Building a Comprehensive Sheet Music Library Application

Klaus Rettinghaus
Enote GmbH, Berlin, Germany
klaus.rettinghaus@enote.com

Alexander Pacha
Enote GmbH, Berlin, Germany
alexander.pacha@enote.com

Abstract

Digital symbolic music scores offer many benefits compared to paper-based scores, such as a flexible dynamic layout that allows adjustments of size and style, intelligent navigation features, automatic page-turning, on-the-fly modifications of the score including transposition into a different key, and rule-based annotations that can save hours of manual work by automatically highlighting relevant aspects in the score. However, most musicians still rely on paper because they don't have access to a digital version of their sheet music, or their digital solution does not provide a satisfying experience. To bring digital scores to millions of musicians, we at Enote are building a mobile application that offers a comprehensive digital library of sheet music. These scores are obtained by a large-scale Optical Music Recognition process, combined with metadata collection and curation. Our material is stored in the MEI format and we rely on Verovio as a central component of our app to present scores and parts dynamically on mobile devices. This combination of the expressiveness of MEI with the beautiful engraving of Verovio allows us to create a flexible, mobile solution that we believe to be a powerful and true alternative to paper scores with practical features like smart annotations or instant transpositions. We also invest heavily into the open-source development of Verovio to make it the gold standard for rendering beautiful digital sheet music.

Introduction

Many parts of our lives have become digital in the last decades, from navigating on the street to buying groceries. This digitization allowed us to do things we were not able to do several years ago, like cars that are driving autonomously. In other areas of our life, it helped to make existing processes more efficient or more convenient.

The digitization of music scores is one of these areas. For over 50 years researchers and companies have been working on typesetting music on a computer [18], digitizing scores [14], or collecting and publishing a digital library of public domain scores [13]. These projects have led to many products that are widely used today (e.g., MuseScore, Finale, Sibelius, and Dorico are used by millions of musicians to typeset music scores). However, when practicing and performing, many musicians still rely on printed music scores for understandable reasons: they are universally available, easy to handle, and do not suffer from technical problems like an empty battery. Even more so, existing digital alternatives are often not a full replacement for paper or fail to unlock the full potential of digital scores, and therefore get rejected quickly. For example, taking a photo of a paper score and opening the image on a tablet is a digital solution that reduces the amount of paper that needs to be carried around but might suffer from a loss of quality. Instead, a machine-readable, symbolic version of the music score can be obtained through Optical Music Recognition (OMR) [2]. Recent research suggests that many challenges of OMR can be solved quite robustly with machine-learning approaches [1, 20, 21]. However, these are never guaranteed to be completely error-free, and even if a flawless digital score can be obtained in the end, e.g., by manually typesetting it in a music score editor, a satisfying solution for musicians must include a way to easily view and interact with that score.

To build such a solution, the start-up Enote GmbH was founded in 2018 by Boian Videnoff, Josef Tufan, and Evgeny Mitichkin to revolutionize digital sheet music. Enote is based in Berlin and has around 30 employees working on a mobile application for iOS that musicians can use to display and interact with our digital music scores. In this paper, we want to give some insights into our solution as well as why and how we use MEI and Verovio for our product.

Digital Music Scores Challenges

There are several major challenges when creating a music score application like Enote, including questions on how to obtain the material that the users will see, how to ensure a high quality, how to store the digital music scores, and how to create a satisfying digital experience on a mobile device? Figure 1 gives an overview of our digitization efforts that are discussed in the following chapters.

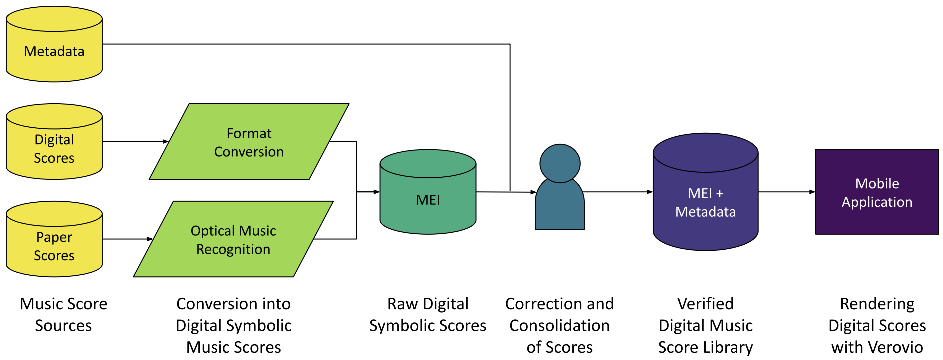


Figure 1: Our digitization efforts in a nutshell: Scores and Metadata are obtained from several public sources. Paper-based scores are converted into digital symbolic scores by an Optical Music Recognition system and manually corrected and enriched with corresponding metadata to build a verified Digital Music Score Library that can be displayed on a mobile device with Verovio.

