A Relational Approach to Musicological Data Management:

How to Create a Coded Database in Obsidian

Erin Lupardus

In this report, I offer a systematic method of tackling the common issue of data management and analysis when it comes to large sets of textual material using the open-source program Obsidian. Though originally intended as a note-taking software, Obsidian allows for the creation of an inter-linked, multi-leveled database, which facilitates the discovery of synchronicities and discrepancies in the sources at hand. I explain how to use Obsidian's built-in features alongside select user-created plug-ins to organize sources, filter targeted data from them, exploratively code connections between them, and visualize these connections in an intuitive manner.

The increasing assimilation of digital humanities concepts into methods of (historical) musicology research has opened new avenues for managing, analyzing, and interpreting historical data, and the case of my own doctoral research is no exception. I am investigating historical perspectives found in treatises and other select textual sources of the 16th–18th centuries on the characterization of the bass violin, an undertaking which currently involves the inspection of passages from more than 100 historical sources, totaling about 60,000 original-language words. Given the large amount of data and the considerable chosen time frame, data management quickly and understandably became an initial hurdle in my analysis, hence the need for an easily manageable and user-friendly database software. With inspiration in mind from the network and relationship graphing functions of the Musikinstrumenten Museum of the Universität Leipzig's musiXplora[1] as well as Niklas Luhmann's famous *Zettelkasten* system[2], a YouTube search for how-to's on Zettelkasten-esque software programs led me to Obsidian, an open-source, free software that saves all files locally to your own device, thus reducing potential copyright infringement concerns resulting from cloud-based storage. Though initially designed as a note-taking or journaling application, given its flexibility, I immediately recognized its potential to be used to create a database. Especially when utilizing some of the many available user-created plug-ins, Obsidian, in my opinion, serves as an effective data management tool in musicological research. In light of the enthusiastic reaction to the presentation of my database at the 2024 symposium of the Junge Musikwissenschaft der ÖGMw[3] and its application to examine textual evidence of 6-string *violoncelli* at the 2025 joint

1 See Josef Focht (ed.), *musiXplora* [2015–2025], <https://musixplora.de/musici/> (03.02.2024).

2 See Johannes F. K. Schmidt, "Der Zettelkasten Niklas Luhmanns" [n.d.], <https://niklas-luhmann-archiv.de/nachlass/zettelkasten> (01.02.2024).

3 Erin Lupardus, "Reapproaching Bass Violin Characterization: A First-Stage Analysis of Bowed Bass Descriptions of the 17th and 18th Centuries", Presentation, Mozarteum University Salzburg,

postgraduate symposium of the Society for Musicology in Ireland and International Council for Traditions of Music and Dance in Ireland[4], I am pleased to offer this brief introduction on creating a database in Obsidian in the hopes of equipping fellow researchers with a method of tackling, organizing, exploring, and analyzing large data sets.

As should be the case with any faithful inductive reasoning approach, familiarization with the data should be guaranteed before proceeding with any analysis or the development of definitions to be used to code and categorize the data.[5] Throughout the process of reading, re-reading, transcribing[6], and translating[7] the descriptions of bowed basses within my chosen corpus, I made note of which characteristics are frequently indicated such as the number of strings, the presence or absence of frets, and given tunings. Simultaneously, I began experimenting with Obsidian's functions and creating database entries therein in order to determine how best to structure my data. From the observations gleaned from this phase of interaction with the data set, I was then able to determine historical criteria used to differentiate double basses, bass viols, and bass violins from one another. These criteria were used to establish 'historically-informed' definitions to be used to sort the varied and numerous bowed basses of these three centuries into broader categorizations as either double basses, bass viols, or bass violins. With these foundational steps realized, it was time to fine-tune and code the Obsidian database.

18.10.2024, <https://oegmw.at/en/junge-musikwissenschaft/nachwuchssymposium/sym-2024> (23.04.2026)

4 Erin Lupardus, "A 6-String Baroque Violoncello? A Report on the Benefits of Database Implementation in Textual Analysis", Presentation, Trinity College Dublin, 17.01.2025.

