



Interacting with Annotated and Synchronized Music Corpora on the Dezrann Web Platform

RESEARCH ARTICLE

Ju[ubiquity press

CHARLES BALLESTER

BAPTISTE BACOT

LOUIS BIGO

VANESSA NINA BORSAN

LOUIS COUTURIER

KEN DÉGUERNE

QUENTIN DINEL

LAURENT FEISTHAUER

KLAUS FRIELER

MARK GOTHAM

RICHARD GROULT

JOHANNES HENTSCHEL

ALEXANDRE D'HOOGHE

DINH-VIET-TOAN LE

FLORENCE LEVÉ

FRANCESCO MACCARINI

IVANA MARIČIĆ

GIANLUCA MICCHI

MEINARD MÜLLER

ALEXANDROS STAMATIADIS

TOM TAFFIN

PATRICE THIBAUD

CHRISTOF WEIß

RUI YANG

EMMANUEL LEGUY

MATHIEU GIRAUD

*Author affiliations can be found in the back matter of this article

CORRESPONDING AUTHOR:

Mathieu Giraud

Univ. Lille, CNRS, Centrale Lille,
UMR 9189 CRISTAL, F-59000
Lille, France

mathieu.giraud@univ-lille.fr

KEYWORDS

music corpora, meta-corpus,
music annotation,
reproducibility, music
representations, web platform,
cross-modal synchronization,
score, audio

TO CITE THIS ARTICLE:

Ballester, C., Bacot, B., Bigo, L.,
Borsan, V.N., Couturier, L.,
Déguernel, K., Dinel, Q.,
Feisthauer, L., Frieler, K.,
Gotham, M., Groult, R.,
Hentschel, J., d'Hooge, A., Le,
D-V-T, Levé, F., Maccarini, F.,
Maričić, I., Micchi, G., Müller, M.,
Stamatiadis, A., Taffin, T.,
Thibaud, P., Weiß, C., Yang, R.,
Leguy, E., & Giraud, M. (2025).
Interacting with Annotated and
Synchronized Music Corpora on
the Dezrann Web Platform.
*Transactions of the International
Society for Music Information
Retrieval*, 8(1), 121–139.
DOI: <https://doi.org/10.5334/tismir.212>

ABSTRACT

Open datasets with annotated corpora are crucial to foster research in music information retrieval (MIR) studies and to disseminate knowledge towards musicians and the general public. Many groups have published corpora, at times accompanied by specific visualization interfaces. However, achieving a cohesive access to such a diverse range of data poses a significant challenge. Dezrann is an open-source web platform to interact with corpora while *sharing music and music analysis* in the form of scores, images, audio files (waveforms), video files, and annotations. We present how we render through this platform ten curated corpora published in the last years in the MIR community, gathering 1500+ pieces and 35000+ annotations. These corpora include works by Bach, Mozart, Beethoven, and Schubert, lieder from the 19th century by female composers (OpenScore Lieder), jazz solo transcriptions (Weimar Jazz Database), piano rolls (SUPRA), Georgian sacred songs (Erkomaishvili dataset), and Slovenian folk song ballads. Showing these corpora with the cross-modal synchronization enabled by the platform improves the way to hear, study, and annotate music. This opens up new possibilities for corpus annotation and analysis in musicology and computer music research, enhances music education, fosters the promotion of music diversity, and facilitates collaborative score editing and correction.

1 INTRODUCTION

1.1 MUSIC ANALYSIS AND ANNOTATION

Music analysis often includes music annotation, which involves adding various elements, such as text, symbols, graphical markings (including colors, highlights, lines, and arrows), to a score or other music representation. In this study, we refer to these elements as *labels*. They may indicate patterns, harmonies, similarities, texture, structure, form, or other musical concepts, serving as a means to document, analyze, and interpret musical content. Some consider that annotating music aligns with what Nattiez (1975) describes as a ‘neutral’ perspective on the score, potentially free from historical, comparative, or aesthetic viewpoints. But it can also be linked to insights on the cultural and historical context in which the music was created, improvised, composed, and/or performed. Music annotation serves a wide range of purposes, including teaching, performing (Winget, 2008), composing, researching, and even facilitating a deeper personal connection to and enjoyment of music.

1.2 MIR CORPUS ANALYSIS

Music information retrieval (MIR) researchers often annotate music corpora for studies in systematic musicology, computational music analysis, and/or music generation. These annotations can be used to *develop and evaluate computational models*. In combination with the raw (audio, video, symbolic) music data, these annotations serve as ‘ground truth’ or reference datasets that are crucial for developing and evaluating computational models. For instance, corpora with harmonic annotations may be used to train and test models for harmony prediction or to help with other tasks, such as music generation based on harmonic constraints (de Berardinis et al., 2023; Gotham et al., 2023b; Le et al., 2025; Wang et al., 2024).

However, the very *process* of annotating and releasing a corpus requires substantial effort and is of fundamental importance to research. Annotating a music corpus requires researchers to make deliberate decisions on how to model and encode annotations in a way that serves their research goals in musicology, MIR, and/or computer science (Gotham and Ireland, 2019; Panteli et al., 2018; Peeters and Fort, 2012; Savage, 2022). This process leads to new insights and questions about the music itself and involves trade-offs between simplicity, accuracy, and expressiveness. For the sake of reproducibility, any annotation work should detail its methodology, individual decisions, and trade-offs, including how it handles ambiguous cases.

For example, working on harmony annotation requires selecting an encoding/taxonomy for chord labels, such as Roman numerals or pitch-class sets. As researchers engage in annotating the corpus, they may encounter challenges, such as a passage that is harmonically ambiguous in ways they had not considered before. These

puzzles raise research questions about the structure and meaning of the music, inspiring the development of new modeling approaches or analytical techniques.

1.3 MODELING ANNOTATIONS AND MUSIC KNOWLEDGE

The approach to modeling music annotation is itself an active area of research. Lewis et al. (2022) proposed a generic framework for annotating music content, organized into three layers arising from musicological needs. The evidence objects layer refers to resources. The musical objects layer serves as an intermediary between resources and musicological work, particularly by modeling how resources are divided or grouped. Finally, the musicological object layer allows researchers to document their work using Web Annotations for recording observations, relationships, and commentaries. Additionally, they explore various methodologies for annotating music, with a specific focus on some of the datasets presented here, including the Schubert Winterreise Dataset (Section 3). In the context of the European funded project TROMPA,¹ Weigl et al. (2021a) focused on interconnecting existing public-domain digital music resources through a knowledge graph and Weigl et al. (2021b) on a model for exchanging annotations using Solid Pods. Other significant efforts towards unifying musical data and knowledge are found in the LinkedMusic Project,² in particular, to create a flexible metadata schema, including the HaMSE ontology (de Berardinis et al., 2022; Lewis et al., 2022; Poltronieri and Gangemi, 2022), and in the Music Meta Ontology and ChoCo meta-corpus (de Berardinis et al., 2023).

1.4 CURATED MIR RESEARCH CORPORA

There have been a number of music corpora published by MIR researchers in recent years. The MIR-datasets repository³ lists more than 100 such datasets, many of which offer multimodal data. They are valuable data for various MIR tasks and, for some of them, also for musicology and ethnomusicology, as reviewed by Panteli et al. (2018) and Savage (2022).

Peeters and Fort (2012) emphasize the importance of clearly defining curated research corpora, particularly through detailed documentation of their methodology. Similarly, Serra (2014) identifies five essential criteria for curated research corpora: purpose, coverage, completeness, quality, and reusability. Reproducibility, which is critical for advancing research, relies on the availability of data and code under open licenses. It also requires effective software development practices, including testing, documentation, packaging, and human organization (McFee et al., 2019).

1.5 VISUALIZING, HEARING, AND INTERACTING WITH CORPORA

Music annotation and analysis have traditionally been carried out using paper scores and pencils, a method

that remains both practical and highly efficient. However, digital tools offer enhanced possibilities, enabling *cross-modal interactions* between scores, audio files, images, annotations, and potentially other types of data.

Synchronizing such different sources of music representation and encoding is a challenge for digital platforms (Garczynski et al., 2022; Thomas et al., 2012). The fields of audio-score alignment and score following are actively researched, both on models and on algorithms (Berndt, 2021; Cancino-Chacón et al., 2023; Müller et al., 2019). Common alignment methods include dynamic time warping algorithms (Ewert et al., 2009; Müller et al., 2021; Thickstun et al., 2020) but also machine learning approaches (Dorfer et al., 2018). These strategies often rely on chroma-based features, which capture tonal content by representing pitch class distributions over time. Chroma features, extracted from either audio signals or symbolic representations, are a versatile tool for aligning and analyzing musical data. Multi-track information allows players to switch between tracks (Zalkow et al., 2018).

It is common for researchers building music corpora to develop *visualization and playing interfaces* in order to allow users to interact with and explore data. Examples of these projects include JazzTube (Balke et al., 2018), the web players coming with the SUPRA (Shi et al., 2019) and the Erkomaishvili (Rosenzweig et al., 2020) datasets, and the E-Laute project for lute music from the German-speaking area between 1450 and 1550 (Schöning et al., 2025).

However, these interfaces are often ‘one-shot’ designs, signifying that they are *specific to a particular corpus*. On the one hand, this approach allows for the development of features tailored to that specific corpus. On the other hand, it can be a limitation, as it may be challenging to adapt the interface to work seamlessly with other corpora. Moreover, maintaining the interface over time becomes a potential challenge as the corpus evolves or as the project, including its funding, is completed.

Is it possible to interact in a unified way with a variety of MIR corpora, potentially from multiple sources, such as score, audio, and analyses? To both enjoy these corpora and conveniently present examples and results from musicological or MIR research, a solution is to develop more *general-purpose platforms* designed for displaying music corpora. These platforms should be agnostic to the specific dataset being used.