Building a Digital Music Score Library

Thousands of music scores have been written in the past. They exist in many forms and can be obtained from several sources including books or existing digital collections like IMSLP, MuseScore.com, or CPDL [10]. Given the legal protection of music (scores), an agreement with the copyright holders must be obtained, or one has to restrict oneself to copyright-free material. For the beginning, Enote settled with the second option, focusing on public-domain music that can be used and distributed freely.

Although many scores can be obtained or purchased, the quality of the material varies significantly, reaching from poor scans of printed music to digitally encoded scores of very high quality [6]. Unfortunately, free, high-quality, digitally encoded music is only sparsely available. Therefore, to ensure the quality of our material, we are developing a highly-accurate Optical Music Recognition (OMR) system that uses state-of-the-art machine-learning approaches and can create a digitally encoded version of the available music scores. Several machine-learning approaches have been shown to work quite well for visually recognizing objects in typeset and handwritten music scores [11, 19]. These approaches have in common that they make heavy use of deep convolutional neural networks. However, even if the visual recognition works well, a complete OMR system also requires a semantical reconstruction stage [16] that builds a suitable musical representation [8, 12] out of these detections, which can be exported into a common digital music format like MusicXML or MEI. The difficulty of building this stage is frequently underestimated and has not received a lot of attention from the research community. Most existing OMR systems attempt to solve this challenge with an extensive set of hand-crafted rules.

Rule-based systems can give quick results but become harder to maintain and expand over time. One reason is that music notation is often used as freely as the music it encodes. Many rules and conventions of Common Western Music Notation [7] have evolved over time and composers as well as editors frequently bent the rules to capture a musical idea with the tools they had at that time (see Figures 2 and 5). This leads to the problem that regardless of how well an OMR system works and how many rules you implement, there are always exceptions that require special treatment. And the more material you collect, the more questionable editorial decisions you will discover (even in many modern editions from esteemed publishing houses). In [3], the authors proposed an interesting idea on how to deal with this problem, by allowing the operator of an interactive system to dynamically enable and disable rules and constraints during reconstruction, allowing the system to handle exceptional cases gracefully while working robustly on most scores with a reduced set of constraints.

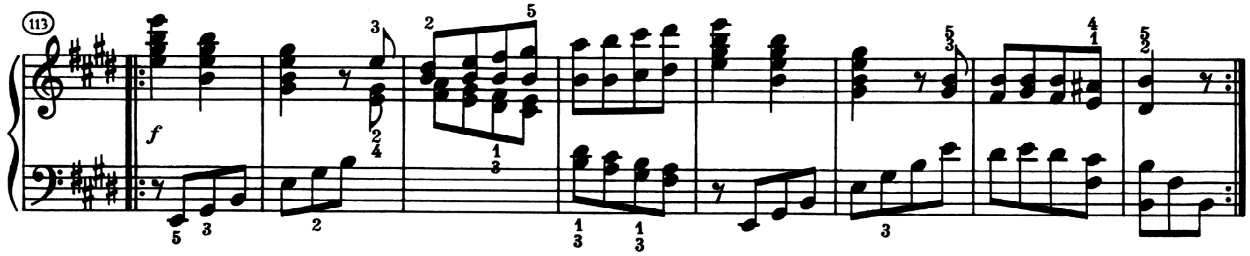


Figure 2: Example of Beethoven's "Alla Ingharese Quasi Un Capriccio" ("Die Wut über den verlorenen Groschen"), Op. 129, where the number of notes in the last measure is missing one eighth note needed to complete the expected number of notes for the given time signature.

Given the current state of the art, any OMR system that aims at producing flawless, digital scores requires manual verification, more often some form of manual correction. Our musicology department is in charge of this process, along with the selection of pieces that are digitized, the consolidation of multiple sources, and advising the software engineers in music-related questions. Scores that have minor errors are manually corrected by our musicologists in a slightly adapted version of the Humdrum Verovio Viewer [17], which is also publicly available.