5 Critically, familiarization with the data is also the first step in Thematic Analysis as outlined in Virginia Braun and Victoria Clarke, "Using Thematic Analysis in Psychology", in: *Qualitative Research in Psychology* 3/2 (2006), pp. 77–101, here pp. 93–94.

6 In order to preserve as much of the original text formatting as possible, I transcribed the entries manually into Word. AI-powered transcription tools proved too unreliable to be useful. I sought to mirror as much of the original formatting as possible since extra-lingual decisions (e.g. text layout on the page) also has implications on meaning. Unfortunately, Obsidian does not offer as many text functionality features as Word, for example paragraph indentations, mixed font sizes, or site-specific placements of images. Therefore, I do not recommend using Obsidian solely to store text-based data sets.

7 I am grateful to Enrico Coden (Italian), Renata Kambarova (French), Amaya Blanco Rad and Beatriz Gonzalez Crespo (Spanish), and Petra Spanos (Latin) for their talents in translating passages neither in German nor English.

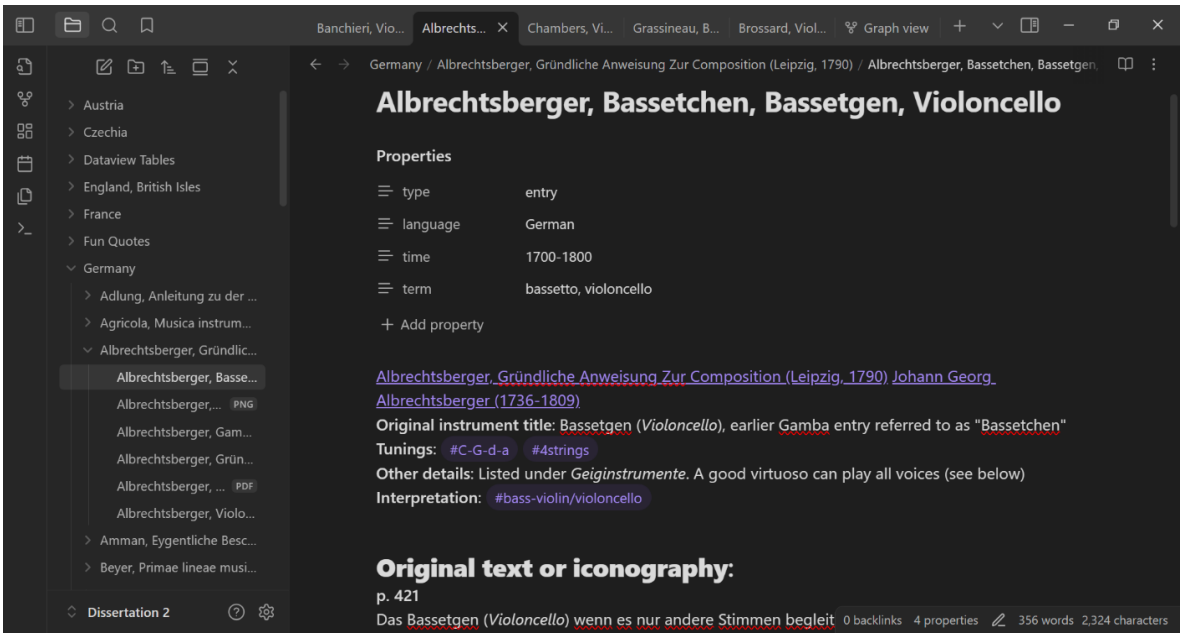


Figure 1: A screenshot showing the organization of Albrechtsberger's (Leipzig, 1790) *Bassetgen* entry within Obsidian.