Systems for digital humanities such as Telemeta⁴ or Dunya (Porter et al., 2013) allow collaborative (ethno)musicology studies, including features for collaborative annotation. Sonic Visualiser⁵ or Audacity (Mazzone and Dannenberg, 2002) provide the capability to place *markers*, which can represent beats, measures, or significant points—and allow further editing of these markers. The *trackswitch.js* web audio player enables

switching between multiple audio tracks while displaying waveforms or other representations such as spectrograms (Werner et al., 2017). The CosmoNote web platform, demonstrated in piano performances, enables interaction with markers on audio data (Fyfe et al., 2022). It visualizes piano rolls and waveforms, but also continuous values like pedals, tempo, or other computed data. The *ianalyse*⁶ software allows users to interact with and annotate electroacoustic music pieces. Recently, Weigl et al. (2023) proposed a web application to compare different recordings and annotate the score extending the Music Annotation Ontology (Lewis et al., 2022).

Despite their strengths, some of these tools lack a notion of (symbolic) musical time or do not give access to diverse music collections. The Dezzrann project aims to provide such a cross-modal web platform for easy music annotation and analysis for a wide range of musicians, analysts, and researchers, including those with limited programming skills and those working with large music corpora.

1.6 CONTENTS AND CONTRIBUTIONS

This work details how we have made ten curated corpora available in a reproducible manner through the Dezzrann platform, encompassing over 1500+ pieces and 35000+ analytical labels (Table 1). The platform presents metadata for corpora and pieces, and shows music in several views, including scores (Lilypond/Verovio) and waveforms/spectrograms, and allows users to interact with labels on any of these views (Section 2). While we do not introduce new annotation data here—these were published in recent years—we have synchronized audio files for certain corpora. Additionally, we have consolidated and converted annotation and synchronization data into a unified format for the platform, enhancing their rendering, reproducibility, and overall usability. The platform itself was refined and adapted throughout this process to support these improvements (Section 3). Such a platform, along with its curated data collection, has the potential to support both MIR research, music pedagogy, and public engagement. While discussing associated challenges and future perspectives, we discuss issues related to corpus selection, diversity in music research, open data, reproducibility, and sustainability (Section 4).

2 BUILDING A WEB PLATFORM FOR INTERACTING WITH ANNOTATED CORPORA

In this section, we highlight key aspects of the Dezzrann platform, focusing on synchronization (Section 2.1) and the management of piece and corpus metadata (Section 2.2).

Corpus, references, licenses	Score annotations	Synchronized audio
Bach Fugues 24 fugues, 1722 ODbL (annotations), CC-BY-3.0 (audio), YT (video) (Giraud et al., 2015)	450 labels fugue structure, cadences, pedals	Two* complete audio/video recordings* K. Ishizaka (Open Well-Temp. Clavier) The Netherlands Bach Soc. (All of Bach)
Mozart Piano Sonatas 54 movements from 18 sonatas, 1774–1789 CC-BY-NC-SA-4.0 and ODbL (scores, annotations), YT (audio) (Couturier et al., 2022; Hentschel et al., 2021)	17800+ labels keys, harmony, cadences, textures	Complete audio recordings* K. Würtz (2006)
Mozart String Quartets 72 movements in 23 quartets, 1770–1790 ? (scores), ODbL (annotations), CC-BY-NC-ND-3.0 (audio) (Allegraud et al., 2019)	2200+ labels sonata form structure, keys, cadences	Recordings* for 8 mvts from 4 quartets Borromeo String Quartet (2009)
First Movements of Classical Symphonies 24 first movements, 1779–1824 Various licences (scores), ODbL (annotations), Public Domain (audio) (Le et al., 2022)	6000+ labels sonata form structure, texture analysis	Recordings* for 6 mvts (Haydn) The Royal Phil. Orchestra (1960)
19th-Century Lieder, Female Composers 170 lieder (OpenScore Lieder), 1780–1920 CC0-1.0 (scores, annotations), YT (audio) (Gotham and Jonas, 2021; Gotham et al., 2023b)	4900+ labels tonality, harmony, phrases on 53 pieces	Recordings* for 25 lieder (Fanny Mendelssohn) L. Kolb, A. Shrut (1992)
Schubert Winterreise 24 lieder, 1827–1828 CC-BY-3.0 (scores, annotations), PDM-1.0, CC-BY-NC- ND-3.0 (audio) (Weiß et al., 2021)	2400+ labels structure, keys, harmony	Two* complete recordings G. Hüsch, H.-U. Müller (1933) R. Scarlata, J. Denk (2006)
SUPRA Piano Roll 456 piano rolls, 1905–1928 CC-BY-NC-SA-4.0 (Shi et al., 2019)	–	Rendered expressive audio
Slovenian Folk Song Ballads 404 transcriptions of ballads, 1819–1995 CC-BY-NC-SA-4.0 (Borsan et al., 2025)	2000+ labels contour, structure, harmony	23 historic recordings*
Weimar Jazz Database 333 transcriptions of jazz solos, 1925–2009 ODbL (scores, annotations), YT (audio) (Pfleiderer et al., 2017)	1200+ labels form, chords, phrases, midlevel units	228 historic recordings (1925–2009)
Traditional Georgian Sacred Music 101 3-voice songs with transcriptions, 1966 CC-BY-NC-SA-4.0 (Rosenzweig et al., 2020)	–	101 historic recordings A. Erkomaishvili (1966) 4 files* per song, mix and sep. voices

Table 1 The first ten open corpora available through the Dezrann platform. Except SUPRA, all corpora have scores, and most of them have some annotations and synchronized audio. For some corpora, the platform only includes a part of the original corpus. In particular, the ‘annotations’ column displays the number of analytical labels visible through Dezrann—usually labels on scores describing patterns, harmony, and/or structure. Datasets with audio components often include additional annotations, such as note-wise onsets (here incorporated into the synchronization) or even frame-wise annotations. New synchronizations were done through the platform (*). All the other data were published elsewhere and were here adapted for the platform. Some corpora have pieces with multiple audio (*). The <https://doc.dezrann.net/status> web page shows the live status of each corpus as well as links to download data. Further documentation to rebuild the corpora is available from <https://doc.dezrann.net/rebuild>.

2.1 SYNCHRONIZING SCORES, AUDIO, AND ANNOTATIONS

On the platform, users can interact with *synchronized* scores (or other visual representations), audio, and labels, all aligned across multiple views (Figure 2). *Musical time*, measured in quarter notes from the beginning of a piece, serves as the central reference for synchronization (Figure 1). Users are generally unaware of this reference, as they directly select points on the score. Moreover, while the selected corpora generally follow a straightforward musical timeline, corpora without such a reference can

instead use *audio time* (measured in seconds) as an alternative. Depending on the corpus, the available data may include:

- Scores, produced from krn/MusicXML/MEI files by music21 (Cuthbert and Ariza, 2010) and Verovio (Pugin et al., 2014) or Lilypond,⁷ or autographs. More generally, the platform can handle any other *graphical document*, as, for example, piano rolls (see, in the next section, the SUPRA dataset). Such sources need synchronization between the musical

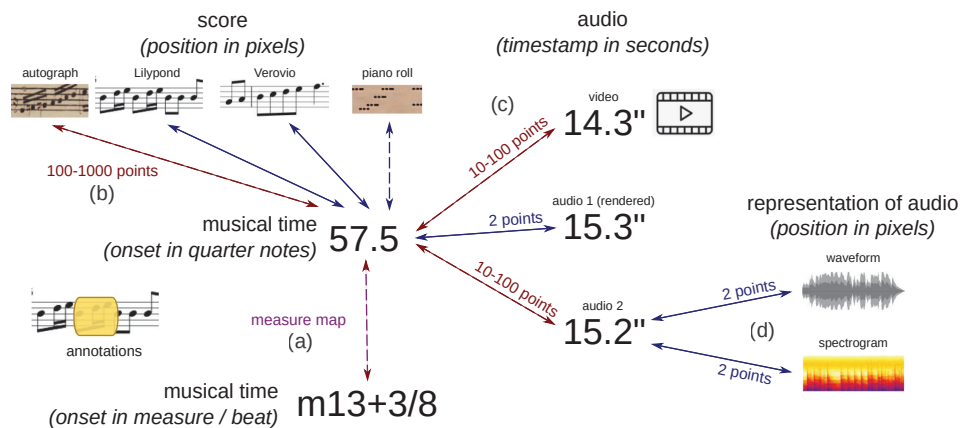


Figure 1 Overview of synchronization types within the Dezrann platform: (a) synchronization between musical time (in quarter notes and measures) (Gotham et al., 2023a), (b) image-to-musical-time alignment, (c) audio-to-musical-time alignment, and (d) image-to-audio-time alignment. Plain arrows indicate straightforward synchronizations with only two reference points (e.g., an audio file with a constant tempo) or those automatically handled by the rendering software. Dashed arrows represent synchronizations that may be derived from explicit data, inferred via Optical Music Recognition (OMR) or score/audio alignment, or manually adjusted using the synchronization editor.

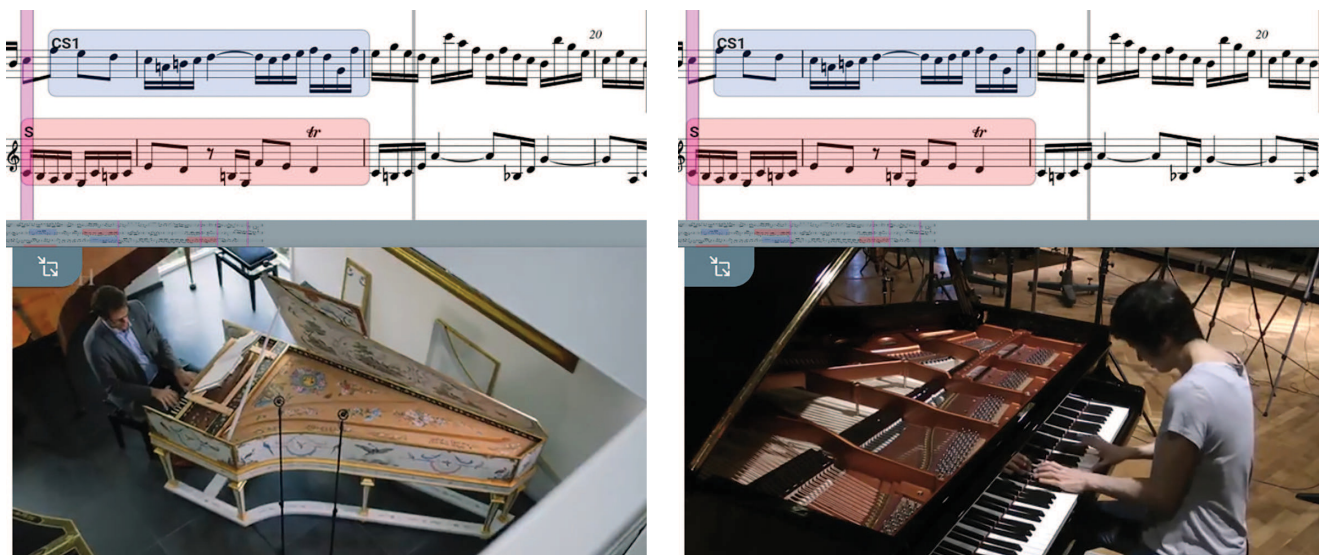


Figure 2 Fugue in Eb Major BWV 852 by J.-S. Bach, <https://www.dezrann.net/~bach-fugues/bwv852>. The score is synchronized with two performances by Pieter-Jan Belder (left, harpsichord, Netherlands Bach Society) and Kimiko Ishizaka (right, piano, <https://welitemperedclavier.org>) as well as annotation labels on fugue analysis (Giraud et al., 2015). As both recordings are synchronized to the musical time, the playback can be switched from one performance to another one.

time and, for a specific source layout, the graphical position (in pixels).