Collecting Metadata

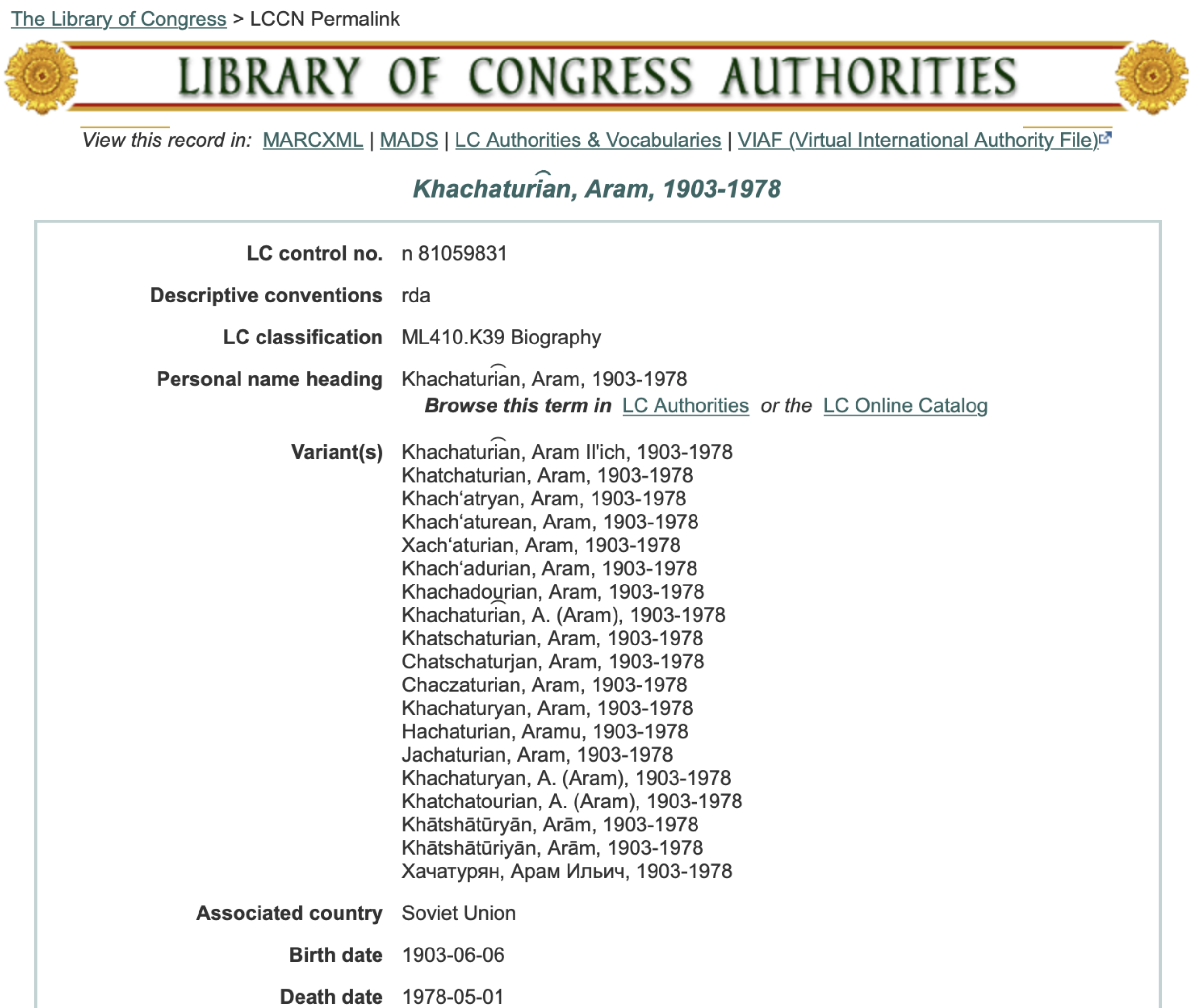
The most complete music library would still be pretty useless if you weren't able to find what you're looking for. To make the library usable and searchable, additional metadata is required. Therefore, we are complementing all of our pieces with extensive metadata that we've collected, such as information about the composer or the instrumentation. Although it is possible nowadays to fetch a lot of metadata directly from sources like IMSLP or Wikidata – that alone is not enough. The collected data can be inconsistent, incomplete, or generally of poor quality. And even if you start an intense search, some information can be genuinely hard to obtain, e.g., information about the duration of a piece, even though several recordings might exist.