Ultimately, in order to maintain an overview of all of the details pertinent to my research, I decided to create individual "notes", which Obsidian saves locally as Markdown text files, for the source itself, for its bowed bass entries according to the original nomenclature used in the source, and for its author. These notes are sorted and organized into nested folders according to the source title and publishing location (see Figure 1).[8] If available, I labeled and attached PDF scans of the manuscript as well as screenshots of relevant diagrams in the sources to the notes for easy reference. Biographies of historical figures such as the authors themselves, their patrons, and their personal and professional contacts are collected from sources such as *Grove Music Online*, *MGG*, or the *Österreichisches Musiklexikon online* and organized into separate notes in a 'Historical Figures' folder. Even with all this collected information, however, notes without codes are merely a different format of my transcriptions and sources, so analysis naturally requires coding and interacting further with the data within the database.

Based on Obsidian's base functions and the available plug-ins[9], I coded the data in several ways so as to be able to display and extract different levels of information during my analysis (see Table 1). Data inherent to the source—the type of source/information the note denotes (treatise, dictionary, an entry therein, etc.), the language the source is written in, in which century it was written, whether the source mentions a bass violin, and if the note is for an individual entry, the original term(s) used by the

8 Obviously certain geographical boundaries were at times far more complex, ambiguous, and fluid than today, but for the sake of having a general sense of where these texts were published, the folders correspond to modern-day countries. The folders are used strictly for organizational purposes and not analytical ones: in contrast, the hashtag-ed location data is organized according to historically defined regions.

9 In order to create a database similar to mine, the plug-ins "Dataview" by Michael Brenan and "Nested tags graph" by drpilman are needed. I also recommend "Tag Wrangler" by PJ Eby to be able to rename and delete hashtags.

author for the instrument[10], etc.—is marked using the "Properties" feature, which essentially is a method of adding metadata to the note. In Figure 1, one can see how the "Properties" are displayed at the topmost part of the note. Similar to the way Wikipedia functions, related notes can be 'linked' to one another, for example, the note for a source author should be connected to the source they wrote via a purple hyperlink, or a source to all of its individual bowed bass entries. The arrangement of these links in the note can be seen in Figure 1 directly underneath the "Properties". I also link notes when one source directly mentions another within the text or in instances where I notice that one source has 'plagiarized' another. Based on the source content or biographies, hashtags are used to keep track of locations (i.e. publishing locations or places of birth and activity), mentioned tunings, numbers of strings, my interpretation of the term according to the inductively constructed definitions, etc. Therefore, hashtags are used not only to code the investigated physical characteristics but also to keep track of my interpretation of how the instrument can be categorized.

These codes can then be used to aid in analyzing, comparing, contrasting, and contextualizing the information found in the selected corpus. For example, by exploratively coding connections between sources and details in their authors' biographies, I can consider not only the explicitly written references but also sociopolitical circumstances and professional and personal relationships in an effort to trace the possible informants of the authors' descriptions. The Dataview plug-in can be used to generate structured tables showing all notes containing queried codes, for example all instances of 6-stringed bass violins, thus simplifying the process of reviewing and comparing specific sources (Figure 2). Besides Dataview tables, Obsidian's flagship "Graph view" illustrates the coded links, both between individual notes and with hashtags, which supports the process of pattern identification (Figure 3 and Figure 4). In this "Graph view", "nodes" are displayed, which represent both notes and hashtags. Hashtags are automatically colored green by Obsidian, but in order to better visualize the data, I also color-coded the other nodes according to their properties: source entries are yellow whereas the sources themselves are various shades of blue and purple according to the type of source, and historical figures are red. When viewing a note, a "Local graph"[11] can be opened to show all incoming and outgoing links (and hashtags, if desired)—the "depth" can also be adjusted to expand the context window of the links (Figure 5). All of these functions have proven to be indispensable tools in my analysis, most recently in decoding the more perplexing uses of nomenclature in each language: the summarized Dataview tables made referencing every instance of a specific term easy and efficient.

10 To not lose track of terms with different spellings or equivalent terms such as *Contrebasse* and *Contrabasso*, I made the decision to standardize spellings in the "Properties". Even after this consolidation of terms, around 80 different terms are used in the corpus for bowed basses.

