

Art Music Curation in Australia Dataset

This dataset contains comprehensive logs, comprising 982,182 records of music presented on three platforms:

- **ABC Classic FM radio broadcasts** (1996-2018)
- **Concert events catalogued by the Australian Music Centre** (2009-2018)
- **Classical genre Spotify playlists entries** (Jan - May 2018)

The raw data has been harvested, transformed and classified at the level of the individual composer whose music was featured in each discrete broadcast or presentation.

Composer attributes recorded for each content item include:

- Name
- Year of birth
- Gender
- Place of activity (including state, country and continent)
- Spotify artist identifier
- Spotify popularity score (as of August 2017)
- iTunes artist identifier

Each discrete content item includes details on:

- The context in which the item was featured (e.g. the radio program, concert program, or Spotify playlist)
- The sequential order in which it was presented
- The date and time of presentation
- The duration of the item

The concert data also includes the following information on each of the performers:

- Name
- Gender
- Year of birth
- Size of the performing ensemble
- Repertoire focus of the performing ensemble

Data Collection

ABC Classic FM

916,906 broadcast records were harvested from ABC Classic FM's published music listings from 1 January 1996 through to 1 November 2018. Each record comprises an individual broadcast of a work on the network.

The listings were compiled from various unstructured HTML and semi-structured XML formats which were previously published on the ABC Classic FM website. The data was parsed, transformed and supplemented with additional composer-level variables.

Concert Event Data

21,939 records were harvested from the Australian Music Centre's (AMC) calendar of events. Each record comprises an individual composer being featured in a public concert event. Each event catalogued by the AMC includes at least one work by an Australian composer. The dataset also includes records on each of the performers responsible for presenting the works in each event.

The data was compiled from an extract provided from the AMC's internal database. The data was transformed and supplemented with additional composer-level variables.

Spotify Playlist Data

43,337 records were harvested from Spotify playlists. Each record comprises an individual work featured on a playlist available on Spotify's digital streaming service. The data was harvested from 34 classical genre playlists over the date range of 1 January 2018 to 31 May 2018. A list of the playlists and the criteria for their inclusion is [listed separately](#).

The data was compiled using the Spotify API. The data was transformed and supplemented with additional composer-level variables.

Data Files

The data is provided in this repository as two tab-separated, gzip compressed, UTF8 text files:

- [amca_dataset_content-items_20200308.zip](#)
- [amca_dataset_performers_20200308.zip](#)

A header row in each file includes field names corresponding to the data dictionaries detailed below.

#1 Content Items

Each row represents a content item featured on one of the three platforms

Filename: [amca_dataset_content-items_20200308.zip](#)

Records: 982,182

Field	Type	Description
platform	Text	The media platform for the content item: one of 'broadcast', 'playlist' or 'concert'.
id	Integer	An integer which uniquely identifies each content item within other content items from the same platform.
sequence	Integer	The order in which the current item appears in its parent context. For radio broadcasts this number increments for each item in the entire dataset. For playlists entries, this number reflects the

Field	Type	Description
		position of the record in the playlist in which it was featured. Sequence data is not available for concert records.
context	Text	The context in which the content item was featured. For radio broadcasts, this reflects the radio program. For playlists items, it is the name of the playlist. For concerts it is a hash which reflects an individual concert program.
date	Date	The date on which the context item was presented. For radio broadcasts, this reflects the broadcast date. For playlist entries it is the date the version of the playlist was published. For concerts it is the year in which the concert program was presented.
time	Time	The time of day at which the context item was presented.
duration	Integer	The duration of the content item in seconds.
raw_composer	Text	The name of the composer of the content item, as it appeared in the raw data source and before matching to a canonical entry.
composer	Text	The canonical name of the composer of the content item, after matching to authoritative data
yob	Integer	Year of birth of the composer of the content item.
gender	Text	Gender of the composer of the content item.
place	Text	Place where the composer of the content item was principally active.
country	Text	Maps the specific location recorded in the place field to a country.
continent	Text	Maps the specific location recorded in the place field to a continent.
state	Text	Maps the specific location recorded in the place field to an Australian state or territory.
spotify_artist_id	Text	The Spotify artist identifier for the composer of the content item.