- *Audio/video* content (rendered through the HTML5 `<audio>` tag, or through a YouTube player), possibly with representations such as waveforms or spectrograms produced by ffmpeg⁸ and librosa (McFee et al., 2015). Such sources need synchronization between the musical time and the audio time (Figure 2).
- *Analytical labels*, based on musical time, describing concepts such as harmony, patterns, texture, or structure. The platform is agnostic to the specific ontology of such labels, accommodating the varying hermeneutics in which different studies model music (Figure 3).

Although we prefer to host corpora having scores, audio/video, and annotations, some of the corpora only have two of the three.

Going back to the three layers proposed by Lewis et al. (2022), the labels in Dezrann function as *observations* within the musicological objects layer, linked to *resources* from the evidence objects layer (such as scores and other sources, with metadata for *references*). The ontology of labels themselves aligns with the musical objects layer but remains largely dependent on the methodology and structure of each corpus. The key difference in Dezrann is that labels are based on musical time, meaning they are anchored to specific beats or measures rather than purely structural or ontological relationships. While this approach may not cover all possible cases, it provides a



Figure 3 Four musical pieces analyzed with different focuses and ontology conventions: (a) Symphony No. 101 'The Clock' by J. Haydn, annotated with texture labels (Le et al., 2022); (b) 'Gondellied' by Fanny Mendelssohn, featuring harmonic analysis (Gotham and Jonas, 2021); (c) 'Die Wetterfahne' by Franz Schubert, analyzed harmonically (Weiß et al., 2021); and (d) 'I Fall in Love Too Easily', a transcription of Chet Baker's trumpet solo from the Weimar Jazz Database (Pfleiderer et al., 2017).

simple and practical way to synchronize annotations with various musical sources.

The front-end interface, as detailed in Garczynski et al. (2022), is built using the Vue.js framework⁹ with *renderless behavior slots*. Such slots provide a flexible

way to inject custom rendering logic into web components, making it easier to adapt the user interface to different use cases. By default, all views remain synchronized, supporting both playback and analysis applications. However, an *edit mode* allows users to modify synchronization data on one source while keeping the others unchanged.

A key design decision was to handle synchronization for all source images using *pixel coordinates*, enabling fast user interaction and ensuring consistent alignment across different visual representations. For rendered scores, these coordinates are derived using Verovio's SVG output, generated with carefully controlled layout parameters, by retrieving the bounding boxes of relevant score elements via their `xml:id`. For other sources, the coordinates are obtained through various (semi-) automated or even manual methods.

2.2 ENCODING AND EXPLORING METADATA AND QUALITY

Ensuring well-structured, accurate, and comprehensive metadata is essential for researchers who may use it in their own studies. To enhance metadata consistency and usability, we extended and refined the metadata of the ten corpora in two key areas:

- *Metadata for individual pieces*: While published corpora often contain structured metadata, integrating them into a unified platform requires additional standardization, particularly in titles and collections.
- *Metadata for each corpus*: This includes historical and musicological context, as well as assessments of data quality and availability. Many MIR corpora are not exhaustive and may lack certain pieces, making transparency about their completeness essential.

The metadata structure varies by corpus and is fully accessible through the platform. Additionally, the platform's corpus and developer documentation (<https://doc.dezrann.net>) provide detailed metadata guidelines and integration instructions. Advice is provided to ensure references to existing publications and databases whenever possible, using identifiers such as RISM numbers or MusicBrainz IDs to precisely identify musical sources, enhancing traceability.

2.2.1 Licenses

Metadata includes licensing information for individual sources, encoded using SPDX¹⁰ strings. Since a score, audio file, and annotation for the same piece may have different licenses, this distinction is explicitly recorded. While open data are preferred, the platform also supports YouTube integration, allowing synchronization with non-free audio/video content.

2.2.2 Quality assessment

Various parameters define the quality of both individual pieces and entire corpora (Figure 4). Quality is evaluated across multiple aspects, including score accuracy, audio fidelity, synchronization precision, annotation accuracy, curation level, and source reliability. These assessments help users understand the strengths and limitations of each dataset.

Ideally, all corpora would maintain high quality across all aspects. However, in practice, quality can vary, and its evaluation is often subjective, depending on the research goals or use case. In some contexts, sub-optimal elements may still be valuable, for example, a readable score lacking dynamics, a bar-level synchronization with audio, a historically significant but noisy recording, or draft or automated annotations that have not yet been curated.

quality:corpus

- | | |
|---|---|
| 0 | individual pieces, no real collection, no corpus.json |
| 1 | work has started, pieces just landed in Dezzrann, first corpus.json |
| 2 | work in progress, should be a nice corpus soon, however for many pieces there are still problems |
| 3 | maintainer identified, at least two “quality sources” (scores, audio/video with good synchronization, images, annotation set). with quality ≥ 3 for most pieces |
| 4 | maintainer identified, at least three “quality sources” for most pieces, published/referenced corpus in an article/database/git, documented steps to build/publish the corpus on Dezzrann |
| 5 | maintainer identified, at least three “quality sources” for most pieces, published/referenced corpus in an article/database/git, fully reproducible steps to build/publish the corpus on Dezzrann starting from a long-term archive |

quality:score

- | | |
|---|---|
| 0 | no score (or score not working through the Dezzrann pipeline) |
| 1 | draft score, for example automatically produced from MIDI, many bugs, not coherent at the end |
| 3 | useable score (even if not beautiful or some strange things in some measures) |
| 4 | near production-quality score (dynamics, typesetting, editing...) |
| 5 | production-quality score, linked to some published critical edition |

quality:musical-time

- | | |
|---|--|
| 0 | no musical time (only audio) |
| 1 | some musical time, but not coherent everywhere, changes of meter or other things are not well described |
| 3 | useable musical time / measure numbers |
| 3 | useable musical time / measure numbers, with a measure map |
| 5 | perfect musical time / measure numbers, with a measure map, linked to some published edition, published in a git |

quality:audio

- | | |
|---|--|
| 0 | no audio |
| 1 | (horrible) generated audio |
| 3 | generated audio with better quality (instrument, dynamics, expression...) or low-value recorded audio |
| 4 | mid-/high-value recorded audio |
| 5 | high-value recorded audio, “professional quality” and/or musico-logical significance, with open licenses |

quality:audio:synchro

- | | |
|---|--|
| 0 | no audio |
| 1 | no synchronization (apart from start/end) |
| 2 | some synchronization, but sometimes very far away |
| 3 | useable synchronization, including tempo changes (at the measure level) |
| 4 | good synchronization (including repeats, at the measure level) |
| 5 | good synchronisation (including repeats, at the beat level), also published in a git |

Figure 4 Some quality criteria applied to corpora (quality:corpus, top block) and individual pieces (bottom blocks). The complete list of quality criteria is available at <https://doc.dezzrann.net/quality>. Additionally, the corpora status page, at <https://doc.dezzrann.net/status>, provides a summary of quality values for each corpus.

Additionally, within a single corpus, quality may be inconsistent across pieces, often due to encoding challenges or annotation complexities.

Rather than striving for uniform quality across all pieces and corpora—an unrealistic goal—we focus on accurately *describing their quality*, while acknowledging that such assessments remain somewhat subjective. This transparency allows researchers to understand corpus limitations and make informed decisions about its suitability for their work.

Quality information is also integrated into the platform’s functionality. For instance, the playback cursor adapts based on the quality of score/audio synchronization, offering measure-level highlighting for lower precision synchronizations and more detailed tracking when higher precision data are available.

2.2.3 Exploring corpora and pieces

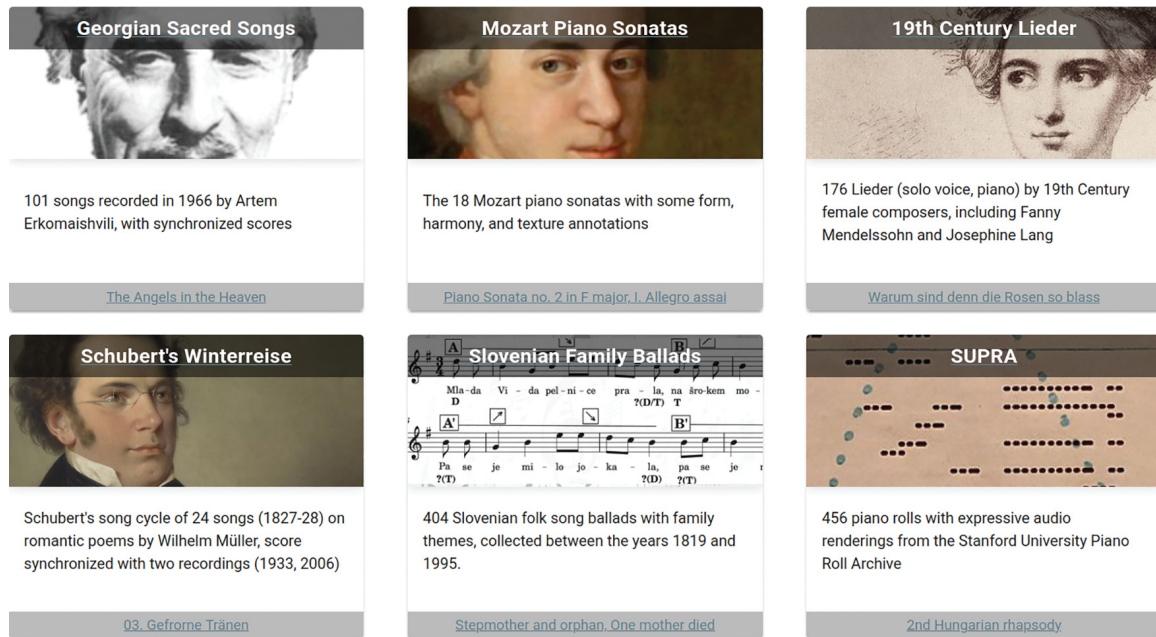
The platform provides an overview of available corpora and individual pieces, incorporating relevant metadata to highlight data availability (Figure 5 and Appendix). The interface and some metadata, including corpus descriptions, are available in simplified Chinese, Croatian, German, English, French, Greek, Italian, and Slovenian, making the platform widely accessible.

