Towards Analysing Big Music Data: Progress on the DML Research Project

Samer Abdallah, Mathieu Barthet, Emmanouil Benetos, Stephen Cottrell, Simon Dixon, Jason Dykes, Nicolas Gold, Alexander Kachkaev, Mahendra Mahey, Mark D. Plumbley, Dan Tidhar, Adam Tovell, Tillman Weyde, Daniel Wolff (*)

(*) authors in alphabetical order

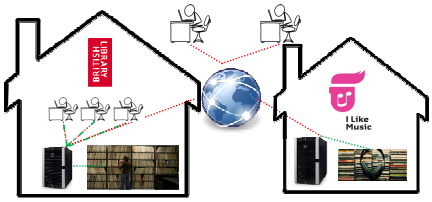
http://dml.city.ac.uk/

Joining Musicology and Data Science

- Systematic musicology research has developed as **"data oriented empirical research"**, which benefits from computational methods. However, this research has so far been limited to relatively small datasets, because of technological and legal limitations.
- In parallel, researchers in Music Information Retrieval (**MIR**) have started to **explore large datasets**, particularly for commercial **recommendation** and playlisting systems (e.g. The Echo Nest, Spotify).
- The **'Digital Music Lab – Analysing Big Music Data' (DML)** project supports music research by bridging the gap between musicology and MIR and enabling access to large music collections and powerful analysis and visualization tools.

An Infrastructure for Large-Scale Music Analysis

- Software **infrastructure** for analysing **large and heterogeneous** music collections.
- Built on **strongly parallelisable software architecture**
- **Short response times** enable exploratory research
- **Integration with existing tools** for music research
- **Accumulate derived data** and enable the deployment of new tools
- **Share** intermediate and final results as **open linked data**



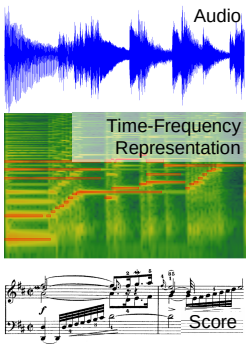
hadoop

SONIC VISUALISER
QUEEN MARY UNIVERSITY OF LONDON

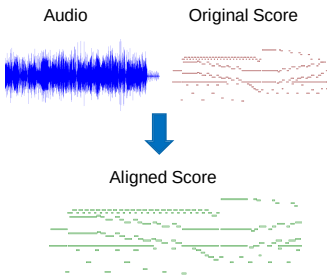
VAMP PLUGINS

Unlocking MIR Methods for Big Data

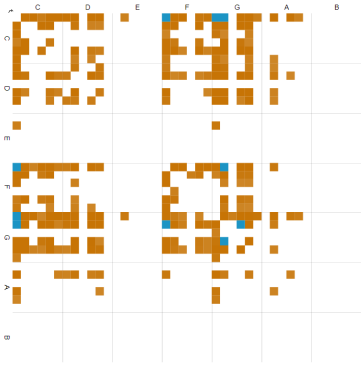
- Use of MIR methods for large-scale quantitative research
- Scalable tools for analysing music **audio**, **scores** and **metadata**
- Combination of **state-of-the-art music analysis** on audio and symbolic data
- Enable **intelligent collection-level analysis**



Automatic music transcription



Audio-to-score Alignment



Chord sequence visualisations

Use and Produce Big Music Datasets

- Access to big datasets: **British Library (>3M tracks)** and **I like Music (1M tracks)**
- Often: Access to audio data restricted by copyright
- **Derived data** can be made freely available and produced on demand for research purposes
- Automatic **transcription and alignment** with scores
- **Annotation** and linking of audio files with metadata and external resources
- Use of **open standards** such as the Music Ontology

Acknowledgements

Arts & Humanities Research Council

The DML consortium includes the Department of Computer Science and the Department of Music at City University London, the Centre for Digital Music at Queen Mary University of London, the Department of Computer Science at University College London, and the British Library Labs. The £560k Big Data Research project is funded by the Digital Transformations Theme of the Arts and Humanities Research Council (AHRC Grant AH/L01016X/1, 2014-2015).