After collecting the metadata, you have to make sure that users can actually find what they are looking for. Although this can seem like an easy task, it has become very tricky in a diverse and open field such as music. Think of terms like *Charakterstück* or *Motet*, which can mean something quite specific or almost anything depending on your point of view. What would you expect to find if you search for *Charakterstück*? Works by Couperin, only works of the 19th century, or even just pieces that contain that specific phrase in their title? Not to mention the many possible spellings of non-western composers, like Aram Khachaturian, Sergej Rachmaninoff, or Tōru Takemitsu, which is a well-known problem in information science, but cannot yet be handled sufficiently by relying on authority controlled data alone. Even public authority records like those from the Library of Congress have substantial gaps in this regard (see Figure 3).

You may clean up and sort some of the metadata automatically, but a rule-based system quickly reaches its limits here. Our musicology department invests extensive brainpower in compiling, completing, and consolidating the metadata to help musicians find what they are looking for in our music library. We do so by comparing our information to the latest available scientific catalogs of works provided by scholars all over the world.

Storing the Musical Data

For storing our digital sheet music we decided to use MEI [9], as it brings some major advantages over several other non-binary, open music encoding formats we evaluated, such as ABC, Guido, Humdrum, or MusicXML.
While rumors have it that MusicXML [5] will more or less reach its end-of-life with the latest version 4.0 release, its designated successor MNX is still only a draft specification. MEI on the other hand is under active development and has seen multiple iterations over the last couple of years. MEI is also quite well documented – at least compared to MusicXML. This is one of the reasons why different implementations of MusicXML diverge substantially from one another and, based on our experience and testing, not a single one seems to be feature-complete. The status of ABC is quite unclear, as the next version to be released is in the making since early 2013, and its documentation website keeps disappearing from time to time.


The Library of Congress > LCCN Permalink

LIBRARY OF CONGRESS AUTHORITIES

View this record in: MARCXML | MADS | LC Authorities & Vocabularies | VIAF (Virtual International Authority File)

Khachaturi͡an, Aram, 1903-1978

LC control no.	n 81059831
Descriptive conventions	rda
LC classification	ML410.K39 Biography
Personal name heading	Khachaturi͡an, Aram, 1903-1978 ***Browse this term in*** LC Authorities *or the* LC Online Catalog
Variant(s)	Khachaturi͡an, Aram Il'ich, 1903-1978 Khatchaturian, Aram, 1903-1978 Khach'atryan, Aram, 1903-1978 Khach'aturean, Aram, 1903-1978 Xach'aturian, Aram, 1903-1978 Khach'adurian, Aram, 1903-1978 Khachadourian, Aram, 1903-1978 Khachaturi͡an, A. (Aram), 1903-1978 Khatschaturian, Aram, 1903-1978 Chatschaturjan, Aram, 1903-1978 Chaczaturian, Aram, 1903-1978 Khachaturyan, Aram, 1903-1978 Hachaturian, Aramu, 1903-1978 Jachaturian, Aram, 1903-1978 Khachaturyan, A. (Aram), 1903-1978 Khatchatourian, A. (Aram), 1903-1978 Khātshātūryān, Arām, 1903-1978 Khātshātūriyān, Arām, 1903-1978 Хачатурян, Арам Ильич, 1903-1978
Associated country	Soviet Union
Birth date	1903-06-06
Death date	1978-05-01

Figure 3: Excerpt of the Library of Congress's authority record for Aram Khachaturian showing several variant spellings – his original Armenian spelling, however, is missing here.

One of the strongest arguments for MEI is the ability to extend the format to fit personal needs in a tailor-made customization, while also allowing us to participate in its active development as part of the community. A huge advantage over similar formats is MEI's 'semantic approach' in encoding musical information. Take the encoding of hairpins as an example: MusicXML connects their start- and endpoints to noteheads. While this sounds like a sensible choice, it can become a problem in certain situations, like a hairpin ending on a barline. In MusicXML, you have to manipulate the hairpin graphically by specifying a relative offset to achieve the desired visual result. This makes encoding of a *messa di voce* messy and will probably result in a broken layout if you compress or stretch the score. In contrast, MEI allows you to set the beginning and ending of hairpins within the flow of the music and thus allows for a truly dynamic layout.

