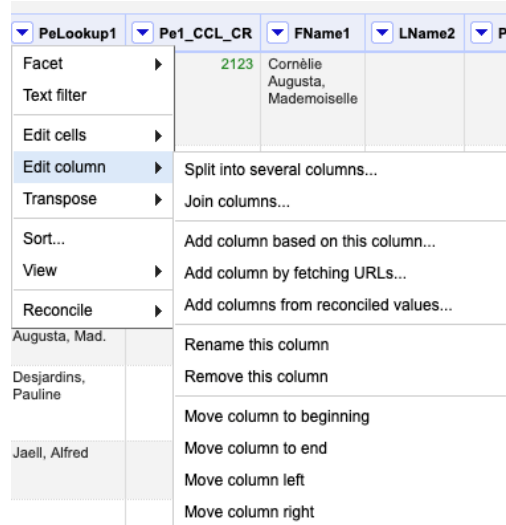
Figure 20: List of column transformations.

I used the split columns feature more than any other when it came to cleaning the actual data. When combined with the join feature, it facilitates numerous cell edits quickly. As seen in Figure 21, OpenRefine presents two ways to split text: first, by a separator (like a comma); and second, by field length (every 5 or 10 characters). I always used a separator, but there are numerous ways to effectively split columns. Some common separators in this project include:

- Semicolon – for situations where multiple entities are grouped together (composers)

- Comma – splitting up last names and first names.

- 1 – separating dates.

- Hyphen – separating date ranges.

- " b. " – using the space, b., and space to separate the abbreviation "born on" from another text string.

- Period – separating sentences into separate fields.

- Open Parentheses – separating nicknames and alternate first names.

Split column Instrumentation into several columns

How to Split Column
- ○ by separator

  Separator [|] ☐ regular expression

  Split into [ ] columns at most (leave blank for no limit)
- ○ by field lengths

  [   ]

  List of integers separated by commas, e.g., 5, 7, 15

After Splitting
- ☑ Guess cell type
- ☑ Remove this column

[OK] [Cancel]
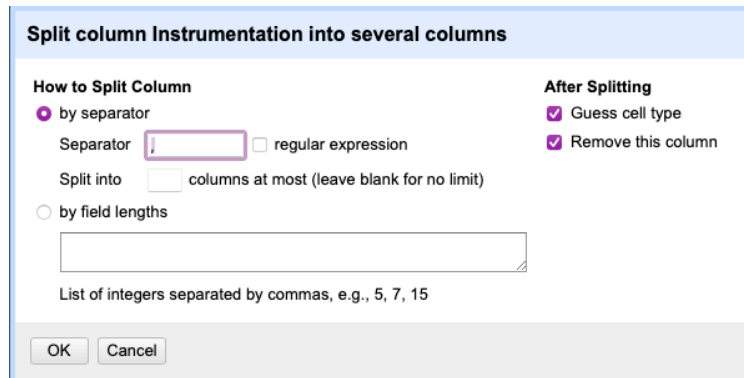
Figure 21: Dialog box for splitting columns.

It is important to consider the number of columns generated by a split while selecting the separator. For instance, one common split I used was to split columns using 1 as the separator, splitting into a maximum of two columns. This effectively put the dates in their own column, after which I could add the one back onto the second column using a cell transformation "1" + value. Limiting the columns to two prevents a date such as "b. 1819" from becoming three columns: b. | 8 | 9; instead, it becomes two columns: b. | 819. After the cell transformation on column two, the two columns would read: b. | 1819. Splitting columns also allows for guessing cell types, which OpenRefine does well. I always left this enabled and used a cell transformation after the fact if I did not like the outcome. The dialogue box also offers the option to delete the column being split. This can be a good idea, but if text is being split while a facet or search is currently enabled, it can create problems for the cells not part of the facet. Essentially, they all get deleted. Fortunately, OpenRefine's robust undo feature can fix this kind of mistake.

Joining columns can be extremely helpful for reassembling data in a different order (See Figure 22). Used in conjunction with splitting columns, this feature can reassemble data in ways that make it accessible to the user. The most common join I used was joining last names and first

98

names to make a new column. To do so, I selected the two columns, arranged them in the order I sought (LName before FName), and set the separator to be ," ." Nulls were replaced with blanks, and I wrote results in a new column called Lookup (or some variation like PeLookup for Performers Lookup). Numerous columns can be joined this way, provided the same separator is desired between each. When I sought different separators, I either performed the join in two steps or used the cell transformation features. For example, "cells["LName"].value + ," " + (cells["FName"].value + " " + (cells["MI"].value" yields LName, FName MI or Anderson, Brian K. as a result.
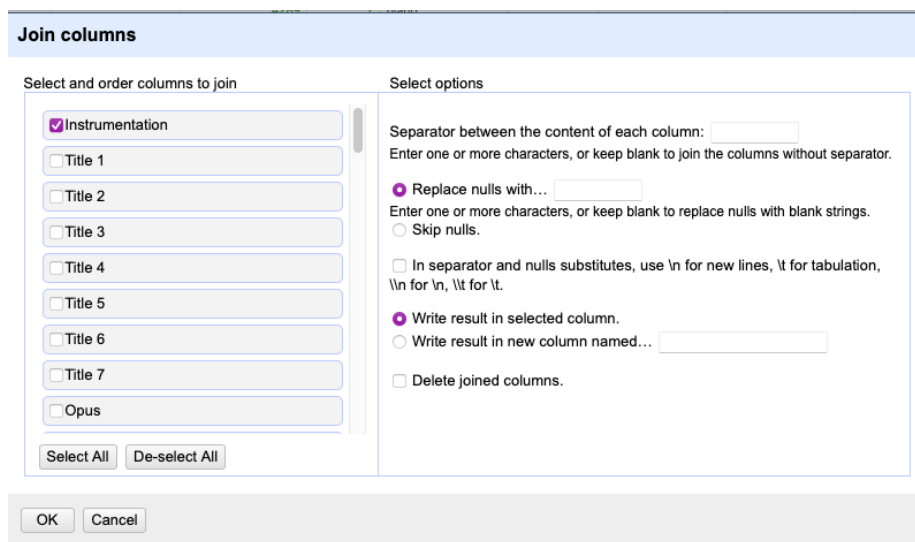
Figure 22: Dialog box for joining columns.

The third and final feature of editing columns to discuss in detail is the "Add column based on this column" feature. The dialogue box is the same as the cell transform box, but instead of writing the results in the current box, the results get written in a new column. This was useful when I wanted a transformation but did not want to alter the current data. In addition to the transformations mentioned above, I want to highlight the cell cross feature. This allows OpenRefine to find an OpenRefine project on the local computer, locate a column, and return a value from another column. The text looks like this: cell.cross("AAA," "BBB")[0].cells

99

["CCC"].value. This formula says to go to table AAA, look through column BBB until a match is found, and trace the row over to column CCC and write the results in that cell into the current project, in this new column. This is the essence of what I call lookup functions throughout this chapter. If multiple cells contain the same value in column BBB, the cell cross stop at the first match and move to column CCC; it does not look to see how many matches there are. I often paired this function with a conditional function to execute lookups that differentiate looking up a blank cell and looking up a cell that has a value but did not find a match.

The other features in the "Edit columns" like renaming, deleting, and moving are self-explanatory. There is a dropdown at the far left of the columns that allows for mass moving and removing of columns, which was my preferred choice for organizing the columns. It is worth reorganizing columns to place the columns needed for a particular search or transformation on the left-hand side of the screen.

Microsoft Excel Common Functions and Actions

Microsoft Excel has many of the same functions as OpenRefine, but in my experience, the program prioritizes viewability of data where OpenRefine emphasizes the data being cleaned/transformed. In this project, I primarily used Excel as a place to scan data before importing it into FileMaker. In the process of scanning and editing data, I did several of the same transformations as above. I used the find function (text filter), split and joined columns (delimit and concatenate), and even occasionally looked up values using a vLookup function (cell.cross). Most of the edits were simply sorting columns and looking for duplicates—particularly as it came to the CREATORS TABLE—before importing data into FileMaker. Excel does have more uses in modeling data with outputs. FileMaker searches that become pivot tables provide a powerful way to model data, and the program remains a top method for organizing data to put into clear,

presentable graphs. But with respect to data entry, Excel played a minimal role.

FileMaker Common Functions and Actions

FileMaker contains several robust functions, especially as concerns relating tables. It is not easy to clean data in FileMaker, and so most of the work done in FileMaker took place outside of the data cleaning process. With that said, a few important processes were useful here.

Search

The most useful is the search function. FileMaker's default search bar naturally peruses a table and looks for any matching text strings in any field. This was useful when going back through data to eliminate duplicates, especially in the BV CONTENTS TABLE. It works differently than the text filter function in OpenRefine, which focuses on one specific column or field, but the search function was more than effective.

Calculations

Calculations are to FileMaker what functions are to Excel, or what transformations are to OpenRefine. This is where more complex code or data manipulations take place. Figure 23 shows the way calculations are designed in FileMaker. The list to the left of the formula box contains the fields present in the database, and the list to the right of the formula box contains functions that can be used in the formula. The formula box in the center can be quite complex, though I did not automate many calculations or scripts in the process of making the database.
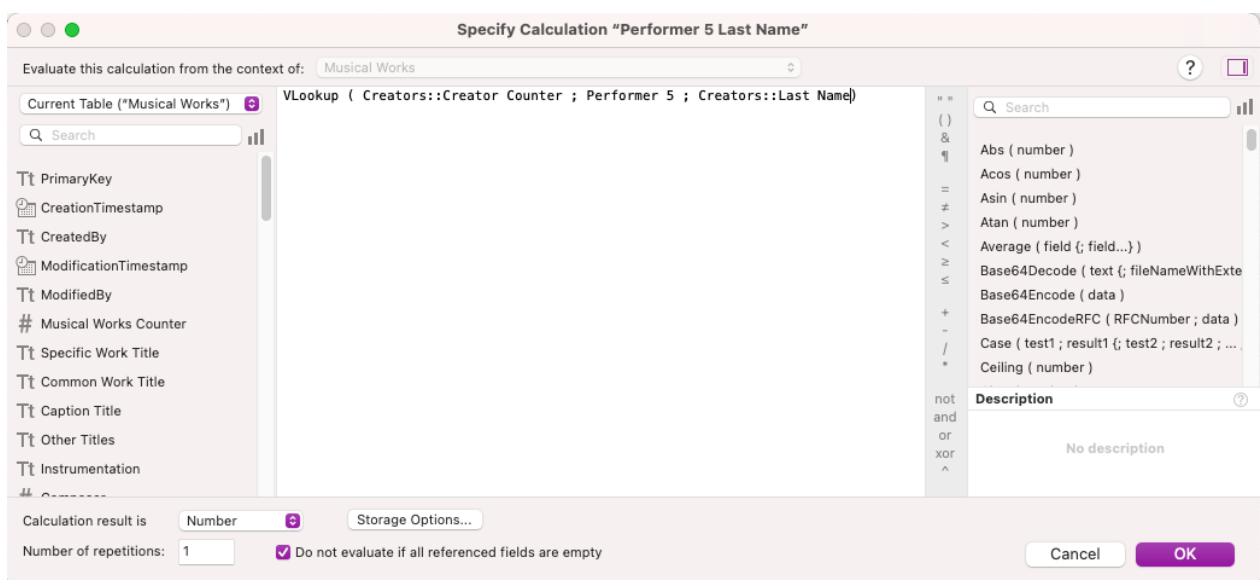
Figure 23: Calculations Dialog Box in OpenRefine.

One calculation I used was nested conditional functions. OpenRefine occasionally simplified serial numbers when I exported data from FileMaker, and I did not catch this the first time until I was putting data back into FileMaker. OpenRefine is capable of fixing the problem,

but nested conditional formulas made it possible to return the data back to its original state. An example is below:

```
If ( Musical Works::Composer < 10;

"CR00000" & Musical Works::Composer;
        If ( Musical Works::Composer < 100;

    "CR0000" & Musical Works::Composer;
      If ( Musical Works::Composer < 1000;
      "CR000" & Musical Works::Composer;
      "CR00" & Musical Works::Composer)
        )

    )
```

This is a relatively straightforward example of using conditional functions to add the beginning part of a serial number. All of my Creator serial numbers begin with CR, and OpenRefine had removed the CR and the 000s before each number. To add CR and the appropriate number of 0s back to the number, I used condition functions to determine the number of digits in the number. A number less than 10 has one digit, less than 100 has two, and so on. Based on that result, I added CR and the appropriate number of 0s to make sure each creator had seven numeric digits.

While the relationships between tables are useful for relating data, looking up data from other tables outside of primary key data can be problematic. I developed a system where several columns indexed the creators table, but the primary key only relates to one secondary key. Most of my solutions centered on building relationships between a parent and child table more efficiently, or in some different way. At one point, I contemplated building several tables, but that also was not the answer.

In general, there are two ways to reference cells in FileMaker: lookups and calculations. Lookups make it possible to fill one cell's values based on using the secondary key to lookup

102

information in another table using relationships. This lookup is quite useful if there is one relationship, but it is limited to one parent/child relationship per table, and joins are hard to execute this way. After some experimentation, I found that the best solution for me was to accomplish the lookup via calculations, and the best way to do this is not through FileMaker itself, but by adding custom functions to the program.[70]

Part 2: The Process

The first part of this chapter was devoted to introducing the tools used in constructing this database. The second portion of this chapter is devoted to the order in which I executed the steps, and pitfalls I came across on the way. I begin with the process of assembling the CREATORS TABLE, then move to PUBLISHERS, MUSICAL WORKS, COMPILERS, BV MASTER, and BV CONTENTS TABLES in that order.

Creators Table

The first list I reviewed and cleaned was composers. I provide more detail on the process that is going to be missing from later steps, which I hope will reduce the repetition in the process. With that said, the process for UNC composers consists of two large steps: (1) generate the list; and (2) execute the lookup.

Generate List of Composers to Start Creators – Method

It took some trial and error to figure out how to analyze and break down the composers

---

[70] Brian Dunning's FileMaker Custom Functions repository is an incredible resource for people looking to build databases in FileMaker as. This website contains a plethora of functions that are already built, many of which have been refined over various iterations of FileMaker to be very robust. FileMaker does not have a variable lookup function native to the program, but the vLookup function on Brian Dunning's FileMaker function website is incredibly powerful. This is the only function I found it necessary to add for my database purposes until now. Doug Staubach, "VLookup (LookupTableAndFieldName ; MatchingValueOrList ; ReturnFieldNames," Brian Dunning's Custom FileMaker Functions (Accessed Jan. 19, 2022): https://www.briandunning.com/cf/1740. I strongly recommend this function.

into usable sets and generate the center of my CREATORS TABLE, and some of these steps took some time as I developed my best practices. I followed the steps below on the list of composers.

Getting the Composers into One Column

First, I copied the list of composers into OpenRefine as a single column and, after trimming white space, delimited the set by semicolon into as many columns as needed. Here I saw that there were six composers listed in the largest set, which changed the number composers' columns I had in my MUSICAL WORKS TABLE later. I trimmed the leading and trailing white space, exported the file to Microsoft Excel, and used the sort functions and delete duplicate functions to place the composers into a single column, in alphabetical order by last name, with each cell being unique. This does not account for typos or minutiae, but it removed many duplicate composers from the list. This list served as the start for the CREATORS TABLE. I then loaded the file back into OpenRefine and trimmed white space.

Dates

To separate out dates from the rest of the information, used the split columns functions again. I first delimited by 1, splitting the data into a maximum of two columns. Because the data set does not extend before 1000 CE, and my research on binder's volumes terminates with certainty before 1950 CE, I was confident that every four-digit date would begin with a 1. After delimiting and erasing that 1, I added the one back onto the dates in the second column. I then trimmed leading and trailing white space.

Next, I sought to isolate the abbreviations of born, died, and flourished. I used a text filter to isolate strings of text at the end of the cell.[71] Regular Expressions allow OpenRefine to search

---

[71] The next three pages are technical in the data field searches, and at times I break the convention of keeping punctuation inside quotation marks to clarify what is part of a text string and what is not.

for a string of text at the end of a string of characters in a cell. I searched for the three characters " b.." The space is important, but it does not guarantee that the search is precise. A composer named "Smith, B." would show up in the search, as would all of the "born on" examples I sought to find. After reviewing the data, flagging relevant outliers, and faceting flags as false, I then copied the single date into a column entitled "Born." After trimming white space, I repeated this process for " d." and formed the column "Died." Finally, I search for " fl." and formed the final column for dates, "Flourished."

With these columns separated, I used a text filter to find any dates with a dash. After reviewing the data to catch any outliers (i.e., ensure examples b. 1815–16 were not missed), I used the dash as a delimiter and split the column into two separate columns. I then merged the first with "Born" dates and the last with "Death," which eliminated all dates remaining in the original dates column. I removed the columns that I had created while faceting data after checking to see that I had secured the information I needed.

The last step was to get information into a number form. Some dates were listed as "1850?," and others were listed as "1815 or 16." If a range was listed, I chose the average for the data set; thus, Robert G. Shrival was born in 1815.5 according to these cells. As time progresses, I will seek the precise answers through additional research. All dates were followed by a period when the data was a text string, but it was subsumed in the editing process and became a decimal when converted into number form. Decimals do not display unless specifically requested, so the period was thus of no further consequence.

Alternate First Names

The next step was to isolate the alternate names in the data set. This was relatively easy to

do especially given the Library of Congress name authorities file.[72] With dates and all relevant

abbreviations removed, the parenthetical name was the last item in a cell, if it existed. After

trimming white space, I split the cell by the open parenthesis character. I then trimmed leading

and trailing white space in both columns. The new column contained a close parenthesis, but it

also often contained an extra period or comma after that character. To account for this, I used a

text filter and searched for the close parenthesis character. I then trimmed the final character

from any cell that had a close parenthesis with the formula value[0,-1]. OpenRefine naturally

reiterates the search for the text filtered character after the action, and it removed any cell that no

longer had a close parenthesis. If the cell did have one yet, I had deleted the period or comma. I

repeated the process until no cells had a close parenthesis. I then trimmed leading and trailing

white space.

Separating First and Last Names

With the following information removed, the only thing remaining in the original cell

was any last or first names present, and some extraneous commas and periods. I first trimmed the

commas. First, I trimmed leading and trailing white space. Then I used a regular expression to

find any cells that had a comma at the end, and used the function value[0,-1] to remove the

commas. Next, I tried the same thing with periods. It was a disaster due to initials interspersed

with extraneous periods (many J.D. first names became J.D), and I had to reinitialize at this place

later.

I next delimited by comma but limited the number of columns to two. This separated last

and first names. I trimmed leading and trailing white space again, and I then exported the file to

---

[72] https://authorities.loc.gov/help/contents.htm. This has not come into play in many instances thus far, but it is a resource when messier data than UNC or UNT will be entered in future portions of the project.

Microsoft Excel. Here I searched for periods at the end of a line, flagged them, and sorted them to the top. I used the function =IF(RIGHT(B2,1) = ".","," "1," "2"). In words, this function says, "If the right-most character in cell B2 is a period, return the value 1; if not, return the value 2." I then went through line by line and eliminated periods that were extraneous. I missed some, but future editing helped me catch subsequent errors. I performed the same check on last names. This gave me a Creators page that was complete with my list of composers.

Duplicates Again

The last step was to do a count of each last name, then sort by last name. I had previously removed many duplicates, but this process exposed more. I went through the list and removed composers that were duplicates. This often happened with composers who were listed by either initials or by first name. Daniel, D.F.E., and Daniel François Espirit were all first names listed for Auber, and this eliminated them as duplicates.[73]

Creating and Executing the Lookup

The list of composers as creators is helpful only insofar as it can be paired with the musical works they created in the database. To this end, I took the Excel file I had with composers, uploaded it into the database, and used FileMaker's auto-generate serial numbers to create my *Creator* numbers. I then exported this list and placed it in OpenRefine. Because larger edits are possible, I did most of the mass edits and lookups in OpenRefine.

I created the table Creators_Lookup in OpenRefine using the most recent table. As the process continued, I archived many versions of Creators_Lookup with a note on when it was

---

[73] Because of this process, I sorted all of the last names by the number of occurrences in the CREATORS TABLE. Smith is the first composer listed (C00002) in the database because there were 14 Smiths with different first names in the initial data batch; the data was sorted this way at the import.

archived and the step I was on at the time. The current version of the table remains

Creators_Lookup, which is useful when performing lookup functions. I then returned to the

original MUSICAL WORKS TABLE master file and began to look up composers.

First, I had to split the composers in the MUSICAL WORKS TABLE to isolate the composers'

last name and first name. I did this with an abbreviated process of the work above. The main

difference is that I knew the dates, alternate names, and other information was saved in a

different place. I simply split off those columns and left them uncleaned.

Next, I joined the last name and first names to create a *Lookup* field. In my join, the last

name and first name were joined with a ," " as the separator and in a file called Lookup. In the

MUSICAL WORKS TABLE, the file was C1_Lookup for Composer 1 Lookup, and other names

proceeded in a logical manner. At the end of this process, there were two people with the same

first and last name that are distinct people: both are Strauss, Johann. They required separate,

individual edits.

With the lookup files created, I used the following GREL formula: if(value==null,"AA" ,

cell.cross("Creators_Lookup," "Lookup")[0].cells["Creator Counter"].value). What this says is:

"If the value in the current cell is blank, or null, please return the value AA. If the value is not

blank, please look at the table "Creators_Lookup" in OpenRefine. Go to the column entitled

Lookup and search for a match to the value in this cell. When a match is found, follow that row

across to the column entitled "Creator Counter" return enter the value of that cell in this current

cell."

The formula has three general outcomes. The first is that it can return the value of AA.

That means the original cell was blank, and that the formula worked. The second outcome is that

the formula can also search for a match and, upon finding one, return the number in question

from the *Creators_Lookup Creator Counter*. This gives the SK to the MUSICAL WORKS TABLE

that allows it to lookup the CREATORS TABLE in the database, building the link. The third

outcome is an error. This happens when the formula searches for a match but fails to find one. I

had errors return blanks, which is why I had the original blanks return "AA" to distinguish them.

This allowed me to facet for blanks and figure out what went wrong. Basically, blanks are errors

and they needed to be corrected before moving to the next phase.

Thus, finding and correcting errors proved a crucial part of the process. Most commonly,

I forgot to remove a period from one table. There were times that I had a match, removed a

period from one table, then realized I forgot to remove it from the other table. I also needed to

correct for names that I changed. Auber, Daniel returned a value, but Auber, D.F.E returned a

blank. This required editing D.F.E. to Daniel so the search would work. Iteratively, I corrected

values in both tables, correcting the database CREATORS TABLE when I changed

Creators_Lookup. The first attempt generated around 85% success, which means about 500

blanks. I repeated the process at least 20 times (possibly more than 50 accounting for errors in

generating the formula) until achieving 100% success.

With Composer 1 complete and a lookup table in place, I moved to Composers 2 through

6 and repeated the same process. This went more smoothly each time. Part of the reason is that

each list was smaller, and part of the reason was that the Creators_Lookup table became more

precise each time. Soon, I had completed the lookups for all composers.

Lyricists and Arrangers (L&A)

With the process of composers complete, the process for arrangers works in a similar

fashion. The process can be thought of in three steps: (1) create a list of creators from the data set

to be imported, isolating the person and identifying information; (2) compare that list to the

existing creators so that only new creators are incorporated into the database; and (3) use the new list to create a lookup for the arranger and assign numbers. This process is very similar for arrangers and lyricists, so I speak of the process just from the lyricist point of view. Arrangers follows exactly.

Create the List

The steps involved here are the same as generating the creators list, identified above. This new list, once complete, overlaps with the existing creators list. While working on it, I preserved the dates and information for subsequent addition to the CREATORS TABLE. Some works listed three lyricists, which means that the list had to be split into columns as composers were. Arrangers listed one person at most.