3 INCLUDING ANNOTATED AND SYNCHRONIZED CORPORA

The Dezzrann platform currently hosts ten curated open corpora (Table 1). The policy for corpus selection and integration is detailed in Section 4.2. Each corpus is here accompanied by a historical, musical, and scholarly introduction, available both in the appendix and on the platform. These corpora originate from seven different research teams/labs worldwide, each employing distinct approaches to modeling and encoding annotations and synchronization. Building on examples from these corpora, we highlight key challenges related to adapting and transforming the original data, integrating annotations and synchronizations into the platform, and addressing data availability and reproducibility.

3.1 GATHERING SCORES AND ANNOTATIONS WEIMAR JAZZ DATABASE (WJD) MOZART PIANO SONATAS

In most corpora, scores were readily available, but a common challenge was selecting the highest-quality scores that rendered correctly in Verovio, sometimes requiring updates. When such updates are needed, the final scores are available in our repositories. Note that updates to upstream sources, or improvements or changes to Verovio, pose challenges for long-term reproducibility. Additionally, integrating existing annotations



Solo pjesme skladateljica 19. stoljeća

Skladbe za glas i klavirsku pratnju napisane u razdoblju glazbenog romantizma imenujemo solo pjesme. U vrijeme kada su žene bile isključene iz brojnih područja javnog života te im nije priličilo bilo kakvo umjetničko stvaralaštvo, je specifična namjena solo pjesama za intimna okruženja dopuštala da se u njoj okušaju i pripadnice drugog spola. K obliku solo pjesme su značajno doprinesle [Fanny Mendelssohn](#), [Clara Schumann](#), [Josephine Lang](#), [Louise Reichardt](#), [Augusta Holmè](#) i druge.

Korpus The OpenScore Lieder obuhvaća preko 1300 solo pjesama dugog 19. stoljeća. Zbirka je dostupna za online reprodukciju na musescore.com, a moguće je i [preuzimanje](#). Za više informacija o zbirci vidi: [\(Gotham i Jonas 2021\)](#) ili [Gothamov članak](#). Ova Dezrannova zbirka je podskup partitura različitih skladateljica, a uključuje i harmonijske analize objavljene na ["When in Rome" meta-korpusu](#), a predstavljene u: [Gotham i suradnici, 2023a](#). Korpus koristi metodu mjerenja razmaka između taktova („measure maps“) [\(Gothama i njegovih suradnika, 2023b\)](#) za poboljšanje interoperativnosti anotacija.

ref	Traži							🔍
Collection	Title	Composer, Artist	Opus	Year	Key	Meter	Sources, Analyses	
6 Lieder, Op.1	🕒 Gondellied	Fanny Mendelssohn	Op.1	1805-1847	A dur	6/8	🎹🎻🎵 imsip mscz wiki	
6 Lieder, Op.1	🕒 Mayenlied	Fanny Mendelssohn, Joseph von Eichendorff	Op.1	1805-1847	G dur	3/4	🎹🎻🎵 imsip mscz wiki	
6 Lieder, Op.1	🕒 Morgenständchen	Fanny Mendelssohn, Joseph von Eichendorff	Op.1	1805-1847	E dur	4/4	🎹🎻🎵 imsip mscz wiki	
6 Lieder, Op.1	🕒 Schwanenlied	Fanny Mendelssohn, Heinrich Heine	Op.1	1805-1847	G mol	6/8	🎹🎻🎵 imsip mscz wiki	

Figure 5 (Top) Extract of the corpora list on <https://dezrann.net/corpora>, where each corpus is presented with a motto text and a link to a showcase piece. (Bottom) Detailed view of the ‘19th-Century Lieder’ from Female Composers corpus on <https://dezrann.net/explore/openscore-lieder>, displaying the corpus description (localized here in Croatian) and a list of included pieces.

often necessitated developing parsers or converters to ensure compatibility with the platform.

Some corpora required special attention. The WJD provides scores and annotations in a database format (Pfleiderer et al., 2017), which was converted using the *melospy* tool provided by the WJD. For the Mozart Piano Sonatas, two separate research groups created annotations in different formats (Couturier et al., 2022; Hentschel et al., 2021). These analysis files were merged, and the platform now allows users to switch between the different analyses seamlessly while displaying the source of the annotations.

3.2 HAVING RELIABLE REFERENCES IN SCORES MOZART PIANO SONATAS 19TH-CENTURY LIEDER FROM FEMALE COMPOSERS TRADITIONAL GEORGIAN SACRED MUSIC

Accurate music analysis requires the ability to unambiguously reference specific points in a score. In two corpora, we addressed this by using measure maps (Gotham et al., 2023a) to encode consistent and reproducible measure numbers. For unmetered music, additional challenges arise, as seen in the Erkomaishvili dataset. To establish a precise reference system, we adopted the ‘quarter note

reference' (Rosenzweig et al., 2020), which counts quarter notes from the beginning of the piece. This approach provides an abstract yet reliable method for pinpointing locations throughout the score.

3.3 DISPLAYING LARGE SCORES

FIRST MOVEMENTS OF CLASSICAL SYMPHONIES

Orchestral music presents unique challenges due to its large scores, which in this corpus contain up to 20 staves (Figure 3a).

To optimize performance, we excluded the full .svg rendering in the score preview used for score navigation, as displaying the entire score at this resolution was not legible or efficient. Instead, the textural data provide sufficient structural information to facilitate navigation through the score. Looking ahead, we are working on enhancing the platform's efficiency in handling such complex and large-scale scores.

3.4 MODELING AND DISPLAYING METADATA AND CONTEXT

19TH-CENTURY LIEDER FROM FEMALE

COMPOSERS

SLOVENIAN FOLK SONG BALLADS

For every corpus, important questions arise regarding metadata and its presentation. Special attention is given to underrepresented corpora, as properly handling their metadata enhances their visibility and scholarly impact.

For instance, female composers have historically received less academic attention compared to their male counterparts, reinforcing the misconception that their works are rare. The OpenScore Lieder corpus (Gotham and Jonas, 2021) provides an important opportunity to highlight compositions by prominent 19th-century female composers (Figure 5). Similarly, the Slovenian Folk Song Ballads corpus (Borsan et al., 2025) represents an ethnomusicological collection, where metadata such as

regional origins and transcribers add valuable context. Additionally, English translations of titles and first verses were included to increase accessibility.

These two corpora illustrate key challenges in accurately describing metadata, particularly for songs within song cycles. The difficulty in locating some recordings highlights broader issues in the preservation and promotion of these musical works. Finally, as with all corpora, the descriptions (see appendix) provide the essential extramusical context, particularly the historical background. The availability of translations in eight languages further enhances visibility and accessibility.

3.5 SYNCHRONIZING OTHER GRAPHICAL DATA

SUPRA PIANO ROLLS

The 456 piano rolls from the Welte T-100 pneumatic player piano system, included in the SUPRA dataset (Shi et al., 2019), offer a fascinating example of image-based music representations distinct from traditional scores (Figure 8). The database contains synchronization data, aligning the horizontal positions of the rolls with their corresponding musical notes—a process that closely parallels the synchronization between scores and musical time.

3.6 SYNCHRONIZING NEW AUDIO/VIDEO RECORDINGS

BACH FUGUES

MOZART PIANO SONATA

MOZART STRING QUARTETS

FIRST MOVEMENTS OF CLASSICAL SYMPHONIES

19TH-CENTURY LIEDER FROM FEMALE

COMPOSERS

Some of the corpora were already published with synchronized audio files, while for the remaining five corpora (marked with * in Table 1), we selected and synchronized audio/video recordings using the platform (Figure 6), prioritizing open data whenever possible.

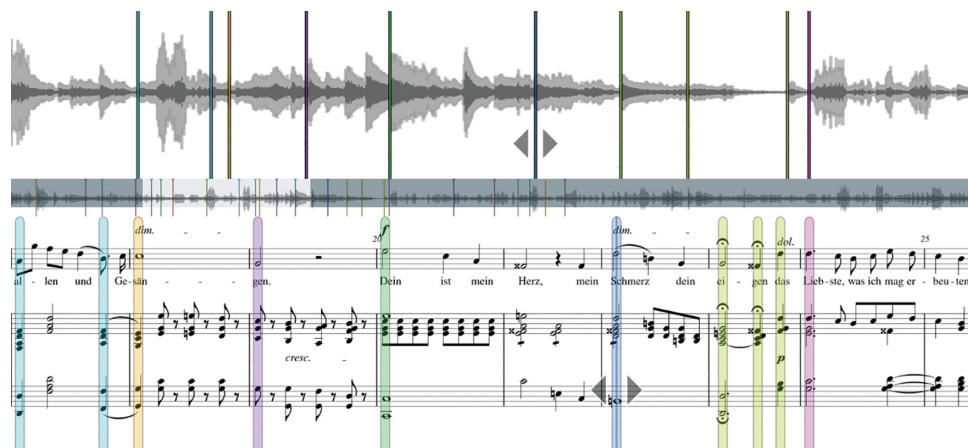


Figure 6 *Dein ist mein Herz* (Op. 7) by Fanny Mendelssohn, the OpenScore Lieder corpus (Gotham and Jonas, 2021), shown in the synchronization editing process. Color coding is used to distinguish corresponding synchronization points. Regular points, such as those on strong beats of each measure, can be added using 'tap' mode. Here, the synchronization is further refined by adding additional points, for example, to account for a slowdown before the cadence.

