

ms3: A parser for MuseScore files, serving as data factory for annotated music corpora

Johannes Hentschel¹ and Martin Rohrmeier¹

¹ École Polytechnique Fédérale de Lausanne, Switzerland ¶ Corresponding author

DOI: [10.21105/joss.05195](https://doi.org/10.21105/joss.05195)

Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [Fabian-Robert Stöter](#)

Reviewers:

- [@TGabor](#)
- [@allorems](#)

Submitted: 17 January 2023

Published: 09 August 2023

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

Digital Musicology is a vibrant and quickly growing discipline that addresses traditional and novel music-related research questions with digital and computational means ([Honing, 2006](#); [Huron, 1999](#); [Urberg, 2017](#)). Research questions and methods often overlap with or draw on those from diverse disciplines such as music theory and analysis, composition history, mathematics, cognitive psychology, linguistics, anthropology, or computer science ([Volk & Honing, 2012](#); [Wiggins, 2012](#)). Corpus research, i.e., the computational study of representative collections of texts (in the case of linguistics) or notated music (in musicology), plays a prominent role in this trans-disciplinary quest to “make sense of music” through scientific models ([London, 2013](#); [Moss, 2019](#); [Shanahan, 2022](#)). ms3 makes scores (symbolic representations of music) operational for computational approaches by representing their contents as sets of tabular files.

Statement of need

Music scores represent relations between sounding events by graphical means. Music notation software therefore is often concerned with problems of layout and easy-to-read rendering of symbols in line with the multitude of notational conventions ([Read, 1979](#); [Ross, 2001](#)); rather than with the explicit encoding of the musical relations themselves. For example, the Free and Open Source Software [MuseScore](#) provides a full-featured yet intuitive interface for engraving music, but its native XML format does not explicitly encode the temporal positions of events such as notes and rests. Hence the need for a parser that extracts the implicit information and stores it in an interoperable format.

Despite being one of the most widespread score encoding formats, current score parsers (e.g., [Cancino-Chacón et al., 2022](#); [Cuthbert, 2010](#); [Pugin et al., 2014](#)), do not handle it without first performing a lossy conversion to the musicXML format¹. The Python library ms3 fills this gap. It loads the XML tree of a MuseScore file into working memory, computes the temporal positions of all encoded elements, and transforms those requested by the user into DataFrames ([Petersohn, 2021](#)). The DataFrames can be used by other Python programs and scripts, or written to Tab-Separated Values (TSV) to enable processing with other software and facilitate version control². The most typical aspects that users extract from a score are tables containing notes, measures (bars), metadata, and text labels, in particular those representing analytical annotations. Moreover, ms3 allows the user to transform scores by removing analytical labels after their extraction or by (re-)inserting annotations from TSV files (whether previously extracted or generated from scratch). This functionality turns MuseScore into a convenient score annotation tool enabling users to graphically insert into a score arbitrary textual labels,

¹For example, musicXML’s implicit encoding of temporal positions is limited to those where a note or rest event occurs. When converting MuseScore XML to musicXML, all score elements occurring between two such events are misplaced.

²Version control is facilitated by the TSV files because, unlike the original XML source, they present score information with timestamps.

to then have `ms3` extract them with their temporal positions for further analysis. It comes with a command line interface that makes its data extraction, transformation, and validation functionalities accessible for productive everyday workflows.

`ms3` has already been used for creating several datasets, namely version 2 of the Annotated Beethoven Corpus (Neuwirth et al., 2018), the Annotated Mozart Sonatas (Hentschel, Neuwirth, et al., 2021), and an annotated corpus of 19th century piano music (Hentschel, Rammos, et al., in press). It has been successful in formatting training and validation data for a chord inference algorithm and for inserting its analytical outputs into the respective scores (McLeod & Rohrmeier, 2021). Moreover, the library is at the heart of a semi-automated annotation workflow running on GitHub (Hentschel, Moss, et al., 2021) and a dependency on the music corpus analysis library DiMCAT (Hentschel, McLeod, et al., in press).

Acknowledgements

Development of this software tool was supported by the Swiss National Science Foundation within the project “Distant Listening – The Development of Harmony over Three Centuries (1700–2000)” (Grant no. 182811). This project is being conducted at the Latour Chair in Digital and Cognitive Musicology, generously funded by Mr Claude Latour.