MEI is also well-suited for OMR applications, as it allows to explicitly model the graphical information of the underlying source, which cannot be done in MusicXML. By using native features (taken from TEI), it is possible to link each semantic musical object exactly to the respective graphical object in the source, which, e.g., allows features like a visual playback overlaid on a scan of the original printed score.

Presenting Scores

To display the sheet music in our mobile application, the Verovio music notation engraving library is used [15]. We decided to use Verovio for score rendition not only because it is the natural choice to engrave MEI files but because it is a fast and reliable cross-platform application. Furthermore, it is free open-source software, which allows us to enhance it with features we need.

Figure 4: Comparison of a small excerpt from Carl Phillip Emanuel Bach's "Prussian Sonatas", rendered in our four different notation styles, using different font faces. From top to bottom: *classic*, *traditional*, *legacy*, and *modern*.

There are already other mobile apps available using Verovio, like *Cantico*, *NomadPlay*, and *Trala*, but we make heavy use of the possibilities to use different SMuFL compliant fonts and styling features. For example, we offer four different styles for displaying the music, which we named *classic, traditional, legacy,* and *modern* (see Figure 4).

Our legacy style, for example, uses a custom-created font named *Legato* [4], which is based on Steinberg's open-licensed Bravura font. We also extended Verovio's own Leipzig font with additional glyphs to cover a broader range of the Standard Music Font Layout.

We strive for a consistent look and feel throughout our scores. That is why apart from simply correcting reconstruction errors, our musicologists and engraving specialists also enforce a consistent style across our library.

In order to improve Verovio with new features that we need, we invest heavily in its development. Basically, all our changes and improvements are sent directly back upstream, so we get a review from the maintainer, and the community immediately benefits from our work. Also, all our development efforts on Verovio are visible in our public fork on Github. So even if changes are not merged back into Verovio's main repository, you can still get those features from our fork.

Figure 5: Example for visual improvements in Verovio. Before (left) the tempo marking was positioned poorly, while now (right) it includes preceding accidentals as proposed in *Behind Bars* [7].

Some of the visual improvements we've made so far include

- automatic cross staff rest positioning,
- rendering of ties on dotted notes,
- alignment of dots in different voices,
- horizontal layout with dots and flags,
- alignment of tempo markings (see Figure 5),
- rendering of octave bracket endings.

We introduced several new options and recently we added support for additive meter signatures (see Figure 6).

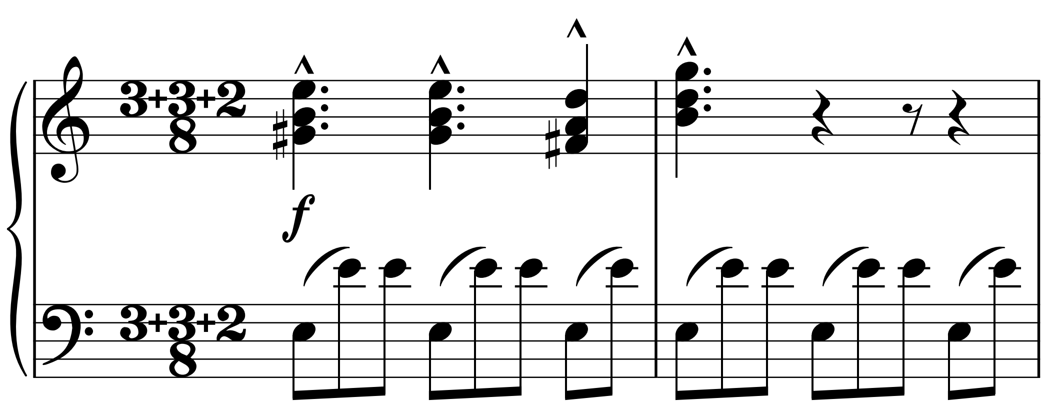


Figure 6: The first two measures *in bulgarian rhythm* from the very last piece from Béla Bartók's *Mikrokozmosz* showing an additive time signature in Verovio's rendering.

One major problem remains: Verovio tries to follow the rules and conventions of music notation, as recently collected and presented by Elaine Gould [7] and others before her. In printed scores, however, you find these rules very often extremely bent or even broken (see Figure 7). But even if you want to address all possible edge cases of notation, Verovio can only produce layouts based on a determined (and finite) set of rules. And trying to reproduce bad or nonsensical layout would contradict our goal of a unified user experience mentioned above.