Figure 7 Editing and filtering capabilities. (Top) Slovenian folk song ‘*Lansko leto sem se vženiv*’ (‘Last Year I Got Married’) with analytical annotations from (Borsan et al., 2023). Each label can be selected and modified. (Bottom) First movement of Mozart’s Piano Sonata K279.1, displaying annotations from two different sources (Couturier et al., 2022; Hentschel et al., 2021), which can be filtered based on the annotation type.

Figure 8 Editing a harmonic analysis at the end of *Träumerei* (Kinderszenen, Op. 15) by R. Schumann, based on a 1905 piano roll from the SUPRA dataset, featuring a recording by Alfred Grünfeld (Shi et al., 2019).

The synchronization process involved 175 audio files across 151 pieces, carried out by 14 contributors. For 83 of these pieces, synchronization times were recorded, showing a time range of 4 to 60 minutes per piece, with an average of 14.3 minutes per piece. Future improvements include enhancing and leveraging audio-to-score alignment algorithms. Additionally, many cases require handling repeats (Fremerey et al., 2010), partial

recordings, or partial scores, which adds complexity to the synchronization process.

3.7 SYNCHRONIZING MULTIPLE AUDIO/VIDEO RECORDINGS

BACH FUGUES

SCHUBERT WINTERREISE

TRADITIONAL GEORGIAN SACRED MUSIC

Each recording is synchronized to the score, with musical time serving as the core reference on the platform. For pieces with multiple audio/video recordings (marked with ‘†’ in Table 1), users can seamlessly switch between different recordings during playback (Figure 2).

3.8 SYNCHRONIZING PARTIAL TRANSCRIPTIONS

WEIMAR JAZZ DATABASE (WJD)

SLOVENIAN FOLK SONG BALLADS

Transcriptions of orally transmitted or improvised music are often incomplete, capturing only selected sections of a performance. In the WJD, transcribed Jazz solos

typically appear in the middle sections of pieces, while in the Slovenian Folk Song Ballads corpus, transcriptions often focus on the first verse. In both cases, the platform enables users to listen to the full recording while following a partial score, which highlights only the transcribed sections. For consistency, we retained the WJD notation convention, where the first fully transcribed bar of a solo is labeled as bar 1.

4 DISCUSSION

4.1 USING THE PLATFORM IN RESEARCH, PEDAGOGY, AND PUBLIC ENGAGEMENT

The Dezrann platform provides a rich set of features that enhance music analysis, education, and public engagement. Key functionalities include metadata search, large-score browsing, label-based navigation, switching between multiple audio/video recordings, and uniform data access and download capabilities. Future perspectives include user trials to further refine and enhance the user experience. In this section, we highlight how these features already support researchers, educators, and the general audience, making the platform a versatile tool for exploring and interacting with annotated music corpora.

4.1.1 Research applications

The integration of formal and harmonic analyses across multiple corpora offers *deeper insights into musical structures*. For instance, in the Schubert Winterreise Dataset, users can identify and annotate motifs in the lieder or engage in interpretative analysis.

For music analysts, musicologists, and MIR researchers, the platform serves as a *comparative tool*, allowing them to evaluate their analyses against existing ones. Additionally, new annotations generated by researchers become a valuable resource for the community, enriching both the dataset and the analytical perspectives available within the platform.

The platform also supports *collaborative annotation*, with three corpora already developed directly within the system. Annotations can be created while viewing or listening to the score (Figure 7). One notable example is the texture annotations for Mozart's piano sonatas, which were systematically annotated by a team of one annotator and two reviewers (Couturier et al., 2022).

Looking ahead, a key goal is to enable anonymous research submissions in MIR, systematic musicology, and computational music analysis. This would allow reviewers to directly assess algorithmic results within the platform, fostering greater transparency and reproducibility in music analysis research.

4.1.2 Pedagogical applications

Beyond research, the Dezrann platform may play a significant role in education, offering a cross-modal approach

that enhances learning for students, educators, and performers. Between 2021 and 2024, the platform was used in one-hour workshops with over 300 school pupils across 15+ classes, focusing on selected pieces and supplemented by educational materials (Sauda et al., 2022). To support flexible teaching, pedagogical sheets have been created for certain works,¹¹ but teachers can freely select music and design their lessons according to their needs. The platform supports different teaching environments:

- In a traditional classroom (with an interactive whiteboard), teachers and students collectively explore the music and its annotations.
- In a computer lab setting, individual students or groups conduct self-guided analyses, which can later be shared and discussed collectively.

By allowing students to interact with synchronized scores, audio, and analytical labels, Dezrann encourages active listening, musical exploration, and independent learning.

4.1.3 Public engagement and cultural heritage

The platform's multilingual interface and diverse corpora contribute to public engagement, making annotated music collections accessible to a global audience. Currently, Dezrann is part of the ANR CollabScore project,¹² within a collaboration with computer science laboratories and institutions dedicated to music preservation, including the Bibliothèque Nationale de France and the Abbaye de Royaumont. A key objective of this initiative is to develop interfaces for browsing historical autographs, synchronized with scores and audio, thereby bridging the gap between archival music collections and digital scholarship (Guillotel-Nothmann et al., 2024). By integrating interactive exploration tools, educational resources, and collaborative research functionalities, Dezrann fosters a deeper connection between historical music traditions, modern analytical methods, and public engagement.

4.2 FROM CORPUS SELECTION TO LONG-TERM SUSTAINABILITY

We conclude by discussing key aspects of corpus selection, integration, and long-term maintenance within the Dezrann platform, but, more generally, for MIR data and software.

4.2.1 Corpus inclusion criteria, curation, and quality

Corpora included in Dezrann should meet the following criteria:

- The corpus is documented in an MIR or musicology publication, explicitly stating its purpose and methodology.

- The corpus ideally contains scores, audio/video, and annotations (or at least two of these), along with quality metadata.
- Data and metadata should be available under an open-data license (or at least partially open, with some content linked to external platforms).
- One or more curators/maintainers should oversee the corpus integration, ensuring that metadata, links, and references remain relevant and up to date.

The original authors of a corpus publication remain responsible for the quality and accuracy of their data and metadata.

4.2.2 Diversity in music research

One goal of Dezzrann is to highlight diverse musical traditions. Historically, musicology and MIR research have focused predominantly on Western classical music, particularly from the Baroque, Classical, and Romantic periods, and primarily on male European composers (Shuker, 2013; Solie, 1993). While the inclusion of folk music, jazz, and works by underrepresented woman composers in Dezzrann represents progress, we acknowledge that these efforts remain incomplete. We encourage both platform maintainers and the broader research community to actively engage with a wider range of musical sources, fostering greater inclusivity in music research and corpus development.

4.2.3 Corpus integration, data formats, and reproducibility

Once a corpus is selected, its integration into Dezzrann requires making the original materials (scores, audio, annotations) accessible, while adhering to the platform's technical and documentation standards, as stated on <https://doc.dezzrann.net/new-corpus>. This process often requires scripting expertise to format and convert data for compatibility.

Several authors, such as Gotham and Ireland (2019) and Devaney (2020), emphasize the importance of uniform encoding formats to facilitate large-scale interdisciplinary research. The standardization of corpus descriptions and annotations into the JSON-based formats aligns with this approach. When developing Dezzrann and this study, we always had in mind the goal of creating a functional system capable of displaying and synchronizing diverse corpora while meeting specific technical constraints, which led us to develop these formats. While no single existing ontology fully met our immediate needs, we acknowledge that interoperability with established ontologies could be improved. We are open to collaborations to explore better integrations.

To enhance reproducibility, the platform also undergoes continuous integration testing. A key objective is to provide scripts for each corpus that can fully rebuild the dataset for the platform.¹³ While this process is

challenging due to the diversity of data sources, it significantly strengthens research reproducibility and promotes open science practices. Several corpora are fully rebuilt from long-term archives hosted on institutional repositories (a requirement for achieving the highest rating in the quality:corpus criteria, see Figure 4), and efforts should continue to further improve the archiving of the complete collection.

Altogether, the platform aligns with the principles of FAIR (Findable, Accessible, Interoperable, and Reusable) data management (Moss and Neuwirth 2021). Implementing and maintaining these processes requires expertise in data handling and software development. For researchers without a computational background, collaboration with computer music specialists is strongly recommended to facilitate data integration and ensure methodological consistency.

Note also that some users likely want to benefit from the platform's features for a few pieces, or even a single piece, without necessarily considering questions of reproducibility. Ongoing work is being done to allow flexible usage and the addition of scores or other music sources by non-technical users.

4.2.4 Open data and accessibility

Openly licensed corpora play a crucial role in advancing MIR research, as they enable other researchers to extend and build upon existing work. A persistent challenge in MIR is the lack of open 'primary data', particularly for large-scale audio collections. Without an open database of millions of recordings, academic MIR research remains constrained.

Currently, approximately half of the Dezzrann corpora includes audio recordings under open licenses, while the remaining recordings are accessible via YouTube. Although linking to proprietary platforms is a common practice in MIR, it raises concerns regarding reproducibility and long-term accessibility.

For example, the WJD contains transcriptions of jazz solos, but the original recordings cannot be openly published. Instead, the transcriptions themselves serve as 'primary data'. The JazzTube project (Balke et al., 2018) linked WJD metadata to YouTube videos, but as of 2025, less than 250 out of the original 329 videos remain available.¹⁴

This situation raises important questions about the sustainability and accessibility of research resources. Beyond ensuring long-term material availability, reliance on for-profit platforms introduces ethical concerns and conflicts with the principles of open science that we strive to uphold. On the one hand, the research community should actively work toward producing and supporting more open-access audio resources. On the other hand, historically significant recordings, while not freely available, remain highly valuable for research and education. Although this is not an ideal solution, maintaining and

regularly testing/updating links to commercial platforms enables us to continue offering meaningful cross-modal interactions with annotated musical content, ensuring broader accessibility while navigating the constraints of proprietary data.