Update the Creator's List to include L&A

For lyricists (and eventually for arrangers), the creators list needs to be updated. This step is not complicated, but it is laborious. The challenge is that there are two lists with the potential for typos that need to be reconciled, and duplicates complicate issues later in the process. Duplicates might happen for a multitude of reasons, including extraneous punctuation (periods and commas in almost all scenarios), a different name listed as first name vs. first name alternate, inconsistent white space (extra spaces on the ends or in between letters, or a lack of spaces between initials), or other minor variance.

Lookup

The first step is to create a lookup row in the Lyricists/Arrangers table OpenRefine by joining the last name and first name in a new column called Lookup. In my system, the most current version of CREATORS TABLE resides in OpenRefine in Creators_Lookup. If questions exist

about whether the table is up to date, return to FileMaker and export again. If the file has been exported again, create the *Lookup* field in Creators_Lookup just as the field was created above.

Once the lookup field is created, use the facet to count. This should be done for every lookup when it is created, and each time it is used after new data is added. Duplicates sneak in through changes in punctuation and the minutiae that can introduce small inconsistencies, and facets are a great way to clean the data. In my data set, Johann Strauss is the only one in my first 2,000 names that I had two identical names in lookup. The use of facets and the counting feature inherent within facets can eliminate duplicates in both the lookup table and the table currently being updated.

Next, use a cell.cross function in OpenRefine to compare the Lookup values in the two tables. In the Lyricists table, I used the following formula: if(value==null,"AA," cell.cross("Creators_Lookup," "Lookup")[0].cells["Creator Counter"].value). Again, this function says: "If the value in the creators cell is blank, change the value to AA. If the value is not blank, compare the value in the current field (Lookup) to the Creators_Lookup table, the *Lookup* field until a match is found. Once a match is found, follow that row across and Return the value for the "Creator Counter" field. If a match is not found, return blank.." This formula will match some names, and those names can be discarded from the table, as this list is meant to locate lyricists not yet in the CREATORS TABLE.

Line by Line

Once the lists are reconciled, the next step is to archive the Creators_Lookup table, export both lists to Microsoft Excel, and go through the data line by line. This takes time, but the easiest way I found to do this is to sort both lists by *LName*, then by *Fname*. Then, set both side by side either on one large screen or two separate screens and compare. This will catch typos in the

111

Creators_Lookup table and in the Lyricists table, and thus tests previous editing. I found

numerous composers who otherwise would have slipped through the cracks due to alternate first

names, periods, or other minutiae. When making changes in the Creators_Lookup Excel file, I

also changed the FileMaker file to make sure the data stays up to date and consistent. I did this

for each addition, and it saved time in the long run.

Updating the List

Once the Lyricists list has gone through the line-by-line update, import the additional

records into FileMaker. FileMaker has an import records function, and it works smoothly. When

importing, make sure the list does not contain duplicates, map fields onto the FileMaker File

(Lname to Last Name, Fname to First Name, born to born…). Not all fields get mapped; I always

leave lookup off, and many fields can (and should) be set to auto-generate. This is good for serial

numbers, date and time stamps, person uploading, and other things. When a field can be auto-

generated, it reduces the chance for human error to interfere with the process, especially with

serial numbers and dates logging when fields were created and last changed. The *Creator*

Counter, which I use as a primary key for the table, is an auto generated serial number that

continues to update as the fields progress. Once the data uploads, check the report from

FileMaker for errors, delete the header column that seems to enter without fail, and the data

should be in the database.

The final step is to export this file again (I usually export as xlsx so I can easily view the

data) and then upload again into OpenRefine. This creates the new list in OpenRefine for

updates. If the previous Creators_Lookup was not archived, I archived it now by renaming the

file with the step it was on: Creators_Lookup_Composers_Only, or Creators_Lookup_CAL (for

Composers, Arrangers, Lyricists). I then uploaded the new file as Creators_Lookup, which then

allowed the formulas I had saved to continue working and created the Lookup field using a join.

Creating and Executing the Lookup

Once the new Creators_Lookup table with the desired lyricists/arrangers is established in

OpenRefine, return to the original MUSICAL WORKS TABLE. This last step works exactly as it did

for composers. Iteratively review the data and implement the Join (create lookups) and Lookup

(if and cell.cross) functions until all lyricists/arrangers are matched with a number in the

Creators_Lookup table. If changes are made to the Creators_Lookup table, change FileMaker

accordingly.

Lithographers and Engravers (LE)

Lithographers and engravers presented an additional challenge that the first three

categories did not. Many entries in this category were not people but were companies. People

were listed in the same format that they were for composers, lyricists, or arrangers. Companies

were listed without commas separating any part of the entity unless the lithographer or engraver

had a name in the company name. The three broad steps above remained the same: (1) create the

list from the current table; (2) update the creator's list; and (3) execute the lookup. With that

said, each step had additional challenges because of the companies.

Create the List

First, this list altered the CREATORS TABLE in general. Originally, my CREATORS TABLE

did not have a place for company name or original. I added these two fields to help account for

lithographers and engravers. Company name became the Lookup value I used when it was clear

that Lname, Fname did not accurately reflect all entries. I also kept the original name as an alternative lookup method for matching back to the MUSICAL WORKS TABLE.

There were some additional steps specific to lithographers that helped speed the process along. First, if the entry had the number 1 in it, it was likely a date. This allowed me to do a text filter search for 1, flag all the rows with a 1 so I could only work with them and go through the process for creating the list described in creators. Next, I searched for a comma and isolated cells without a comma. Looking through the data, most of the original cells without commas were companies; people generally have commas as a last name/first name. This still required reviewing, but it did make the list smaller and the review faster. Finally, I dealt with the cells with a comma, but without a number 1. This required a review to see if any companies had slipped through the cracks, then dealt with them as I did composers starting at the Alternate Name location.

Update the Creator's List to Include LE

This step was similar to the process used for lyricists & arrangers. The main difference is the *Lookup* field for lithographers and engravers. I created the *Lookup* field using two steps. First, I did the join described before for any cells with a Lname and/or a Fname value. Once this was done, I then filtered for blanks (still in OpenRefine) and then used a cell transformation: cells["Company Name].value. This filled the company name in any blank cells – that is, cells who did not have a person associated with them. At this point in the process, the only blanks were company names, and after I confirmed that no blanks remained. Once the *Lookup* field was generated, I went through the process outlined in Update the Creator's List to include L&A.

Creating and Executing the Lookup

This step was similar to the one outlined in Creating and Executing the Lookup for

lyricists and arrangers (L&A), but I did add one thing that sped the process. While I did create

the lookup field, I also used the original column to match up *Creator Counter* numbers with

fields. This was an experiment at this stage, but I used this extensively to match performers from

that table to the original due to the complexity in the Performers descriptions.

Performers

      Performers, perhaps more than any other category, present a particular challenge to

efficient parsing. Table 9 shows some examples of the entries in the performers category. These

provide a sampling of some, but not all of the structures contained in the category. The three

large steps remain the same, but the first—generating the list of creators from the performers

entry—required a different setup.

Table 9: A short sampling of performer descriptions to illustrate the challenges in pulling key information from text.

- As sung at the principal concerts by Mr. H. Russell
- As played by the Germania Society
- As performed in all his concerts in the United States by Henri Herz
- As performed by various bands at balls in Newport and the great cities
- Adapted to the celebrated Spanish dance La Cachuca
- As performed by the Westpoint Band

Create the List

      Table 9 demonstrates clearly that the previous structures for creators of content do not

apply here. First, the text is in a sentence. The typical punctuation parsers have largely been

abandoned. Entries do not typically end with a period, and the text provides the context. In this

situation, prepositions serve as the delimiters. The parsing was slow, and required white space

trimming and collapsing; however, by using words such as " as," " at," and " at," most of the

performers can be identified. Where the process was more methodical for parsing information in

other tables, this table required much more patience.
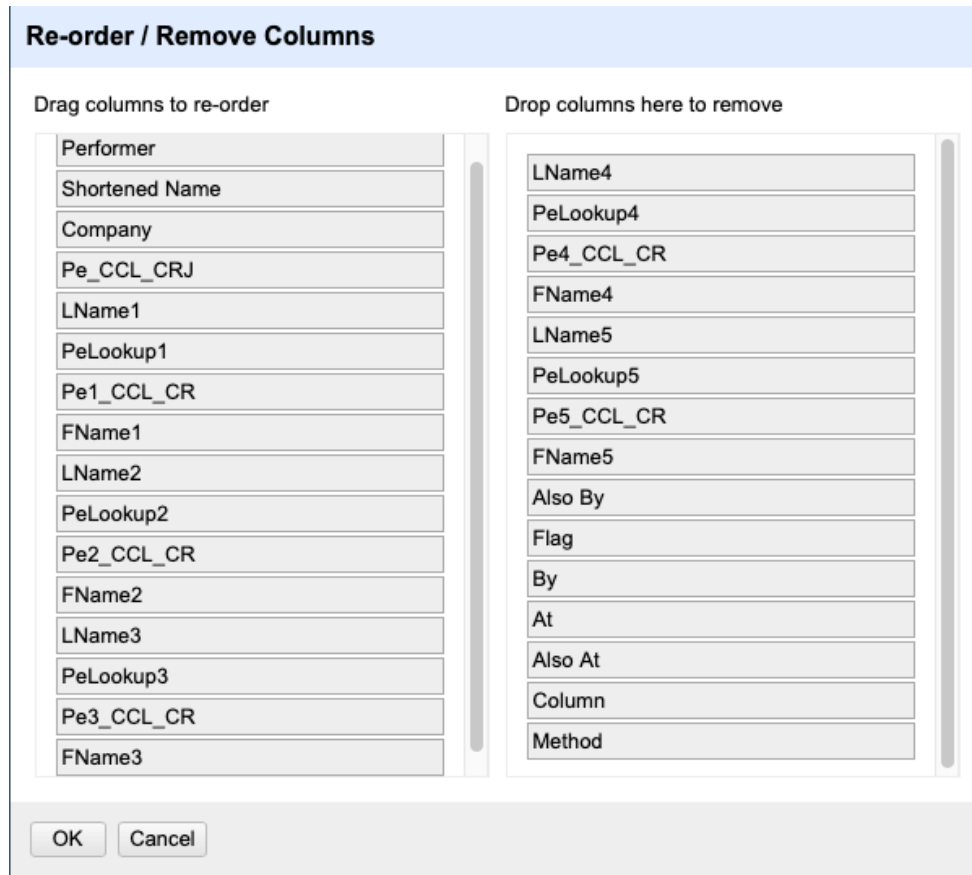
Figure 24: Fields for the Performers_Lookup Table. The list is in two columns and are organized from left to right as Performer to Fname3, then Lname 4 on to Method.

Once names were parsed out, they had to be searched to see if it was an ensemble name

or a person's name. The Westpoint Band was categorized as a company, while Herz, Henri was

listed by Lname, Fname. The solution I used was to create an intermediate table called

Performers_Lookup. In this table, I separated out the performers, but also preserved the original

names and did not delete duplicates. The fields (columns) list in Figure 24 shows that the

intermediate table preserved last names and first names, a company/ensemble name (which then

was added to last name as appropriate, fields for looking up *Creator Counters* (blank until the

final step of this section), and additional information about the performance description. This

additional information could include locations (up to two) and methods of performance. I then

made a second table that was the list culled from the intermediate performers table by exporting the table, placing all first and last names into one set of fields (as opposed to the five in the intermediate table, and eliminating redundant entries.

Update the Creator's List to Include Performers

This step is very similar to the previous updating the creators lists. The biggest challenge is dealing with many of the performing groups, who are represented in different ways. Christy's Minstrels are the same group as The Christy Minstrels, and these listings take some time to eliminate. Eliminating the duplicates is an ongoing process with editing, as some variants of names inevitably slip through at this stage. After this list is culled to reduce duplicates and to find and remove the duplicates with the existing creators list, I performed the same steps as before: upload to FileMaker, reexport to Excel, review the list, upload the list to OpenRefine, archive the previous Creators_Lookup list (in this case as Creators_Lookup_CALLE to indicate that it contains Composers, Arrangers, Lyricists, and Lithographers and Engravers), and rename this Creators_Lookup. Finally, I created the *Lookup* field by joining last and first name with the ," " separator.

Creating and Executing the Lookup (Performers)

With the new Creators_Lookup list in place, this step has two parts. First, I went to the Performers_Lookup table and executed the joins and lookups to find all performers. At this time, I created and populated the Pe1Lookup field (Join Lname1 and Fname1, separate by ," ," create new column called Pe1Lookup). I then used the lookup field and the new Creators_Lookup table to create and populate Pe1_CCL_CR, which stands for Performer 1, Creator Counter Lookup, all creators (to differentiate from a prior iteration where the list was incomplete). The GREL Function used to create the column was in the "add column based on this column option, and is:

117

if(value==null,"AA," cell.cross("Creators_Lookup," "Lookup")[0].cells["Creator

Counter"].value), which is the same as before. Once again, I checked for blanks, corrected any

errors (which were the cause of the blanks), and repeated the process until no blanks remained.

During the process, I often created columns such as Pe1_CCL_CR1 … Pe1_CCL_CR5…etc.

until all blanks were accounted for, and I could make sure that no numbers were changing

between lookups in unexpected ways. At the conclusion of this process, I deleted the spare

columns and renamed the final column to Pe1_CCL_CR. This process then happened for the

subsequent columns Pe2Lookup and Pe2_CCL_CR on through Pe5_CCL_CR.

　　　　After looking up the Creator Counter fields in the Performers_Lookup table, I used the

original names of the performer columns to add the files back into the MUSICAL WORKS TABLE,

where the lookups had been happening up to this point. To do this, I went to the performers

column in the MUSICAL WORKS TABLE in OpenRefine. This column contains the unaltered

performer name. I then added a column based on this one and used the formula

if(value==null,"AA," cell.cross("Performers_Lookup,"

"Performer")[0].cells["Pe1_CCL_CR"].value). This formula tells OpenRefine to: "If the current

cell is blank, input value "AA" and move to the next cell. If the current cell is not blank, go to

the OpenRefine Project entitled Performers_Lookup, and find the column "Performer." Once

there, compare values in that table with the current cell until a match is found. When a match is

found, trace the row over to the column entitled "Pe1_CCL_CR" and input the value of that cell

here. Once done, move to the next row in this field and continue to complete the task until the

column has been looked up." The title of the new column became Pe1_CCL_CR, and this moved

the column from Performers_Lookup to the main MUSICAL WORKS TABLE. Because of sorting and

eliminating blanks, a function like this is necessary. I then moved several other columns this

way, changing the "retrieved" value from Pe1_CCL_CR to the other performers creator counter values, and retrieving the places and method of performance for the future.

Publishers Table

With respect to creating the PUBLISHERS TABLE, the process is similar to the initial CREATORS TABLE. The main difference is that I sought to ensure that different addresses remained separate, and so before reconciling different publishers, there was an extra step to ensure that different addresses for the same publisher remained distinct. The three broad steps above remained the same: (1) create the list from the current table; (2) update the creator's list; and (3) execute the lookup. With that said, each step had additional challenges because of the companies.

Create the List

As with the CREATORS TABLE, I first went through and separated data out into columns. Luckily, most of the information already existed in columns: publisher name and location were separate columns, and there was much less material to disentangle in the publishers column. The process was straightforward, and the only additional step was preserving different locations. I did so by joining clustering for obvious typos, then merging publishers and addresses before making the lookup. By merging publishers and addresses when creating the lookup list, I ensured that instances like Firth, Hall (Philadelphia) and Firth Hall (NY) do remain distinct.

Creating and Executing the Lookup

This step was similar to the one outlined in other Creating and Executing the Lookup sections. With the PUBLISHERS TABLE lookup, it was important to match both name and address, but a join column in both the lookup and the MUSICAL WORKS TABLE made that process

119

straightforward. Once publishers were created as a table, the lookup was straightforward. Some errors had to be corrected, but they were sporadic, especially when compared to the initial CREATORS TABLE lists and the MUSICAL WORKS lists, which had more clusters and other minutiae to reconcile.

Musical Works Table

Through the aforementioned process, I assigned reference numbers to every creator and publisher in the UNC data and associated those numbers with the musical works in the main UNC data file. The next step became uploading MUSICAL WORKS TABLE itself into FileMaker, and ensuring the data became properly situated.

I needed the backup copy several times as I would import, then realize I made a mistake, then import again, then revert. Some mistakes were simple, such as not assigning columns correctly in the import records step. Other mistakes were more complicated. The two biggest mistakes were (1) not fully understanding the mechanics of related tables and lookups in the complicated case I created for this database, and (2) making mistakes in eliminating duplicates within the file. Before describing the process, I want to describe these pitfalls and the process of working through them.

Lookups

When it comes to executing lookups, the problem with relations and multiple secondary keys referencing a primary key was one that took me some time to solve. The solution—using the vLookup—eluded me largely because I was looking for an answer within FileMaker's functionality instead of outside it. Having detailed the vLookup process earlier, I do not elaborate further here.

Duplicates

The OpenRefine cleaning process was just as important here as it was with each individual table, but at this point several editorial questions began to arise as part of the database formation. Broadly categorized, these questions fall into two buckets: (1) editing titles and (2) reevaluating creator designations. I take a moment to address each before going on to discuss the process itself, but first I want to specify my fundamental framework for making these decisions.

The database undertaken as a part of this dissertation is itself a platform for hosting other materials. As such, I view the role here as creating a space for information to be available, but the curating and editing of that information will be relatively minimal. There are some steps that I take to reconcile the data and facilitate searchability, but the full work of going through each volume is work left to librarians at individual sites. Should this database expand into a digital collection, it will at that point serve the dual functionality of being the primary host of the data and an interface for public-facing availability of binder's volumes.

Editing Titles

The first question that I ran across was the degree to which I would edit titles. Given my editorial philosophy, I focused my efforts on quick ways to edit data and to remove inconsistencies that might be considered typographical in nature. Some examples include commas as opposed to colons between lines of text, capitalization, and other such instances. My guiding principle was that if both the composer and publisher were the same, it was worth editing the title to represent these works as the same. If the composer was the same but the publisher was not, I tended to use a bit more discretion. Finally, I did little editing (if any) to works that had similar titles but different composers and publishers. Edits in this particular instance only happened after reviewing the available digital manuscripts.

121

To accomplish this editing, I focused on OpenRefine's editing tools, especially the ability to cluster and edit. OpenRefine offers many different types of clustering and editing tools. I went through each of the clustering methods and reviewed groupings. Some were obviously not clusters to be reconciled. Series of works that share the same title but feature different works. This includes serial collections like the colorfully titled "Scraps from the Opera." I began with the most restrictive clustering methods and worked my way through to clusters that tended to be much broader. In cases of punctuation or capitalization as the only area of disagreement, I accepted clustering to one version of the title. In cases where the issue was a bit more complex (an article included, punctuation involving exclamation points as opposed to question marks, minor differences in wording), I opened the record to apply the publisher/composer standard previously outlined. Whenever the metadata was unclear, I returned to the original manuscript as much as possible.

Even after this clustering, numerous titles appeared repeatedly. This is because I treat the printing plate as the object cataloged in this table to the greatest degree possible. In cases where a composer is different, two otherwise very similar objects are categorized as different musical works. This is also true with publishers, but the issue gets a bit murkier when the plate number is the same. I explore this issue in greater depth with my case study on "Home, Sweet Home" in Chapter 5.

Creator Designations

As I edited titles, I came across several instances where a title would be listed with the same composer across multiple publishers, and one instance of the same title being listed with a different composer. As I reviewed such cases, I tried to take time when I saw these atypicalities

to examine each case. Some were separate pieces with the same name.[74] Others were the same

piece with the publisher having denoted a different composer to what is readily identified as the

same tune. Still others involved a cataloging error (few and far between, but they did happen). In

such instances, I corrected situations as I saw them, but no systematic effort was undertaken to

review the data from the catalog.

Creating the List

With these principles in mind, I set about to narrow down the list of titles to a single title

for each piece. The first step was clustering and eliminating duplicates using the clustering

principles above, and correcting errors as I saw them. Next, I used the list that was reduced

through clustering and eliminating duplicates and exported it to FileMaker via Excel, using

FileMaker to assign a serial number to each musical work. With that done, I exported the list out

of FileMaker and back into OpenRefine and began the process of working backwards to ensure

that my lookup would fine each song in the BV contents page. To do this, I created a lookup

column in the MUSICAL WORKS TABLE and in the BV contents page and used this process to

compare the works.

Editing the List

Most of the MUSICAL WORKS TABLE list was edited through lookups as they happened

during the creation of the CREATORS and PUBLISHERS TABLES. Each time a column/field was

incorporated into the CREATORS TABLE, it was then tested by using a lookup function to replace

the column in the MUSICAL WORKS TABLE. Thus, the bulk of the work was done in creating the

first two tables: CREATORS and PUBLISHERS.

---

[74] All three instances are discussed in Chapter 5, part 2 (Home, Sweet Home!)

Eliminating duplicates took some time. The processes above were tedious for 5,000

entries, and an early attempt went awry as I did not account for works by the same composer

with different publishers. Once duplicates were eliminated, the final steps of the MUSICAL WORKS

TABLE also intersect with the BV CONTENTS TABLE, and I discuss the process in detail there.

Compiler's Table

This table was a product of information present in the data itself, and such data was

scarce. Instead of it being imported person by person, it was much easier to type in data by hand.

This will be the case at times, and it is worth acknowledging here. Sometimes large-scale data

manipulation via OpenRefine is the way to go, but there will be instances where a single volume

is being added in the future, or a new compiler needs to be introduced. In such cases, mostly

manual entries are often the easiest solution. Even with larger data projects, keeping it simple is

an important principle, and it applies here.

BV [Binder's Volumes] Master Table

When it comes to creating the binder's volumes list, the process was much more

straightforward than for the previous lists. Here, I simply copied the columns from the "Volume"

column and broke the column into multiple fields, including series, institutional designation, and

volume number. For UNC Chapel Hill, I used the volume number and series number together as

the institutional designation; for other repositories like UNC, the call number is also the

institutional designation. Other information is not available in the metadata, and I will collect it

on site during a research visit at some point in the near future.

BV [Binder's Volumes] Contents Table

The process here is quite similar to the one for creating the MUSICAL WORKS TABLE by

aligning creators and publishers. The basic process involves three tables: the original BV

CONTENTS TABLE, the MUSICAL WORKS TABLE, and the BV MASTER TABLE.

Set up Tables in OpenRefine

At this point multiple versions of each table exist, so having a reasonable method to track

revisions and versions of tables is essential. In creating the MUSICAL WORKS TABLE, a step in my

workflow involved backups at a point where all of the creators and publishers had been added to

the table, but none of the duplicates had been removed from any of the rows. This is the point

where I picked up the BV CONTENTS TABLE, and I used this as the first table from which the other

lookups could take place. The musical works lookup table and the binder's volume master

lookup table both came from FileMaker. I exported both and used them to find the respective

counters. The BV MASTER TABLE was quite convenient; with minimal manipulation, it was

possible to get the unique identifiers isolated for each work, and to add a column containing my

reference number alongside the existing "Volume" column. After changing Roman numerals to

Arabic numerals, the lookup process was smooth.

The MUSICAL WORKS TABLE required more editing for the process to function. One result

of reconciling titles in an earlier step is that multiple cases exist where editing the title caused the

work to need edited. The process I employed was:

1. Create the lookup, and resolve any typos in the lookup columns.
   a. The first issue was that many numbers transferred with a (.0) at the end of the joined lookup between title and composer.
2. Executed Lookup
   a. Work through some of the failed lookups (facet by blank to find them)
   b. If changing OpenRefine BV Contents, proceed
   c. If changing MW_Lookup, also change FileMaker File
3. After 50 edits, preserve MW_Test_# as version control
4. Reinitiate Steps 1-3

Creating the Lookup

To begin, I took the reduced MUSICAL WORKS TABLE created by eliminating duplicates, and I created a Lookup row by joining Title, Composer Reference Number, and Publisher Reference Number into one column. The first step was to undo any transformations to cell type and restore serial numbers to their original type.[75] Once that was done, I attempted a merge and looked to see if any serial numbers for composers or publishers contained a decimal at the end. If so, I used a replace to trim the .0 from the join. Once the join was created in the MUSICAL WORKS TABLE, I then created the join in the BV CONTENTS TABLE to make the same lookup of Title, Composer Reference Number, and Publisher Reference Number.[76]

Executing Lookup

Once the lookup rows in each OpenRefine file were created, I then compared the two as a way to assign MW numbers to the BV Contents page. I directed OpenRefine to compare the BV Contents Lookup to the Musical Works MW_Lookup column, find a match, and then place the Musical Works reference number (MW_Counter or Musical Works Counter) into a separate column in OpenRefine.[77] The formula also contained a conditional function to ensure that blanks would not return the same value as failed matches. Thus, blanks returned a value of "AAAAA," successful searches returned a match of the appropriate MW number (the first found), and failed searches returned a blank cell. I called this columns MW_Test_#, based on the iteration of the process. The first time, the column was MW_Test_1.

