

Enabling Music Search and Analysis: A Database for Symbolic Music Files

Yaolong Ju, McGill University

Emily Hopkins, McGill University

Gustavo Polins Pedro, McGill University

Julie Cumming, McGill University

Cory McKay, Marianopolis College

Ichiro Fujinaga, McGill University

Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)

Music Encoding Conference 2019

2019.5.30

Two questions

(1) How to store symbolic music files?

Example: Beethoven's Third Symphony — a musical work

- Different movements
- Different formats
- Different encoding settings
- Different sources

(2) As the database continues to grow, how to search music?

Introduction

Symbolic music files are invaluable resources for music research:

- Harmonic analysis (Condit-Schultz et al. 2018)
- Composer attribution (McKay et al. 2017b)

Using automated feature extraction software (e.g., jSymbolic), statistical analysis, and machine learning, symbolic music data allows us to study large quantities of music

We need numerous, high-quality symbolic music files!

Existing symbolic music databases

Classical Archives

Musescore

ChoralWiki

The SEILS dataset

Kern Scores

Musedata

The Josquin Research Project

Introducing the SIMSSA* DB

[SIMSSA Database](#)

[About](#)

[Browse](#) ▼

[Upload](#) ▼

[Login](#)



[English \(en\)](#) ▼

SIMSSA

Database

Search

Powerful content based search for scores in symbolic notation with metadata.

Download

Download the files you need.

Upload

Upload your symbolic music files!

*: Single Interface for Music Score Searching and Analysis

Introducing the SIMSSA* DB

Modelling Bibliographic Metadata for Music

Tracking Provenance

Searching Musical Content

Archiving Research

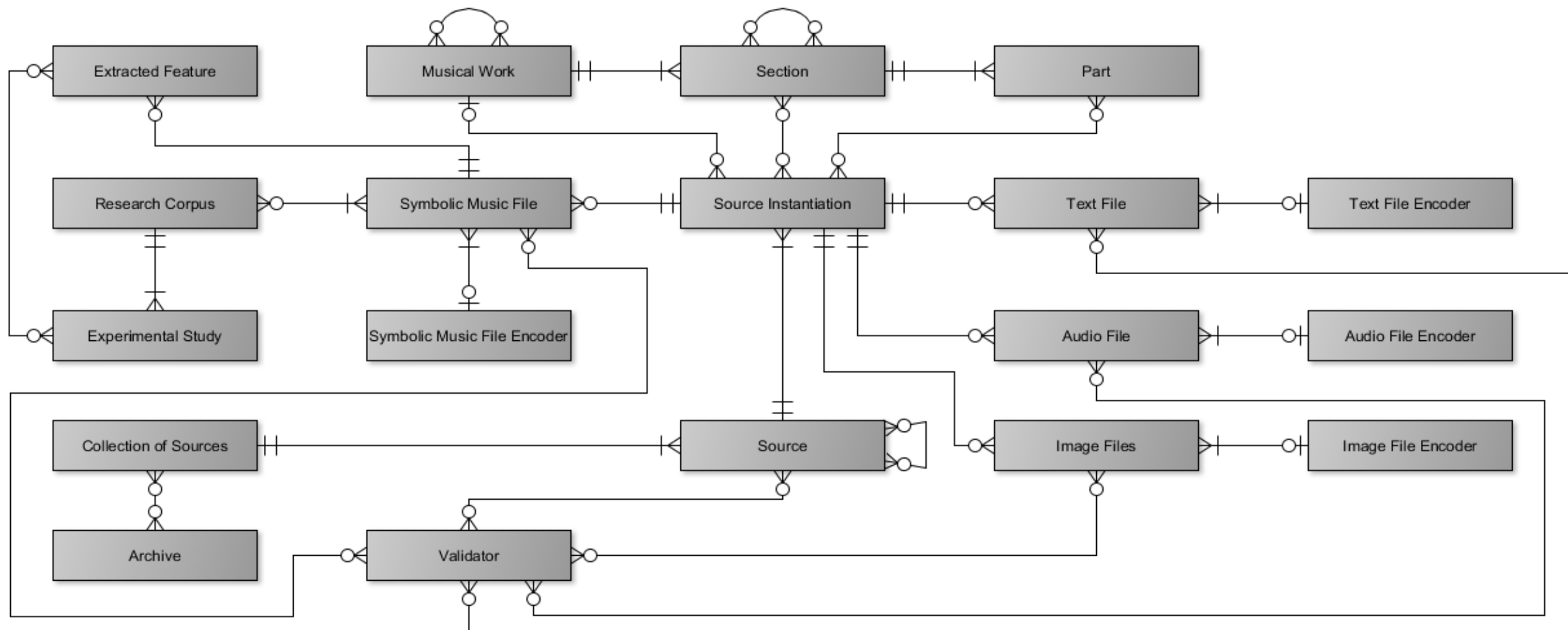
*: Single Interface for Music Score Searching and Analysis

