

# Digital Music Lab

An AHRC Digital Transformations Project

## The DML datasets

Dan Tidhar



CITY UNIVERSITY  
LONDON



Queen Mary  
University of London



# Dataset overview

- ▶ **British Library (~50K)**

- ▶ World and traditional
- ▶ Classical



- ▶ **CHARM (~5K)**

- ▶ Classical

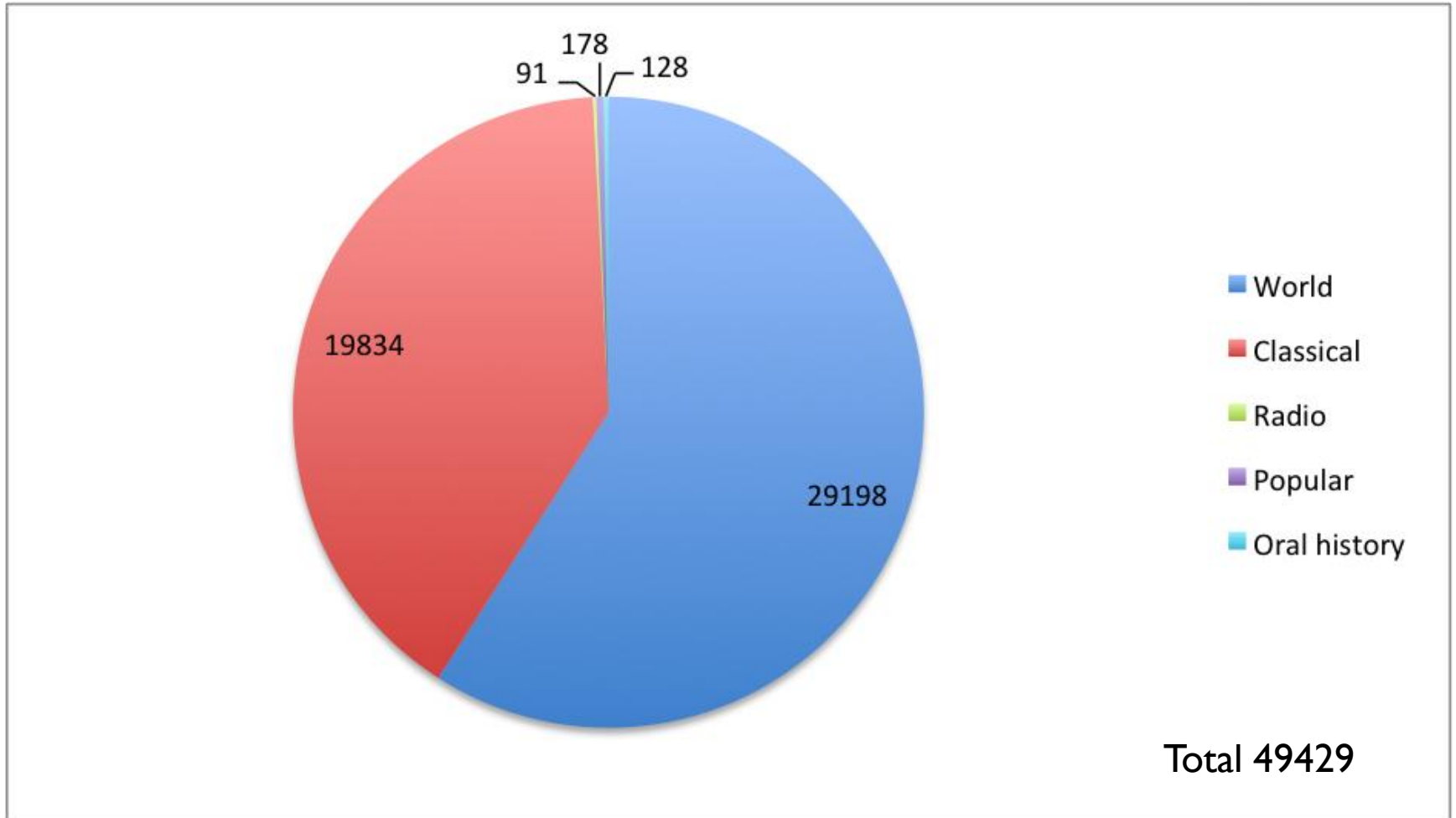


- ▶ **I Like Music (> 1M)**

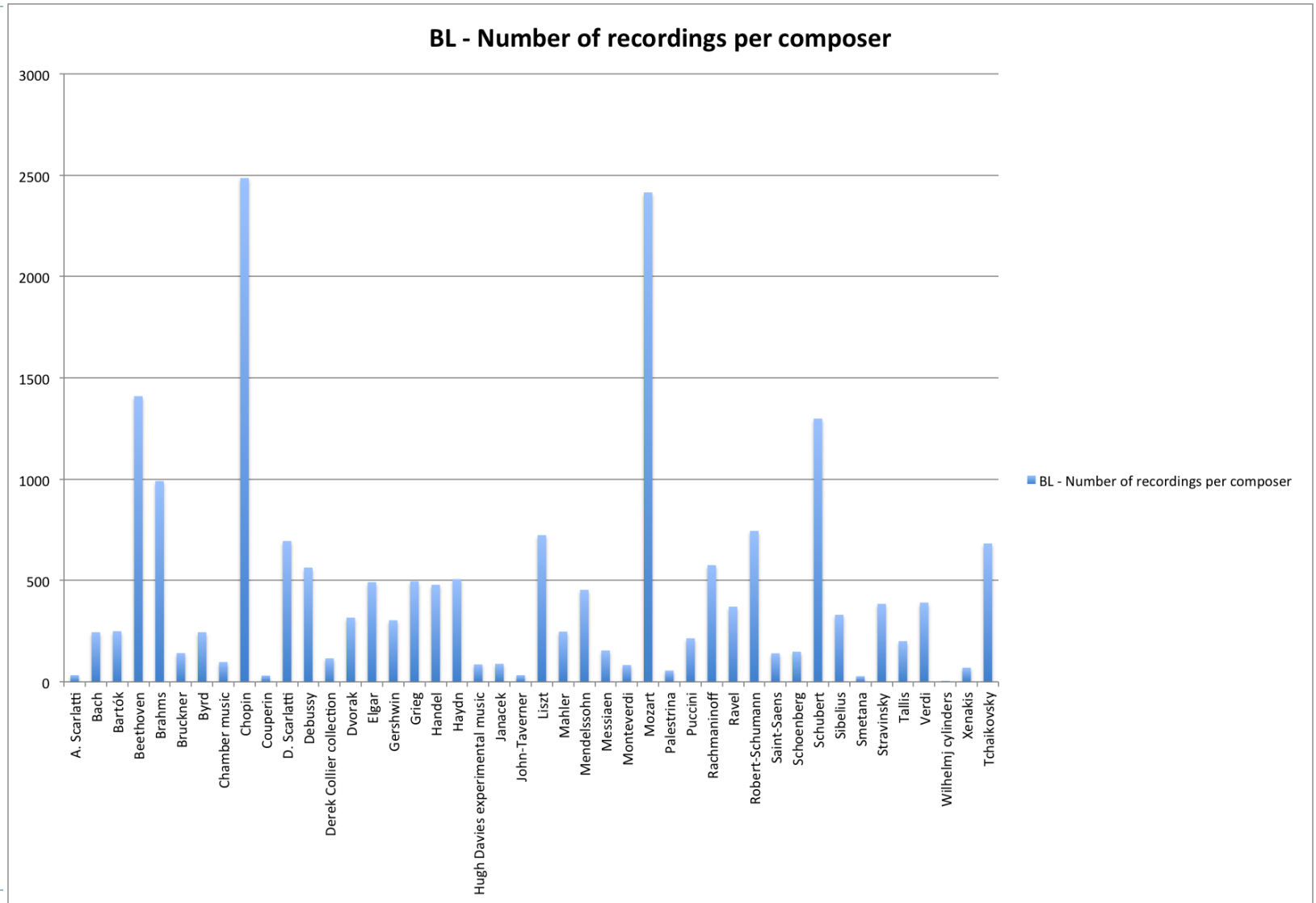
- ▶ Popular
- ▶ Classical
- ▶ World