---

[75] grel:if(value < 10, "P00000" + value, if(value < 100, "P0000" + value, if(value< 1000, "P000" + value, "P00" + value)))

[76] Join functions in the cell transformations.

[77] if(value==null,"AAAAA" , cell.cross("MW_Lookup," "MW_Lookup")[0].cells["Musical Works Counter"].value)

After each search, I began to parse blank cells to see why the searches failed. These failed searches were almost always a product of minor edits made to cells in the cleaning process earlier. After each search, I worked through approximately 50 failed matches at a time to resolve the problem. If the change needed to be made to the BV Contents page, no further updates were needed; the BV contents page was the one under construction. If the edit needed to be made to the MUSICAL WORKS TABLE, I made a point of editing both the MUSICAL WORKS TABLE in OpenRefine and the FileMaker table.

Repeating the Process

After 50 or so entries, I remade the merge in the MUSICAL WORKS TABLE and made a new merged column in the BV CONTENTS TABLE (both in OpenRefine). After making the two new merged columns, I created a new MW_Test column to see what progress the changes made to those 50 entries yielded in the table. After creating a new MW_Test, I faceted the old one as not being blank and the new one as being blank as a way to check and make sure that the changes I made in the past were not breaking matches that I had previously made, giving different numbers than previously looked up, or that I knew why matches in the lookups were being broken. I then went through 50 examples (or thereabouts) and fixed errors, then remade the Musical Works lookup, the BV Contents Lookup, and made another column.

After going through about 5 iterations (200–250 entries), I archived the MUSICAL WORKS TABLE, updated the FileMaker table with any changes I needed to make that involved salvaging entire entries that I had previously merged, and created a new MUSICAL WORKS TABLE. The old one was archived, and I reinitiated the process above. As before I created lookups, created the MW_Test, and paid particular attention to any matches that were not made in this MW_Test that had occurred in previous lookups. This process continued until the process yielded no blanks.

Once finished, I saved the file, and exported it into FileMaker, providing the final table for the initial institution.

Removing Unused Records

One data cleaning step remains after building the lists: purging unused records. While these records are few and far between, cleaning up the list is necessary as more and more institutions are added. The process to clean the list is just a count for how many times each lookup is used. I performed the search in excel so I could more easily scroll through the numbers, then went back to FileMaker to see if the record was indeed missing or if the record was simply unused.[78] If unused, I searched for a similar text string to see what might have replaced this record and deleted the record. I began the process working from BV CONTENTS to MUSICAL WORKS, then from MUSICAL WORKS to PUBLISHERS, and finally from MUSICAL WORKS to CREATORS. Often a creator might not be listed as a performer at all, but instead as a performer or arranger only so CREATORS did take several checks. To ensure I had a record of any entry that I might want later, I saved several versions during the record purge.

Adding Additional Repositories

A central purpose of this database is to provide the process for establishing the basic process, and to clarify how to add new institutions into a single database. As mentioned in the introduction to this chapter the details and rationale presented in Chapter 3 are meant as a companion to Appendix B, which describes in a step-by-step process the best practices as

---

[78] In Excel, create a list of numbers 1 through the end of the serial numbers in column A, then =CountIf(OriginalSheet!Column$row,A1) in column B. I then sorted column B for values of 0 (sort low to high), and manually checked the list against FileMaker. It is possible to do a second lookup using this list against the table in question to confirm the blanks, but the lists were short enough that I felt manually looking numbers up was most beneficial.

currently understood for integrating new databases into the BVRI. Each step will require

flexibility; each data set is going to have its own quirks and information to be sorted. With this

caveat, the system outlined in Appendix B, taken together with the details on how to implement

each step in this chapter, are a clear roadmap for expanding the BVRI going forward.

This concludes the technical portion of the dissertation. Chapters 2 and 3 were focused on

outlining the processes used to create the BVRI, the rationale behind the metadata schema, the

steps used in actually building the BVRI as it currently stands, and best practices for expanding

into new data sets. Chapters 4 and 5 turn to results. Chapter 4 focuses on large-scale analysis and

general questions, where Chapter 5 focuses on smaller-scale questions and case studies.

CHAPTER 4

FROM METADATA TO HISTORICAL CONTEXT

Chapters 2 and 3 have largely been concerned with the detailed methodology of how the database is created. For such a database to be effective for research queries, the underpinnings of data additions need to be clear at the outset, including the editorial practices used in its creation. This chapter of the dissertation moves beyond questions of methodology to consider some results from the preliminary iteration of the BVRI. In presenting these results, this chapter is also designed to explain which questions and topics I had in mind at the outset, and how those goals evolved. My method seeks to privilege the materials themselves, with the goal of considering in aggregate what this collection of objects and the data attached to it might tell us about the context for collecting and binding sheet music. While this research relies on work by other scholars—whether on publishers in specific locations, or on previous case studies of binder's volumes and their collectors—it also builds on such work by taking the materials present in volumes at a large scale to seek new insights into music publishing and consumption in mid–nineteenth-century America.

This chapter pairs with Chapter 5 to demonstrate the BVRI's research potential. Chapter 4 focuses on metadata-based questions while Chapter 5 focuses more on how the data can support inquiries into more localized, specific questions. When I conceived this project, I wrote several research questions that pertain to the metadata as well as to individual songs and volumes. In developing and using the database, I have found that some of my initial questions have proven aspirational, while others have developed into meaningful queries. The current scope of the database limits the questions that can be defensibly answered at this stage, although this will likely change as the BVRI grows. For example, the mean and standard deviations for

binder's volumes beginning and end dates are useful for a set of 125 volumes, and I discuss this

information in detail; however, as the BVRI grows to encompass thousands of volumes, such

details will become more nuanced, and questions can then be asked based on regions. Similarly,

the publishers' locations are useful and worth discussing in detail, but more volumes will

increase the useful impact of the data.

With these caveats, these are a few of the core questions that served as an impetus for

engaging in the project, organized with a move from more specific questions stemming directly

from the data to broader questions that engage with the social history of these objects.

- When do volumes come from?
  - What is the typical span of dates in the musical works table?
  - What is the typical span of dates in a binder's volume?
- Where do volumes come from?
  - What is the geographical spread of publishers incorporated in each binder's volume and how does it change over time?
  - Is a Westward migration of publishing over time noticeable in the data?
  - What regional variance in songs can be found in BVs? How does this change over time?
- Who is included in the volumes?
  - Which composers are included? What is the top 10 list?
  - How does the popularity of a composer relate to their death date?
- How does this help us?
  - How can the BVRI provide a look at music consumption in nineteenth-century America?
  - How can the BVRI enhance our understanding of women in nineteenth century American music?

This chapter addresses these questions in the order above, starting with those that

establish a broad chronology. To do so, I use data from the BVRI about both individual pieces

and collected volumes. The following section addresses where the volumes' contents come from.

Here I focus less on the provenance of the compiler and more on the materials: where pieces

were published, who published them, and then where the publishers are located. I then turn to trends in the data, specifically the migration of publishing in America as observed in binder's volumes and the regional variance found within a single volume.

After addressing questions of place, I turn to which composers are included. First, I establish who specifically is included, and how often. Instead of working through all of the composers, I generate a top 10 list of sorts, and discuss some of the lesser-known composers briefly. I then turn to the question of when composers are included. This section uses the composers' death date in relation to the date of the work published to determine the extent to which binder's volumes seem to engage with composers of the day versus those from previous generations. Essentially, I am seeking to answer whether composers were popular in their day, and whether their work endured long after their passing.

The final section turns to broad questions that are more challenging to answer directly with data, but to which an awareness of metadata context can yield some insight. More specifically, I consider the ways in which earlier questions might be recontextualized to help us develop a clearer understanding of domestic music consumption and gender. Again, these are difficult questions to answer, but the beginnings of answers presented here will direct future research efforts.

<div align="center">When Do Volumes Come From?</div>

This first large question is a complex question; I therefore find it useful to break it into parts. The first relates to the musical works table. When were various works published? Many of the decisions surrounding musical publishing rely on knowing approximately how the span of music divides so that I can break locations and publisher timings into sufficiently large and

representative chunks. Thus, the first question I ask focuses on the musical works themselves, and not yet on the binder's volumes.

The second questions about dates concern how the dates of musical works group into the volumes. There are several relevant questions that arise from this line of thinking, and I extract dates that give useful context to each individual volume. These dates include the earliest work, latest work, range between them, mean, median, and quartiles of works with known dates, and percentage of works with known dates. In this way, each volume has a set of notable dates that describes relevant temporal information about the volume and provides context for some of the other questions that I attempt to answer later in the chapter. I go through each of these questions below.

What is the Typical Span of Dates in the Musical Works Table?

The term average is by its nature unclear, in large part because it can represent numerous statistical measures. The two primary measures I am interested in are the median and mean of the samples, and their respective metrics for describing range and variance. The mean and standard deviation give an idea of the center and how broad the divergence is. The median, quartiles, and range provide a more discrete picture of distribution of works in a binder.

For the UNC binder's volumes, the average (mean) date is 1852.3 based on the volumes that have dates in the metadata. This correlates with my expectations; from the outset, I imagined the study as focusing on the 1800s, but most likely from about 1830–1880. The earliest dated work in a volume work comes from 1815, but this does not necessarily indicate the first volume date as some works are undated. The standard deviation is 20.7 years, showing that most dates (in this case 93.6%) of dates fall within the two standard deviations of 1811–1893. This weight confirms my expectation that many of the volumes fall in the range of 1830 to 1870,

if not a bit later. In a perfect bell curve, 68% of samples fall in one standard deviation. Here, the weight of pieces falls more within the first standard deviation (74.4%), with more pieces than modeled in a bell curve existing outside two standard deviations, all on the later end of the data. This is to say that the data fits a bell curve well that highlights a rise in popularity through the 1850s, followed by a gradual tapering of the practice. Some works and practitioners exist very late in the process, even if the 1945 work is a notable outlier.[79]
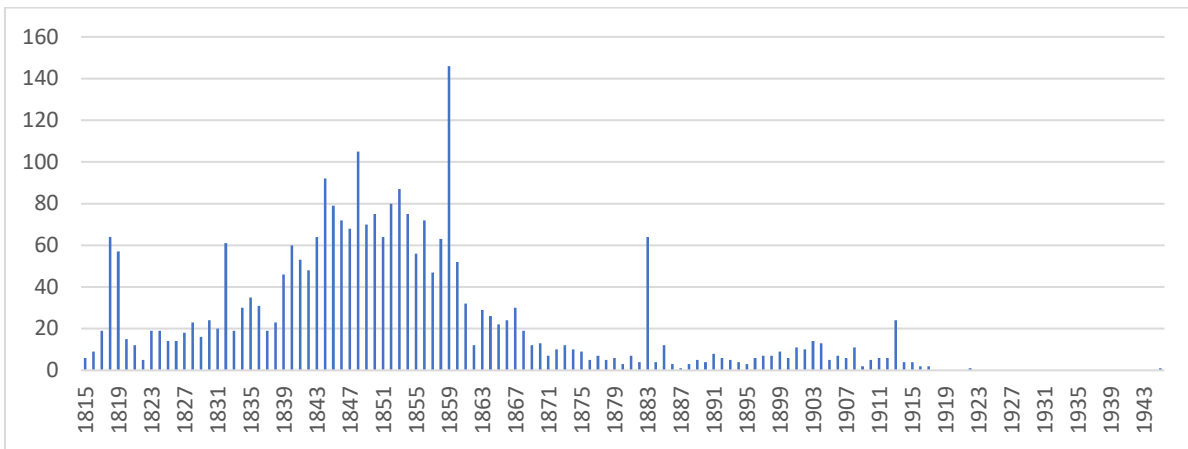
Figure 25: Musical Works per year in the BVRI data set from UNC.
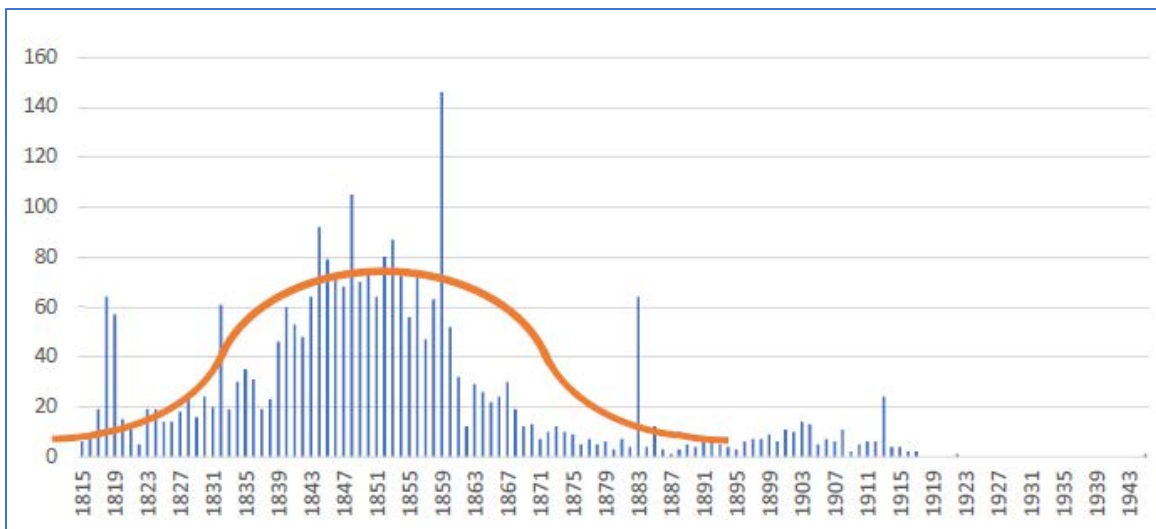
Figure 26: Prototypical bell curve imposed on Figure 25.

---

[79] Such a gradual tapering aligns with and can thus confirm the histories of print and of binder's volumes outlined in Chapter 1: more printed materials existed later, although the trend of collecting and binding appears to have waned by the nineteenth century's end.

The median and quartiles show a spread throughout the 1800s, with a strong focus in the center. Just as with the mean, the emphasis is on the 1850s, with the median being 1850 and the quartiles 1841 and 1859. Figure 27 shows just how much 1945 lies outside the data, and otherwise the data models that split of binders matching the mean quite well. The box and whisker plot displays this data succinctly. The median is the line going through the middle of the box (1850), and the first and third quartiles define the edges of the box. The whiskers extend 1.5 times the IQR, with outliers plotted above and below the whiskers as needed.[80]
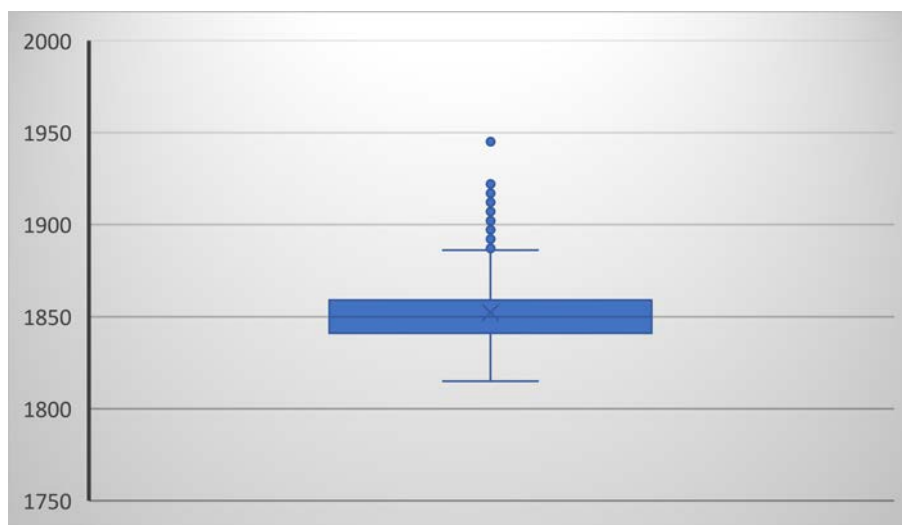
Figure 27: Box and Whisker chart for Musical Works per year.

What is the Typical Span of Dates in a Binder's Volume?

Having looked at the dates for musical works contained in the BVRI, it is worth considering what the typical binder's volume might look like. This section explores that question

---

[80] In a box and whiskers chart, the interquartile range (IQR) is the range between the 25th percentile and the 75th percentile (the box of the box and whiskers). An outlier is a number that lies more than 1.5 times the interquartile range below the first quartile or above the third quartile. For example, in Figure 27, the first quartile is 1841 and the third quartile is 1859. The interquartile range is 18, and 1.5 times the IQR is 27 years. An outlier, or a number not contained in the line portion of the box and whiskers graph, is a number that is 27 larger than the third quartile (1859) or 27 smaller than the first quartile (1841). Thus, a musical work published after 1886 or before 1814 would be considered an outlier. As the earliest work is 1815, there are no outliers on the early end of the box and whiskers graph, but there are outliers on the upper end of the graph (works published after 1886), with 1945 being the most notable outlier.

as it relates to dated works within the binder's volume. When it comes to the typical binder's volume, the range is incredibly wide. Within this set, there are 126 volumes, with the fewest number of works in a volume being 9 works and the largest 113 works. Very few volumes come close to containing 113 works—the next highest in this set is 84—and these are clearly outliers as compared to the rest of the data. The median number of works per volume is 38, and the interquartile range spans 27–50 works per volume. The mean and standard deviation track closely with the median, with a slight skewing upwards due to the high-end outliers. The mean is 40.6 work per volume, and the standard deviation is 17.7. Thus, most volumes contain between 23 and 58 works (1 standard deviation), as reflected in the median. Outside of outliers above ~80 works, the data is not particularly skewed.
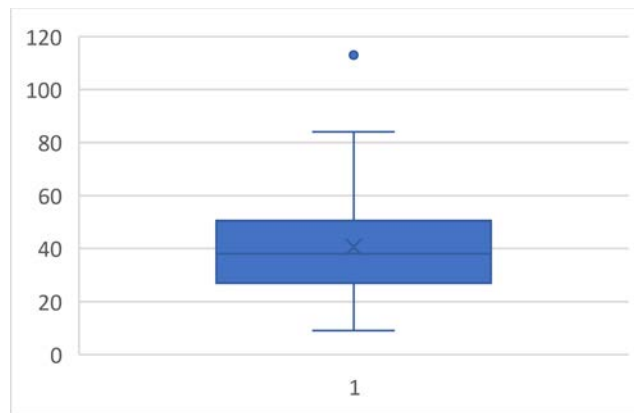
Figure 28: Number of works in a volume. 113 works is the outlier volume (Volume 100)
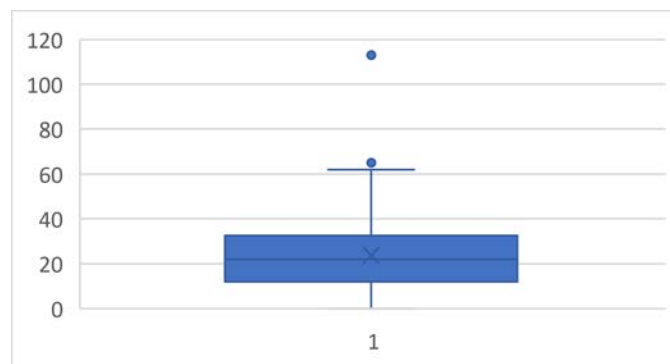
Figure 29: Number of dated works per volume. 113 works is the outlier volume (all works in Volume 100 are dated).

A little over half of the works in each binder's volume have recorded dates. Both the median and mean comparisons of dated works to total works show that about 58% of works would be dated (57.9% according to the medians and 58.4% according to the means). While some volumes have no dates at all (possibly due to cataloging; Chapter 3 notes that we do not correct or update cataloging at this point in the BVRI project), most volumes have at least 40% of works dated. The interquartile range of dated works is 40-80% of volumes, with some volumes also being completely dated in the catalog.[81]
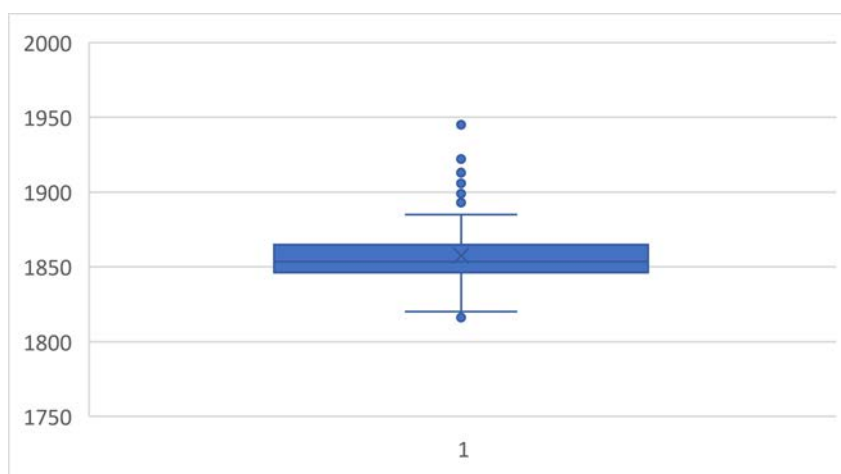
Figure 30: Box and Whisker Chart showing the earliest possible date each binder currently in the BVRI could be bound.

The typical range of binder's volumes' dates is a challenging question to explore, but for this portion I focus on the terminus post quem as defined by the assumption that, for the volume to include a particular dated work, it must be bound in the year the dated work was published or after. That is, what is the last dated work in a volume? This data largely tracks with the MUSICAL WORKS TABLE, but with the box and whiskers portion of the table pushed back five years. The interquartile range shifts from 1841–1859 to 1846.25–1864.75, which shows that the earliest possible date at which a volume could have been bound usually falls approximately five years

---

[81] BV 100 has 113 works, all of which are dated. I've included the volumes.

after the average date of publication of its dateable contents. The range of outliers is also noteworthy; even with the focus being on the latest work, the data shows that volumes bound after 1885 remain rare, and that the bulk of works fits the original hypothesis of this dissertation as centering on the middle of the nineteenth century.

The grouping of dated works is useful in the context of the typical binder's volumes, especially as it relates to clusters of dates. Going into the project, I anticipated that binder's volumes would contain a tight grouping of published dates, which I predicted because of the role of volumes as part of the dowry. Thus, I anticipated works being published between a young woman's fifth and eighteenth birth dates, or some approximation of the range between when a young woman starts playing piano and when she is married. This ended up being true, but with a few additional caveats. The first caveat is that outliers can make ranges look much larger than they are. To account for this disparity, I also looked to the interquartile range to help reduce the impact of outliers on the range. Before breaking down volumes specifically, it is worth looking at a specific case to see the range disparity in practice.