Introducing the SIMSSA DB

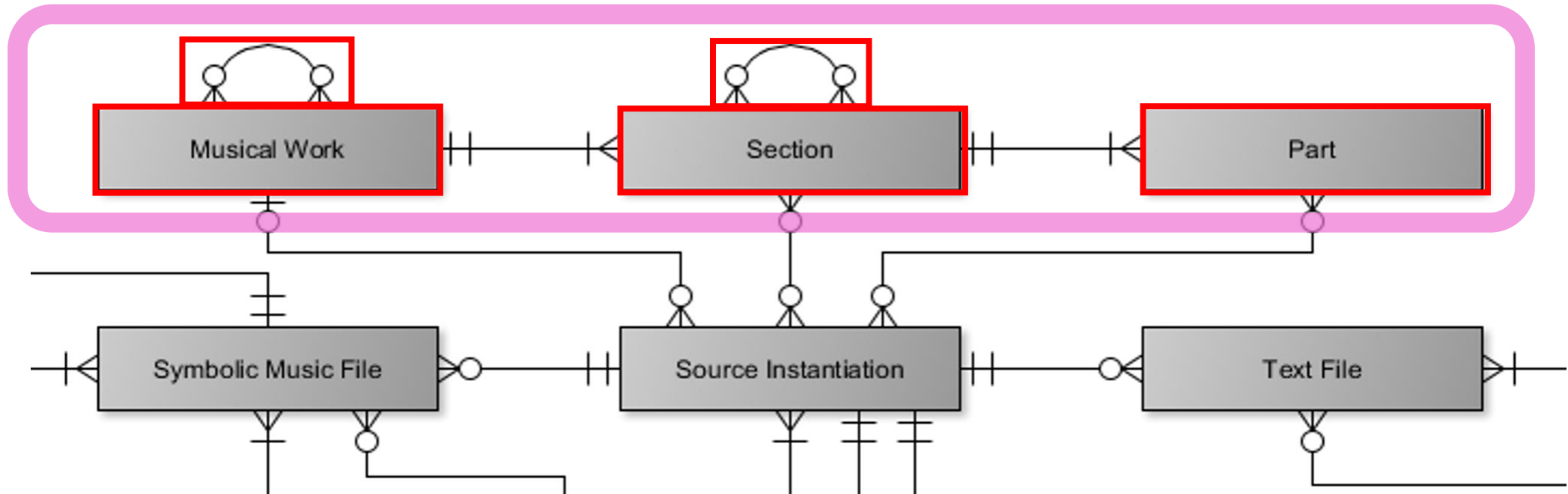
Modelling Bibliographic Metadata for Music

- Allows modelling of complex relationships

Data model for complex relationships (McKay et al. 2017a)



Data model for complex relationships (McKay et al. 2017a)