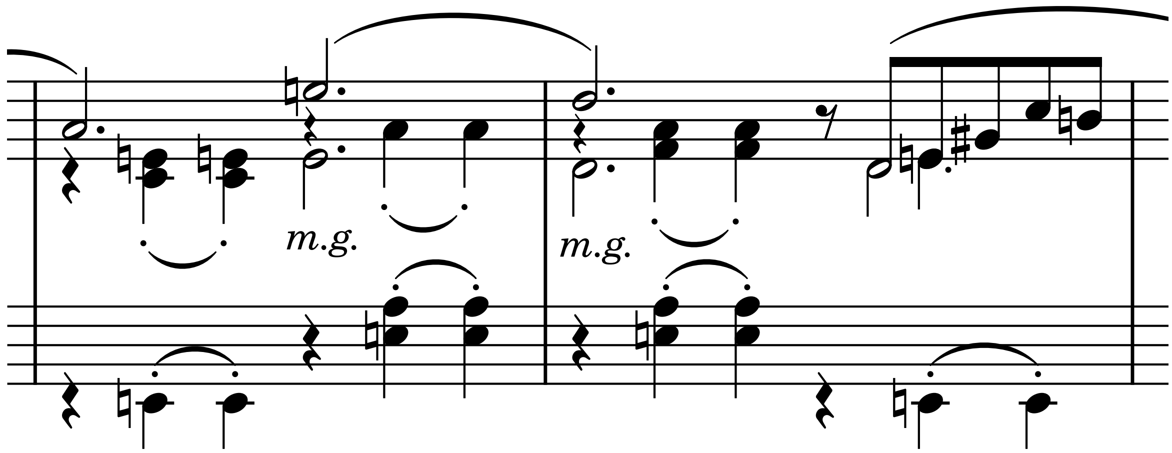


Figure 7: In this excerpt of Chopin's Ballade No. 1 in G minor, Op. 23, the editor decided to squeeze the quarter rests in the upper system between the noteheads by making it smaller than other rests, although there would be other (more consistent) ways to notate this. Also, it is not totally clear why rests are shifted vertically in the lower staff.

It is yet unclear how to deal with such cases, but you certainly have to deal with them. Interestingly, editors tend to stick with erroneous layouts when they prepare new editions of scores, even if there would be solutions that better comply with the standard rules of notation and are more pleasant to the musician's eye. Unfortunately, to clear up problematic situations, appropriately handle dense scores, or automatically do a complete re-layout, a computer would need to have a deep understanding of the music – maybe even deeper than their original editors.

Conclusion

Digital symbolic sheet music offers several advantages compared to paper-based scores. However, to offer musicians a true alternative, more than just a collection of PDF documents is needed. Our efforts are directed towards building a comprehensive sheet music library application, made (not only) for musicians. We are collecting and digitizing thousands of music scores, enriching them with vital metadata, and offer our users a digital solution that unlocks the full potential of digital sheet music. In building a comprehensive sheet music library application, we see a viable business model, but equally importantly a step towards the preservation of our cultural heritage. We also see an opportunity to break with the habit of constantly copying poor or faulty notation. Our artificial intelligence plays a key role in our efforts by automatically reporting situations that require additional manual work so we can bring the music of our ancestors to musicians worldwide.

Works Cited

[1] Baró, Arnau, Pau Riba, Jorge Calvo-Zaragoza, and Alicia Fornés. “From Optical Music Recognition to Handwritten Music Recognition: A Baseline” *Pattern Recognition Letters* 123 (2019), 1–8, https://doi.org/10.1016/j.patrec.2019.02.029.

[2] Calvo-Zaragoza, Jorge, Jan Hajič Jr., and Alexander Pacha. “Understanding Optical Music Recognition” *ACM Computing Surveys* 53, no. 4 (2020), 1–35, https://doi.org/10.1145/3397499.

[3] Chen, Liang, and Christopher Raphael. “Optical Music Recognition and Human-in-the-loop Computation” in *Proceedings of the 1st International Workshop on Reading Music Systems* (WoRMS 2018), 11–12, https://sites.google.com/view/worms2018/proceedings.

[4] Enote GmbH. ”Legato Music Font”. https://github.com/eNote-GmbH/Legato (accessed January 12, 2022).

[5] Good, Michael. “MusicXML: An Internet-Friendly Format for Sheet Music”, 2001, http://citeseerx.ist.psu.edu/viewdoc/summary?-doi=10.1.1.118.5431.