The exemplar for date range disparity in this particular case is Volume 65 (from the UNC collection), which is one of the larger volumes in the collection. There are 62 dated works in the volume, which contains 84 works in total. The earliest work in the volume, "Oh Don't You Remember" by Nelson Kneass, was published in 1848, but most works in the volume were published after 1900 (the median is 1900 exactly). In fact, this volume contains 48 of the 62 dated works as published within 10 years of each other: 1897–1906. Additionally, 1906 was the final year of any published work in the volume and happens exactly once. This leaves 14 works yet unaccounted for, and the range of dates for these works is much larger than the rest of the volume. Counting backward from our last 1896 work, the range is 1895, 1894, four works in

1891, 1874, 1869, 1856, 1854 (twice), 1853, 1849, and 1848. The works from the 1890s make

sense, but the works from 1848 are generations off. In this case a generational hypothesis could

offer an explanation: works from 1869–74 predate the nexus of works contained in this volume

by about 25 years and works from 1848–56 predate the 1869 cluster by another 20 years. This

could indicate a gift from a parent or an older relative. Instead of thinking of this volume as a

particularly large range of works, with a spread of years, it makes more sense to look at the

clusters of works and examine where those lie and compare those clusters to the rest of the

volume.

The type of detailed analysis as it pertains to Volume 65 largely lies outside the scope of

this chapter; here I want to continue my focus on larger analyses. With that said, some context on

the particulars is both helpful and necessary, and Volume 65 helps provide a framing for the

following observations. The median of the date range of a binder's volume is 17 years, and the

mean range of dated works contained within a binder's volume is 20.2 years. Within that range,

there tends to be a much narrower range of dates once outliers are factored out of the equation.

The average interquartile range of works is less than 6 years (5.7), and while some volumes have

a large range generally, this does not appear to correlate strongly with the interquartile range of

works. In other words, outliers are just outliers here. The data does not model a bell curve well

due to the strong skew towards the latter half of the date range. Indeed, the average distance

between the earliest dated work and first quartile is only .14 years smaller than the average span

of the first quartile to maximum (8.86 years to 9 years). To say that another way, if a binder's

volume spans the average of about 18 years, one quarter of the works are in first nine years, and

the last three quarters of dated works are from the last nine years.[82]

---

[82] In practice, this makes sense and one can imagine how this situation arose. Anecdotally, I played piano growing

Looking at Volume 65, I see the evidence of the same type of musical practice literally bound in place. These types of temporal distribution by ten years, 20 years, or longer all highlight the nature of taste, but also of families sharing a habit and practice. Binder's volumes are themselves a document that testifies to the shared communal experience of music and how it was passed between families, and the dates are one way to see this in action.

Where Do Volumes Come From?

Having addressed some questions of time with respect to binder's volumes, this section turns to where music is created. The BVRI offers a powerful way to explore publishing and how it changed over the nineteenth century. The database can show which publishers were incorporated into binder's volumes at various times. In its current form, the BVRI can help document regional variance. As the BVRI expands, I intend to include more repositories and augment the depth and scope of these answers. For now, the data clearly shows the migration of publishing westward, which is one aspect I foresaw at the beginning of the project. I did not, however, expect the growth to be quite as pervasive from the Northeastern US to the South, but this was just as significant as the Westward expansion of publishers contained in the UNC volumes.

---

up (still do as often as I can!), and most of my piano collection was made of music from a few different places. The largest segment of my sheet music was, outside of pedagogy books like the Faber series, music that I wanted to purchase and to play. A few of the standalone songs included "Superman" by Five for Fighting (2000), "Drops of Jupiter" by Train (2001), some Matchbox 20 from the early 2000s, and Coldplay. There is a strong cluster of music that I played as a young pianist from when I was in high school. The second cluster of music is gifts from relatives – primarily my mother and grandmother. In my case, my grandmother taught music in a small town and had the larger collection, but the sheet music I received from them was decidedly older. One example, Mr. Bojangles, comes from around 1970, which reflects the generational gap. I also received some swing music of the 1940s and 50s, and some 1950s polkas. The music clustered around a few groups though: my music of the late 1990s and early 2000s, my mom's music of the 1970s and 1980s, and my grandmother's music of the 1940s and 1950s. Her collection was a bit more diverse—she played and taught music before she passed away—but the trend held. There is not space to differentiate my experience from that of a binder's volumes compiler, but this anecdote does show how generational collecting took place for me and might have for some compilers.

140

More speculatively, I predict that more data will reveal that publishers and binder's volumes will fall into geographic groupings. Two competing forces shape these historical trends. The first is the increasing ease of publishing and the aforementioned rise of publishers throughout the United States. With more publishers comes a trend of more music being published closer to home—a trend mirrored in many industries throughout the world. The second factor is advancements in mass production, which made it easier for large publishers to print even more copies of a work than before, and for a cheaper price. I hypothesize that the increase in the number of local publishers will outweigh mass production, but I do not yet know the degree to which the BVRI will support these claims as it expands, or the strength with which I will be able to argue either case based on distribution of samples. At the moment, research into provenance for each binder's volume is ongoing, and that is a part of the scaling up of the BVRI that I discuss in Chapter 6. As I continue to pursue these answers, I revised the research question to work within the bounds of the current data, which yielded the questions below.

What is the Geographical Spread of Publishers Incorporated in Each Binder's Volume and How Does It Change Over Time?

To examine the regional spread of publishers, I took the musical works' median (1850) and looked at publishers before and after. In particular, I took note of the publishing city listed first and how the publishing city changed over time.[83] In the pre-1850 set, there are 1329 examples and in the post-1850 set 1401. A work needed to have both a location and date to be included in this example. The samples are roughly equal size, and so I looked at the change in works from before to after 1850, and a few clear trends emerged.

---

[83] Chapter 2 discussed the publishing for cities, and the first city was the one that I consistently used as the anchor for a publisher.

First, a few cities dominated the pre-1850 data set, as shown in Figure 31. Of the 1,329 works, 1,292 (97.2%) were from only four cities: New York (576), Philadelphia (400), Boston (183), and Baltimore (133). I had expected a very strong concentration of works, but this was notably stronger than I had expected. The next largest contributor of works is St. Louis with 7 total works. Only 19 cities published a work dated before 1850 that found its way into a binder's volume, reflecting the vast centralization of publishing in the northern United States. The remaining 14 cities are Charleston, SC (3); Cincinnati, OH (4); Cleveland, OH (1); Hamburg, Germany (1); Kentucky (city unlisted, 1); London, England (1); Louisville, KY (6); Maryland (city unlisted, 2); New Orleans, LA (1); Newark, NJ (1); Paris, France (1); Providence, RI (3); Richmond, VA (3); Toronto, Canada (1); and Washington D.C. (1). Also of note is the paucity of European publishers (at least with dates). 12 cities in what is today Germany did have works published there that later found their way into binder's volumes, but nine of the 12 cities' works were undated.

Locations of works published before 1850

St. Louis, MO
0.5%

Philadelphia, PA
30.1%

Paris, France
0.1%

Baltimore, MD
10.0%

Boston, MA
13.8%

Cincinnati, OH
0.3%

Louisville, KY
0.5%

New York, NY
43.3%
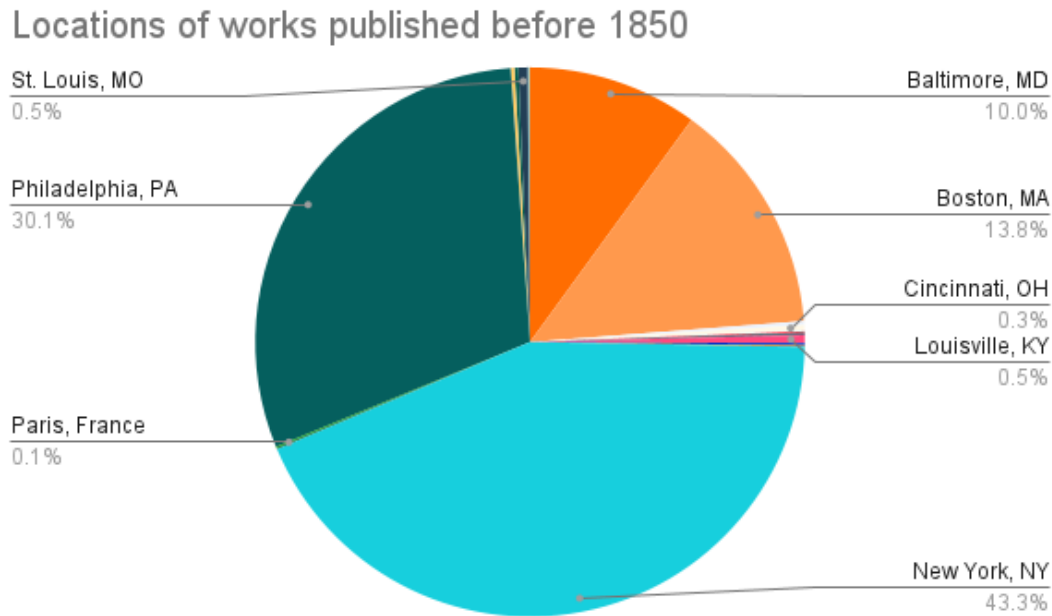
Figure 31: Locations of works published before 1850. Note in particular the prevalence of four cities.

The publishing shifts away from the aforementioned publishing centers in a significant way after 1850, as demonstrated in Figure 32.. Now the four relative powerhouses of publishing only account for 78.7% of included works, as shown in the breakdown of Figure 5. New York now accounts for 472 works, Philadelphia 200, and Baltimore 84. Boston is the only publisher in this set to see its portion rise notably, and rise it does (from 183 to 347, a 90% increase). These four publishers still account for a sizeable portion of the included works, but several other publishers start to rise. St. Louis in particular surpassed Baltimore in included works at 113—an astronomical leap from the 7 works predating 1850.
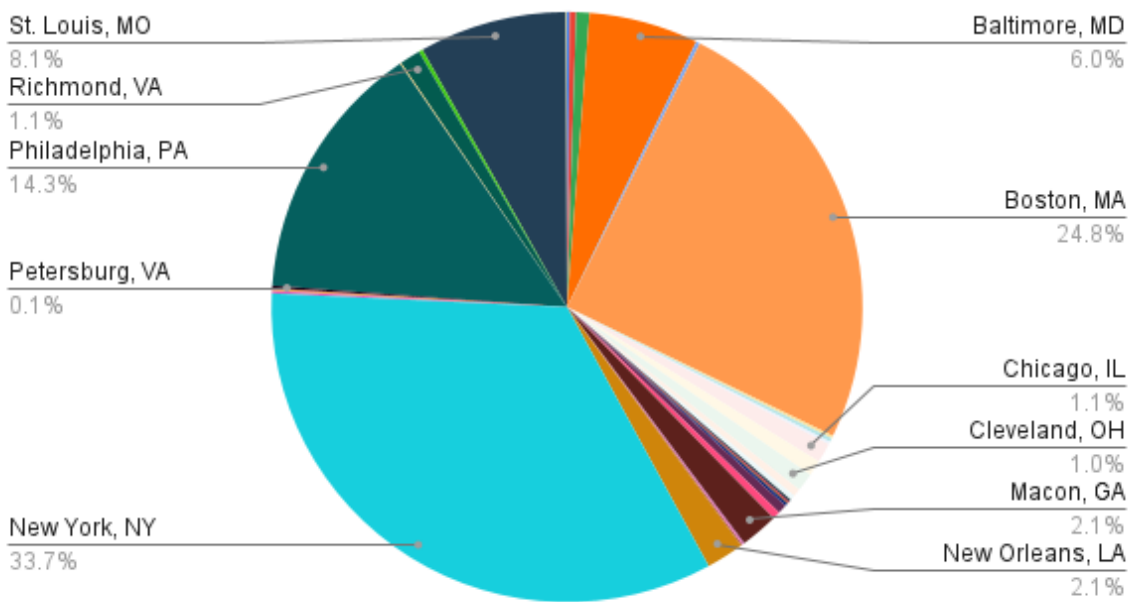
Figure 32: Locations of works published after 1850. In particular, note the inclusion of several more publishing centers, including St. Louis.

Within this rise of included cities is also a diversification of publishers representing a broader cross-section of American cities. Now 36 cities have included works (nearly double the 19 previously), and the South and Midwest see much greater representation in their inclusion. These trends obviously have some bias towards where volumes were collected, but they are a

143

noteworthy starting place for examining the rise of publishers as industry expanded beyond the East coast and older presses became available for repurchase. Georgia, which had no works pre-1850, now has 39: 29 from Macon and 10 from Augusta. Virginia had 3 works out of Richmond before 1850, and now 18 works afterwards (16 from Richmond and 2 from Petersburg). New Orleans, a major publishing hub in the South, saw 30 total works included in UNC volumes, and of those 30 only one was published before 1850. South Carolina also sees increased inclusion, though not quite as significant a rise as several other Southern states.

The Midwest presents a tale of even more rapid growth in publishing. St. Louis's meteoric rise in publishing has already been noted,[84] and Mexico, MO's lone work after 1850 moves Missouri in total form 7 works to 114 – a 1,529% rise. Ohio, Michigan, and Illinois also saw dramatic increases in numbers of published works included in volumes. Five works were published in Ohio before 1850, and that number increases to 26 after 1850.[85] Chicago is the only city in Illinois to see works included in a volume housed at UNC, and all 15 works were published after 1850. Michigan only had 8 works, but like Illinois saw all 8 works published after 1850. Even a state like New York saw publishing centers rise outside of New York City itself. Before 1850, the only works published in New York state were cited as published in New York City. While 104 fewer works were published in New York City, 11 were published outside of New York City in places like Albany (5), Buffalo (3), and Northport (1).[86]

Multiple forces likely drive the spread of publishing to the South and Midwest through the nineteenth century. One such force is industrialization, or the spread of industry through

---

[84] And is studied in detail in Krohn's book

[85] 14 in Cleveland, 12 in Cincinnati.

[86] Two were published in Brooklyn, which is now a part of New York City (and has been since 1898) despite retaining its own status.

places that previously did not have such technology. The South was particularly agrarian before 1850, and the decline of slavery (hastened by the Civil War) taken with mechanization, and the ability for fewer hands to accomplish more work, together helped to drive a rise in the population of cities in the South. With cities came the rise of a middle class, who with the upper class pursued leisure activities (like music) and could support local publishing that catered to local taste. This increase in demand coincided with the rise of print culture, and the ability for local printers to meet said demand profitably. Future research may help disentangle the magnitude of each force, but both undoubtedly played a role in the spread of music publishing.[87]

The Midwest would have experienced similar growth in population, though no such centuries-old agrarian system existed through the Midwest. The area obviously supported farming (and still does), but cities rapidly grew in places like St. Louis and Chicago, and such cities supported industries, just as was the case for Southern cities. Thus, the dual drivers of publishing—industrial expansion and a growing population of consumers—helped fuel Midwestern printing. The details are seen below.

Is a Westward Migration of Publishing Over Time Noticeable in the Data?

Throughout the data listed above, two strong trends present themselves: a move from larger centers to smaller ones (noted above), and a move from the north Atlantic to the West and South. The four cities are nicely grouped along the East coast, and nicely spaced apart to service this area with music. The distance from New York to Philadelphia is 97 miles; Philadelphia to

---

[87] For more on publishing in various places, see Ernst C. Krohn, *Music Publishing in St. Louis*, Ed. J. Bunker Clark (Warren, MI: Harmonie Park Press, 1988), Christa Pehl, "Men, Women, and Their Manuscripts: Musical Culture in Pennsylvania in the 1790s" (PhD diss., Princeton University, 2017), and Frank W. Hoogerwerf, *Confederate Sheet-Music Imprints*, 21 Vol. (Brooklyn, NY: 1984).

Baltimore 98 miles, and New York to Baltimore about double that at 215 miles.[88] Just as

Baltimore is 215 miles southwest of New York, Boston is about 215 miles northeast of New

York. The four cities are nearly colinear, and all four have easy access to the Atlantic ocean via

waterways, if not a port on the ocean itself. It would be difficult to choose four cities at random

and so efficiently service the Northeastern United States. Looking at the state-level geomapping,

the weight of the states in the Northeastern US is clear. The four contiguous states demonstrate

how great the weight of publishing was in one corner of the US, and many of the lighter states

are included for publishing only a handful of works (as shown in Figure 33).
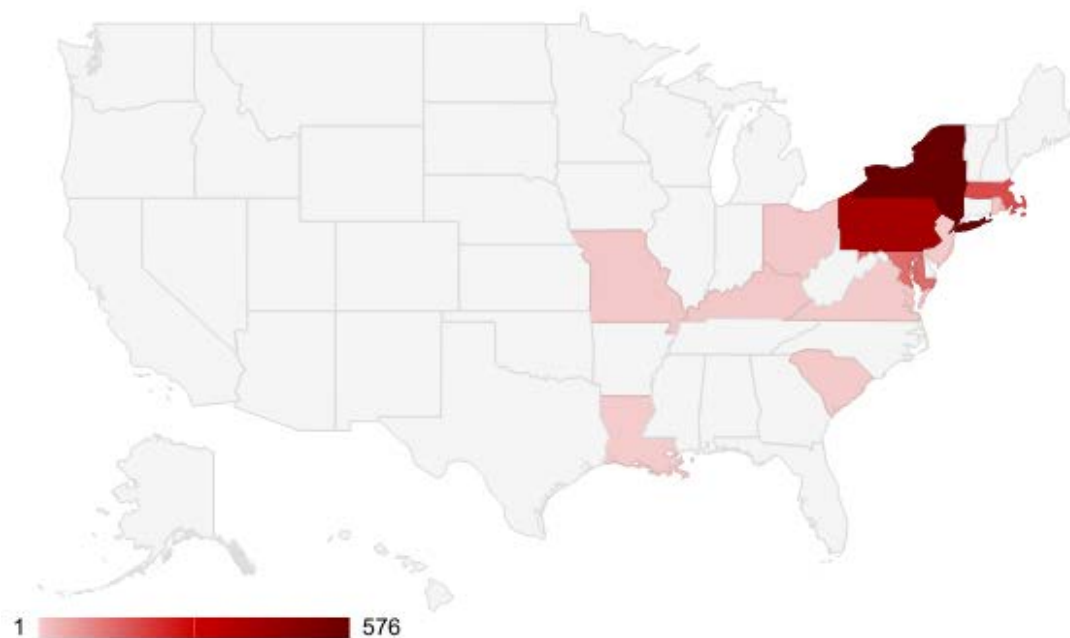
Figure 33: Publishing locations of works published before 1850 by state. The gradient is difficult to see, but the cluster in the Northeast is much darker than other locations.

The examples after 1850 show how the migration occurs and where new publishers are

established, as shown in Figure 34.[89] At a state level, it is clear how much New Orleans, Macon,

---

[88] Apple Maps

[89] Figures 33 and 34 are made possible using Google tools, for which I am very grateful.

and St. Louis in particular stand out in this set as publishing hubs moving West and South. I anticipated New Orleans and St. Louis, but I was surprised by the presence of Macon as a predominant publisher, and I look forward to seeing how increasing my sample size impacts Chicago's representation in binder's volumes going forward. These are a few questions I look forward to exploring as I continue to pursue westward expansion through additional repositories.
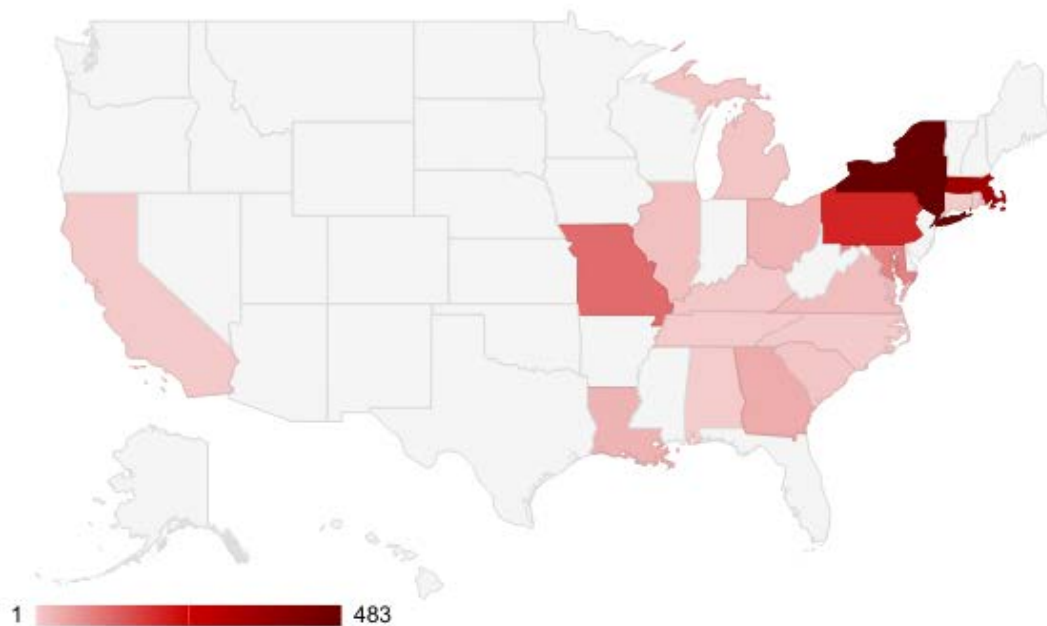
Figure 34: Publishing by location after 1850.

What Regional Variance in Songs can be Found in BVs? How Does this Change Over Time?

With these two questions, we begin to consider not just individual works, but also collections of works, which is a major benefit a database on binder's volumes can provide in contextualizing music consumption. First, volumes tend to draw on a small number of locations for their works. Of the 125 volumes, 113 draw on works published in more than 10 locations, and thirteen is the maximum here, but most drew on about 5-6 locations. The mean number of locations is 5.72 and the interquartile range is 4–7 locations per volume. The locations that

factor prominently into the volumes are the same ones as mentioned before: 56 volumes feature

New York most frequently, 33 Philadelphia, 13 Boston, and 12 Baltimore. Saint Louis and

London are the next most common modes of works published per volume, and each occurs in

two of the 126 volumes contained in the UNC repository.

One factor I hoped to pursue was where the locus of an average work was published, and

this proved to be skewed by the European publishers and dramatic distance of the Atlantic.

Despite the prevalence of American publishers, the pull of Europe places the mean latitude and

longitude in the Atlantic Ocean, just off the coast of Boston. Figure 35 shows the mean latitude

and longitude on Google Maps for works published before 1850. The location is not far from

Boston, but it is not representative of America's publishing. Factoring out Europe's publishing,

the center of American publishing remains on the East coast but does move inland.
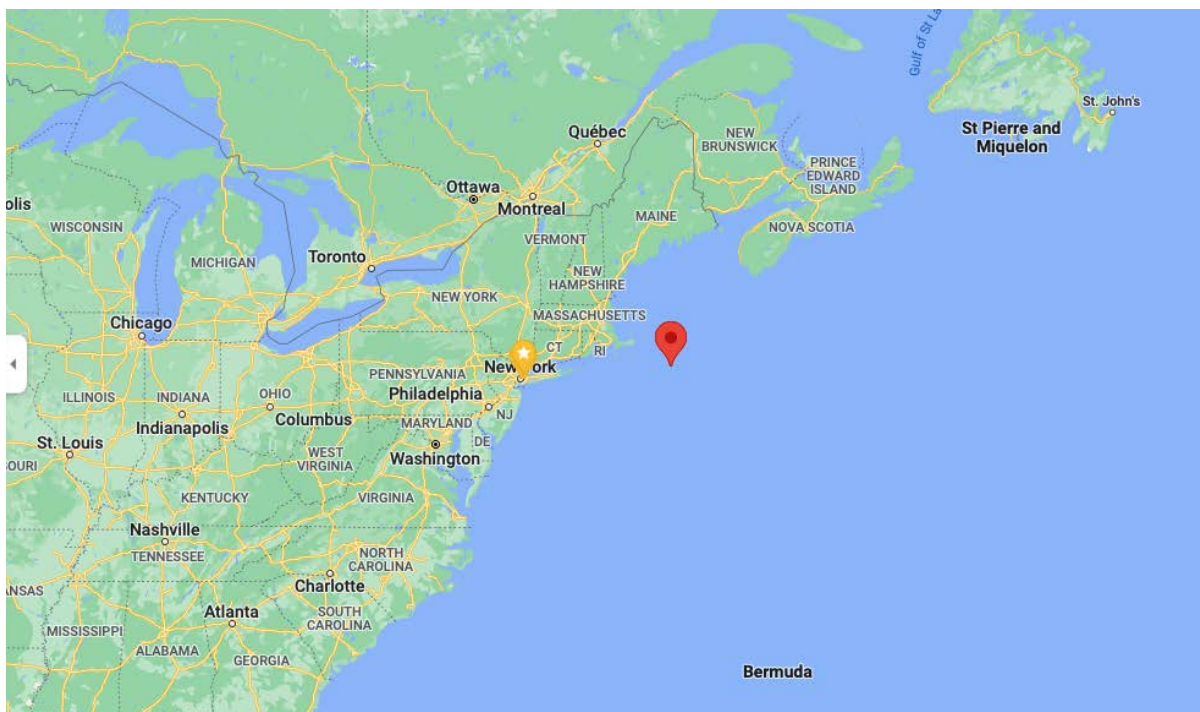
Figure 35: Geolocation plotting the center of publishing in the United States before 1850. While it is in the Atlantic Ocean, the prevalence of American publishing is clear.

