## **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, 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



**W** McGill

Social Sciences and Humanitie Research Council of Canada

Centre for Interdisciplinary Research in Music Media and Technology



compute | calcul canada



Schulich School of Music École de musique Schulic

sciences humaines du Canada

Canada



canada



Search amor Genre (Type of Work)	9 results for " <b>amor</b> " 11 files match the feature search parameters. Only <mark>highlighted</mark> files match all search parameters.		Please note that features only appl to valid MIDI, Music XML and MEI files, and will exclude file formats from Sibelius, Finale, etc. For an	
<ul> <li>Frottola(1)</li> <li>Madrigal(8)</li> </ul>	Amore amor quando io speravo	+	explanation of all features, please consult the jSymbolic Manual.	
Genre (Style) • Renaissance(9) Composer • Festa, Sebastiano(4) • Pisano, Bernardo(4) • Tromboncino, Bartolomeo(1) Instrument/Voice	Composer(s): Pisano, Bernardo 14901548 Section(s):		Chords and Vertical Interval Features Dynamics Features Instrumentation Features Melodic Interval Features Musical Texture Features Average Number of Independent Voices:	
	• Amore amor quando lo speravo Genres (Type of Work): Madrigal			
	Genres (Style): Renaissance			
Voice(9)  Sacred or Secular      Sacred(9)	Collection(s) of Sources: Florence, Italy, Biblioteca Nazionale Centrale, MS Magliabechi XIX.164-167			
<pre>File Format</pre>	File(s) Holding Complete Musical Work: • sibelius			
	o <mark>xml</mark>		Contrary Motion:	_
	File(s) Holding an Individual Section:		0.1578 - 0.1924	
	Hor vedi Amore che giovinetta donna		0 - 1	
	Composer(s): Pisano, Bernardo 14901548	+	Maximum Number of Voices:	Independent
	Amor se vuoi ch'io torni al giogho antio	ho	1 - 4	
	Composer(s): Festa, Sebastiano 14951524	+	Oblique Motion:	

A keyword search for "amor" combined with a search for contrary motion in the features menu. Here, metadata facets are shown on the left, results in the middle, and feature sliders on the right. Feature value slider limits are calculated dynamically according to search results.

<b>Title</b> What is the title of the work? Click t applicable (e.g., Op. 55, D960, BWV
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
Users can enter varian roles for contributors, improvisor, and perfor
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	$\mathbf{v}$
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).



#### UPLOAD

reen button to add variant titles or nicknames. Please include opus number or catalogue numbers if


o-down menu to choose between different kinds of contributions. Add more contributors with the green button.

it titles or nicknames, and can choose from different such as composer, arranger, author of text, transcriber, rmer.