11 Unfortunately, no color-coding of nodes is possible in the "Local graph" view.

	properties (metadata)	links	hashtags
Searchable using the search bar?	X		X
Nestable?			X (with the "Nested tags graph" plug-in)
Displayed in Dataview tables (with the Dataview plug-in)?	X		X
Displayed visually in "Graph view"?	not in the graph itself, but can be used to color-code nodes	X	X (but can be untoggled)

Table 1: Functions of Coding Methods in Obsidian

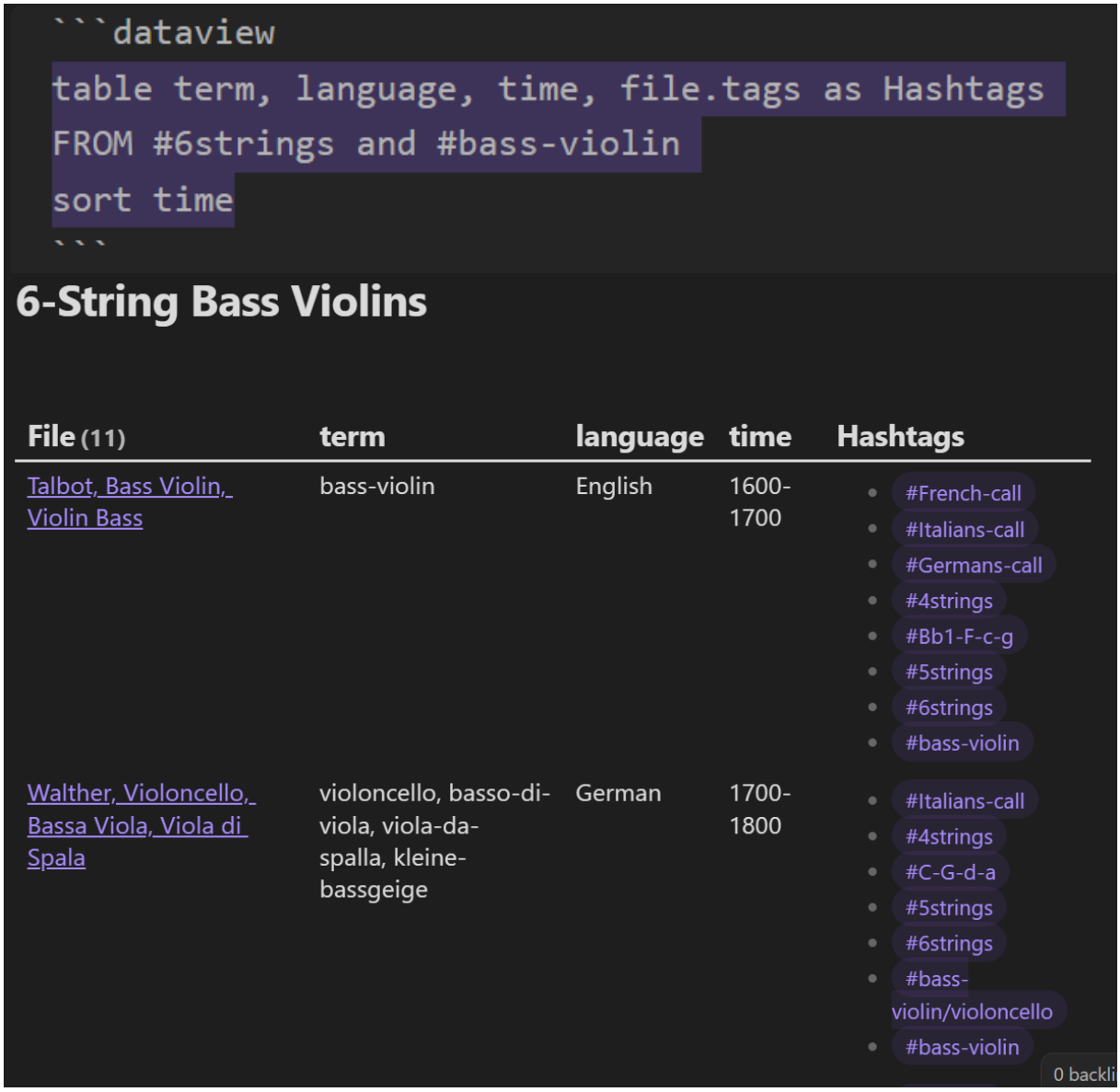


Figure 2: A screenshot demonstrating how to instruct the Dataview plug-in to create a chronologically-sorted table showing all the notes in the database that mention 6-string bass-violins and listing the terms used, the language the source is in, the century it was written in, and the hashtags coded in the note (above) results.