The map in Figure 36 shows that the center of publishing in American publishers of binder's

148

volumes from before 1850 moves notably inland, toward Philadelphia. Here the standard

deviation is also interesting, which for latitude is approximately 4.5 (206 miles at the mean

longitude of 40.16296) and the longitude is 2.21 degrees (133 miles at the mean latitude of

75.03951). All of this is to confirm the locus of publishing as the Northeast US, with a strong

clustering within New York, Boston, Baltimore, and Philadelphia. Boston is the only city that is

not within one standard deviation of Boston, and yet Boston would be within that one standard

deviation if Europe was considered for the data set.
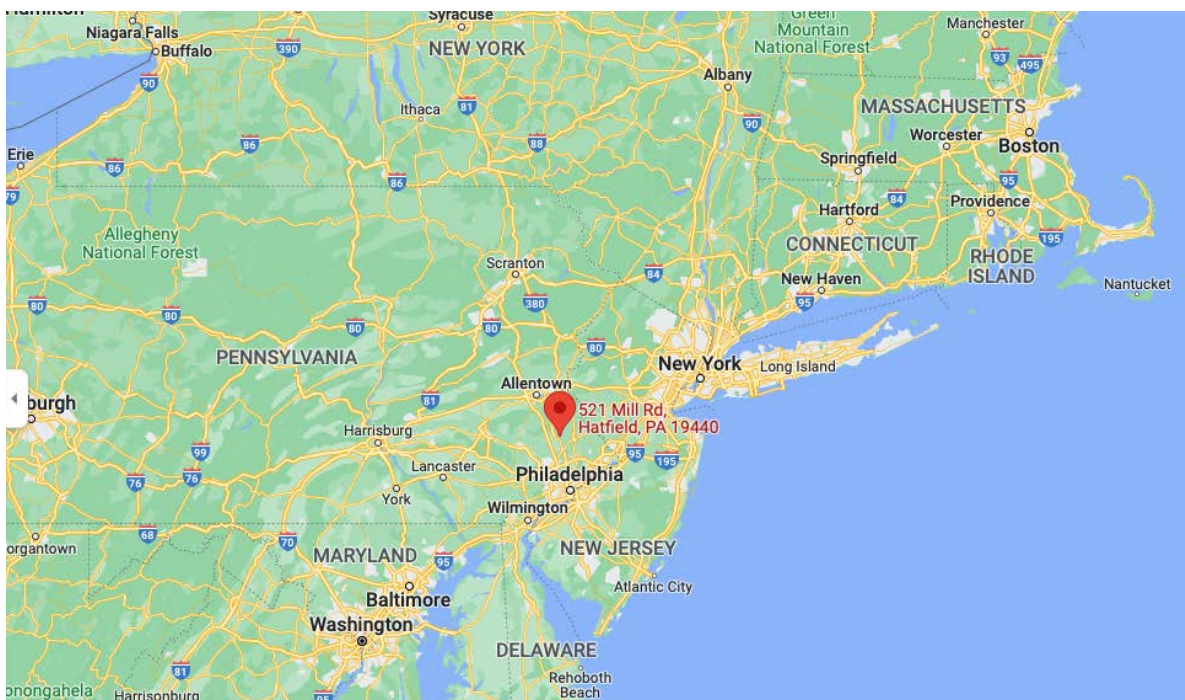
Figure 36: The center of publishing before 1850, with European cities excluded from geolocation. This map, combined with the small standard deviation, demonstrates the center of American publishing and the small geographic location before 1850 and industrialization.

To examine the change in publishing over time as it relates to the geography of a binder's

volume, I split the data in half, which for the maximum dates of the volume splits between 1853

and 1854 for the current data set. The split is particularly noticeable in publishing. The median

number of works in a volume bound no earlier than 1853 is 4 works, and the interquartile range

is 4–5 works. This means that 75% of dated works found in volumes published after 1853 were

published in less than 5 places. Remembering the prevalence of the four publishing centers mentioned above; these four locations are almost always New York, Philadelphia, Boston, and Baltimore. The thing that surprised me in this set is that volumes seem to draw on most of these four, drawing on at least four different places more than 75% of the time.

This expands in the volumes bound in or after 1854 (at the earliest). The average number of locations is just over 7, which nearly doubles the pre-1854 average. The interquartile range is 5–9 locations per volume, further indicating the diffusion of publishers over time. Where 80% of volumes whose earliest bound date is 1853 or before had at most five publisher locations, 85% of volumes bound after 1853 have at least 5 publisher locations included in the work. This figure may be the most telling in terms of the spread of works being included in binder's volumes and the proliferation of music publishing throughout the United States.

Who is Included in the Volumes?

The next logical question is which composers are included in the volumes. In attempting to answer this question, I have selected two questions to focus on. The first asks who is included in the volumes, and here I offer a preliminary "Top Ten" list of sorts, with the obvious caveat that such a list is admittedly quite likely to change as more volumes are added. With the BVRI's expansion, such lists will undoubtedly evolve, and it will eventually be possible to focus on quarters of the 1800s, if not decades. Additionally, a growing dataset will likely also reveal regional variance in popularity. For now, the first section previews the question. The second section addresses when a composer was popular. The nineteenth-century was a time of artistic canonization in Europe, with a historical move by publishers and musicologists who began to anthologize, and to celebrate (if not deify) deceased composers like Mozart and Beethoven. Here I interrogate such questions in binder's volumes.

Which Composers were Included in Binder's Volumes?

In answering this question, I found it most appropriate to refer to the table of composers. Table 10 shows the composers most frequently included in binder's volumes at UNC, and the number of instances for each composer based on the order of their listing as cataloged. The composers that just missed the list were Stephen Foster (46), Henry Bishop (who comes up in Chapter 5, 44 works), Michael Balfe (also 44), Johann Strauss and Giuseppe Verdi (both at 43), and Henri Herz at 42.

Table 10: Top 10 most frequent composers in binder's volumes of the UNC data set. Columns first, second, and third refer to the order that composers were listed on the work.

Last Name	First Name	Born	Died	First	Second	Third	Total
Bellini	Vincenzo	1801	1835	77	5	0	82
Donizetti	Gaetano	1797	1848	73	2	1	76
Grobe	Charles	1817	1879	63	0	0	63
Blessner	Gustav	1808	1888	61	0	0	61
Auber	Daniel	1782	1871	53	2	0	55
Beethoven	Ludwig	1770	1827	52	0	0	52
Russell	Henry	1812	1900	51	0	0	51
Glover	Stephen	1813	1870	48	2	0	50
Wallace	William	1812	1865	49	0	0	49
Beyer	Ferdinand	1803	1863	48	0	0	48

A few things about the table stand out immediately. First, about half of the composers are still recognized as canonical or otherwise of significant historical importance. Beethoven stands out in this regard, but so too do Donizetti and Bellini in relation to Italian bel canto, and D.F.E Auber is still known for his contributions to French opera. The other six are largely absent from our typical histories; outside of these four, none of the other six appear in the index of Norton's *A History of Western Music,* 10th edition.[90] Ferdinand Beyer was discussed in Chapter 2, but the

---

[90] J. Peter Burkholder, Donald Jay Grout, and Claude V. Palisca, *A History of Western Music*, 10th ed. (New York: Norton, 2019).

other five composers on the list may be unknown to readers. Because of the general lack of information otherwise available about these composers, it is worth taking a moment to give a brief sentence about the rest of the top 10 list.

William Wallace (1860–1940) was Scottish, and with a death date of 1940, his inclusion on this list is surprising.[91] Grobe is a German composer who eventually settled in Delaware.[92] Both Wallace and Grobe have OMO articles, which is more than can be said for Glover and Blessner. At the moment, the only thing I know of Glover is that he had compositions published in England sometime in the middle 1800s, Blessner is another European composer who published in England, and he and eventually moved to America. More is known about Henry Russel, and I discuss him in Chapter 5.[93]

The purpose of this top 10 list is not to claim that these 10 composers are the most influential in an American history of the 1800s; that is almost assuredly false. I do seek to raise questions about the general lack of information available on some of the above composers (particularly Gustave Blessner) and their collective absence from a history of American music. Without adding more repositories, this section remains a speculative analysis of a subset of American composers, but it is a useful time to lay an initial marker and to map the evolution of such a list as the data set expands.

---

[91] Duncan J. Barker, "Wallace, William," *Grove Music Online.*2001; Accessed 21 Oct. 2022, https://www-oxfordmusiconline-com.libproxy.library.unt.edu/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000029839.

[92] Ann L. Wilhite, and Charles S. Wilhite, "Grobe, Charles," *Grove Music Online,* 2001; Accessed 21 Oct. 2022. https://www-oxfordmusiconline-com.libproxy.library.unt.edu/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000011804.

[93] John A. Stephens, William McClellan, and Andrew Lamb, "Russell, Henry." *Grove Music Online*. 31 Jan. 2014; Accessed 21 Oct. 2022. https://www-oxfordmusiconline-com.libproxy.library.unt.edu/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-1002258090.

How Does the Popularity of a Composer Relate to Their Death Date?

A consistent theme in the nineteenth century, moving into the twentieth century, is the beginning of music's canonization. That is, music starts to become a practice that develops a sense of historical consciousness driven by the idea of the timeless, great composer that publishers attempt to market to consumers, and that audiences in turn take to heart as a paradigm for their judgments of taste. This marketing occurs in a manner that simultaneously aligns with this historical thrust of canonization and saves publishers money by allowing them to reuse existing materials.[94] Thus, there are two parts to the question: do publishers continue to publish works long after a work is completed, and do collectors purchase works after they are used? James Deaville has done excellent work answering questions about the most popular composers, particularly Liszt.[95] This approach offers a methodology to contextualize the question more broadly.

Method: Establishing a Metric for Composer Death and Binding Date

Within the data set, there are significant limitations on how one might answer either question. It is nevertheless worth exploring the extent to which the data might address such inquiries. As a caveat, I would emphasize that this section becomes the most technical in this chapter as I seek to preserve the distinction between what I claim with the data on the one hand, and the inferences that I hope to highlight more tentatively along the way.

The first step is to establish the date for the work being used, and this date is framed by the earliest published date on the one hand, and the earliest date a work could be bound on the

---

[94] Peter Mondelli writes about this phenomenon in his forthcoming book, particularly in the first two chapters.

[95] James Deaville, "Publishing Paraphrases and Creating Collectors: Friedrich Hofmeister, Franz Liszt, and the Technology of Popularity," in *Franz Liszt and His World* (Princeton: Princeton University Press, 2006). https://ebookcentral.proquest.com/lib/unt/detail.action?docID=617585.

other. With respect to the date of publishing, we look to the title page. The title page of a musical work is commonly the marketing page, and it changed frequently: much more often than plates did. Title pages were not often engraved; it was much easier to adjust information for a specific occasion. Thus, the copyright date of a work is updated as per the title page and the new publisher's name (see the Firth, Pond, and Hall family referenced in Chapter 2), and can be a metric for the time a work is used and sold. I say can, because this only happens in some cases, and U.S. Copyright Law is inconsistent in its writing and its application when it comes to musical works.[96] Some title pages would be used for years, while others certainly present as ephemeral in their reference. Thus, the published date provides the earliest date a work is produced and sold; however, this is not a complete picture; it is one side of the frame.

The next side of the frame that I attempt to construct as a means of answering this question involves taking the last published date of a work in a volume and using it to establish a rough date when the work was assuredly in use. There is good reason to focus on this method. For one, the published dates of works in the volume tend to cluster towards the last date in the volume, as one might expect from a pianist moving beyond their learning stages into more difficult sheet music. The last published work in a binder indicates a time when works are being added, and the clustering effect mentioned above strongly suggests that multiple works are being added at this time. It is not a terminus ante quem, as a proper frame might hope. It is, however, a useful marker for when the works had been collected.

In this instance the volume functions as an anthology. It is a collection of works published before this point of binding that become a unit together. Considering the volume as an

---

[96] Music was not protected by copyright in the United States until 1831, and the concept of fair use did not become codified until 1976, leaving for more than a century of murkiness in the middle. For more information on music copyright, see "Historically Significant U.S. Copyright Legislation," *Music Library Association*, Accessed 20 September, 2022, https://www.musiclibraryassoc.org/mpage/copyright_hsl#1831.

anthology for this query provides useful context on how a composer's work is absorbed and repurposed, approximately when that reabsorption occurs, and where it fits in its current context. Once placed in a volume, the work has framing of other works and is set in a time and place, and the volume's terminus post quem (the latest date in the volume) is a separate frame. It faces forward, as the first frame did, but the space between the publishing and the earliest binding provides a context for when composer's work was used. In some instances, the gap is narrow: a work is published in 1886 and bound into a volume as early as 1886 or 1887. As is seen below, this frame is at times 30 years wide, which shows the space between which it is produced and first might have been bound.

The composer's death date is taken here as a very rough approximation of the end of musical productivity, and this date provides the metric to gauge usage within the frame established above. Given life expectancy, most composers were working within the last 10 years of their lives, but some did have fifty-year careers, or composed closer to the first portion of their life than their death. I considered taking the composer's birth date, or birth date plus 20 years, as a place to measure productivity but ultimately decided against this for a few reasons. First, more death dates are known than birth dates, providing a richer field of dates. Second, composers achieve popularity at varying points in their life, just as they compose closer to or further from their deaths. Taking the birth date as a starting place, whatever the appropriate distance to count after their birth still presents some of the same problems with a window for analysis, and I find the problems greater with the birth date than death date. Finally, I contemplated taking both dates and creating a window-type of approach, but this also fell to challenges. Too few composers have both a birth and death date known, and it would shrink the sample pool considerably. The shrinking would then bias famous composers who naturally have works that might live beyond

them, skewing the data. Ultimately, I concluded that using the death date as the metric for both comparisons was the logical step in studying the degree of canonization in consumption through binder's volumes.[97]

Having framed the question and offered some observations on methodology, I must clarify the critical limits of what I can and cannot claim in this analysis. This method is an approximation for the space in which a composer's work is being produced and used, both in how the publisher produces it and the binder consumes it. A work might sit unused for a generation before it is included in a volume by a daughter (or in the case of Volume 65, apparently a granddaughter). Even in the case of a work that sat so long, it is still in circulation, and the subsequent generation still found value enough in the work to bind it, and so I consider it relevant here. It is also possible that a pile of music sits idly for five or ten years before it is bound. A young girl may step away from the piano and purchase no music while away, and then the work is bound after she has been gone a decade. The volume would still reflect the time the music was in use, though perhaps in a different way than initially proposed. The information is still worth pursuing in this manner, and the approximation holds. Probably most concerning is the potential for outliers in the maximum dates, or in the composer's dates. Outliers are the bane of statistical analysis, and the best way to combat them is to add more data. This is the mission of the BVRI, and an analysis with 10 repositories will assuredly yield more information on this question than the analysis I present here. Nevertheless, the methodology here is worth understanding and worth establishing now so that changes in the data with more volumes and more repositories can be fully explored, and any change in the answer due to outliers or simply

---

[97] This is an estimate, which is helpful and accurate, but is also subject to revision in specific cases. Generalizations are important and useful but should always be subject to updates and revision as more information becomes available and as the general gives way to the specific.

anomalies in this data set can be interrogated. Even with all the problems addressed, this analysis remains an attempt to set rough markers on the relationship between a composer and publisher on the one hand, and the composer and binder on the other. Within these markers lies the probable space for active use of the works, and the query loosely binds the space for later exploration.

Making the Comparisons

Thus, the three figures that I use to examine composer activity as compared to their time of productivity are (1) the composer death date, (2) the date published, and (3) the earliest date the work could have been bound. The discussion below compares the composer death date to the date published and earliest bound date. Along the way, I discuss sample sizes and other data that contribute to the analysis.

First, I compared the composer death date and the date published. The average gap between published date and death date is about 23 years, meaning that composers published a work on average 23 years before they died. The mean difference is 21.9 years, the median 24.2 years difference on average. This defied my expected hypothesis that the average composer death date was going to align with the published date on average.[98] The sample size is small, to be sure. Only 9.9 works per volume have a death date and a published date, meaning the sample is only about 1,500 works. I hope to explore this trend further, But based on preliminary results, most composers were alive when the works in these binders were published.

With this information, I turn to the space between composer death and earliest bound date. While the difference between the average published date and death date is worth noting for

---

[98] My rationale was based on canonization narratives in which the presence of eighteenth-century repertory in nineteenth century collections might skew the average.

the publishing standpoint, I am also interested in the space between the composer's date and the last date a published work was added to the volume. This figure contributes more to an understanding of usage and consumption than to production. My standing approximation for this is using the last date a work was published as the terminus post quem for the volume. Thus, I examined composer dates with respect to volume dates, which references when the composer was alive to the earliest date a volume could be bound. While the measurement changes exactly what is being assessed, the change is dramatic.

Before elaborating on the change, it is worth exploring in a bit more detail how the methodologies in the two questions yield different types of results. With composers' dates vs published dates, I set up my analysis by restricting the cases to works where both the published date and composer death date were known. This ruled out several works, and even entire binders, where some dates and some composers are known but a work does not have both known.[99] By contrast, the second measure of composers to end dates is going to be looser. This metric considers every composer date in the volume, regardless of whether the published date of the work is known, and it compares that information to the binder's last known published date: the terminus post quem. Thus, this measurement includes many more dates, compares the average of each volume to the last date of publication, and becomes a looser measure. The first metric includes less than 10 works per volume, or only about 25 percent of the total works. The second metric includes closer to 90% of all works, which expands the question but also radically shifts the results.

With this context, the dates for composers' deaths as compared for their potential for

---

[99] Binders 104 and 105 both have several composers and dates known, but in both cases, no single work has both dates identified and a composer identified.

inclusion in a volume shift dramatically from the first measure. Instead of being published 23 years before the composer died, the works are bound about 11 years after the composer's death. The mean is 9.3 years after the composer died, the median 13.3 years. The dates are spread over a slightly larger distribution in this model, with the standard deviation being 15 instead of 13.9 and the overall range being 103.4 years instead of 99.8 years.

In these two ranges of comparing composer dates to publishing dates, what is really being measured? The first range compares the composer death to pieces by year, and thus becomes a measure of when the publisher first produces a particular manifestation of the composer's work. This does not guarantee that it is the earliest manifestation of the work created by the publisher, but it is a marker that the composer's work is being produced by the publisher. This measurement is narrow, and as a narrow comparison will benefit from more instances to become a robust measurement. I interpret this metric as a relationship between the publisher and composer, as the work is generated.

The second metric, which compares all death dates to the terminus post quem of the volume, is a looser measure, but also one that focuses more on the volume and consumer relationship to the composer than that of composer to publisher. By identifying composers as compared to the latest date of a binder, this gets closer to identifying how the work was used and consumed. In this instance, two things are simultaneously true: first, there is a great deal of ambiguity in the metric; and second, this is a very useful approximation of the time between a work's production and binding, given a recognition of the caveats. The date of binding is approximate, and it could be a few years after the last published work date. This is the most substantial caveat, and it is possible that a work is bound more than a decade after its last published date, which would mean that the approximation of a work being bound about 10 years

after the composer's death is most likely an early approximation. There are other factors; the work is assuredly purchased and used before binding, which to some extent counterbalances the potential for binding years after the last work, while simultaneously loosening the range effectiveness somewhat. With these factors loosening the range, comparing the average composer date to the final does provide a useful metric for the age preference of a bound volume.

While further research is needed, I nevertheless believe this data indicates a trend of music being printed and reprinted long after it was composed. Deaville's tables at the end of the chapter her wrote for *Franz Liszt and His World* offer a detailed indication of when certain copies of Liszt's works were printed as compared to their initial print run. This information, taken with Liszt's biography it is immediately evident that large print runs took place well after his death.[100] The question that I contextualize is this: was Liszt a one-time phenomenon, or was his fame the driver of these late runs? The information that has arisen thus far in the BVRI suggests that many composers' works had life spans that existed far beyond their deaths. Parlor music has a reputation for its ephemeral nature, but this data shows a longevity in binder's volumes and in consumption that belies assumptions about fame and canonicity driving longevity. Further research is indeed needed, and I look forward to the BVRI functioning as one tool with which to answer these questions.

How Does This Help Us?

This is the real question. Up to this point, I have addressed questions that are specific in nature, and in some way quantifiable. None of the numbers matter (at least, not to musicologists generally) if they cannot help address the larger questions of music and culture at the time. This

---

[100] Deaville. Tables 1–10 show print runs.

final section addresses some of the ways that binder's volumes quantitatively can help address music research on the nineteenth century broadly. First, I address issues of gender in the nineteenth century before turning to nineteenth-century America more broadly.

How Can the BVRI Enhance Our Understanding of Women in Nineteenth Century American Music?

This project focuses on collections of music bound primarily by women, and the past half-century of research has thoroughly demonstrated the overrepresentation of male figures in music research. From ideas such as the (male) genius and "woman" composer to deeper questions of embodied performance,[101] music research has long given voice to a male perspective at the expense of a holistic, balanced examination of music in the past. Columnists in papers were frequently men; composers (and especially published composers) are most often men; publishers are very frequently men. By contrast, binder's volumes offer a glimpse of women and their tastes, views, and preferences. To be sure, male preference factors into the binder's volumes. Some works are gifts from fathers and friends, and some few volumes did belong to men.[102] With this caveat, the project assuredly accomplishes the goal of starting a process that sheds needed light on the actual music women use.

Details on the gender of collectors is something that I plan to add as the BVRI progresses outside of the dissertation project, but right now the focus has been on the works part of the data and metadata, and less on the provenance of the women who composed. This is a question that can be answered by the BVRI in the future. As the database grows to incorporate the genres of

---

[101] Suzanne Cusick, "Feminist Theory, Music Theory, and the Mind/Body Problem," *Perspectives of New Music* 32, no. 1 (Winter, 1994): 8–27.

[102] Jeff Kallberg explores the question of male projection and "female" compositions in his work "The Harmony of the Tea Table: Gender and Ideology in the Piano Nocturne," *Representations*, No. 39 (Summer, 1992): 102–133.

men and women, differences in collections will become increasingly obvious, and if enough volumes bound by men become available, any differences will become statistically significant.[103] Additionally, the repertoire will grow in scope to the point that it will serve as a comparable foil to other records of music consumption at the time (sales records, concerts performed, newspaper reviews, etc.). With that said, research on provenance lies outside the scope of the project at the present time.

How Can the BVRI Provide a Look at Music Consumption in Nineteenth-Century America?

Binder's volumes offer material evidence of the consumption and use of music. Research into music history tends to focus on the genesis of sheet music as the object of performance, due in no small part to this being the way we today encounter music of the past. The entire experience of music in the literate tradition from before 1890 is mediated by a combination of paper and performance. Other examples of musical recordings do grow through the beginning of the twentieth century, but live performance or a paper-documentation of that performance via music or written description remained the default musical experience for anyone born before 1900. Chapter 1 discusses the transition between experiencing live performance as the primary means of musical interaction and reading about it in print, but for us today the live performance of these pieces in their contemporary setting is obviously not an option. This leaves three primary ways that scholars come to understand the music before recordings were made, and these distinctions are worth interrogating to understand how we understand music. I group them into three categories: sheet music, treatises, and reviews.