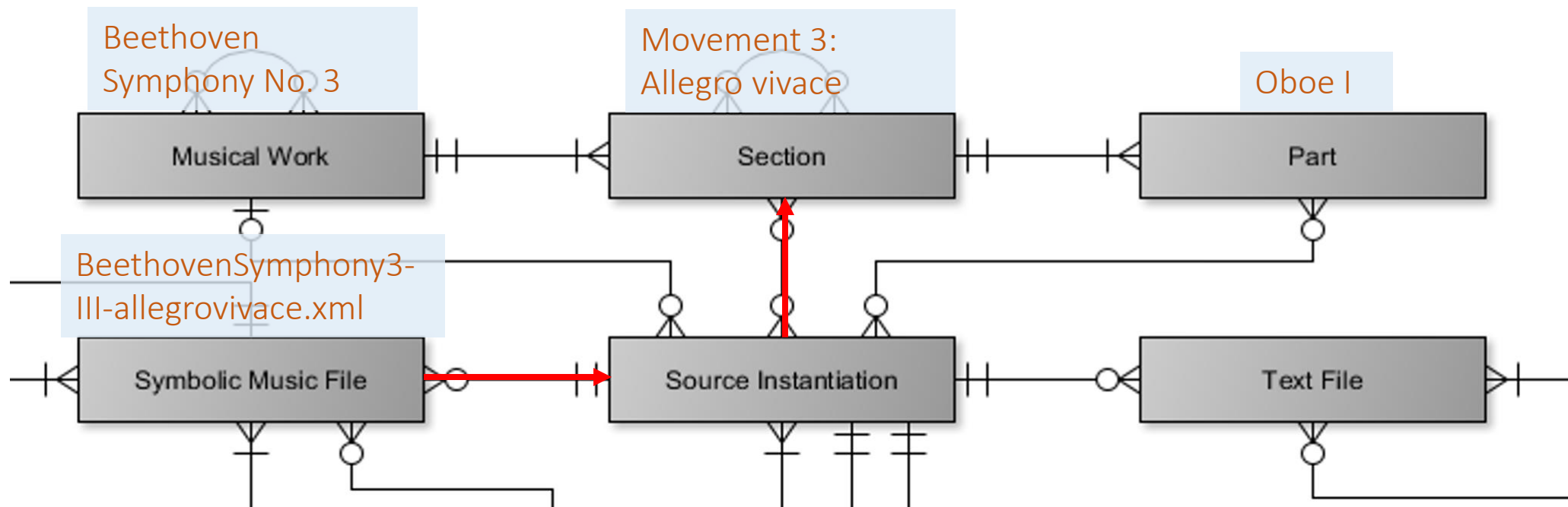


We drew inspirations from FRBR* and IFLA-LRM**

*:Functional Requirement for Bibliographic Records

** : International Federation of Library Associations and Institutions-Library Reference Model

Data model for complex relationships (McKay et al. 2017a)



Uses authority control and controlled vocabulary

To ensure the quality of metadata:

- We use [VIAF](#) (Virtual International Authority File) entries for composers
 - Each entry is provided with linked data URIs (Uniform Resource Identifiers), which unambiguously identifies a resource
- We also use controlled vocabulary for genres and instrumentation

This helps us:

- Guard against typographical errors
- Manage variant spellings of fields
- Increase interoperability
- Allow for the semantic query of data on a larger scale

Auto-suggest

Genre(s)

What type of piece is this? (e.g., song, symphony, motet)

- Madrigal
- Motet
- Create "M"

Bibliographic metadata for music & Upload

Contributions

Who created the work? Use the drop-down menu to choose between different kinds of contributions. Add more contributors with the green button.

Name:

Role

- Composer
- Arranger
- Author of Text
- Transcriber
- Improviser
- Performer

Unknown

Location:

Date:



Tracking provenance

Specifies where the symbolic files are coming from

Where did these files come from?

Please indicate the provenance of the file you are uploading. If the music is part of a larger collection with multiple works (e.g., a complete edition of The Well-Tempered Clavier I, a Fake Book, or an online collection), please include the title of this collection below. If the music you are submitting is from a stand-alone work (e.g., a score of Bach's "Prelude and Fugue No. 2 in C minor"), then "Title" will be the same as the musical work, but you should still fill in the other fields.

Title of Collection *:

Collection URL (if applicable):

Archive/Library where this source can be found (optional):

Portions:

from the original manuscript (grandparent source)

- ...

Search

Example: retrieve all pieces with:

- “Missa ave” in the title (search for title)
- Midi format (search for additional metadata)
- **With vertical tritone (search for musical content)**

Search for title

Filter results by

Search

Composer
 Pierre de La Rue(34)
 des Prez Josquin(20)

Sacred/Secular
 Sacred(54)

Attribution
 Certain(54)

File Format
 .mid(25)
 .xml(8)
 .krn(7)
 .pdf(7)
 .sib(7)

FILTER

54 results for "missa Ave"

Missa ave maris stella
File Type: .krn
Source: JLSDD 156
File: symbolic_music/Josquin_Missa_Ave_maris_stella_-_Agnus_II_mbi0lxx.krn

Missa ave maris stella
File Type: .krn
Source: JLSDD 157
File: symbolic_music/Josquin_Missa_Ave_maris_stella_-_Benedictus_qSiN6i0.krn

Missa ave maris stella
File Type: .krn
Source: JLSDD 158
File: symbolic_music/Josquin_Missa_Ave_maris_stella_-_Qui_venit_BxES0Jn.krn

Please note that features only apply to valid MIDI, Music XML and MEI files, and will exclude file formats from Sibelius, Finale, etc.