4.2.5 Sustainability and long-term maintenance

The long-term maintenance of Dezzrann faces challenges similar to those of corpus-specific platforms. While the technical and scientific leads of the platform hold permanent positions, they are also engaged in other projects, which may impact sustained development efforts. However, we firmly believe that a few general-purpose platforms featuring diverse corpora, especially when built as open-source initiatives, are ultimately more efficient and sustainable than multiple isolated, corpus-specific platforms. To support this development, we actively seek funding for open science projects. Additionally, comprehensive documentation on platform installation and corpus integration enables other research groups to deploy and maintain their own instances of Dezzrann, further enhancing its long-term viability.

The reuse of existing corpora in research, artistic, and educational contexts, combined with the transparent sharing of technical insights, breathes new life into these datasets and strengthens their long-term preservation. More broadly, we advocate for the sustained maintenance of an integrated ecosystem that includes both corpora and software tools. Our vision is to see the emergence of agnostic MIR/music platforms that facilitate the seamless exchange of diverse musical data, fostering greater interoperability and collaboration across the field.

ACKNOWLEDGEMENTS

We extend our gratitude to all users, researchers, teachers, pupils, and students who have engaged with Dezzrann, in the Algomus team and elsewhere, with special thanks to those acknowledged at <https://doc.dezzrann.net/credits>. This project was supported by CPER MAuVE (ERDF, Hauts-de-France), ANR CollabScore (ANR-20-CE27-0014), and the LesSciencesInfusent program, in collaboration with Académie de Lille. MM was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant No. 500643750 (MU 2686/15-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS. The large language model ChatGPT was used as a writing and translation aid; however, the human authors fully controlled the writing process and endorse the final text.

AUTHOR CONTRIBUTION

Project supervision: MG, EL. Web development: CB, QD, MG, EL, TT. Corpus integration and curation: CB, VNB, LC, QD, KF, LF, MG, MGo, JH, DVTL, EL, FM, GM, CW. Usage reports: BB, LB, VNB, LC, KD, LF, MGo, RG, JH, AH, GM, PT, RY. Synchronizations: BB, LB, VNB, LC, KD, LF, MG, AH, DVTL, FL, FM, IM, PT. Translations: VNB, MG, JH, FM, IM, AS, RY. Draft paper and corrections: BB, LB, VNB, LC, MG, RG, KD, FL, FM, MM, PT. All authors approve the final text.

COMPETING INTERESTS

Several authors of this paper were involved in the original studies related to the corpora presented on Dezzrann. Their appearance on the platform naturally reflects and promotes their earlier work. The authors declare that they have no other competing interests.

CODE AND DATA ACCESSIBILITY

Dezzrann is open-source, licensed under the GPL version 3 or any later version. The code and data are available from <https://gitlab.com/algomus.fr/dezzrann/>, and from the long-term archive <https://archive.softwareheritage.org>. The page <https://doc.dezzrann.net/rebuild> provides instructions to rebuild the corpora and further lists long-term archives where some corpora are published or republished. See Section 4.2.3.

NOTES

1. <https://trompamusic.eu/>
2. <https://linkedmusic.ca/about/>
3. <https://www.ismir.net/resources/datasets/> and <https://github.com/ismir/mir-datasets>
4. <https://parisson.github.io/Telemeta/>
5. <https://www.sonicvisualiser.org/>
6. <https://logiciels.pierrecouprie.fr/>
7. <https://lilypond.org>
8. <https://www.ffmpeg.org>
9. <https://vuejs.org/>
10. <https://spx.dev/>
11. <https://algomus.fr/edmus/> (in French)
12. <https://collabscore.cnam.fr>
13. <https://doc.dezzrann.net/rebuild>
14. <http://mir.audiolabs.uni-erlangen.de/jazztube/>


All these URLs, as well as the URLs in the main text, were accessed on April 21, 2025.


AUTHOR AFFILIATIONS

Charles Ballester

Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Baptiste Bacot  <https://orcid.org/0009-0006-2883-3874>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Louis Bigo  <https://orcid.org/0000-0002-9865-2861>
Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

Vanessa Nina Borsan  <https://orcid.org/0000-0001-8421-4409>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Louis Couturier  <https://orcid.org/0009-0000-5226-9157>
Université de Picardie Jules Verne, MIS, UR 4290, F-80000 Amiens, France


Ken Déguernel  <https://orcid.org/0000-0001-7919-3463>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Quentin Dinel
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Laurent Feisthauer  <https://orcid.org/0000-0002-4256-9989>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Klaus Frieler  <https://orcid.org/0000-0002-6055-377X>
Max Planck Institute for Empirical Aesthetics, Frankfurt a. M., Germany

Mark Gotham  <https://orcid.org/0000-0003-0722-3074>
Department of Digital Humanities, King's College London, UK


Richard Groult  <https://orcid.org/0000-0002-2704-6629>
Univ Rouen Normandie, INSA Rouen Normandie, Université Le Havre Normandie, Normandie Univ, LITIS UR 4108, F-76000 Rouen, France

Johannes Hentschel  <https://orcid.org/0000-0002-1986-9545>
École Polytechnique Fédérale de Lausanne, CH


Alexandre d'Hooge  <https://orcid.org/0000-0003-1634-3406>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France


Dinh-Viet-Toan Le  <https://orcid.org/0000-0001-6991-4079>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Florence Levé  <https://orcid.org/0000-0002-6571-5096>
Université de Picardie Jules Verne, MIS, UR 4290, F-80000 Amiens, France; Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Francesco Maccarini  <https://orcid.org/0000-0002-0219-5898>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Ivana Maričić  <https://orcid.org/0000-0002-7114-5245>
ZRC SAZU, Institute of Musicology, Ljubljana, Slovenia


Gianluca Micchi  <https://orcid.org/0000-0002-3826-4744>
Universal Music Group, London, UK


Meinard Müller  <https://orcid.org/0000-0001-6062-7524>
International Audio Laboratories Erlangen, 91058 Erlangen, Germany


Alexandros Stamatiadis  <https://orcid.org/0009-0009-4032-4759>

Université de Picardie Jules Verne, MIS, UR 4290, F-80000 Amiens, France


Tom Taffin
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Patrice Thibaud  <https://orcid.org/0009-0006-1996-2201>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Christof Weiß  <https://orcid.org/0000-0003-2143-4679>
Center for Artificial Intelligence and Data Science, Universität Würzburg, Würzburg, Germany

Rui Yang  <https://orcid.org/0009-0007-6079-1480>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Emmanuel Leguy  <https://orcid.org/0000-0002-9415-4961>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

Mathieu Giraud  <https://orcid.org/0000-0003-2741-8047>
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

REFERENCES

- Allegraud, P., Bigo, L., Feisthauer, L., Giraud, M., Groult, R., Leguy, E., and Levé, F. (2019). Learning sonata form structure on Mozart's string quartets. *Transactions of the International Society for Music Information Retrieval*, 2(1): 82–96.
- Balke, S., Dittmar, C., Abeßer, J., Frieler, K., Pfeleiderer, M., and Müller, M. (2018). Bridging the gap: Enriching YouTube videos with jazz music annotations. *Frontiers in Digital Humanities*, 5: 1.
- Berndt, A. (2021). *The music performance markup format and ecosystem*. In *International Society for Music Information Retrieval Conference (ISMIR 2021)* (pp. 50–57).
- Borsan, V. N., Giraud, M., Groult, R., and Lecroq, T. (2023). *Adding context to content improves pattern matching: A study on Slovenian folksongs*. In *International Society for Music Information Retrieval Conference (ISMIR 2023)* (pp. 474–481).
- Borsan, V. N., Kovačič, M., Giraud, M., Pisk, M., Pesek, M., and Marolt, M. (2025). Introducing the digitised dataset of Slovenian Folk Ballads. *Ethnomusicology Forum*, in press.
- Cancino-Chacón, C., Peter, S. D., Karystinaios, E., Foscari, F., Grachten, M., and Widmer, G. (2023). *Partitura: A python package for symbolic music processing*. In *Music Encoding Conference (MEC 2022)*.
- Couturier, L., Bigo, L., and Levé, F. (2022). *A dataset of texture annotations in Mozart piano sonatas*. In *International Society for Music Information Retrieval Conference (ISMIR 2022)*.