[6] Gotham, Mark, Peter Jonas, Bruno Bower, William Bosworth, Daniel Rootham, and Leigh VanHandel. “Scores of Scores: An OpenScore Project to Encode and Share Sheet Music” in *Proceedings of the 5th International Conference on Digital Libraries for Musicology* (DLfM 2018), 87–95, https://doi.org/10.1145/3273024.3273026.

[7] Gould, Elaine. *Behind Bars: The Definitive Guide To Music Notation.* London: Faber Music, 2011.

[8] Hajič Jr., Jan, and Pavel Pecina. “The MUSCIMA++ Dataset for Handwritten Optical Music Recognition” in *Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition* (ICDAR 2017), 39–46, https://doi.org/10.1109/ICDAR.2017.16.

[9] Hankinson, Andrew, Perry Roland, and Ichiro Fujinaga. “The Music Encoding Initiative as a Document-Encoding Framework” in *Proceedings of the 12th International Society for Music Information Retrieval Conference* (ISMIR 2011), 293–298, http://ismir2011.ismir.net/papers/OS3-1.pdf.

[10] Ornes, Rafael. “Choral Public Domain Library”. https://www.cpdl.org/ (accessed January 12, 2022).

[11] Pacha, Alexander, Kwon-Young Choi, Bertrand Coüasnon, Yann Ricquebourg, Richard Zanibbi, and Horst Eidenberger. “Handwritten Music Object Detection: Open Issues and Baseline Results” in *Proceedings of the 13th IAPR International Workshop on Document Analysis Systems* (DAS 2018), 163–168, https://doi.org/10.1109/DAS.2018.51.

[12] Pacha, Alexander, and Jan Hajič Jr. “The Music Notation Graph (MuNG) Repository”. https://github.com/OMR-Research/mung (accessed January 12, 2022).

[13] Project Petrucci LLC. *International Music Score Library Project*. 2006, https://imslp.org (accessed January 12, 2022).

[14] Pruslin, Dennis Howard. “Automatic Recognition of Sheet Music” Massachusetts Institute of Technology, Cambridge (Mass.), 1966. PhD dissertation.

[15] Pugin, Laurent, Rodolfo Zitellini, and Perry Roland. “Verovio: A Library for Engraving MEI Music Notation Into SVG” in *Proceedings of the 15th International Society for Music Information Retrieval Conference* (ISMIR 2014), 107–112, https://archives.ismir.net/ismir2014/paper/000221.pdf.

[16] Rebelo, Ana, Ichiro Fujinaga, Filipe Paszkiewicz, Andre R. S. Marcal, Carlos Guedes, and Jaime S. Cardoso. “Optical Music Recognition: State-of-the-art and Open Issues” *International Journal of Multimedia Information Retrieval* 1 (2012), 173–190, https://doi.org/10.1007/s13735-012-0004-6.

[17] Sapp, Craig Stuart. “Verovio Humdrum Viewer”. https://github.com/humdrum-tools/verovio-humdrum-viewer (accessed January 12, 2022).

[18] Smith, Leland. “SCORE – A Musician’s Approach to Computer Music” in *Journal of the Audio Engineering Society* 20, no. 1 (1972), 7–14.

[19] Tuggener, Lukas, Ismail Elezi, Jürgen Schmidhuber, and Thilo Stadelmann. “Deep Watershed Detector for Music Object Recognition” in *Proceedings of the 19th International Society for Music Information Retrieval Conference* (ISMIR 2018), 271–278, http://ismir2018.ircam.fr/doc/pdfs/225_Paper.pdf.

[20] Vigliensoni, Gabriel, Alex Daigle, Eric Liu, Jorge Calvo-Zaragoza, Juliette Regimbal, Minh Anh Nguyen, Noah Baxter, Zoé McLennan, and Ichiro Fujinaga. “From Image to Encoding: Full Optical Music Recognition of Medieval and Renaissance Music” presented at the *Music Encoding Conference* (MEC 2019), University of Vienna, Austria, May 29–June 1, 2019, https://music-encoding.org/conference/abstracts/abstracts_mec2019/vigliensoni19from%20camera%20ready.pdf.

[21] Wick, Christoph, and Frank Puppe. “Automatic Neume Transcription of Medieval Music Manuscripts Using CNN/LSTM-Networks and the Segmentation-Free CTC-Algorithm” *Preprints* 2020010149 (2020), https://doi.org/10.20944/preprints202001.0149.v1.