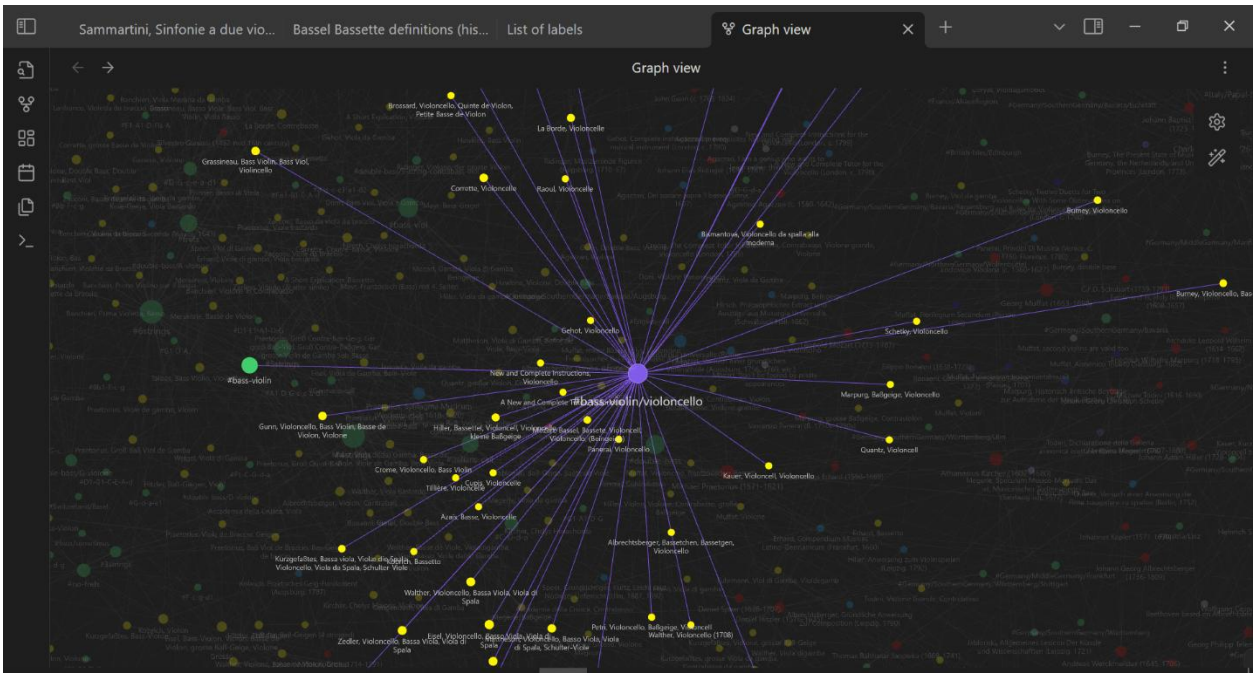


Figure 3: A screenshot of "Graph view" with the mouse hovered over a specific node, highlighting all of the node's connections. Here, for example, are all of the notes and nested hashtags connected to #bass-violin/violoncello, i.e. all the instruments in the database interpreted to be a violoncello.

