



SIMSSA DB: A Collaborative Musicological Research Database

Cory McKay

Marianopolis College, Canada

*Digital Musicology Study Group @ 2019 Digital Humanities Conference
Utrecht, Netherlands*

Introduction to the SIMSSA DB

- **Collaborative database infrastructure** for holding **symbolic music files**
 - e.g. Music XML, MEI, MIDI, Sibelius, Finale, etc.
- **Populated by:**
 - **Now:** Datasets we have constructed
 - **Soon:** Datasets musicologists, libraries, etc. have constructed for their own purposes
 - **Long-term:** Auto-population via (verified) OMR
- **Focused (for now) on early music**
- **Current status:**
 - Undergoing **internal user-testing** and **consultation** with domain experts

Data quality

- Focusing on **high-quality** data
- Quality of individual documents especially important in early music:
 - **Few extant sources**, so limited training/testing data will ever be available
 - Individual **detail** very important to domain experts
 - e.g. a single cadence or even a single note
- There is a need for a central infrastructure making such information **accessible** and **discoverable** to musicologists

Search

- **Content-based** search based on global features
 - Automatically extracted with **jSymbolic**
 - Extracts almost 1500 feature values
 - Complete feature sets can also be downloaded for direct use in **machine learning** or **statistical analysis**
- Global and faceted metadata search
 - Contributor
 - Composer, arranger, author of text, transcriber, etc.
 - Sacred, secular, etc.
 - Instruments / voices
 - Genre / type of work
 - e.g. madrigal, motet, etc.
 - Etc., etc., etc.

Provenance

- Each symbolic file linked to direct parent source(s)
 - Digital or physical
- Sources in turn can have their own parents: **chains of provenance**
 - e.g. MEI < printed score < copied manuscript < composer's manuscript
- Sources can also be linked to **collections of sources and archives**

Authority control and controlled vocabularies

- Should be able to automatically match **differing but equivalent metadata**
 - e.g. “Stravinsky” and “Stravinski”
 - e.g. “Le Sacre du printemps” and “The Rite of Spring”
- The SIMSSA DB uses **authority control** and **cataloguing standards** to reduce ambiguity and redundancy (and increase consistency) as much as possible
 - The DB is currently using **VIAF** authority files
 - Populates fields with **URIs** and uses **linked open data** practices when possible
- Metadata tags are **auto-suggested** as users type based on these authority files
 - e.g. composer name, genre name, etc.

Flexible data model

- Music can be divided into **abstract works, sections and parts**
 - Symbolic files sometimes contain whole pieces, and sometimes only parts of pieces
- Symbolic music files are separate but linked to these abstract entities
- Makes it possible to **keep track of complex musical relationships**
 - e.g. a movement of one mass might be reused in another mass
 - e.g. an orchestral score and a keyboard reduction of it have different parts, but they are also different versions of the same abstract work

Archiving research dataset

- Want to promote **repeatability of research**
- Specific datasets (and their extracted features) used in specific studies can be archived on the well-established **Zenodo** open research repository
- These can then be linked to directly from the SIMSSA DB

Long-term goals

- Optical music recognition (**OMR**) integration
- **Melodic** and **harmonic** queries
 - As distinct from feature-based queries
 - David Garfinkle and Yaolong Ju have started work on this
- Store **linked multimodal data** (not just symbolic music files)
 - Images of scores or manuscripts
 - Musical texts
 - Audio files

Consultation

- We would be very grateful for any ideas, wants or needs you may have:
 - Is there anything you would especially like the SIMSSA DB to be able to do?
 - Do you have any music you would like us to host and make discoverable?

Thanks for your attention!

- **E-mail:** cory.mckay@mail.mcgill.ca
- **SIMSSA DB:** <http://db.simssa.ca>
- **Contributors:** Emily Hopkins, Gustavo Polins Pedro, Yaolong Ju, Andrew Kam, Julie Cumming, Ichiro Fujinaga



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

