


# SIMSSA DB: Symbolic Music Discovery and Search

 **Emily Hopkins** (emily.hopkins@mcgill.ca),  
Yaolong Ju, Gustavo Polins Pedro, Cory McKay,  
Julie Cumming, Ichiro Fujinaga

**INTRO**  
SIMSSA DB is designed to integrate with the optical music recognition workflow developed as part of the SIMSSA (Single Interface for Music Score Searching and Analysis) Project (simssa.ca). However, the popularity of specialized projects such as the Josquin Research Project, KernScores, and the SEILS dataset implied the need for a general-purpose database of symbolic files for researchers to discover, access, and contribute files in a range of formats for any genre or era of music. In development since summer 2017, we aim for a full public release in summer of 2020.

- DATA MODEL**
- Inspired by DIAMM, RISM, and the IFLA-LRM
  - Describing provenance: users can enter a source (e.g., IMSLP) as well as a chain of “parent” sources (e.g. Breitkopf & Härtel, 1862)
  - Capture information about software and digitization workflows
  - Describe relationships: arrangements, parts, prints, subsections
  - Support for symbolic files with plans to support audio and scanned images for OMR


- SEARCH & DOWNLOAD**
- Content search based on Cory McKay’s jSymbolic, which extracts musical features (e.g. range) on upload (MIDI, MusicXML,s MEI)
  - Users will be able to batch download search results and use these filters to develop research corpora

- METADATA**
- Harvesting metadata whenever possible; e.g. from VIAF (Virtual International Authority File) and Library of Congress (Medium of Performance Thesaurus) reduces errors and makes things easier
  - Users can also add new values (as in RISM’s Muscat)
  - Developing workflows with RISM to contribute these new values to RISM (just as they contribute data back to VIAF)
  - User expertise can be shared and preserved beyond this database

- LINKED DATA**
- Building in linked-data compatibility using external resources that assign URIs (e.g. VIAF, Library of Congress Linked Data Service)
  - Unite variant spellings, extra names, and different dates with URIs
  - Enables semantic queries
  - Future work: linked data quads to track metadata provenance
    - Triple: Composer—composed—symphony
    - Quad: Composer—composed—symphony—according to—historian

- RESEARCH ARCHIVING & REDISTRIBUTION**
- Stable data sets required for machine learning and repeatability
    - Problem: Users can edit their own contributions, so files may not remain identical over time
  - Integrate with external repositories (e.g. Zenodo, Dataverse) to store complete, static version of dataset and link back to DB
  - Include licensing metadata for proper redistribution and attribution

SIMSSA DB improves discovery and access for symbolic music files by combining metadata and musical content search.

 Take a picture for link to poster & paper



## SEARCH

Search

amor

Genre (Type of Work)

Frotolia(1)

Madrigal(8)

Genre (Style)

Renaissance(9)

Composer

Festa, Sebastiano(4)

Pisano, Bernardo(4)

Tromboncino, Bartolomeo(1)

Instrument/Voice

Voice(9)

Sacred or Secular

Sacred(9)

File Format

mid(8)

sibelius(8)

xml(8)

FILTER

9 results for "amor"

11 files match the feature search parameters. Only highlighted files match all search parameters.

Amore amor quando io speravo

Composer(s): Pisano, Bernardo 1490–1548

Section(s):

Amore amor quando io speravo

Genres (Type of Work): Madrigal

Genres (Style): Renaissance

Collection(s) of Sources: Florence, Italy, Biblioteca Nazionale Centrale, MS Magliabechi XIX.164-167

File(s) Holding Complete Musical Work:

sibelius

mid

xml

File(s) Holding an Individual Section:

Hor vedi Amore che giovinetta donna

Composer(s): Pisano, Bernardo 1490–1548

Amor se vuoi ch'io torni al gioglio anticho

Composer(s): Festa, Sebastiano 1495–1524

Chords and Vertical Interval Features

Dynamics Features

Instrumentation Features

Melodic Interval Features

Musical Texture Features

Average Number of Independent Voices:

1 - 3.804

Contrary Motion:

0.1578 - 0.1924

Importance of Loudest Voice:

0 - 1

Maximum Number of Independent Voices:

1 - 4

Oblique Motion:

0.5066 - 0.6374

Please note that features only apply to valid MIDI, Music XML and MEI files, and will exclude file formats from Sibelius, Finale, etc. For an explanation of all features, please consult the [jSymbolic Manual](#).

## UPLOAD

Title

What is the title of the work? Click the green button to add variant titles or nicknames. Please include opus number or catalogue numbers if applicable (e.g., Op. 55, D960, BWV 202)

Title \*

+ -

e.g. Symphony No.3 Op. 55

Variant Titles:

+ -

e.g. Eroica

Sections:

+ -

e.g. I. Allegro con brio

Contributions

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

Users can enter variant titles or nicknames, and can choose from different roles for contributors, such as composer, arranger, author of text, transcriber, improviser, and performer.

Genre(s)

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

× symphony

What style is this piece? (e.g., classical, jazz)

× classical

Sacred Or Secular:

Not Applicable

Medium of Performance

Please enter the instruments or voices below.

Instruments:

SUBMIT

Genre as in “Symphony” or genre as in “Classical”? We have separated the most common understandings of the term into two fields. Medium of Performance auto-completes from the LC Medium of Performance Thesaurus.

## ACKNOWLEDGEMENTS

This research is part of the SIMSSA Project, which is funded by a Partnership Grant from the Social Sciences and Humanities Research Council of Canada (SSHRC 895-2013-1012) and also funded by the Fonds de recherche du Québec-Société et culture (FRQSC).