- Cuthbert, M. S., and Ariza, C.** (2010). *music21: A toolkit for computer-aided musicology and symbolic music data*. In *International Society for Music Information Retrieval Conference (ISMIR 2010)* (pp. 637–642).
- de Berardinis, J., Carriero, V. A., Meroño-Peñuela, A., Poltronieri, A., and Presutti, V.** (2022). *The music meta ontology: A flexible semantic model for the interoperability of music metadata*. In *International Society for Music Information Retrieval Conference (ISMIR 2023)* (pp. 859–867).
- de Berardinis, J., Meroño-Peñuela, A., Poltronieri, A., and Presutti, V.** (2023). Choco: A chord corpus and a data transformation workflow for musical harmony knowledge graphs. *Scientific Data*, 10(1): 641.
- Devaney, J.** (2020). Using note-level music encodings to facilitate interdisciplinary research on human engagement with music. *Transactions of the International Society for Music Information Retrieval*, 3(1).
- Dorfer, M., Henkel, F., and Widmer, G.** (2018). *Learning to listen, read, and follow: Score following as a reinforcement learning game*. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*.
- Ewert, S., Müller, M., and Grosche, P.** (2009). *High resolution audio synchronization using chroma onset features*. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)* (pp. 1869–1872).
- Fremerey, C., Müller, M., and Clausen, M.** (2010). *Handling repeats and jumps in score-performance synchronization*. In *International Society for Music Information Retrieval Conference (ISMIR 2010)* (pp. 243–248).
- Fyfe, L., Bedoya, D., and Chew, E.** (2022). Annotation and analysis of recorded piano performances on the web. *Journal of the Audio Engineering Society*, 70(11).
- Garczynski, L., Giraud, M., Leguy, E., and Rigaux, P.** (2022). *Modeling and editing cross-modal synchronization on a label web canvas*. In *Music Encoding Conference (MEC 2022)* (pp. 61–71).
- Giraud, M., Groult, R., Leguy, E., and Levé, F.** (2015). Computational fugue analysis. *Computer Music Journal*, 39(2).
- Gotham, M., Hentschel, J., Couturier, L., Dykeaylen, N., Rohrmeier, M., and Giraud, M.** (2023a). The ‘Measure Map’: An inter-operable standard for aligning symbolic music. In *Digital Libraries for Musicology (DLfM 2023)* (pp. 91–99).
- Gotham, M., and Ireland, M.** (2019). *Taking form: A representation standard, conversion code, and example corpus for recording, visualizing, and studying analyses of musical form*. In *International Society for Music Information Retrieval Conference (ISMIR 2019)* (pp. 693–699).
- Gotham, M., Micchi, G., López, N. N., and Sailor, M.** (2023b). When in Rome: A meta-corpus of functional harmony. *Transactions of the International Society for Music Information Retrieval*, 6(1): 150–166.
- Gotham, M. R. H., and Jonas, P.** (2021). *The OpenScore Lieder corpus*. In *Music Encoding Conference (MEC 2021)* (pp. 131–136).
- Guillotel-Nothmann, C., Rigaux, P., Coüasnon, B. B., Giraud, M., and Lemaitre, A.** (2024). The Collab-Score project – from optical recognition to multimodal music sources. In *Workshop on Reading Music Systems (WoRMS 2024)* (pp. 33–37).
- Hentschel, J., Neuwirth, M., and Rohrmeier, M.** (2021). The annotated Mozart sonatas: Score, harmony, and cadence. *Transactions of the International Society for Music Information Retrieval*, 4(1): 67–80.
- Le, D.-V.-T., Bigo, L., Herremans, D., and Keller, M.** (2025). Natural language processing methods for symbolic music generation and information retrieval: A survey. *ACM Computing Surveys*.
- Le, D.-V.-T., Giraud, M., Levé, F., and Maccarini, F.** (2022). A corpus describing orchestral texture in first movements of classical and early-romantic symphonies. In *Digital Libraries for Musicology (DLfM 2022)* (pp. 22–35).
- Lewis, D., Shibata, E., Saccomano, M., Rosendahl, L., Kepper, J., Hankinson, A., Siegert, C., and Page, K.** (2022). A model for annotating musical versions and arrangements across multiple documents and media. In *Digital Libraries for Musicology (DLfM 2022)* (pp. 10–18).
- Mazzoni, D., and Dannenberg, R. B.** (2002). A fast data structure for disk-based audio editing. *Computer Music Journal*, 26(2): 62–76.
- McFee, B., Kim, J. W., Cartwright, M., Salamon, J., Bittner, R. M., and Bello, J. P.** (2019). Opensource practices for music signal processing research: Recommendations for transparent, sustainable, and reproducible audio research. *IEEE Signal Processing Magazine*, 36(1): 128–137.
- McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O.** (2015). *librosa: Audio and music signal analysis in python*. In *Python in Science Conference (SciPy 2015)* (pp. 18–24).
- Moss, F. C., and Neuwirth, M.** (2021). FAIR, open, linked: Introducing the special issue on open science in musicology. *Empirical Musicology Review*, 16(1): 1–4.
- Müller, M., Özer, Y., Krause, M., Prätzlich, T., and Driedger, J.** (2021). Sync Toolbox: A Python package for efficient, robust, and accurate music synchronization. *Journal of Open Source Software (JOSS)*, 6(64): 3434:1–4.
- Müller, M., Arzt, A., Balke, S., Dorfer, M., and Widmer, G.** (2019). Cross-modal music retrieval and applications: An overview of key methodologies. *IEEE Signal Processing Magazine*, 36(1): 52–62.
- Nattiez, J.-J.** (1975). *Fondements d’une sémiologie de la musique*. Dufrenne.
- Panteli, M., Benetos, E., and Dixon, S.** (2018). A review of manual and computational approaches for the study of world music corpora. *Journal of New Music Research*, 47(2): 176–189.
- Peeters, G., and Fort, K.** (2012). Towards a (better) definition of annotated MIR corpora. In *International Society for Music Information Retrieval Conference (ISMIR 2012)* (pp. 25–30).

- Pfleiderer, M., Frieler, K., Abeßer, J., Zaddach, W.-G., and Burkhart, B.** (Eds.). (2017). *Inside the Jazzomat – New Perspectives for Jazz Research*. Schott Campus.
- Poltronieri, A., and Gangemi, A.** (2022). The HaMSE ontology: Using semantic technologies to support music representation interoperability and musicological analysis. In *Multisensory Data and Knowledge (MDK 2021)* (pp. 62–76).
- Porter, A., Sordo, M., and Serra, X.** (2013). *Dunya: A system for browsing audio music collections exploiting cultural context*. In *International Society for Music Information Retrieval Conference (ISMIR 2013)* (pp. 101–106).
- Pugin, L., Zitellini, R., and Roland, P.** (2014). *Verovio: A library for engraving MEI music notation into SVG*. In *International Society for Music Information Retrieval Conference (ISMIR 2014)* (pp. 107–112).
- Rosenzweig, S., Scherbaum, F., Shugliashvili, D., Arifi-Müller, V., and Müller, M.** (2020). Erkomaishvili dataset: A curated corpus of traditional georgian vocal music for computational musicology. *Transactions of the International Society for Music Information Retrieval*, 3(1): 31–41.
- Sauda, A., Giraud, M., and Leguy, E.** (2022). Soutenir en classe l'écoute active, l'autonomie et l'échange en analyse musicale avec la plateforme web Dezrann. In *Sound and Music Computing / Journées d'Informatique Musicale (SMC 2022 / JIM 2022)* (pp. 617–623).
- Savage, P. E.** (2022). An overview of cross-cultural music corpus studies. In *The Oxford Handbook of Music and Corpus Studies*. Oxford University Press.
- Schöning, K., de Valk, R., Weigl, D. M., Kyriazis, I., Janjuš, O., Burghoff, H., and Steindl, C.** (2025). A collaborative digital edition of 15th- and 16th-century german lute tablature: The E-LAUTE project. *Journal of New Music Research*, 1–16.
- Serra, X.** (2014). Creating research corpora for the computational study of music: The case of the CompMusic project. In *AES International Conference on Semantic Audio*.
- Shi, Z., Sapp, C., Arul, K., McBride, J., and Smith III, J. O.** (2019). SUPRA: Digitizing the stanford university piano roll archive. In *International Society for Music Information Retrieval Conference (ISMIR 2019)* (pp. 517–523).
- Shugliashvili, D.** (2014). *Georgian Church Hymns (Shemokmedi School): Transcriptions of Artem Erkomaishvili Recordings* (2nd ed.).
- Shuker, R.** (2013). *Understanding Popular Music Culture*. Routledge.
- Solie, R. A.** (Ed.). (1993). *Musicology and Difference: Gender and Sexuality in Music Scholarship*. University of California Press.
- Thickstun, J., Brennan, J., and Verma, H.** (2020). Rethinking evaluation methodology for audio-toscore alignment. *arXiv preprint arXiv:2009.14374*.
- Thomas, V., Fremerey, C., Müller, M., and Clausen, M.** (2012). Linking sheet music and audio – challenges and new approaches. In *Multimodal Music Processing* (Vol. 3, pp. 1–22).
- Wang, L., Zhao, Z., Liu, H., Pang, J., Qin, Y., and Wu, Q.** (2024). A review of intelligent music generation systems. *Neural Computing and Applications*, 36(12): 6381–6401.
- Weigl, D. M., Crawford, T., Gkiokas, A., Goebl, W., Gómez, E., Guti, N., Liem, C. C. S., and Santos, P.** (2021a). FAIR interconnection and enrichment of public-domain music resources on the web. *Empirical Musicology Review*, 16(1): 16–33.
- Weigl, D. M., Goebl, W., Baker, D. J., Crawford, T., Zubani, F., Gkiokas, A., Gutierrez, N. F., Porter, A., and Santos, P.** (2021b). Notes on the music: A social data infrastructure for music annotation. In *Digital Libraries for Musicology (DLfM 2021)* (pp. 23–31).
- Weigl, D. M., VanderHart, C., Rammner, D., Pescoller, M., and Goebl, W.** (2023). Listen here! a web-native digital musicology environment for machine-assisted close listening. In *Digital Libraries for Musicology (DLfM 2023)* (pp. 109–118).
- Weiß, C., Zalkow, F., Arifi-Müller, V., Müller, M., Koops, H. V., Volk, A., and Grohgan, H. G.** (2021). Schubert Winterreise Dataset: A multimodal scenario for music analysis. *Journal on Computing and Cultural Heritage*, 14(2).
- Werner, N., Balke, S., Stöter, F.-R., Müller, M., and Edler, B.** (2017). *trackswitch.js: A versatile webbased audio player for presenting scientific results*. In *Web Audio Conference (WAC 2017)*.
- Winget, M. A.** (2008). Annotations on musical scores by performing musicians. *Journal of the Association for Information Science and Technology*, 59(12): 1878–1897.
- Zalkow, F., Rosenzweig, S., Graulich, J., Dietz, L., Lemnaouar, E. M., and Müller, M.** (2018). A web-based interface for score following and track switching in choral music. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*. late-breaking demo.

APPENDIX: CORPORA DESCRIPTIONS

On Dezrann, these descriptions are translated into each one of the eight languages. We release these texts as public domain (CC0-1.0).

BACH FUGUES

The 24 annotated fugues from Book I of The Well-Tempered Clavier, with some video recordings from Kimiko Ishizaka and the Netherlands Bach Society.

The two books of J. S. Bach's *Well-Tempered Clavier* are a particularly consistent corpus in their exploration of the 24 major and minor tonalities in 48 pairs of preludes and fugues. The Italian word *fuga* is related to the Latin words *fugere* (to flee) and *fugare* (to chase): The themes (*subject* and *counter-subjects*) are played by each voice, one following the other in succession. They occur either in their initial form or, more often, altered or transposed, building a complex harmonic texture.