- Chords and Vertical Interval Features
- Dynamics Features
- Instrumentation Features
- Melodic Interval Features
- Musical Texture Features
- Pitch Statistics

Search for additional metadata

Filter results by

Search

Composer

- Pierre de La Rue(17)
- des Prez, Josquin(8)

Sacred/Secular

- Sacred(25)

Attribution

- Certain(25)

File Format

- .mid(25)

FILTER

25 results for "missa Ave"

Missa ave maris stella

File Type: .mid
Source: JLSDD 189
File: symbolic_music/Josquin_Missa_Ave_maris_stella_-_Agnus_Il_pgmcyHL.mid

Missa ave maris stella

File Type: .mid
Source: JLSDD 190
File: symbolic_music/Josquin_Missa_Ave_maris_stella_-_Benedictus_8ctKxV0.mid

Missa ave maris stella

File Type: .mid
Source: JLSDD 191
File: symbolic_music/Josquin_Missa_Ave_maris_stella_-_Qui_venit_wrlzJ3H.mid

Missa ave sanctissima maria

File Type: .mid
Source: JLSDD 412
File: symbolic_music/La_Rue_Missa_Ave_Sanctissima_Maria_-_Crucifixus-2-2_KZcurow.mid

Please note that features only apply to valid MIDI, Music XML and MEI files, and will exclude file formats from Sibelius, Finale, etc.

Chords and Vertical Interval Features
Dynamics Features
Instrumentation Features
Melodic Interval Features
Musical Texture Features
Pitch Statistics Features
Rhythm Features
Rhythm and Tempo Features

These musical contents (features) are automatically extracted with jSymbolic (McKay et al. 2018)

Search for musical content

Filter results by

Search

Composer

des Prez Josquin(2)

Sacred/Secular

Sacred(2)

Attribution

Certain(2)

File Format

.mid(2)

FILTER

9 results for "missa Ave"

Missa ave sanctissima maria

File Type: .mid

Source: JLSDD 430

File: symbolic_music/La_Rue_Missa_Ave_Sanctissima_Maria_-_Crucifixus-2-2.mid

Missa Ave maris stella

File Type: .mid

Source: RenComp7 1252

File: symbolic_music/Jos0301c-Missa_Ave_maris_stella-Credo.mid

Missa Ave maris stella

File Type: .mid

Source: RenComp7 1268

File: symbolic_music/Jos0301d-Missa_Ave_maris_stella-Sanctus.mid

Missa Ave maris stella

File Type: .mid

Source: RenComp7 1294

File: symbolic_music/Jos0301a-Missa_Ave_maris_stella-Kyrie.mid

Missa Ave Maria

File Type: .mid

Source: RenComp7 1522

File: symbolic_music/Rue1004d-Missa_Ave_Maria-Sanctus.mid

Please note that features only apply to valid MIDI, Music XML and MEI files, and will exclude file formats from Sibelius, Finale, etc.

Chords and Vertical Interval Features

Variability of Number of Simultaneous Pitches:

0.1725 - 1.499

Most Common Vertical Interval:

0 - 9

Second Most Common Vertical Interval:

0 - 8

Distance Between Two Most Common Vertical Intervals:

1 - 9

Prevalence of Most Common Vertical Interval:

0.1444 - 0.3025

Vertical Tritones:

0.00082 - 0.0819

Vertical Perfect Fourths:

Research archives

The work introduced above encourages high-quality data input

Once we use the database for research, how to archive it for:

- Our own future use
- For others to reproduce our results or conduct their own studies

Research archives

We use Zenodo to include a static dump of the music files as studied

- Zenodo is an open-access platform for “release quality” datasets
- Generate a DOI (Digital Object Identifier) for a stable dataset for citation

SIMSSA DB is great for:

- Finished corpora
- Store and search for metadata and musical contents

Research archive example (Zenodo)

Preview

JLSDD.zip

- Clean Sibelius Templates
 - .DS_Store 6.1 kB
 - Clean Sibelius Templates.zip 295.1 kB
 - bb_template_2_1.sib 32.5 kB
 - sb_template_2_1.sib 32.7 kB
 - sb_template_3_1.sib 32.6 kB
 - ss_template_2_1.sib 32.4 kB
 - st_template_2_1.sib 32.7 kB
 - st_template_3_1.sib 32.7 kB
 - tb_template_2_1.sib 32.7 kB
 - tb_template_3_1.sib 32.7 kB
 - tt_template_2_1.sib 32.6 kB
- Josquin (secure)
 - KRN
 - Josquin Credo De tous biens playne - Et in spiritum.krn 956 Bytes
 - Josquin Missa Ave maris stella - Agnus II.krn 1.3 kB
 - Josquin Missa Ave maris stella - Benedictus.krn 722 Bytes
 - Josquin Missa Ave maris stella - Qui venit.krn 807 Bytes

Files (15.7 MB)

License (for files):

[GNU General Public License v3.0 or later](#)

Versions

Version v1.1 10.5281/zenodo.2635499 Apr 10, 2019

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.2635498](https://doi.org/10.5281/zenodo.2635498). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Share



Cite as

Cumming, Julie E., McKay, Cory, Stuchbery, Jonathan, & Fujinaga, Ichiro. (2019). Josquin La Rue Secure Duos Dataset (JLSDD) (Version v1.1) [Data set]. Zenodo.