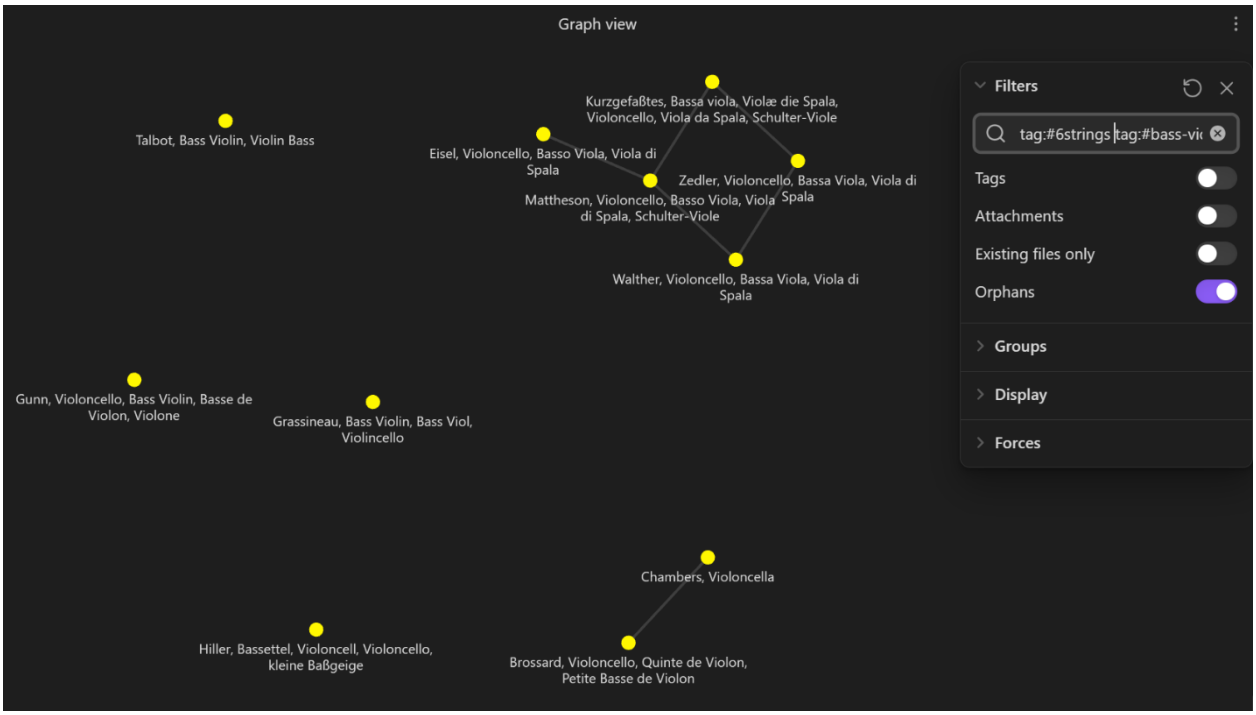


Figure 4: A screenshot of search filters applied in "Graph view". Here, hashtags are hidden and all notes mentioning 6-string bass-violins are displayed. Lines between nodes illustrate links.