Bach's Well-Tempered Clavier has been extensively studied, and systematic analyses of Bach's fugues have been published by Prout (1910), Tovey (1924), Keller (1965), and Bruhn (2014). [Giraud et al. \(2015\)](#) published an annotation dataset detailing the 24 fugues of the first book, together with algorithms for fugue analysis. The Dezzann corpus contains these 24 annotated fugues, with scores synchronized to open recordings by [Kimiko Ishizaka](#) as well as performances recorded by the [Bach Netherlands Society](#).

MOZART PIANO SONATAS

The 18 Mozart piano sonatas with some form, harmony, and texture annotations

Wolfgang Amadeus Mozart (1756–1791) was a composer of the 18th-century Classical style period, recognized as one of the three principal figures of the *First Viennese School*, alongside Joseph Haydn and Ludwig van Beethoven. He expressed his versatile music ideas in a large palette of genres. His *piano sonatas*, published over a 15-year period, were composed for various purposes, including educational material and private commissions from aristocrats. The classical sonata (typically for a solo keyboard instrument) is composed of usually three movements, of which the first generally follows *sonata form*. Mozart's sonatas are well known to have a remarkable structural and textural composition. The corpus consists of complete scores of all 18 sonatas with form, harmony, and cadence annotations ([Hentschel et al., 2021](#)). Sonatas 1 (K279), 2 (K280) and 5 (K283) also have texture annotations ([Couturier et al., 2022](#)). Some movements also have synchronized audio. The corpus uses measure maps ([Gotham et al., 2023](#)) to improve annotation interoperability.

MOZART STRING QUARTETS

23 Mozart string quartets with some key, cadence, and form annotations.

Wolfgang Amadeus Mozart composed 23 *string quartets* spread over his whole life. These pieces denote the evolution of his musical style and the construction of his melodies, from the *galant style* of the first quartet to the *classical style*. Mozart uses them as an experimentation field where he can explore the *sonata form* in hybrid configurations and spicy modulations. The corpus here shows 72 out of the 86 movements. Cadence and key and form annotations are provided for some of these movements (mainly first movements, in sonata form), as published in [Allegraud et al. \(2019\)](#) and [Feisthauer \(2021\)](#).

ORCHESTRAL TEXTURE

The first movements of Classical and Early-Romantic Symphonies (Haydn, Mozart, Beethoven), with orchestral texture analysis.

Symphony is a music genre that developed from the late 18th to the 20th century. It is usually composed for a large instrumental ensemble, e.g., orchestra, and conventionally consists of four movements, which commonly share some musical ideas. The *classical symphony* was developed by the so-called *First Viennese School*, that is Joseph Haydn (1732–1809), Wolfgang Amadeus Mozart (1756–1791), and Ludwig van Beethoven (1770–1827). Its first movement is usually in *sonata form*. The corpus includes the first movements of 24 symphonies composed between 1779 and 1824: the last six Haydn Symphonies (99–104), three Mozart Symphonies (38–40), and all nine Beethoven Symphonies. These movements are analyzed with textural annotations by ([Le et al., 2022](#)), and some are synchronized with recordings by the Bamberger Symphoniker (Mozart) and by the Royal Philharmonic Orchestra ([Haydn, Beethoven, 1960–61](#)).

SCHUBERT'S WINTERREISE

Schubert's song cycle of 24 songs (1827–28) on romantic poems by Wilhelm Müller, score synchronized with two recordings (1933, 2006).

The song cycle *Winterreise* by [Franz Schubert](#) (1797–1828) includes 24 art songs or *lieder*, which were composed in two parts during the last two years of Schubert's life. The *lieder* were set to selected poems by his contemporary, [Wilhelm Müller](#) (1794–1827). The romantic era brings illustrious descriptions of individual emotions or states, focusing especially on loneliness and isolation. This is sensed not only in Müller's poems but also well captured by Schubert's music in both voice and piano parts. His expertise in word painting releases the piano from a mere chordal or arpeggiated accompaniment. Musical emphasis on certain words embodies '*leitmotifs*', which reoccur in selected songs throughout the cycle. The Schubert *Winterreise* Dataset (SWD, [Weiß 2021](#)) contains, for all of the 24 *lieder*, scores, harmonic and formal analyses, as well as synchronized recordings. The free recordings with Gerhard Hüsch and Hanns-Udo Müller (1933) and Randall Scarlata and Jeremy Denk (2006) are available through Dezzann.

19TH-CENTURY LIEDER

176 Lieder (solo voice, piano) by 19th-century female composers, including Fanny Mendelssohn and Josephine Lang.

Romantic *lieder* consist of solo songs for voice and piano accompaniment. These songs are typically set to poetry or lyrics and are known for their expressive and emotional qualities. In the 19th century, a period when women were often excluded from many areas of music, *lieder* was one genre where female composers made significant contributions. [Fanny Mendelssohn](#), [Clara Schumann](#), [Josephine Lang](#), [Louise Reichardt](#),

Augusta Holmès, and other women added depth and diversity to the world of lieder, expanding its horizons and contributing to the rich tapestry of 19th-century music. The OpenScore Lieder corpus consists of over 1,300 songs from the long 19th century. The collection is available to play online at musescore.com and is also [available for download](#). For more on the score collection see (Gotham and Jonas, 2021) or [this magazine piece](#). This Dezrann collection presents a subset of the scores by women composers, including harmonic analyses published on the ‘When in Rome’ meta-corpus reported in Gotham et al. (2023a). The corpus uses measure maps (Gotham et al., 2023b) to improve annotation interoperability.

SUPRA

456 piano rolls with expressive audio renderings from the Stanford University Piano Roll Archive.

At the start of the 20th century, piano rolls were common in bars and other public spaces. They were one of the main ways to distribute music to the society. Among them, the Welte T-100 rolls are often called *red Welte rolls* due to their red paper. They keep information of the notes, but also of several parameters controlling dynamics and pedals. The [Stanford University Piano Roll Archive](#) is a research portal for some rolls digitized from the Stanford Libraries’ collection of 15,000+ piano and organ rolls. SUPRA contains 456 Welte T-100 piano rolls from the years 1905–1928 with rendered ‘expressive’ audio, taking into account dynamics and tempo information. The Dezrann corpus shows here these 456 piano rolls aligned to the audio files.

WEIMAR JAZZ DATABASE

330+ high-quality jazz transcriptions, some of them with synchronized audio, from the Jazzomat project.

[Improvisation](#) is the essence of jazz. It includes the spontaneous creation of solo lines or accompaniment parts, often performed over a chord grid. Listening, transcribing, analyzing, and sometimes playing great jazz solos are invaluable to better understand jazz music and to contribute to its perpetual dynamism. Started at the University of Music in Weimar, the [Jazzomat](#) project studied the jazz repertoire, in particular, by transcribing and analyzing 400+ solos and aligning them to recordings. The Dezrann corpus contains 330+ of these high-quality jazz transcriptions, with chords, sections, and form annotation, from which 200+ are with synchronized audio.

SLOVENIAN FAMILY BALLADS

404 Slovenian folk song ballads with family themes, collected between the years 1819 and 1995.

Family songs are *folk song ballads*, i.e., songs with relatively short but repetitive melodic sections through which

the singer narrates a longer story through the lyrics. The melody and lyrics are known not to be fully dependent on each other, meaning the same melody could be adapted for another lyric and vice versa. Thematically, they fall into the category of those that sing about *family destinies*, including several motifs from Slovenian, Slavic, and broader European themes. The content often focuses on describing the natural course of family life, from courtship to childbirth. The themes and motifs of family ballads (feudal, rural environment, the time of Turkish invasions, pilgrimages, etc.) revolve around socio-legal relations and both immediate and broader family matters in historical periods from which individual ballads originate.

The collection of Slovenian folk song ballads contains transcribed field material collected by Slovenian ethnologists, folklorists, ethnomusicologists and various colleagues of [Glasbenonarodopisni inštitut ZRC SAZU](#) spanning from the years 1819 to 1995. Categorized thematically as family ballads, this collection features 404 folk songs, and includes the initial verse of the lyrics, extensive metadata, and musical analysis, encompassing contours, harmony, and song structure (melody and lyric) (see (Borsan et al., 2023) and (Borsan et al., 2025)). Twenty-three of these songs have historical recordings.

GEORGIAN SACRED SONGS

101 songs recorded in 1966 by Artem Erkomaishvili, with synchronized scores.

Georgian Orthodox liturgical chants feature three-part harmonies with distinct dissonances, specific temperaments, and ornamentations. The tradition is very old. Some songs can be traced back to neume transcriptions from the 10th century (Graham, 2015). The *master chanters* (sruligalobelni) were trained for several years in both music and theological subjects. They memorized hundreds of harmonized melodies. Georgian religious traditions, including sacred chants, faced suppression during Soviet rule and are now being studied once again.

The Erkomaishvili dataset consists of historic tape recordings of three-voice Georgian religious songs performed in 1966 by the master chanter [Artem Erkomaishvili](#). Successive overdubbing recordings were done for each song: top voice, then top and second voice, then the three voices together. These recordings have been digitized, curated, and analyzed by [Rosenzweig \(2020\)](#) for computational musicology research. The dataset includes audio material, scores based on the transcriptions by [Shugliashvili \(2014\)](#), synchronizations, and F0 annotations. The Dezrann corpus contains all 101 songs of the Erkomaishvili dataset, with scores synchronized with the audio files.

TO CITE THIS ARTICLE:

Ballester, C., Bacot, B., Bigo, L., Borsan, V.N., Couturier, L., Déguernel, K., Dinel, Q., Feisthauer, L., Frieler, K., Gotham, M., Groult, R., Hentschel, J., d'Hooge, A., Le, D-V-T, Levé, F., Maccarini, F., Maričić, I., Micchi, G., Müller, M., Stamatidis, A., Taffin, T., Thibaud, P., Weiß, C., Yang, R., Leguy, E., & Giraud, M. (2025). Interacting with Annotated and Synchronized Music Corpora on the Dezzann Web Platform. *Transactions of the International Society for Music Information Retrieval*, 8(1), 121–139. DOI: <https://doi.org/10.5334/tismir.212>

Submitted: 30 August 2024 **Accepted:** 12 May 2025 **Published:** 30 May 2025

COPYRIGHT:

© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <https://creativecommons.org/licenses/by/4.0/>.

Transactions of the International Society for Music Information Retrieval is a peer-reviewed open access journal published by Ubiquity Press.