The first, and most common way that people today interpret musical experience is via the

---

[103] The ambiguity of this question is one reason for the project. There are binder's volumes by men, but how many remains an open question that will take years, or decades, to properly answer.

printed instructions encoded in an agreed-upon format. This document is a prescriptive document; it tells someone else what to do to execute a performance, but on its surface, it is not a descriptive tool of what happened when someone attempted to play the work. As noted in Chapter 1, the rise of printing before the advent of recorded technology makes this particularly true in the nineteenth century, and the rise of streaming and video have all but ensured that print will not be the primary method of interacting with music again. Typically, printed instructions encoded in an agreeable format denotes sheet music of some sort, but lead sheets and paleographic musical notation would be included in this category.[104] Other examples of such instructions would be figured bass and continuo parts. In short, I consider this category of documents to be any piece of paper that attempts to directly convey a means of realizing a musical act. This is the substance of binder's volumes and is what a typical person means when using the term "music" to describe paper.

The other two types of paper that aid in musical meaning also align with either prescriptive or descriptive accounts of music. When I say descriptive, I mean a category that denotes interaction with the music as performed, not just as it *should be* performed. Prescriptive documents include guidance on how to realize the previous category of music. Theory textbooks and pedagogy are probably the most common item here, but musical treatises and other types of instruction neatly fit here and indicate the way to play music. Such instructions are vital as a means for understanding the way music should be performed. Notation itself falls into this category. Perhaps the easiest way to imagine this is to think of common but often unwritten

---

[104] This is an intentionally awkward description of the sheet music experience, but it is worth highlighting the distance between what music aims to do and what it is. Gary Tomlinson explores the outer limits of notated instruction in his work *A Million Years of Music: The Emergence of Human Modernity* (Brooklyn: Zone Books, 2015): 177–187. His discussions on musicking more broadly are particularly relevant here, but are intriguing nonetheless.

performance practices: swinging the eighth notes, or French Baroque *notes inégales*, or starting

Classical trills on the auxiliary, and so forth. The important thing to keep in mind about treatises

and guidebooks is that they are prescriptive and bound to taste; they represent one person's ideal

and describe music as it actually was with varying degrees of accuracy. At times these texts are

in part campaigns to convince people to change behavior and as such contribute to the

understanding of how music should be understood, but do not serve as a sole guide.

The counterpart to understanding such prescriptive texts is the descriptive texts that

explain what music sounded like to a person. Again, such texts are subjective, and are not

necessarily based solely on the music.[105] With that said, descriptive reviews are useful resources

for understanding what music is valued and how examples are being received by audiences.

Newspaper accounts provide insight into the ways audiences (and the reviewers themselves)

enjoyed a performance, what was lacking, and to some degree how often works were being

performed. This descriptive information that reveals the actual use of a work is crucial to

understanding what music was performed for audiences.

Binder's volumes provide crucial insight into all three of these categories, but especially

the sheet music and the descriptive documents. In terms of sheet music, binder's volumes are

obviously bound collections of music. The BVRI documents the material itself and provides a

mechanism for researchers and performers to look through collections and see tens of musical

examples in each volume, and thousands of examples throughout the BVRI database. This

research will increase access to sheet music and provide a way to pursue and peruse primary

---

[105] An example today is review bombing on Rotten Tomatoes. Historically, Rotten Tomatoes recalls the French press's ire with Wagner and (to some greater or lesser degree) *Tannhauser* in the 1860s. For a historically grounded discourse on political animus and bias as it pertains to nineteenth-century music, see Margaret Notley, "Brahms as Liberal: Genre, Style, and Politics in Late Nineteenth-Century Vienna," *Nineteenth-Century Music,* Vol. 17, No. 2 (Oct. 1993): 107–123.

sources of works that otherwise might be difficult to find.

The power of the BVRI is providing a look at the music as it was used and consumed. While not precise, the relationship between producers and consumers is in some ways analogous to the prescriptive and descriptive documents of the past. In describing this relationship, I do not mean to map the exact categories onto producers and consumers of music, but some interesting perspectives arise when one pushes this perspective. Publishers and studies of publishers tend to focus on what is being produced, which is a way of studying what publishers think people should buy.[106] For example, if Disney were to create 5,000,000 Lion King soundtracks, there would be a hope and expectation that they can convince consumers to buy these and (ideally) listen to them. This is slightly different than a study of what people actually purchase, although the two often correlate. This is similar to the treatise that tells musicians what they should do (which is often what musicians do in large part) but is not necessarily what is actually happening in music. A publisher producing music is often in part a statement of what the public should buy, and this is part of the reason that so many music publishers also publish journals about music.[107]

The consumption side of this comparison—which I identify as analogous to the descriptive side—is just as vital as prescriptive/ producer metrics when it comes to understanding musical practice, if not more so. In today's world, we have access to this data in droves. YouTube publishes view counts that update in real time, and Spotify/Apple playlists are

---

[106] This is an idea that Peter Mondelli explores at the end of his chapter "Parisian Opera between Commons and Commodity, ca. 1830," in *Consuming Music: Individuals, Institutions, Communities, 1730–1830* (Rochester: University of Rochester Press, 2017).

[107] Several music publishers also supported music journals to support their own works, and to advocate for them. In some cases, this functioned as a practical way to increase demand for works where the supply existed and allowed publishers to at least attempt to match products with markets they help generate. French examples include Schlesinger (*Revue et gazette musicale*), Heugel (*Le Menestrel*), and Escudier (*La France musicale*). Even when it wasn't a publisher supporting the journal, these music journals certainly interacted with and impacted musical activities as well as describing them.

updated on a near-hourly basis. When it comes to analyzing music of the past, the amount of data

available rapidly decreases. Music consumers and stores are much further apart, and information

about orders becomes harder to grasp. Print runs, a production metric, provide an indication of

the publisher's expectations that a work will sell, and subsequent reruns (or a lack thereof) are

often an indication of how that prediction is fulfilled. However, none of that substitutes for a way

to understand the way music is consumed and used.

Binder's volumes offer a space to explore publishing information, consumption

information, and the space between these two forces in the music marketplace. For examining

publishers, the BVRI stores information on the volumes, the music, and everything that came

from the original work: the documented MUSICAL WORKS TABLE information that captures the

manifestation. In doing so, the BVRI thus "finds" data that had been hidden through dispersion

across many collections and archives. The database also houses data on what people do with

manifestations (the printed paper) by noting how the document changes and what marginalia is

contained therein. Binder's volumes provide a consistent mechanism to associate multiple works

with one consumer. As the collection of binder's volumes grows over time, so too does the

potential to answer questions concerning consumption, some of which I explored earlier in the

chapter. The ability to link a consumer with their purchases, a task that is difficult to consistently

achieve outside of special cases, provides insight into consumer taste, what things people like

and do not like, the span of works in a volume, and other information from the volumes

themselves.

In addition, individual binder's volumes provide a chance to see how musicians

interpreted the music contained in their volumes. Marginalia can provide insight into how the

works are being played, how much the work is used, and how each work is different.[108]

Currently, it is difficult to know where different copies of a single song are housed, but the BVRI promises to locate several versions of the same song, allowing for comparisons of editions by different publishers. These comparisons can entail both the manifestation level of comparing various publishers and print runs as well as item-level comparisons between the same work. All of these components of marginalia in binder's volumes—the fingerings, written-out ornamentations, edits to make a work harder or easier—contribute to our understanding of how people using the music actually used the music. They help performers today as we work to understand who used music and how; or to say it a different way, how a work might be more fully understood in the differences between the publisher's prescriptive notes and edits on the one hand, and how different persons understood the work and altered it on the other. Looking at different editions of a work across time and place in binder's volumes also provides a way to understand reception history across people, regions, and over time. The BVRI will continue to facilitate such comparisons and makes them easier to find and analyze.

More Questions

The BVRI poses the opportunity for many more queries, especially as repositories are added. The metadata available here and the ability to explore the relationships herein present exciting new ways to employ quantitative musicology, but such analysis should always exist in relation to the specific and qualitative questions. Chapter 5 explores some of these relationships.

---

[108] Morabito, "Rehearsing the Social" provides a clear instance of this practice in action.

CHAPTER 5

FROM METADATA TO CASE STUDIES

Having considered some of the broader context that metadata can provide for binder's

volumes, I now turn to ways the BVRI can enhance more targeted historical analyses. Although

the overarching goals of this project tend to highlight larger scale analyses, this section focuses

on two smaller case studies. The first looks at a single volume to describe both the works inside

it, the provenance, the span of works, and other information besides. In doing so, I compare the

metadata context to the contents of the volume itself to discuss what is typical and atypical. The

second case study moves from the volume to the music. "Home Sweet Home" by Henry Bishop

is the most prevalent work in the database; it appears 24 times, and so I talk about the publishers,

composer, and others listed in attributions via arrangements. In each case, I use Chapter 5 to

show how the BVRI is not just about large numbers and queries, but about how to understand a

story, a person, or a song in the quantitative framework that data can provide.

Volume 38 is one of the volumes from the University of North Carolina collection. The

volume is one of the more filled-in volumes with respect to metadata, and it is slightly larger

than the average volume. For both reasons, I chose this volume as an exemplar for the basic

information that statistics can provide context for. This volume does have its provenance known,

and while that was not a contributing factor in selecting the volume, it did allow for contextual

clues that I suspected to be at least partially verified. In placing the volume in context, I discuss

the dates contained in the volume, the regions where works come from, and the various

publishers involved in producing works for the binder's volume.

The song "Home, Sweet Home" is one that stood out to me because of its presence in

numerous vocal works and arrangements. In total, I counted 19 instances of the work or a direct

derivative, which places the song in almost 1 out of every six volumes. The song's popularity was certainly driven by Jenny Lind, and in this section, I examine the song and its context, look at the composers credited with the work, and explore how the BVRI can indicate the prevalence of the work both before and after Lind's American tour.

In both cases, the BVRI is not answering broad questions, but specific ones. Each case study is about one song, or one volume, but the data allows for each case to carry additional weight. For Volume 38, the power lies in the single volume, but also in the ability to characterize the volume in the context of other stories. For "Home, Sweet Home," the song is important, but so too is the presence of the work before Lind's tour. This study offers one quantifiable way to measure Lind's impact on a published work and her ability to transform the narrative of a song into an exemplar of material practices shaped by a social phenomenon. In reading these two case studies, I urge a bifurcated focus that balances the specific on the one hand with the context of the larger picture on the other.

Placing Volume 38 in Its Metadata Contexts

In this numbering system, Volume 38 contains 58 works spanning 1819 to 1846. The binder belonged to Sara Burton of North Carolina, and I discuss more about her and the volume itself below. First, it is worth looking at the list of the works in the volume. The appendices have all of the information about volume 38, but here I've narrowed the information to Title, Composer, and Date Published (see Table 11). The works appear in order here as they do in the volume. The discussion that follows touches on how this volume was selected before taking the basic information about this binder and placing it into a historical context. After the groundwork for the volume is established, my focus is when and where, as well as composers and other information from the volume.

Table 11: Volume 38 Contents

Title	Composer First Name	Composer Name	Date Published
I'm afloat! I'm afloat!	Henry	Russell	1841
Miss Myrtle, the charming woman	Helen Selina Blackwood, Baroness	Dufferin and Clandeboye	
There's a charm in spring : ballad	John	Hullah	1840
The Gipsie's festival, or, Come to the old oak tree	B. S.	Barclay	1844
Twas ten o'clock			1845
Thy cheek is pale, thy lip is mute	W. C.	Peters	1844
O! take the rose, 'tis blighted	Charles F.	Heuberer	1844
Flirtation, answer to Fanny Grey			1839
Would'st thou inherit paradise'?	J. R.	Ling	1846
I remember how my childhood fleeted by	E., Mrs.	Fitzgerald	
Aileen Mavourneen	Alexander D.	Roche	
The washing day			1834
To the gay tournament	T.	Cooke	1830
Oh! Take me back to Switzerland	Caroline	Norton	1840
The Blind Boy	William R.	Dempster	
The harper's song	T. V.	Wiesenthal	
I have riches, thou hast beauty	Gaetano	Donizetti	1840
Morning's ruddy beam	George	Linley	1844
Oh! share my cottage gentle maid	Robert G.	Shrival	1843
They have given thee to another: a favorite ballad	Henry	Bishop	1836
Oh what full delight, the celebrated finale	Michael	Balfe	1845
Kathleen Mavourneen : Irish ballad	F. Nicholls	Crouch	1842
Flow gently sweet Afton	Jonathan E.	Spilman	

Title	Composer First Name	Composer Name	Date Published
Oh! Come to me	Thomas Haynes	Bayly	1840
The old elm tree.	John Hill	Hewitt	1842
The forget me not There is a flower a lovely flower	H.	Kleber	1846
The Ingle side	T. V.	Wiesenthal	
The ivy green	Henry	Russell	
Mister Hill, pray be still	Daniel	Auber	1843
Tis now the may time a much admired song	George	Linley	1840
The Gipsy girl's dream I dreamt that I dwelt in marble halls	Michael	Balfe	1819
The moon's on the lake, or the Macgregors Gathering	Alexander	Lee	1819
The Puritans' mistake. 'A church adorned with bishops, - a state without a king.'			1844
Farewell! If ever fondest prayer	J. Dodsley	Humphreys	1839
The popular song Miss Lucy Long			1842
The pilgrim's legacy : a church without a bishop, a state without a king			1844
Tis said that absence conquers love	E.	Thomas	1833
The Rose of Allandale	S.	Nelson	1831
Then you'll remember me	Michael	Balfe	1840
Mrs. B.	William	Ball	1840
The banks of the blue Moselle	G. Herbert	Rodwell	1831
The Lords of Creation		J. S. R.	
I have come from a happy land the celebrated dancing girls song	Robert Archibald	Smith	1838
Rock'd in the cradle of the deep	Joseph Philip	Knight	1840
Underneath thy lattice love	J. G.	Osbourn	1844
The song of the Gypsey's bride	Michael	Balfe	1845
The dream I dreamt that I dwelt in marble halls	Michael	Balfe	1840
Let me rest in the land of my birth farewell to the home of my childhood	J.	Harroway	1840

Title	Composer First Name	Composer Name	Date Published
The gypsies wild chant	Alexander	Lee	1819
Long Ago	Thomas Haynes	Bayly	1840
The Coronach	John, Dr.	Clarke	
Then you'll remember me'	Michael	Balfe	
There was a time	Vincenzo	Bellini	
List to the convent bells	John	Blockley	
Ah! Believe love's holy flame	Adolphe	Adam	1840
In early childhood's smiling morn	H. J., Mrs.	St. Leger	1840
I never have been false to thee.	Charles	Horn	1842
The old arm chair	Henry	Russell	

The Selection Process

I selected Volume 38 for this case study because it is both statistically typical in many of BVRI's metrics, it provides enough data to warrant more substantial yet specific historical claims. In choosing the volume, I looked at the total number of works and dated works. I wanted a volume with more than 20 works, and a high percentage of works that were dated. I also glanced at the mean publication date (1839) just to make sure I did not pick a binder that existed too far outside either side of the interquartile range.

The volume belongs to Sarah (who also went by Sallie) Burton of North Carolina. Born on Sept. 22, 1827, she married a prominent businessman on December 2, 1846. As Meyer-Frazer documents, Sarah was very active in matters of business, which does mark her as unusual for her time. Her husband was killed in the Civil War, and she eventually became a businesswoman and landowner, looking after property in North Carolina and Texas until her death in 1912. The information we have on Burton and her family is very useful in providing some additional context to her volume of music.

The Basic Information in Context

As mentioned, the volume has 58 works, 44 of which are dated. Both values land in the larger quarter of binder's volumes in the sample; the third quartiles of works and dated works are 50 and 32, respectively. One standard deviation above these means would be 58 and 41, indicating that the size of the volume is comparatively large, and the number of dated works is larger yet compared to the database's mean number of dated works. The percentage of dated works reinforces this; the average volume has 60% of works dated, with the third quartile and one standard deviation above the mean both indicating 80% of the works in a volume would be dated. Here 76% of works are dated, which again approaches the 75% mark. In all ways, this

173

volume approaches the upper edge of what is typical. Focusing on volumes before 1853, the volume stands out more, as the works before 1853 tended to be smaller—about three works smaller, and with less information per volume. This is still around the third quartile range, but in considering only works before 1853, this pushes the typical envelope even more.

Dates

The range of dates in the volume spans 1819 to 1846, but there is a story to tell with the dates. Only 3 works come from 1819, and the next earliest works come from 1830. In total, only nine more works were created in the 1830s, leaving 32 of the 44 dated works from the 1840s. The drastic difference in the first few dates is apparent looking at the box and whiskers chart below. The median is 1840, the interquartile range 1840–44, and the last work was published in 1846. Fortunately, in this case it is possible to document a bit more about Sarah, or Sallie as she sometimes went by, and use this to frame our dates.
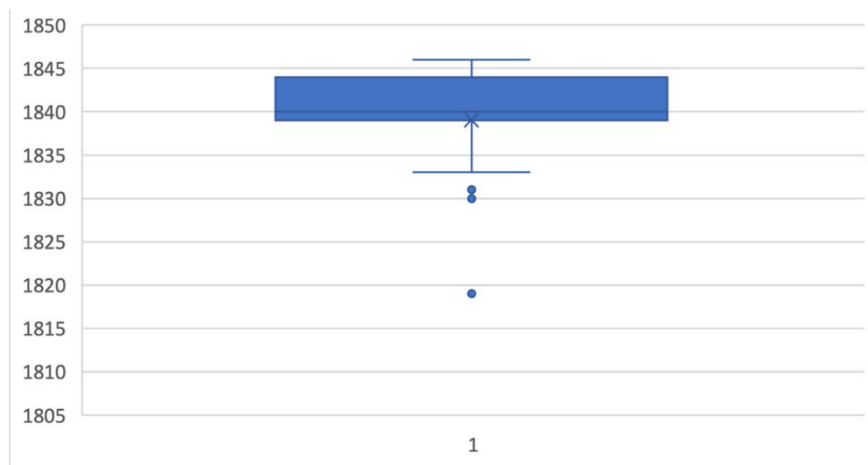

Figure 37: Box and Whiskers graph of dates from Volume 38.

Alfred Moore Burton married Elizabeth Fullenwider on June 1, 1811, and together the two had 9 children: Robert (1812–76), Elizabeth (1813–99), Mary L. (1815–1900), John W. (1820 – 1900), Frances C. (1820–96), Sarah (1827–1912), Julia (1828–29), Anna Maria (1830–

33), and Alfred Jr. (1833–33).[109] Of these 9 children, Sarah is the focus for our binder's volume.

With just this framing, we can reexamine the dates from the binder's volume. The works from 1819 are probably not hers originally; they might have belonged to her mother, Elizabeth Fullenwider, who was married in 1811. Sarah Burton was one of nine siblings, and several of her older siblings would have studied piano, as she did, and the volumes purchased in 1819 could have also belonged to Robert or Elizabeth at first, but such works were probably beyond the abilities of a seven-year-old child. In any case, the music from 1839 to 1846 very likely did belong to Sarah, and the sudden end to the volume in 1846, combined with Sarah's marriage to Robert in December of that year, strongly indicates that this volume was bound as part of her dowry. This was mentioned as a possible scenario for other volumes in Chapter 4, but here archival documents confirm that this distribution of dates could indicate a generational gap between mother and daughter.

Regional Differences

The works in Sarah's volume all come from the four most prominent publishing cities mentioned in Chapter 4: New York, Philadelphia, Boston, and Baltimore. The actual numbers based on location strongly favor Philadelphia, which has more than half of the total works from this time (34 of the 58 volumes, as seen in Figure 37). Given that this volume exists around the first quartile, it is unsurprising in this context that all the works come from the Northeast, even

---

[109] Verified using historical records of Elizabeth Fullenwider's and Sarah Burton's families on My Heritage, a family history website. https://www.myheritage.com/names/elizabeth_fullenwider and https://www.myheritage.com/research/record-1-258410241-1-502656/sarah-virginia-young-born-burton-in-myheritage-family-trees. This information also led to the North Carolina Digital Collections papers from the Burton family. Burton Family Papers, *N.C. Family Records Collection*, https://digital.ncdcr.gov/digital/collection/p15012coll1/id/39712/rec/1. Meyer-Frazier also discusses this volume in her dissertation (13). I had picked the volume independent of rereading her selection, but she highlights much of the same history I do.

though Burton is from North Carolina.
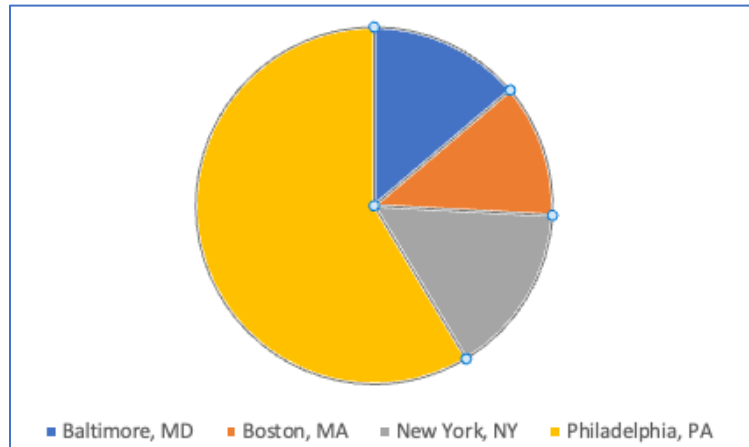
Figure 38: Location of works by city.

Does Burton Seem to Frequent a Particular Publisher or Seller? Does This Change Over Time?

With the large bias towards Philadelphia composers, Burton-Young's volume in particular is interesting when it comes to the numbers of publishers as compared to the places. Figure 2 shows the numbers of works by publisher, and the large number of publishers is a bit surprising given the small number of places. A couple of factors could be at play here, and further research is necessary to disentangle the degree to which each of these is possible. These are speculative, and each merits further research, but these are a place to start; it is worth outlining these possibilities and the BVRI's contribution to understanding each. First, Philadelphia had an extremely active publishing presence at the time; several publishers were competing in Philadelphia, including some of those listed in Figure 38. The BVRI provides a mechanism to show which publisher is active, and how much of their music makes it into binder's volumes year-after-year. Second, Raleigh may be a city that imports from Philadelphia, and this could influence the number of works in Sarah Burton's volume. Alternatively, her family may have had roots in Philadelphia that my research has not uncovered. Here, the BVRI provides a way to contextualize the likelihood of either possibility. If the former is true, and

Raleigh (or even North Carolina) has a connection regionally, then the numbers of Philadelphia works would be roughly equal across many volumes during this time. Alternatively, if the number of Philadelphia volumes is an outlier in Burton's volume, this phenomenon is much more likely to be driven by a more personal connection.[110]
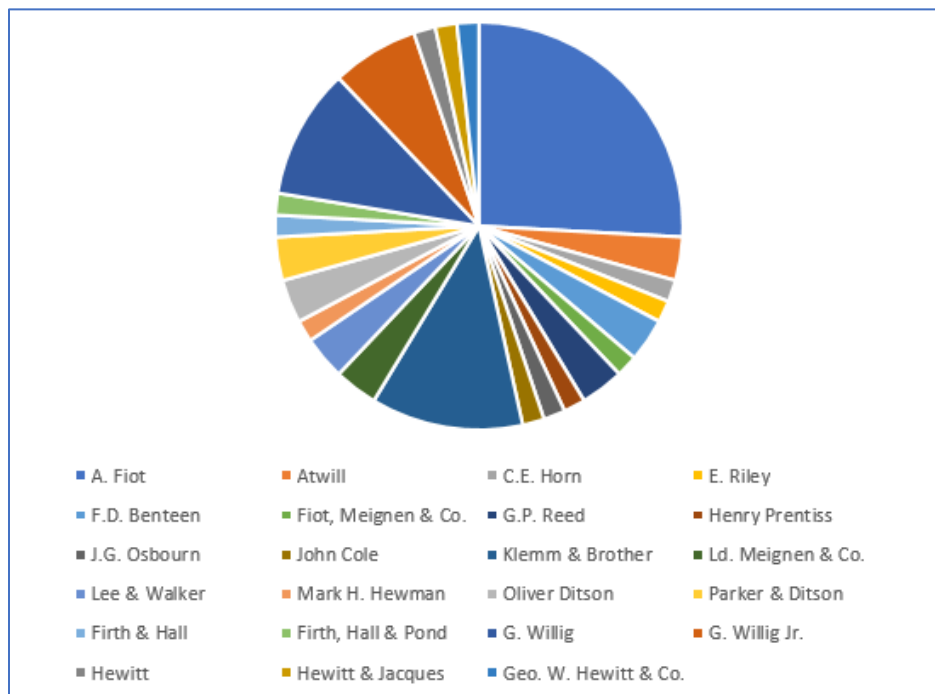
Figure 39: Number of works per publisher in Sarah Burton's volume.