Musical Works (700)

« 1 2 3 4 5 6 7 ... 28 »

Ave Maria (1596)

7 Sections

Composers: Palestrina, Giovanni Pierluigi da 1525–1594

Ave Maria ... virgo serena

2 Sections

Composers: Josquin, des Prez 1440–1521

Ave Maria a4

1 Sections

Composers: Victoria, Tomas Luis de 1548–1611

Ave Maris Stella

7 Sections

Composers: Victoria, Tomas Luis de 1548–1611

Conclusion

SIMSSA DB: A high-quality database for symbolic music files search and analysis

- Permits the storage and distribution of a wide range of music in various symbolic formats
- Provides meaningfully structured metadata
- Uses authority control to ensure the quality of metadata
- Offers content-based search
- Emphasizes the provenance of resources
- Facilitates the archiving of research experiments
- Enables users to upload symbolic music

Future Work

Batch download and upload

Add more high-quality symbolic files, for example:

- Bach, Praetorius (Condit-Schultz et al. 2018) and Schutz chorales (721)
 - 721 chorales, in Kern and musicXML formats
- Late Medieval Liturgical Offices by Andrew Hughes (~5900)
 - About 5900 chant melodies
 - Translated and converted into MEI format (Ju and Helsen 2018)

Incorporate linked data

Future Work

Official release, made publicly accessible

Project URL: <http://db.simssa.ca>

Contact us: @simssaproject on Twitter!

We need your contribution!

Enabling Music Search and Analysis: A Database for Symbolic Music Files



Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada

Canada

SIMSSA | Single Interface for Music
Score Searching and Analysis



McGill



Schulich School of Music
École de musique Schulich

DDMAL

DISTRIBUTED DIGITAL MUSIC
ARCHIVES & LIBRARIES LAB



Centre for Interdisciplinary Research
in Music Media and Technology

Fonds de recherche
Société et culture

Québec



References

- Condit-Schultz, Nat, Yaolong Ju, and Ichiro Fujinaga. 2018. “A Flexible Approach to Automated Harmonic Analysis: Multiple Annotations of Chorales by Bach and Prætorius.” In *Proceedings of the 19th International Society for Music Information Retrieval Conference*, 66–73.
- Cumming, Julie E., and Cory McKay. 2018. “Revisiting the Origins of the Italian Madrigal.” Presented at the Medieval and Renaissance Music Conference, Maynooth University, Maynooth, Ireland.
- Cumming, Julie E., and Cory McKay, Jonathan Stuchbery, and Ichiro Fujinaga. 2018. Methodologies for creating symbolic corpora of Western music before 1600. In *Proceedings of the International Society for Music Information Retrieval Conference*, 491–8.
- Garfinkle, David, and Peter Schubert. 2018. “Computer-Assisted Corpus Analysis Finds a Signature Progression in Willaert and Palestrina.” Presented at the Medieval and Renaissance Music Conference, Maynooth University, Maynooth, Ireland.
- Ju, Yaolong, and Kate Helsen, “The LMLO goes MEI: An Exercise in Melodic Encoding Translation” Presented at the *Music Encoding Conference*, 2018.
- McKay, Cory. 2018. “Performing Statistical Musicological Research using jSymbolic and Machine Learning”. Presented at the Anatomy of Polyphonic Music around 1500 International Conference, 34–5.
- McKay, Cory, Andrew Hankinson, Julie Cumming, and Ichiro Fujinaga. 2017a. “A Database Model for Computational Music Research”. Presented at the International Workshop on Digital Libraries for Musicology.
- McKay, Cory, Tristano Tenaglia, Julie Cumming, and Ichiro Fujinaga. 2017b. “Using Statistical Feature Extraction to Distinguish the Styles of Different Composers.” Presented at *the Medieval and Renaissance Music Conference*, Prague, Czech Republic.
- McKay, Cory, Julie Cumming, and Ichiro Fujinaga. 2018. “JSYMBOLIC 2.2: Extracting Features from Symbolic Music for Use in Musicological and MIR Research.” In *Proceedings of the 19th International Society for Music Information Retrieval Conference*, 348–54.