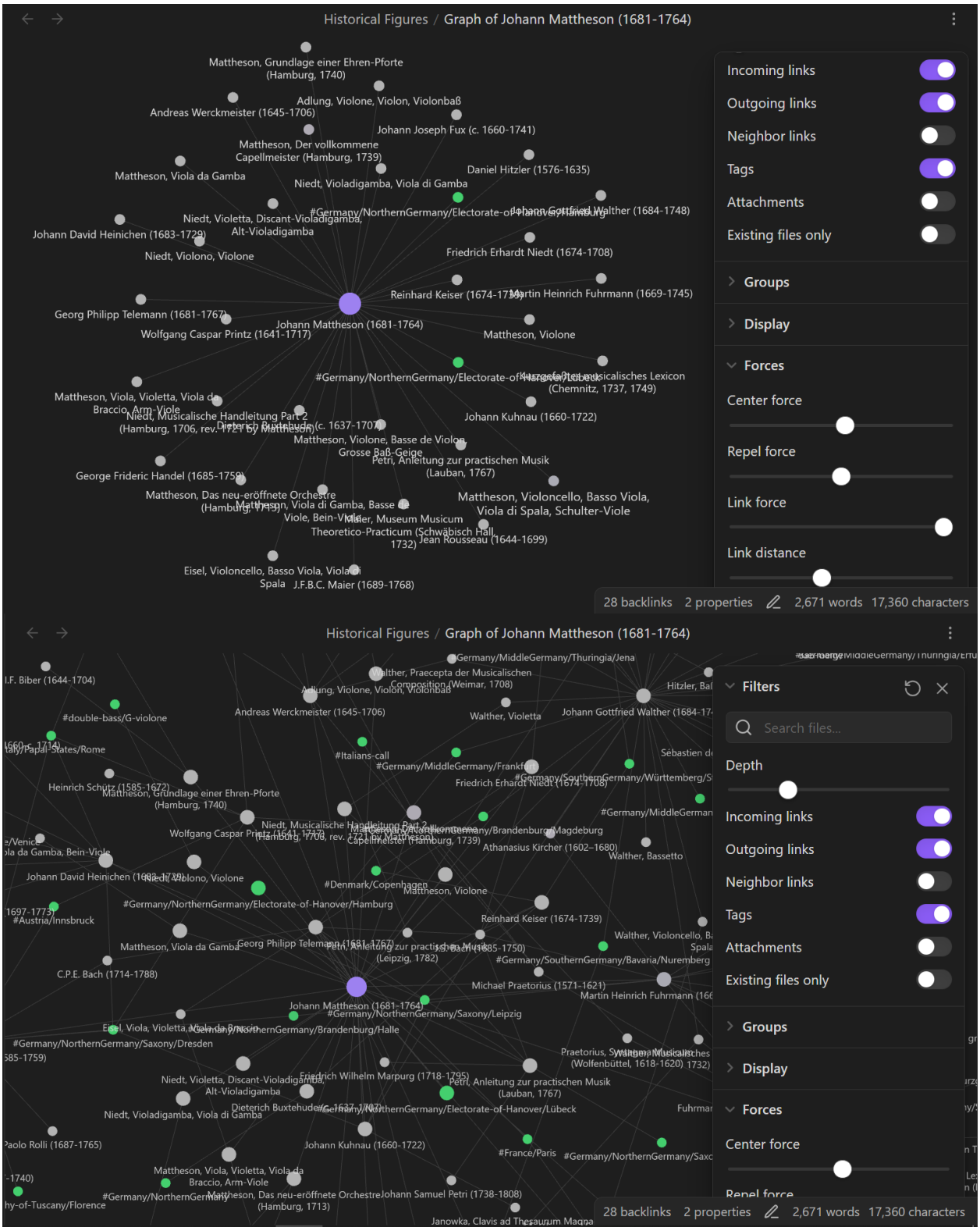


Figure 5: In the screenshot above, the one-level context window for all notes and hashtags linked to Johann Mattheson are displayed in the “Local graph”. In the screenshot below, a small section of the two-level context window “Local graph” is shown.

Notably, more qualitatively descriptive data, e.g. described sound quality, aesthetics, and sociocultural associations, is not coded within my Obsidian database. Using Inductive Thematic Analysis, I will apply a second round of coding to the texts with the help of a second program, MaxQDA. This two-phased,

'two-software' approach not only allows greater ease in finding synchronicities and discrepancies in the primary sources but also maintains a degree of separation between clearly specified details in the texts and semantics that are more open to speculation—this separation is necessary in order to keep track of my own understanding of these sources as well as to be able to retrace the interpretations of other musicologists to determine which of their explanations are more plausible than others. Through these processes, I can examine the textual *and* contextual dimensions of these instrument descriptions and therefore aim for a greater understanding of bass violin characterization than has until now been achieved.

•••••••

This report emerged from my eager desire to share the database method I developed in the framework of my doctoral research, "The Mythical Monolith: Reapproaching Bass Violin Characterization According to 17th- and 18th-Century Treatises". I am indebted to my first supervisor, Dr. Lars-Edvard Laubhold at the Anton Bruckner University of the State of Upper Austria, who originally encouraged me to create such a database, for his tireless support throughout my studies.