References

- Cancino-Chacón, C. E., Peter, S. D., Karystinaios, E., Foscari, F., Grachten, M., & Widmer, G. (2022). Partitura: A python package for symbolic music processing. *Proceedings of the Music Encoding Conference (MEC2022)*.
- Cuthbert, M. S. (2010). Music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data. *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR*, 637–642. ISBN: 978-90-393-5381-3
- Hentschel, J., McLeod, A., Rammos, Y., & Rohrmeier, M. (in press). Introducing DiMCAT for processing and analyzing notated music on a very large scale. *Proceedings of the 24th International Society for Music Information Retrieval Conference, ISMIR*.
- Hentschel, J., Moss, F. C., Neuwirth, M., & Rohrmeier, M. A. (2021). A semi-automated workflow paradigm for the distributed creation and curation of expert annotations. *Proceedings of the 22nd International Society for Music Information Retrieval Conference, ISMIR*, 262–269. <https://doi.org/10.5281/ZENODO.5624417>
- Hentschel, J., Neuwirth, M., & Rohrmeier, M. (2021). The Annotated Mozart Sonatas: Score, Harmony, and Cadence. *Transactions of the International Society for Music Information Retrieval*, 4(1), 67–80. <https://doi.org/10.5334/tismir.63>
- Hentschel, J., Rammos, Y., Neuwirth, M., Moss, F. C., & Rohrmeier, M. (in press). An annotated corpus of tonal piano music from the long 19th century. *Empirical Musicology Review*.
- Honing, H. (2006). On the Growing Role of Observation, Formalization and Experimental Method in Musicology. *Empirical Musicology Review*, 1(1), 2–6. <https://doi.org/10.18061/1811/21901>
- Huron, D. (1999). *The New Empiricism: Systematic Musicology in a Postmodern Age* (Music and Mind: Foundations of Cognitive Musicology No. 3). The 1999 Ernest Bloch Lecture.
- London, J. (2013). Building a Representative Corpus of Classical Music. *Music Perception*, 31(1), 68–90. <https://doi.org/10.1525/mp.2013.31.1.68>

- McLeod, A., & Rohrmeier, M. A. (2021). A modular system for the harmonic analysis of musical scores using a large vocabulary. *Proceedings of the 22nd International Society for Music Information Retrieval Conference, ISMIR*, 435–442. <https://doi.org/10.5281/zenodo.5624433>
- Moss, F. C. (2019). *Transitions of Tonality: A Model-Based Corpus Study* [PhD thesis, École Polytechnique Fédérale de Lausanne]. <https://doi.org/10.5075/epfl-thesis-9808>
- Neuwirth, M., Harasim, D., Moss, F. C., & Rohrmeier, M. (2018). The Annotated Beethoven Corpus (ABC): A Dataset of Harmonic Analyses of All Beethoven String Quartets. *Frontiers in Digital Humanities*, 5(July), 1–5. <https://doi.org/10.3389/fdigh.2018.00016>
- Petersohn, D. (2021). *Dataframe Systems: Theory, Architecture, and Implementation* [PhD thesis]. University of California.
- Pugin, L., Zitellini, R., & Roland, P. (2014). Verovio: A Library for Engraving MEI Music Notation into SVG. *Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR*, 107–112.
- Read, G. (1979). *Music notation: A manual of modern practice* (2d ed). Taplinger Pub. Co. ISBN: 978-0-8008-5453-9
- Ross, T. (2001). *The art of music engraving & processing*. npc Imaging. ISBN: 978-0-9706231-1-9
- Shanahan, D. (2022). What the history of computational musicology can tell us about the future of corpus studies. *Future Directions of Music Cognition Virtual Speaker Series*. <https://doi.org/10.17605/OSF.IO/CB26R>
- Urberg, M. (2017). Pasts and Futures of Digital Humanities in Musicology: Moving Towards a “Bigger Tent.” *Music Reference Services Quarterly*, 20(3-4), 134–150. <https://doi.org/10.1080/10588167.2017.1404301>
- Volk, A., & Honingh, A. (2012). Mathematical and computational approaches to music: Challenges in an interdisciplinary enterprise. *Journal of Mathematics and Music*, 6(2), 73–81. <https://doi.org/10.1080/17459737.2012.704154>
- Wiggins, G. A. (2012). Music, mind and mathematics: Theory, reality and formality. *Journal of Mathematics and Music*, 6(2), 111–123. <https://doi.org/10.1080/17459737.2012.694710>