To summarize, more research is needed on many fronts to answer the types of questions I pose above, and some will forever remain based on conjecture. With that said, the BVRI is a way to ground deductive reasoning in larger patterns, making outliers easier to spot. As I continue my research on binder's volumes, I plan to continue incorporating case studies from repositories as a way to frame larger analyses, and I hope to find clearer answers to many of the issues raised here.

Already some answers are taking shape. The general trends of collecting established in

---

[110] Bailey explores some of these outlier connections in her book *Unbinding Gentility*, showing how not all surprising publishing stories are outliers.

Chapter 4 will undergo continued evolution through iterations of volume uploads, but each iteration's swing in data will likely become smaller. For example, the number of works in a volume seems unlikely to change drastically, and this number is extremely useful in giving context to the size and scope of volumes. One number that may shift much more is the geographical and publishing information. Right now, the focus has been on one repository, and expansion will likely include geographical blips of a sort. The BVRI will allow each input to be studied separately, and for the information to come together in tandem. Thus, while I continue to push for more research and my own expansion in this world of cataloging binder's volumes, the information thus far is enriching and exciting, and I am excited to continue with future case studies.

<div align="center">"Home Sweet Home"</div>

"Home, Sweet Home" is the most common work to appear in the data set. In this case study, I begin by discussing the background of the work, before moving into musical features. I conclude the section by going through many of the editions available for the work and their dates, including the variations and how those variations factor into the Lind visits to America.

Background

"Home, Sweet Home" was composed by Henry Rowley Bishop (1786–1855). Bishop was born in London, and by the age of 13 was writing and selling music with his cousin, Charles Wigley.[111] Three years later, Bishop received a musical education at the expense of Thomas Panton.[112] His studies continued, and he had a few musical successes in 1809,[113] and he

---

[111] Richard Northcutt, *The Life of Sir Henry Bishop* (London: The Press Printers, 1920): 2.

[112] Ibid., 2

[113] Ibid., 10.

produced several operas throughout the subsequent decades. Some were mostly original compositions, while others were notably reworkings of Mozart and Rossini operas.[114] Copyright protections from the 1710 Statute of Anne did not cover musical works until 1842, and such practice was legal at the time. In his original and derivative works, Bishop was a primary driver of opera in London throughout his lifetime.

Despite Bishop's active compositional output, the work he is most famous for today is "Home, Sweet Home." The ballad started as a melody, a "Sicilian Air" (1821), which was then used in the opera *Clari, the Maid of Milan* (1821) as an overture.[115] In *Clari*, Bishop used the text by John Howard Payne, and the ballad achieved a fame far beyond the opera itself. Amidst the myriad of works that borrow this tune for one reason or another, Donizetti's *Anna Bolena* (1830) stands out. The tune reached its zenith in the 1850s; Jenny Lind incorporated the song into her standard repertoire and Bishop launched the song as a parlor song. During the Civil War, the song was a favorite of both Confederate and Union troops, with bands playing the song on both sides. Eventually, the work was banned by Union commanders for fear that it would make men feel too attached to home to properly fight.[116] Today, the song exists as a common American folk tune.[117]

Musical Features

At this point, it is worth looking at the work itself and discussing some of the musical features about this manifestation, and about the other manifestations contained in the BVRI.

---

[114] Northcutt, 10.

[115] Ibid.

[116] Abe, Ernst L. "'Home, Sweet Home': A Civil War Soldier's Favorite Song." *America's Civil War*, May 1996.

[117] My daughter's toy piano plays a slightly simplified version of this melody as one of the 10 presets that come with her piano.

Figure 40 shows the first 4 pages of the manuscript. The melody is short and simple, in an aabb

verse, followed by a two-phrase chorus. Each line alternates between an imperfect authentic

cadence on scale degree 3 for odd lines, followed by a PAC on even lines. The written range of

the song is an octave, and the volumes of the BVRI contain setting in Eb major, F major, and G

major.[118] All harmonies are either tonic, dominant, or predominant (ii^6 more often than IV), and

over half the song sets a pleasant melody over a I chord. In describing the song so simply, I do

not mean to detract anything from its charm or overall affect. Indeed, the simplicity and

authenticity of the melody is exactly why Donizetti used the melody to evoke nostalgia in his

mad scene of Anna Bolena. The essence of this song is a pleasant, charming melody that remains

simple but effective throughout.

Editions

Within the BVRI, 31 musical works have the words "Sweet" and "Home" in the title, not

necessarily connected. Of those, 23 are in some way tied to the aforementioned "Home, Sweet

Home" by Bishop. That is, 23 works would fall within the broad bubble of being some

manifestation of an expression of the work that is "Home, Sweet Home" as per the FRBR model.

Ten are vocal arrangements of "Home, Sweet Home": eight are for piano and voice, two for

guitar and voice. The remaining 12 are instrumental arrangements of, or variations on, the

original tune. Both the ornamentation present on some manifestations of the song—including the

vocal part of the one above—and the variations of the piano versions are facilitated in no small

part by the open and unrestricted nature of the melody. This may contribute to the fact that

---

[118] Only one setting is in G major, and it is for guitar. This may be to give the guitarist a friendlier key (the other guitar manifestation is in E major), although Benedict's arrangement notes that Ms. Lind sings the song in the key of G major.

binder's volumes contain more piano examples of the work than voice-instrument

manifestations.

2

The First and Third Verses.

Mid pleasures and pal-aces though we may roam, Be it
How sweet 'tis to sit, 'neath a fond Father's smile, And the
pia

ev---er so hum--ble there's no place like Home! A
cares of a Moth---er to soothe and be--guile, Let

charm from the skies, seems to hal--low us there, Which
oth---ers de-light mid new plea--sures to roam, But

seek thro' the world is ne'er met with else where;
give me oh! give me the pleasures of Home

182

Figure 40: Pages 1-4 of "Home Sweet Home"

While I have spoken thus far about the work itself, the primary use of the BVRI is giving context to how this work, and other works, are realized in various manifestations. I begin this analysis with the listed composer. Of the 10 vocal works that possess Payne's lyrics, six list Henry Bishop as the composer. Two do not list a composer, and two list different composers as the point of origin. One lists Julius Benedict, and the other an E. Guilbert, and these two composers' versions are worth investigating more deeply.

Benedict was a German composer who worked in England for most of his career. During the late 1840s, Benedict conducted Jenny Lind's first appearance in an oratorio (Mendelssohn's *Elijah*), and in 1850 Lind invited him to participate in her concert tour throughout America. One of her most common closing songs was "Home, Sweet Home," and this version was produced and sold in conjunction with the tour. The work itself was published in 1851, and the title in its entirety reads "Farewell Songs Of Jenny Lind In America. With New Accompaniments By Jules Benedict. 5. Home Sweet Home." Despite the advertisement of a "new accompaniment" for this song, the piano part is remarkably true to the one above. Figure 41 shows Benedict's contributions. Outside of placing the song in Eb instead of F major as the original (or Lind's purported key of G major), the accompaniments are quite similar; so much so that this has every appearance of being the same song rereleased for exclusively commercial reasons. By contrast, Guilbert's arrangement is notably different and idiomatic to the guitar. The melody is the same, but the accompaniment is clearly built around open G, Am, C, and D7 chords: the same tonic, dominant, and predominant chords present in Figure 40, but transposed up one step so that they are open chords on a guitar. Guilbert demonstrates much greater compositional agency due to the change to another instrument and the active rearrangement necessitated by making the transition idiomatic for picking individual notes.

Figure 41: Benedict's "new arrangement" of "Home, Sweet Home."

The difference in compositional role vs notated role is one worth exploring in the piano

arrangements as well. Seven composers are identified for the piano arrangements and variations

in the BVRI: W.R. Coppock, Louis Moreau Gottschalk, J.H. Slack, Sigismond Thalberg,

Thomas Valentine, and Stephen Collins Foster. Several of these are listed multiple times, but I

track them separately in the MUSICAL WORKS TABLE because different publishers or dates make

each one a different manifestation of the work. Each of the 12 piano examples is a set of

variations, although not all the variations are the same. Three volumes contain Valentine's

variation sets, and one is notably in E♭ major and very different from the other two F major variations. Others are the same for both sets of variations (like the two Valentine F major variations), but because the publisher is different, they remain different manifestations.

Turning to publishers, several publishers are involved with the song, but several of these are part of the same family. Specifically, the Firth/Hall/Pond/Ditson family outlined in Chapter 2 contains 9 of the 23 publishers. They are grouped separately here, but it is worth acknowledging the relationship between the publishers. Other publishers figure into the volumes prominently, most notably George Willig's publishing company.

There are a few additional things that should be noted on "Home, Sweet Home" in the BVRI. First, all the manifestations of "Home, Sweet Home" come from different volumes. While it makes sense that no duplicates are in volumes, I was expecting at least one person to have a vocal copy and piano variations. There also isn't a strong correlation for where the work falls in a particular volume; the standard deviation is 28%, and the work falls in the 42% of the order. That is to say, the work falls in the middle (roughly), but two standard deviations cover all the placements in the order except dead last. One additional note is the dates of the works. While some of the songs are copyrighted in 1824 or 1830, the majority are copyrighted in 1851 or after, which coincides with Lind's trip and the reintroduction of the song to American audiences. Some would have known the song from *Clari,* and the volumes demonstrate this. Of the volumes containing versions of "Home, Sweet Home" published before 1835, all were likely bound before 1850; they all have the final published work date as at least a few years before 1850, and most are in the 1830s. Thus, the BVRI tracks two waves of popularity. One exists around the time of *Clari* and in the decade that followed, and the larger wave rises after Lind's immensely popular tour revitalizes the work.

Conclusion

The two case studies here, taken with the larger corpus studies of Chapter 4, are in many ways are the beginning of what a data-driven approach to binder's volumes can highlight. Many more possibilities for case studies remain, and Chapter 6 details future plans for this research and how it might be upscaled moving forward. As the corpus of known and cataloged binder's volumes continues to grow, the BVRI will become an increasingly powerful tool for looking at the larger patterns in music purchasing and consumption throughout the United States, and perhaps someday beyond.

CHAPTER 6

CONCLUSION AND NEXT STEPS

Many things are set outside the scope of the research questions for the project, and to understand what those tasks and questions will be, it is worth looking back at what this project has done before exploring where the BVRI goes from here. To recap, Chapter 1 set the stage for why the project exists, what it can contribute, and how it fits into the current scholarship. In this chapter, I established the state of research on binder's volumes, recent digital projects, material history, and situated this research at the juncture of these three categories.

Chapters 2 and 3 focused on the structure of the BVRI and the data inside it. Chapter 2 established the framework for the BVRI, including the metadata schema and the logic behind the choices I made in establishing the database structure. Chapter 3 was focused on the mechanics of how to move data en masse from other sources and systems of organization into the BVRI, what I have done thus far, and the choices I made in doing so. Taken together, these two chapters accomplish two broad tasks. First, they put forth the logic behind how data is included and where, and I attempted to do so in as straightforward and transparent a way as possible. Quantitative research runs the risk of opacity, and these two chapters provide rationale and clarity in a format designed to be relevant to an experienced statistician and relevant for a humanist more familiar with traditional methodologies. Second, they provide a look at one possible path for other digital humanities projects to embark upon. I intend my metadata schema (and rationale) to be a document that others might use as a framework for the way data is structured. Similarly, Chapter 3 is intended to provide ideas for digital humanists who might not be familiar with ways that data can be cleaned, reorganized, and uploaded into another platform. The two chapters vary between historical and technical, but on balance they should provide

information for scholars interested in new methodologies and for technically oriented scholars looking at the rationale behind these types of decisions.

Chapters 4 and 5 are focused on how this data can be used, and why it matters. Chapter 4 focuses on the larger questions: what can be gained from the large–scale analysis of metadata? Chapter 5 turns to how large-scale analyses might support and expand more focused and nuanced questions, and I attempt to address the possibilities here with two case studies: one on a single binder's volume, and one on a song. In these chapters, I scratch the surface of what is possible with metadata, and these two chapters provide a launching point for what comes next with the BVRI.

The recap of what has been done leads to the obvious question: what are the next steps for quantitative research in binder's volume? First, the goal is to expand the sample size, specifically by expanding repositories included. Data analysis only gets better with more data, and several other repositories have been at least partially explored, and now merit fuller consideration. The UNT volumes will be fully integrated quite soon; I have used the data for test cases and for experimenting with the design, but their inclusion will happen shortly after this document is submitted. Karen Stafford's work on the Library of Congress volumes makes them a likely next candidate, and then Harvard's volumes would be a strong fourth candidate. From there, Middle Tennessee State and other, smaller collections are likely repositories that should be included. The aforementioned collections will offer regional diversity and increase the sample size of the project, and each set will provide a test for the BVRI's ability to expand, testing the methodology put forth in Chapters 2 and 3.

The second step is completing information on provenance and compilers. This information is important, but thus far my focus has been on the published work and collected

items. Expanding the BVRI to address provenance more properly will extend the types of case studies possible, and will support research into case studies.

The third step in expanding this project is to expand the uses of the data. On a large scale, this will be driven by expanding sample size and new data sets to explore. Some portions of the research will focus on comparing new data to old data (regional differences), while other research questions may update findings presented in Chapter 4.

The last major question is access and increasing the ability for other researchers to engage with the database in a meaningful way. One future step is creating a space for researchers to submit volume information. There would need to be an intermediate step to ensure data validity, but the goal is for information to be logged and shared. The second, and more important goal, is for users to be able to search and use the BVRI. The tables from the BVRI at the time of the dissertation's conclusion are available in Appendix A, but the data also needs to be available online. The most immediate goal for this project is creating the online space for researchers to engage with volumes and the data contained therein.

In all of this research, the goal remains what it was at the start: to peer into the music libraries of the past, and to facilitate this view on an ever-larger scale of nineteenth-century American music. Data is a powerful tool for analysis, and this dissertation is built on that premise, but for all that data can offer, I want to close by taking a step back and remembering that this research is about telling the story of our past through the lens of music. By focusing on the material history of printed music, the BVRI provides insight into the social, the political, the historical, and hopefully into the lives and decisions of the past. Data is a means to this end, but the endpoint of this research project remains unpacking and analyzing shared musical experiences.

APPENDIX A

BVRI DATA TABLES AND FIELD TYPES

CREATORS TABLE		
Field Title	**Description**	**Data Type**
Last name	Standardized Spelling of Creator's Last Name	Text
First Name	Standardized Spelling of Creator's First Name	Text
Alternate Spellings, Last Name	List of alternate spellings of a creator's last name, found in the BVRI	Text
Alternate Spellings, First Name	List of alternate spellings of a creator's first name, found in the BVRI	Text
Birth Date	Four-digit year creator was born, if known	Number
Flourished	Years when a composer was known to be active	Number
Death Date	Four-digit year creator died.	Number
*Creator Counter	Automatically generated number, serves as Primary Key	Integer
Company name	Name of company (if creator is a company). This primarily applies to performers.	Text
Original name	Name originally in cell (pre-edits). This is useful if a name is disputed or if data somehow becomes corrupted.	Text
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

PUBLISHERS TABLE		
Field Title	**Description**	**Data Type**
Generic Publisher Name	Publisher name for general set of publishers. This name is broad and repeats across publishers to denote a family (Firth, Pond, etc.).	Text
Specific Publisher Name	Specific name of publisher. Given unique addresses, this name does not repeat.	Text
Address	Address for publisher	Text
City	First city listed for publisher, corresponds to address if given	Text
State	State or territory where city resides as of January 1, 2020 (if relevant or given), particularly in U.S.	Text
Country	Country (January 1, 2020 boundaries) where publisher resides	Text
City 2	Second city listed with publisher	Text
State 2	Second state/territory listed with publisher	Text
Country 2	Second country listed with publisher	Text
City 3	Third city listed with publisher	Text
State 3	Third state/territory listed with publisher	Text
Country 3	Third country listed with publisher	Text
Start Date	Earliest known date publisher at this address: 4-digit year.	Number
End Date	Last known date publisher at this address: 4-digit year.	Number
Plate number start	First plate number for publisher at this address, if known.	Text
Plate number end	Final plate number for publisher at this address, if known.	Text
Related Publishers	Generic name of related publishers.	Text
Related Publishers 2	Generic name of related publishers (if needed twice)	Text
*Publisher Counter	Automatically generated number, serves as Primary Key	Integer
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

COMPILERS TABLE		
Field Title	**Description**	**Data Type**
Last name	Standardized Spelling of Compiler's Last Name	Text
First Name	Standardized Spelling of Compiler's First Name	Text
Middle Name	Standardized Spelling of Compiler's Middle Name or Middle Initial	Text
Last Name Alternate	Piped List of Alternate Spellings of compiler's Last Name	Text
First Name Alternate	Piped List of Alternate Spellings of compiler's First Name	Text
Birth Date	Four-digit year compiler was born, if known	Number
Death Date	Four-digit year compiler died, if known.	Number
Address	Address for compiler, if known	Text
City	First city listed for compiler, corresponds to address if given	Text
State	State or territory where city resides as of January 1, 2020 (if relevant or given), particularly in U.S.	Text
Country	Country (January 1, 2020 boundaries) where compiler resides	Text
City 2	Second city listed with compiler	Text
State 2	Second state/territory listed with compiler	Text
Country 2	Second Country listed with compiler	Text
*Counter	Automatically generated number, serves as Primary Key	Integer
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

MUSICAL WORKS TABLE		
Field Title	**Description**	**Data Type**
*Musical Works Counter	Automatically generated number, serves as Primary Key	Integer
Specific Work Title	Statement of work title as printed.	Text
Common Work Title	Commonly known title of work.	Text
Caption Title	Caption listed on work	Text
Other Titles	Titles listed as other on the work; Second Title	Text
Title 3	Third Title listed, originally part of main title	Text
Title 4	Fourth Title listed, originally part of main title	Text
Title 5	Fifth Title listed, originally part of main title	Text
Title 6	Sixth Title listed, originally part of main title	Text
Title 7	Seventh Title listed, originally part of main title	Text
**Composer	First Composer listed on sheet music (SK, Creators)	Integer
Opus	Opus and work number of work	Text
**Composer: 2nd	Second Composer listed on work or composer of larger work (SK, Creators)	Integer
Relationship to work (2nd)	Statement of how 2nd composer is related to work	Text
Larger Work: Related Work	Name of work that is a related to specific work at hand.	Text
**Composer: 3rd	Third Composer listed on work or composer of larger work (SK, Creators)	Integer
Relationship to work: 3rd	Statement of how 3rd composer is related to work	Text
**Composer: 4th	Third Composer listed on work or composer of larger work (SK, Creators)	Integer
**Composer: 5th	Third Composer listed on work or composer of larger work (SK, Creators)	Integer
**Composer: 6th	Third Composer listed on work or composer of larger work (SK, Creators)	Integer
**Lyricist: 1st	First Lyricist listed (SK, Creators)	Integer
Translator: 1st	Is the first lyricist a translator?	Boolean
**Lyricist: 2nd	Second Lyricist listed (SK, Creators)	Integer
Translator: 2nd	Is the second lyricist a translator?	Boolean
Language	Language or Languages listed in performance (Listed in English)	Text

MUSICAL WORKS TABLE		
Field Title	**Description**	**Data Type**
**Lithographer	Lithographer listed (SK, Creators)	Integer
Lithograph Design	Short description of lithograph	Text
**Lithographer 2	Second Lithographer listed (SK, Creators)	Integer
**Lithographer 3	Lithographer listed (SK, Creators)	Integer
**Performer 1	Performer advertised with piece (SK, Creators)	Integer
**Performer 2	Second Performer advertised with piece (SK, Creators)	Integer
**Performer 3	Third Performer advertised with piece (SK, Creators)	Integer
**Performer 4	Fourth Performer advertised with piece (SK, Creators)	Integer
**Performer 5	Fifth Performer advertised with piece (SK, Creators)	Integer
Performer Method	Description of how the performer executed the performance	Text
Performer Full	Full text of the performer description	Text
Performer By Long	Full Text String that captures performers contributing to the work (for later disambiguation if needed)	Text
Performer By Short	Shortened Text String that captures performers contributing to the work (for later disambiguation if needed)	Text
Performer Also By Short	Text String that captures performers contributing to the work (for later disambiguation if needed)	Text
**Arranged by	Arranger of this particular edition, if specifically listed as arranger (SK, Creators)	Integer
Instrumentation	Description of voice parts and overall instrumentation	Text
**Publisher: 1st	First publisher listed on work (SK, Publishers)	Integer
Plate numbers Publishers Numbering	Plate numbers (listed as range)	Text
Date Published	Date published, or average of range of dates published	Number
Date Margin of Error	Number of years that the date might be off by	Number
Number of Pages	Number of pages, including title page	Number
**Publisher :2nd	Second publisher listed on work (SK, Publishers)	Integer
**Publisher: 3rd	Third publisher listed on work (SK, Publishers)	Integer
Location Listed	Location listed for the work publisher	Text
First Line of Text	First line of text	Text
First Line of Chorus	First line of chorus	Text
Edition	Edition of Work	Text

MUSICAL WORKS TABLE		
Field Title	**Description**	**Data Type**
Topic	Topics listed as associated with work	Text
Inscription	Printed Inscription on Work	Text
Full Title	Full Title as Listed	Text
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

BV [Binder's Volume] Master Table		
Field Title	**Description**	**Data Type**
Institution	Current Location of BV	Text
Call Number	Call number at Institution	Text
Institution Designation	Number or way institution denotes particular volume	Text
Series Collection	Name of Series or Collection the volume belongs to	Text
**Compiler	Compiler of BV (SK, Compilers)	Text
Date Bound	Date volume was bound, or average of date range that volume was bound (if known).	Number
Date Margin of Error	Margin of error for date bound	Number
**Location Bound	Location Volume was bound, if known (SK, Locations)	Integer
Color	Main color of the outside of the binder |colors of bindings	Text
Physical Description	Description of the outside of the volume	Text
Notes	Relevant information from the BV (sections delimited by vertical pipe)	Text
*Counter	Automatically generated number, serves as Primary Key	Integer
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

BV [Binder's Volume] Contents Table		
Field Title	**Description**	**Data Type**
**Work Information	Description of manifestation of musical work (SK, Musical Works)	Integer
**BV Master	Binder's Volume containing the work (SK, BV Master)	Integer
**Seller	Person who sold work to compiler, if known (SK, Publishers)	Integer
Location Address	Address for seller, if known	Text
City	First city listed for seller, corresponds to address if given	Text
State	State or territory where seller resides as of January 1, 2020 (if relevant or given), particularly in U.S.	Text
Country	Country (January 1, 2020 boundaries) where seller resides	Text
Order in Binder	Number in binder	Number
Notes	Relevant information from the work (sections delimited by vertical pipe)	Text
*Counter	Automatically generated number, serves as Primary Key	Integer
Content dm number	An in-house repository code that I track when applicable	Number
Previous owner	Information on the previous owner, if available	Text
Marginalia	Notes on marginalia contained within a piece.	Text
Dedication	The dedication and/or dedicatee of the work	Text
Inscription	Inscription on the work	Text
Price	Price of the work, if given.	Text
Defects	Defects noted on the work.	Text
Reference URL	URL that links to where the work is hosted, if given.	Text
Notes 2	Additional notes on the work.	Text
OCLC Number	WorldCat Reference number.	Text
Primary Key	Auto entered FileMaker Calculation, cannot be changed. (alphanumeric number)	Text
Creation Timestamp	Timestamp at time of creation	Timestamp
Created by	Person who created entry	Text
Modification Timestamp	Timestamp of last modification.	Timestamp
Modified by	Person who made last modification.	Text

APPENDIX B

BVRI DATA IMPORT BEST PRACTICES

1.) Organize data import into one or more tables, import table into OpenRefine

2.) Identify fields containing Creators, Publishers, Compilers, and Musical Works information.