Field	Type	Description
spotify_popularity	Integer	The popularity score for the composer of the content item, as obtained from Spotify's API in August 2017.
spotify_followers	Integer	The number of followers for the composer of the content item, as obtained from Spotify's API in August 2017.
itunes_artist_id	Integer	The iTunes artist identifier for the composer of the content item.
spotify_track_id	Text	The Spotify track identify for the content item work. Only available for playlist content items.

#2 Performers

Each row represents a content item featured on one of the three platforms

Filename: [amca_dataset_performers_20200308.zip](#)

Records: 34,908

Field	Type	Description
content_item_platform	Text	References the platform field of the Content Items file. Together with the id field, this defines a foreign key relationship.
content_item_id	Integer	References the id field of the Content Items file.
name	Text	The name of the performer.
gender	Text	The gender of the performer.
yob	Integer	The year of birth of the performer.
performer_size	Text	A classification of the size of the performer entity. One of: Accompanist, Composer-as-performer, Soloist, Small [2-4 players], Medium [5-11 players], Large [12+ players].
performer_repertoire_focus	Text	A classification of the general repertoire focus of the performer entity. One of: Heritage, Mixed, Contemporary.

Spotify Playlists

The Spotify playlists included in the dataset are based on a selection of classical genre playlists published between 1 January 2018 and 31 May 2018.

Playlists were selected by querying Spotify's API for classical genre playlists, with the additional criteria that each playlist must:

- have been updated at least three times in the five month data collection period
- feature music by more than a single composer

Playlist Name	Publisher	Category	# Playlisted Works	# Followers ¹
Best of Australian Classical Music	ABC Classic	Specialist	966	84
Classical Music for Concentration and Study	ABC Classic	Task	850	727
Classical Music for Mindfulness	ABC Classic	Task	728	301
Classical Sleep	ABC Classic	Task	742	269
New on ABC Classic	ABC Classic	General	502	138
Swoon: Music for Sheer Relaxation	ABC Classic	General	1,019	2,186
The Best Classical Music of All Time	ABC Classic	General	802	328
The Greatest Music in the Movies	ABC Classic	Specialist	904	1,155
Women in Music	ABC Classic	Specialist	388	65
BBC Classical (BBC Radio 3)	BBC Music Playlists	General	2,324	2,355

Playlist Name	Publisher	Category	# Playlisted Works	# Followers¹
Chillout Piano	Deutsche Grammophon	Mood	1,100	2,014
Classical New Releases	Deutsche Grammophon	General	1,658	3,650
Famous Classical Works	Deutsche Grammophon	General	201	3544
Great Voices in Classical Music - Choir Edition	Deutsche Grammophon	Instrument	244	311
Piano Masters	Deutsche Grammophon	Instrument	900	10,601
Piano Melancholy	Deutsche Grammophon	Mood	1,231	2875
Reflections	Deutsche Grammophon	General	300	4288
Piano Chill	Filtr Canada	Mood	1,958	64431
Soundtrack for Study	Filtr Canada	Task	2,061	359,021
Classical Music for Reading	Filtr Legacy Sweden	Task	1,200	131,514
Relaxing Classical	Filtr UK	Mood	1,315	231,460
Study With Classical	Filtr US	Task	1,300	5,440
Focus Classical & Electronic	HITS	Task	853	21,379
New & Now Classical	NAXOS	General	270	1,112
NMC curates: new contemporary classical releases	NMC Recordings	General	216	53

Playlist Name	Publisher	Category	# Playlisted Works	# Followers¹
Classical New Releases: Spotify Picks	Spotify	General	1,065	416,546
Mørketid	Spotify	Mood	4,760	36,313
Top Latin Classical	Spotify	Specialist	3,400	19,874
Brain Fuel: Music for Studying	unCLASSIFIED	Task	2,486	55,444
Chill Piano	unCLASSIFIED	Mood	2,525	1,389
Listen to This: New Classical Music	unCLASSIFIED	General	1,440	1,846
Movie Music	unCLASSIFIED	Specialist	804	28
Need to Know: Romantics	unCLASSIFIED	Specialist	527	505
Perfect Piano	unCLASSIFIED	Instrument	2,298	78

¹ As at 1 January 2018