# British Library

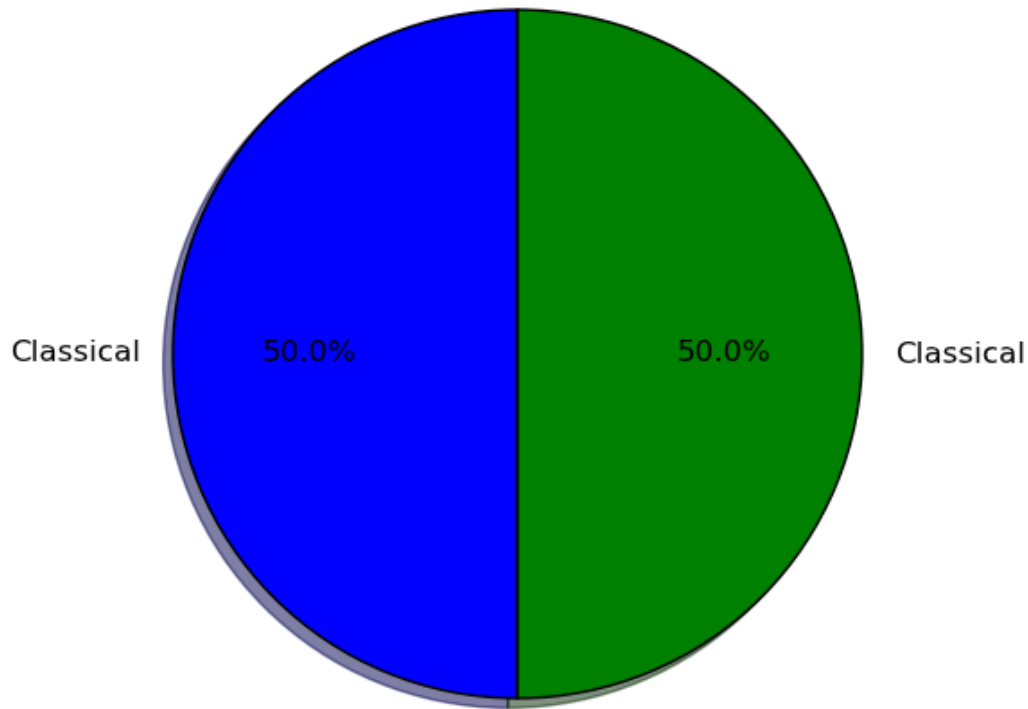


# British Library – Classical Part



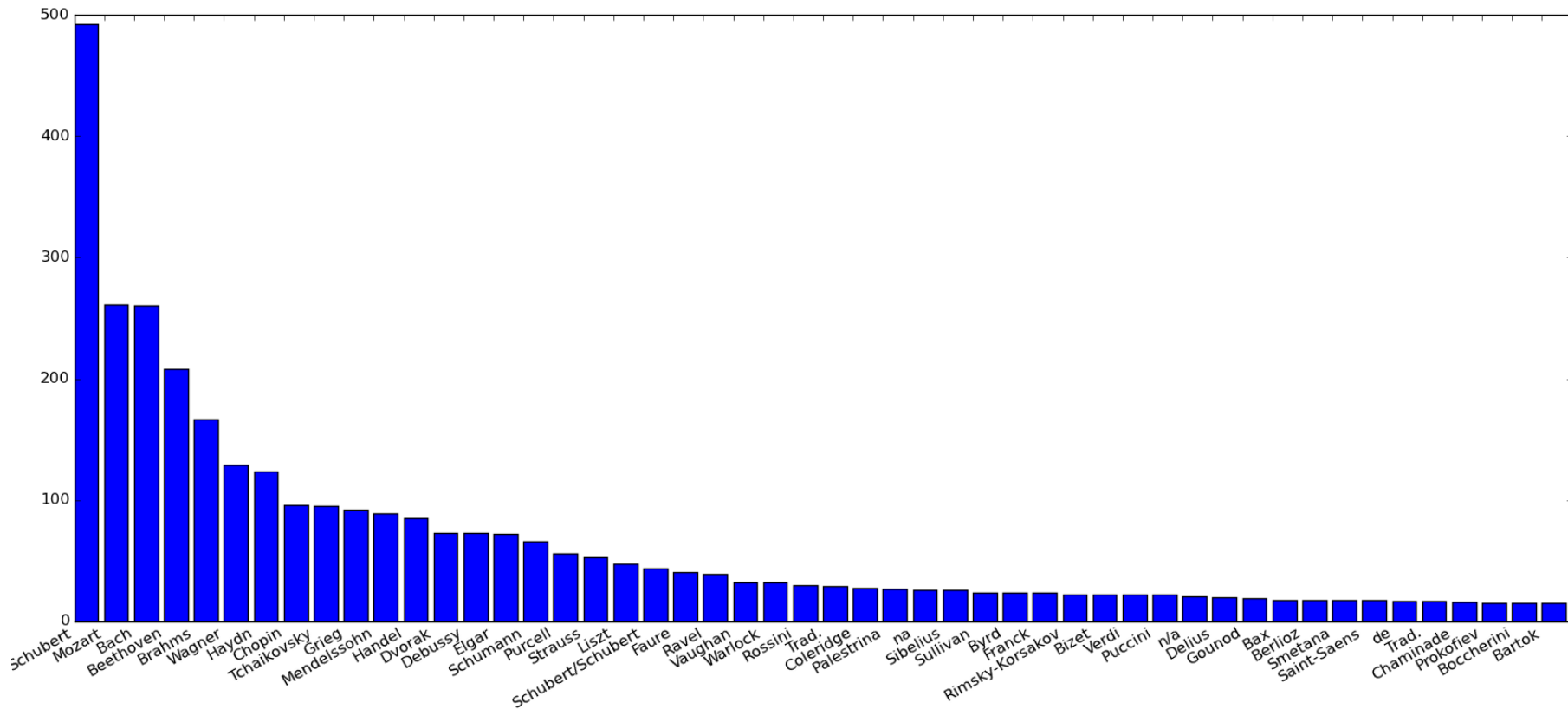


Charm genre distribution



# CHARM - Composers

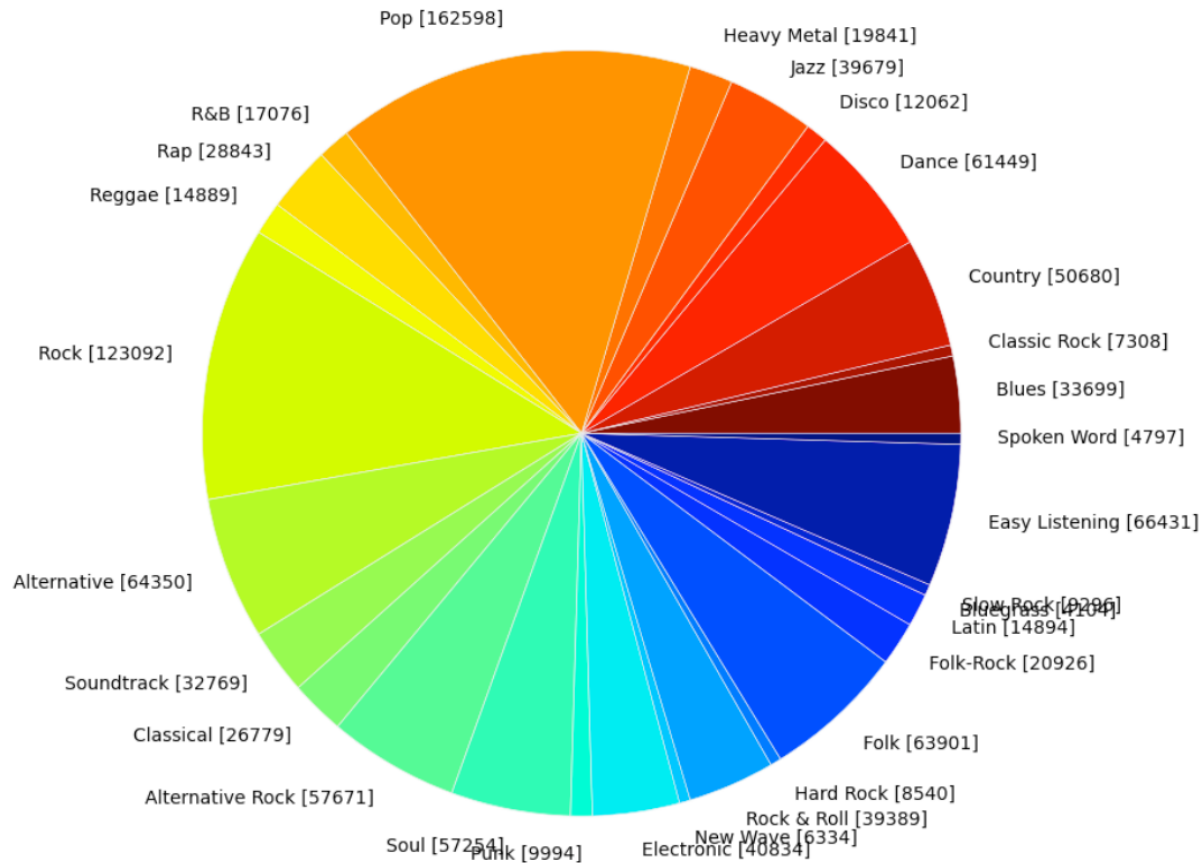
---



# I Like Music

## ILM Commercial Music Across Genres

- Alternative
- Alternative Rock
- Bluegrass
- Blues
- Celtic
- Classic Rock
- Classical
- Country
- Dance
- Disco
- Easy Listening
- Electronic
- Folk
- Folk-Rock
- Gospel
- Grunge
- Hard Rock
- Heavy Metal
- Jazz
- Jazz-Funk
- Latin
- New Age
- New Wave
- Other
- Pop
- Punk
- R&B
- Rap
- Reggae
- Rock
- Rock&Roll
- Slow Rock
- Soul
- Soundtrack
- Southern Rock
- Spoken Word
- World Music



I Like Music

# Dataset Demarcation

Wolff, D., et al. **Incremental dataset definition for large scale musicological research.** In: DLFM 2014

- ▶ Tags (e.g. Genre, style, period) are key to collection exploration
- ▶ **Cost** renders **annotation** of large datasets **difficult**
  - ▶ **Automatic Classification needed** (e.g. instrumentation)
  - ▶ But: training of **classifiers** requires ground truth
- ▶ Incremental annotation and training of automatic classifiers
  - ▶ **Only** the **weakest classifications** presented to human annotators
- ▶ Experiments on CHARM:
  - ▶ 87% reduction of data input for < 2.38% classification error





# Collaborative Tagging Interface



Online game with  
CASIMIR framework  
(Wolff et al 2014)