3.) Import Creators

    a.) Composers

        i.   Pull information about composers into a separate table in OpenRefine.

        ii.   Separate the composers list so each composer is a separate column

        iii.   Separate Creators information from new set of Creators into Last Name, First Name, Date Born, Dates Flourished, and Date Died with all extraneous punctuation and spaces removed.

        iv.   Standardize format for new Creators by creating a Lookup column that joins relevant information.

        v.   Compare new composers with existing creators list. The goal is to only import new entries into the database. Existing entries should be matched with existing musical works numbers.

            a.   Delete from list any creator that already exists in BVRI

            b.   Compare alternate names for creators that did not come up in original search.

        vi.   Update the list

            a.   Add new creators to list.

            b.   Do a count of Creators in lookup (last name, first name) to ensure no duplicates arose in adding creators.

            c.   Import new list into BVRI database.

        vii.   Use updated lookup field to assign each composer a creator number.

            a.   Check for errors as necessary

            b.) Lyricists and Arrangers

                i.   Repeat Steps i-vii of composers for lyricists and arrangers.

            c.) Lithographers and Engravers

                i.   Repeat Steps i-vii of composers for lithographers and engravers.

> ii. In step 3.a.vi., double check company names for duplicates as well as individual names.

  d.) Performers

>   i. Use linguistic demarcations to isolate performers, locations, and methods of performance

>   ii. Repeat steps i-vii for performers

>   iii. As with lithographers and engravers, double check company names for duplicates as well as individual names in step 3.a.vi.

4.) Import Publishers

  a. Identify columns containing publisher information, including name, address (if given), and location. Place information into a separate table in OpenRefine.

  b. Separate the publishers list from the new data set so each publisher is a separate column.

>   i. Multiple publishers are far less common than multiple composers or arrangers, but they do happen.

  c. Separate publisher information into the columns publisher name, address, and location with all extraneous punctuation and spaces removed.

  d. Standardize format for new Publishers by creating a Lookup column that joins relevant information.

  e. Compare new publishers with existing publishers list. The goal is to only import new entries into the database. Existing entries should be matched with existing musical works numbers.

>   i. Delete from list any publisher that already exists in BVRI.

>   ii. Compare alternate names for creators that did not come up in original search.

  f. Update the list.

>   i. Add new publishers to list.

>   ii. Do a count of publishers in lookup (last name, first name) to ensure no duplicates arose in adding creators.

>   iii. Import new list into BVRI database.

  g. Use updated lookup field to assign each publisher a publisher number.

>   i. Check for errors as necessary.

5.) Import Musical Works

    a. Identify columns containing musical works information (see Appendix A table). Place information into a separate table in OpenRefine.

    b. Separate musical works information into relevant columns.

    c. Standardize format for new musical works by creating a Lookup column that joins relevant information (in this case musical work title, composer, and publisher).

    d. Compare new musical works list with existing musical works list. The goal is to only import new entries into the database. Existing entries should be matched with existing musical works numbers.

       i. Delete from list any musical work that already exists in BVRI

      ii. Compare alternate names for musical works that did not come up in original search.

         a. This is where editing titles and clustering will be necessary.

    e. Update the list

       i. Add new musical works to list.

      ii. Do a count of musical works in lookup (last name, first name) to ensure no duplicates arose in adding creators.

      iii. Import new list into BVRI database.

    f. Use updated lookup field to assign each musical work a musical work number.

       i. Check for errors as necessary.

6.) Compilers Table

    a. This table is largely manual entry. As it increases in size, the process for matching creators with preexisting creators may be implemented later, but repetition in compilers is much less likely.

7.) Binder's Volumes Master Table

    a. Identify columns containing relevant information about binder's volumes (use Appendix A as a guide for relevant information). Place information into a separate table in OpenRefine.

    b. Import into OpenRefine to generate BV Master numbers to serve as primary keys.

    c. Use compilers list and lookups to link compilers imported in previous step to their respective volumes.

8.) Binder's Volume Contents Table

    a. Identify columns containing relevant information about binder's volumes contents (use Appendix A as a guide for relevant information). Place information into a separate table in OpenRefine.

    b. Use the updated Musical Works Table to match Musical Works numbers for the lookup.

    c. Use the updated Binder's Volume Master Table to match BV Master numbers for the lookup.

    d. Check for duplicates in steps 8b and 8c by using count functions with the lookups, but this should be completed in step 5.f.ii. for musical works, and binder's volumes duplicates are extremely unlikely.

    e. Import new list into BVRI database.

WORKS CITED

Archives and Special Collections

Library of Congress: Binder's Volumes Collection (M1.A15, particularly volumes 1–120).

University of North Carolina, Chapel Hill: Binder's Volume Collection.

University of North Texas Binder's Volumes: Binder's Volume Collection.

Secondary Sources

Abe, Ernst L. "'Home, Sweet Home': A Civil War Soldier's Favorite Song." *America's Civil War*, May 1996.

Allen, Robert C., and Douglas V. Morr. *Horrible Prettiness: Burlesque and American Culture*. Chapel Hill: The University of North Carolina Press, 2006.

Atkinson, David and Steve Roud, ed. *Street Ballads in Nineteenth-Century Britain, Ireland, and North America: The Interface between Print and Oral Traditions.* Farnham: Ashgate, 2014.

Bailey, Candace. "Binder's Volumes as Musical Commonplace Books: The Transmission of Cultural Codes in the Antebellum South." *Journal of the Society for American Music* 10, no. 4 (2016): 446–69.

———. *Charleston Belles Abroad: The Music Collections of Harriet Lowndes, Henrietta Aiken, and Louisa Rebecca McCord.* Columbia, SC: University of South Carolina Press, 2019.

———. *Music and the Southern Belle: From Accomplished Lady to Confederate Composer*. Carbondale, IL: Southern Illinois University Press, 2010.

———. *Unbinding Gentility: Women Making Music in the Nineteenth-Century South*. Urbana: University of Illinois Press, 2021.

Balleisen, Edward J. "Vulture Capitalism in Antebellum America: The 1841 Federal Bankruptcy Act and the Exploitation of Financial Distress." *Business History Review* 70, no. 4 (1996): 473–516.

Barker, Duncan J. "Wallace, William." *Grove Music Online.* 2001; Accessed 21 Oct. 2022. https://www-oxfordmusiconline-com.libproxy.library.unt.edu/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000029839.

Barnard, William Harrison. "Significant Objects: American Literature and Consumer Culture, 1850–1900. PhD diss., University of Michigan, 2001.

Beerli, Aunción and Josefa Martín Santana. "Design and Validation of an Instrument for Measuring Advertising Effectiveness in the Print Media." *Journal of Current Issues and Research in Advertising* 21, no. 2 (Fall 1999): 11–30.

Bekiari, Chryssoula, Martin Doerr, Patrick Le Bœuf, and Pat Riva, editors. "Definition of FRBR$_{OO}$: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism." *International Federation of Library Associations and Institutions.* Netherlands: IFLA, 2016.

Bloechl, Olivia A. "Protestant Imperialism and the Representation of Native American Song." *The Musical Quarterly* 87, no. 1 (2004): 44–86.

Board of Music Trade of the United States of America. *Complete Catalogue of Sheet Music and Musical Works, 1870*. Ed. New introduction by Dena J. Epstein. New York: Da Capo Press, 1973.

Bode, Carl. *The Anatomy of American Popular Culture, 1840–1860*. Berkeley: University of California Press, 1960.

Boyd, Ryan L. "Psychological Text Analysis in the Digital Humanities." In *Data Analytics in the Digital Humanities*, edited by Shalin Hai-Jew, 161–190. eBook: Schirmer, 2017.

Brooks, Jeanice. "Musical Monuments for the Country House: Music, Collection, and Display at Tatton Park." *Music & Letters* 91, no. 4 (2010): 513–35.

Brown, Richard D. *Knowledge is Power: The Diffusion of Information in Early America, 1700–1865*. New York: Oxford University Press, 1989.

J. Peter Burkholder, Donald Jay Grout, and Claude V. Palisca, *A History of Western Music*, 10th ed. (New York: Norton, 2019).

Cairns, Zachary. "Switching the Backbeat: The Quick Flip and Polymetric Pogo in 1980s-era Rock Music." *Music Theory Online* Vol. 28, No. 1 (March 2022). https://mtosmt.org/issues/mto.22.28.1/mto.22.28.1.cairns.html

Cavicchi, Daniel. *Listening and Longing: Music Lovers in the Age of Barnum.* Middletown, Conn: Wesleyan Univ. Press, 2011.

Charosh, Paul. "Studying Nineteenth-Century Popular Song." *American Music* 15, no. 4 (1997): 459–92.

Corey, Alexander William. "Beyond the Blues: Music, Gender, and Black Modernisms." PhD diss., University of Colorado, 2017.

Crawford, Richard. *The American Musical Landscape: The Business of Musicianship from Billings to Gershwin.* Berkeley: University of California Press, 2000.

Crouthamel, James L. "The Newspaper Revolution in New York: 1830–1860." *New York History: Quarterly Journal of the New York State Historical Association*, Vol. 45, No. 2 (April 1964): 91–113.

Cusick, Suzanne. "Feminist Theory, Music Theory, and the Mind/Body Problem." *Perspectives of New Music* 32, no. 1 (Winter, 1994): 8–27.

Davis, Lauren. "Beyond the Archive: The Seward Family Digital Archive as Research, Instruction, and Community Engagement." Paper presented at the annual meeting of the Nineteenth Century Studies Association, Rochester, NY, March 2021.

Deaville, James. "Publishing Paraphrases and Creating Collectors: Friedrich Hofmeister, Franz Liszt, and the Technology of Popularity." In *Franz Liszt and His World*. Princeton: Princeton University Press, 2006. https://ebookcentral.proquest.com/lib/unt/detail.action?docID=617585.

Devriès, Anik and François Lesure. *Dictionnaire des éditeurs de musique française*, Vol. 2 (Geneva: Minkoff, 1979).

Dichter, Harry, and Elliott Shapiro. *Early American Sheet Music: Its Lure and its Lore, 1768–1889* New York, Bowker, 1941.

Eichengreen, Barry J. *Globalizing Capital*. 2nd ed. Princeton, NJ: Princeton Univ. Press, 2008.

"Exploring Facets," *OpenRefine Documentation: User Manual and Technical Reference.* July 2022.

Faucett, Bill F. and Michael J. Budds. *Music in America 1860–1918: Essays, Reviews, and Remarks on Critical Issues.* Hillsdale, N.Y.: Pendragon Press, 2008.

Feist, Leonard. *An Introduction to Popular Music Publishing in America*. New York: National Music Publishers' Associations, 1980.

Ferris, Marc. *Star-Spangled Banner: The Unlikely Story of America's National Anthem.* Baltimore, MD: Johns Hopkins University Press, 2014.

Feyel, Gilles. "Les transformations technologiques de la presse." In *La Civilisation du journal: Historire Culturelle et littéraire de la presse française au XIXe siècle*. Ed. Dominique Kalifa, Philippe Régnier, Marie-Ève Thérenty, and Alain Vailland. Paris: Nouveau Monde, 2011.

Finson, Jon W. *The Voices that Are Gone: Themes in Nineteenth-Century American Popular Song.* New York: Oxford Univ. Press, 1994.

Foster, Daniel H. "Sheet Music Iconography and Music in the History of Transatlantic Minstrelsy." *Modern Language Quarterly* 70, no. 1 (2009): 147–61.

Goehr, Lydia. *Imaginary Museum of Musical Works: An Essay on the Philosophy of Music*. New York: Oxford University Press, 1992.

Gold, Matthew K. and Lauren F. Klein. "Introduction." In *Debates in the Digital Humanities 2019,* edited by Matthew K. Gold and Lauren F. Klein, ix–xiv. Minneapolis: University of Minnesota Press, 2019.

Goodman, Glenda. *Cultivated by Hand: Amateur Musicians in the Early American Republic*. New York: Oxford University Press, 2020.

Grier, Katherine C. *Culture & Comfort*. Washington, DC: Smithsonian Inst. Press, 1997.

Hai-Jew, Shalin. "Parsing Related Tags Networks from Flickr® to Explore Crowd-Sourced Keyword Associations." In *Data Analytics in the Digital Humanities*, edited by Shalin Hai-Jew, 191–214. eBook: Schirmer, 2017.

Hamm, Charles. *Music in the New World*. New York: W. W. Norton and Company, 1983.

Hankins, Laurel. "What the Folk Printed: Verse Culture and the Black Press in 1865 New Orleans." *African American Review* 45, no. 4 (2012): 527–40.

Haveman, Heather A., Jacob Habinek, and Leo A. Goodman. "How Entrepreneurship Evolves: The Founders of New Magazines in America, 1741–1860." *Administrative Science Quarterly* 57, no. 4 (2012): 585–624.

Higginbotham, Barbra Buckner. *Preservation in American Libraries at the Turn of the Century: A Portrait of Thought and Activity in the Period 1876 to 1910.* D.L.S. Thesis, Columbia University, 1988.

Hijar, Katherine. "The Pin-Up, the Piano, and the Parlor: American Sheet Music, 1840–1860." *Imprint* 30, no. 2 (2005): 7–21.

"Historically Significant U.S. Copyright Legislation." *Music Library Association.* Accessed 20 September, 2022. https://www.musiclibraryassoc.org/mpage/copyright_hsl#1831.

Hoogerwerf, Frank W. *Confederate Sheet-Music Imprints*. 21 Brooklyn, NY, 1984.

Horowitz, Daniel. *The Morality of Spending: Attitudes Towards the Consumer Society in America, 1875–1940.* Baltimore, Md.: Johns Hopkins Univ. Press, 1985.

Horwitz, Howard. *By the Law of Nature: Form and Value in Nineteenth-Century America*. New York: Oxford University Press, 1991.

Hunter, David Chalmers. "English Opera and Song Books, 1703–1726: Their Contents, Publishing, Printing, and Bibliographical Description." PhD diss., University of Illinois at Urbana-Champaign, 1989.

Jumonville, Florence M. "Set to Music: The Engravers, Artists, and Lithographers of New Orleans Sheet Music." *Proceedings of the American Antiquarian Society* 105, no. 1 (1995): 127.

Kaestle, Carl F. and Helen Damon Moore. *Literacy in the United States: Readers and Reading Since 1880.* New Haven: Yale University Press, 1991.

Kallberg, Jeffrey. "The Harmony of the Tea Table: Gender and Ideology in the Piano Nocturne." *Representations*, No. 39 (Summer, 1992): 102–133.

Karpf, Juanita. "'In an Easy and Familiar Style:' Music Education and Improvised Accompaniment Practices in the U.S., 1820–80." *Journal of Historical Research in Music Education* 32, no. 2 (2011): 122–44.

Key, Susan. "Sound and Sentimentality: Nostalgia in the Songs of Stephen Foster." *American Music* 13, no. 2 (1995): 145–66.

Koza, Julia Eklund. "Music and the Feminine Sphere: Images of Women as Musicians in "Godey's Lady's Book," 1830–1877." *The Musical Quarterly* 75, no. 2 (1991): 103–29.

———. "Music Instruction in the Nineteenth Century: Views from Godey's Lady's Book, 1830–77." *Journal of Research in Music Education* 38, no. 4 (1990): 245–57.

Krohn, Ernst C. *Music Publishing in St. Louis*. Ed. J. Bunker Clark. Warren, MI: Harmonie Park Press, 1988.

Kwolek-Folland, Angel. "The Elegant Dugout Domesticity and Moveable Culture in the United States, 1870–1900." *American Studies* 25, no. 2 (1984): 21–37.

Laderman, Gary. *The Sacred Remains: American Attitudes Towards Death, 1779–1883*. New Haven, Conn.: Yale Univ. Press, 1996.

Lears, Jackson. *Fables of Abundance: A Cultural History of Advertising in America.* New York: Basic Books, 1994.

Levine, Lawrence W. *Highbrow/Lowbrow: The Emergence of Cultural Hierarchy in America.* Cambridge: Harvard University Press, 1988.

Lott, Eric. *Love and Theft*. 20th-anniversary ed. Oxford: Oxford Univ. Press, 2013.

Loughridge, Deirdre. "Making, Collecting, and Reading Music Facsimiles before Photography." *Journal of the Royal Music Association* (May 2016): 27–59.

Love, Timothy M. "Irish Nationalism, Print Culture and the Spirit of the Nation." *Nineteenth-Century Music Review* 15, no. 2 (2018): 189–208.

Martin, Edgar W. *The Standard of Living in 1860: American Consumption Levels on the Eve of the Civil War.* Chicago, IL: Univ. of Chicago Press, 1942.

McGaw, Judith A. *Most Wonderful Machine: Mechanization and Social Change in Berkshire Paper Making, 1801–1885*. Princeton: Princeton University Press, 1987.

McMullin, B.J. "Machine-Made Paper, Seam Marks, and Bibliographical Analysis." *The Library: The Transactions of a Bibliographic Society* 9, no. 2 (March 2008): 62–88.

Meyer-Frazier, Petra. "American Women's Roles in Domestic Music Making as Revealed in Parlor Song Collections: 1820–1870." PhD diss., University of Colorado, 1999.

———. "Music, Novels, and Women: Nineteenth-Century Prescriptions for an Ideal Life." *Women and Music: A Journal of Gender and Culture* 10, no. 1 (2006): 45–59.

Meyer-Frazier, Petra and Michael J Budds. *Bound Music, Unbound Women: The Search for an Identity in the Nineteenth Century*. Missoula, Montana: College Music Society, 2015.

Mondelli, Peter. "The Sociability of History in French Grand Opera: A Historical Materialist Perspective." *19th-Century Music,* Vol. 37, No. 1 (Summer 2013): 37–55.

———. "Parisian Opera between Commons and Commodity, ca. 1830." In *Consuming Music: Individuals, Institutions, Communities, 1730–1830*. Rochester: University of Rochester Press, 2017.

Morabito, Fabio. "Rehearsing the Social: Beethoven's Late Quartets in Paris, 1825–1829." The Journal of Musicology Vol. 37, Issue 3, pp. 349–382.

Nagao, Isaac. "Beyer, Ferdinand." *Oxford Music Online*. 2001; Accessed 20 Jan. 2022. https://www-oxfordmusiconline-com.libproxy.library.unt.edu/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000051838.

Northcutt, Richard. *The Life of Sir Henry Bishop*. London: The Press Printers, 1920.

Notley, Margaret. "Brahms as Liberal: Genre, Style, and Politics in Late Nineteenth-Century Vienna." *Nineteenth-Century Music,* Vol. 17, No. 2 (Oct. 1993): 107–123.

"OCLC Research Activities and IFLA's Functional Requirements for Bibliographic Records." *OCLC*. https://www.oclc.org/research/activities/frbr.html.

Orvel, Miles. *The Real Thing: Imitation and Authenticity in American Culture, 1880–1940*. Chapel Hill: University of North Carolina Press, 1989.

Pehl, Christa. "Men, Women, and Their Manuscripts: Musical Culture in Pennsylvania in the 1790s." PhD diss., Princeton University, 2017.

Pendergast, Tom. "Consuming Questions: Scholarship on Consumerism in America to 1940." *American Studies International; Washington* 36, no. 2 (Jun 1998): 23–43.

Ping, Nancy Regan. "Music in Antebellum Wilmington and the Lower Cape Fear of North Carolina." PhD diss., University of Colorado, 1979.

Rumbley, Erica Joy. "From Piano Girl to Professional: The Changing Form to Music Instruction at the Nashville Female Academy, Ward's Seminary for Young Ladies, and the Ward-Belmont School, 1816–1920." PhD diss., University of Kentucky, 2014.

———. "Ornamental Music and Southern Belles at the Nashville Female Academy, 1816–1861." *American Music* 33, no. 2 (2015): 219–50.

Samuels, Shirley. *The Culture of Sentiment: Race, Gender, and Sentimentality in Nineteenth-Century America*. New York: Oxford University Press, 1992.

Sanjek, Russell. *From Print to Plastic: Publishing and Promoting America's Popular Music (1900–1980)*. Brooklyn, N.Y: Institute for Studies in American Music, Conservatory of Music, Brooklyn College of the City University of New York, 1983.

———. *American Popular Music and Its Business, II, From 1790 to 1909*. 1988.

Sellers, Charles Grier. *The Market Revolution: Jacksonian America, 1815–1846*. New York: Oxford University Press, 1994.

Sengupta, Oishani "'Images of Various Species': Remediating Visual Data in the William Blake Archive." Paper presented at the annual meeting of the Nineteenth Century Studies Association, Rochester, NY, March 2021.

Slobin, Mark, James W. Kimball, et al. *Emily's Songbook: Music in 1850s Albany*. Middleton, WI: A-R Editions, 2011.

Stoker, David. "John Marshall, John Evans, and the Cheap Repository Tracts, 1793–1800." *The Papers of the Bibliographical Society of America* 107, no. 1 (2013): 81–118.

Staubach, Doug. "VLookup (LookupTableAndFieldName ; MatchingValueOrList ; ReturnFieldNames." *Brian Dunning's Custom FileMaker Functions.* Accessed Jan. 19, 2022): https://www.briandunning.com/cf/1740.

Stephens, John A., William McClellan, and Andrew Lamb. "Russell, Henry." *Grove Music Online*. 31 Jan. 2014; Accessed 21 Oct. 2022. https://www-oxfordmusiconline-com.libproxy.library.unt.edu/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-1002258090.

Tawa, Nicholas E. *Sweet Songs for Gentle America: The Parlor Song in America, 1790–1860*. Bowling Green, Ohio: Bowling Green University Popular Press, 1980.

-----. *A Music for the Millions: Antebellum Democratic Attitudes and the Birth of American Popular Music*. New York: Pendragon Press, 1984.

Thompson, James, 1951. *Models of Value: Eighteenth-Century Political Economy and the Novel*. Durham: Duke University Press, 1996.

Thornton, Tamara Plakins. *Handwriting in America*. New Haven, Conn.: Yale Univ. Press, 1996.

Tomlinson, Gary. *A Million Years of Music: The Emergence of Human Modernity.* Brooklyn: Zone Books, 2015.

Urberg, Michelle. "Past and Futures of Digital Humanities in Musicology: Moving Towards a Bigger Tent," *Music Reference Services Quarterly,* Vol. 20, no. 3–4 (2017): 134–150.

"The U.S. Census Bureau Fast Facts of the 1850s." *United States Census Bureau*. https://www.census.gov/history/www/through_the_decades/fast_facts/1850_fast_facts.html

"The U.S. Census Bureau Fast Facts of the 1860s." *United States Census Bureau*. https://www.census.gov/history/www/through_the_decades/fast_facts/1860_fast_facts.html

Ward, Keith C. and Michael J. Budds. *For The Parlor and The Concert Stage: A Guide to Recent Collections of American Piano Music from the Classic and Romantic Eras.* Hillsdale, N.Y.: Pendragon Press, 2010.

"What Is a Marc Record, and Why Is It Important?" *Library of Congress* information booklet, 2009. https://www.loc.gov/marc/umb/um01to06.html.

Wilhite, Ann L., and Charles S. Wilhite. "Grobe, Charles." *Grove Music Online.* 2001; Accessed 21 Oct. 2022. https://www-oxfordmusiconline-com.libproxy.library.unt.edu/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000011804.

Wilkinson, Christina. "The Musical Legacy of Maspeth Firth's Family." *The Juniper Berry,* March 6, 2020. https://junipercivic.com/juniper-berry/article/the-musical-legacy-of-maspeths-firth-family.