

# The DML Research Project: Digital Music Lab – Analysing Big Music Data

Mathieu Barthet, Emmanouil Benetos, Stephen Cottrell, Simon Dixon, Jason Dykes, Nicolas Gold, Mahendra Mahey, Mark Plumbley, Dan Tidhar, Tillman Weyde, Daniel Wolff (\*)

(\*) authors in alphabetical order

dml@city.ac.uk

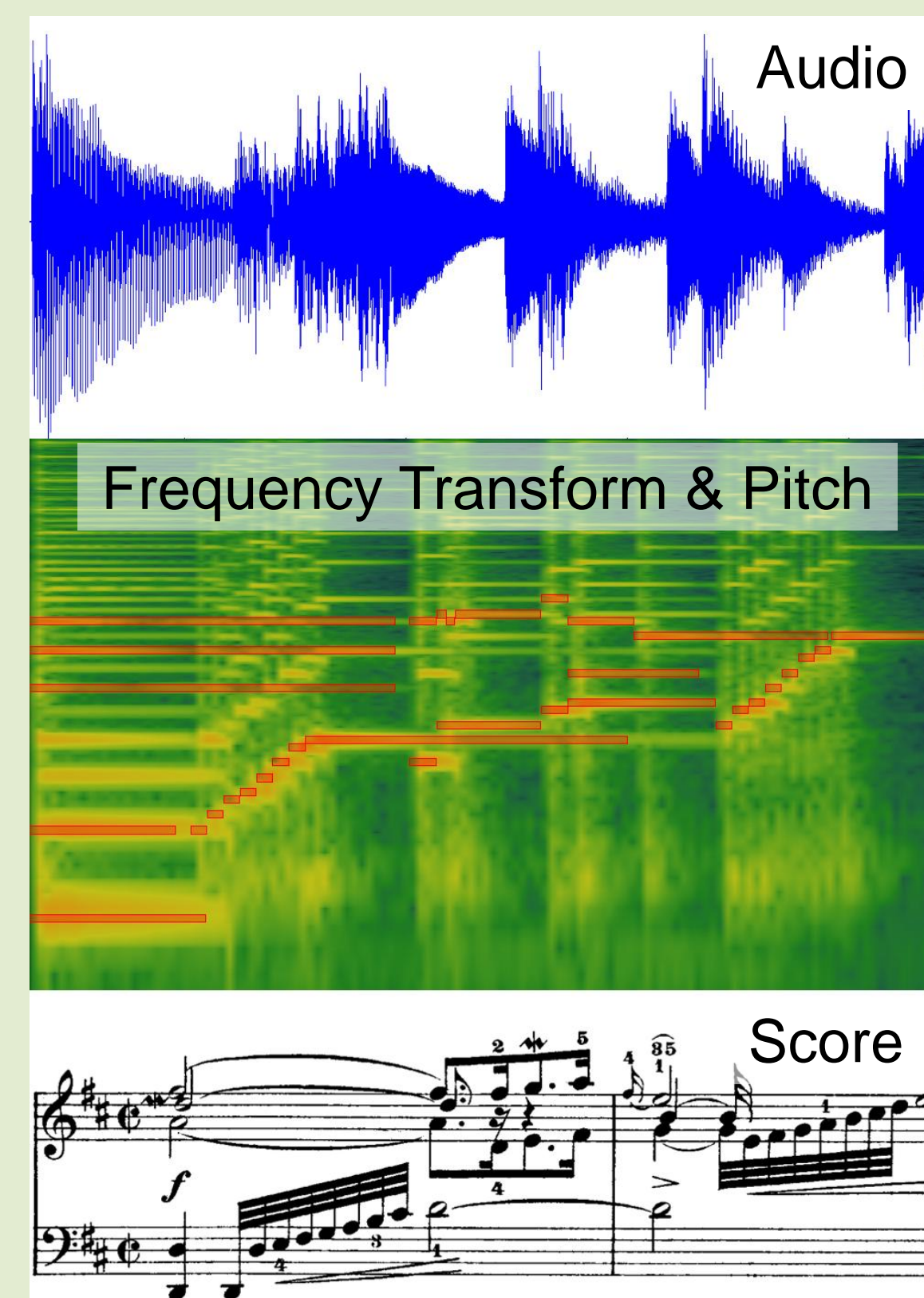


## Joining the Forces of MIR and Musicology

- Music research, particularly in systematic musicology, ethnomusicology, or music psychology, has developed as “**data oriented empirical research**”, which benefits from computational methods. In ethnomusicology there has been a recent growing interest in computational musicology. Traditional analysis methods designed for **symbolic music representations** can be complemented with analysis on **audio data**. However, this music research has so far been limited to relatively small datasets, because of technological and legal limitations.
- On the other hand, Researchers in Music Information Retrieval (MIR) have started to **explore large datasets**, particularly in commercial **recommendation** and playlisting systems (e.g. The Echo Nest, Spotify).
- The Digital Music Lab will support music research by **bridging the gap to MIR** and enabling access to large music collections and powerful analysis and visualization tools.

## Unlocking MIR Methods for Big Data

- Use of MIR methods for large scale exploration and 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**

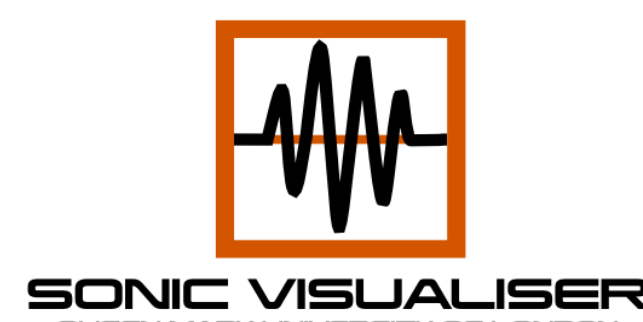
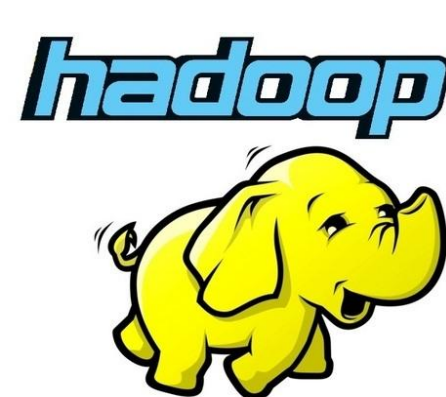


## Example Data Types and Analysis



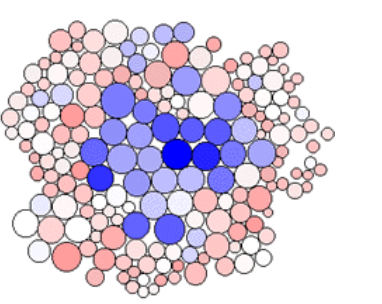
## An Infrastructure for Large-Scale Music Analysis

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



## Large-Scale Computational Musicology

- Highly **interactive visual interfaces**
- Enable study of **characteristics on collection level**
- Musical **timing and structure** in **piano, folk, and world music**
- Relating **similarities** on audio and metadata.



## Acknowledgements



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 project is funded by the Digital Transformations Theme of the Arts and Humanities Research Council (AHRC Grant AH/L01016X/1, 2014-2015).

## Use and Produce Big Music Datasets

- Access to big datasets:  
**British Library (>3M)** and **I like Music (1M)**
- Often: Access to audio data restricted by copyright
- Derived data** can be made freely available and produced on demand
- 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

