



What the Data Reveals: Diversity and Inequality in Music Streaming

Insights from the Fair MusE Project's Analysis of Listening Behaviours (D4.1)

Authors: Antoine Henry (ULILLE), Daniel Bedoya (ULILLE), Jannick Kirk Sørensen (AAU)

While today's streaming platforms promise infinite access to music, the reality of what users actually hear tells a more complicated story. As part of the Fair MusE project, researchers analysed millions of listening data points from users across Europe, comparing them with radio playlists to understand the true shape of digital music consumption. Their findings expose significant disparities in exposure, reveal systemic metadata blind spots, and raise urgent questions about how algorithms and platform design influence fairness and diversity in the music we discover.

Streaming platforms have transformed the music landscape, offering unprecedented access to vast catalogs and reshaping how we listen, discover, and share. But behind the façade of unlimited choice lies a more complicated reality: how fair and diverse is the music that actually reaches our ears?

To explore this question, the **Fair MusE** project undertook a comprehensive investigation into streaming behaviours, drawing on anonymised listening data from 235 users across Europe, alongside curated playlists from streaming platforms and broadcasts from 73 major radio stations in eight countries. This research aims to uncover hidden patterns, structural imbalances, and the algorithmic mechanisms that influence who gets heard—and who doesn't—in today's digital music economy.

A Deep, Yet Uneven Dataset

The user-contributed streaming data covers up to 14 years of listening history per person and includes track-level details such as titles, artist names, timestamps, or skip rates. The overwhelming majority of data (96.3%) came from Spotify, with smaller contributions from Apple Music (3.6%) and YouTube Music (0.1%). Demographically, most participants were in their 20s and 30s, aligning with known patterns of streaming usage.

In total, the dataset spans more than 2 million unique tracks by nearly 400,000 artists from 237 countries¹. Despite its richness, the dataset has significant blind spots. One major limitation is that—for Spotify and YouTube Music— we need to approximate whether a song was selected by

¹ The research has used [ISO-3166](#) standard, which contains 249 countries



This project is funded by the European Union's Horizon Europe Framework Programme, under the Grant Agreement no: 101095088



the user or recommended by an algorithm. Only Apple Music includes this field explicitly per listening event. Even more problematic is the lack of accessible metadata related to key attributes such as release year, genre, label, gender of the artist, or rights ownership.

This highlights a longstanding problem: music metadata remains fragmented, inconsistent, and often hidden from public view. Without it, attempts to evaluate fairness or diversity are hampered. Still, the longitudinal depth of the data offers a valuable lens into listening behaviours and platform dynamics.

Comparing Streaming with Radio: More Choice, But Is It More Diverse?

To understand how streaming compares to traditional broadcast formats, Fair MusE researchers also analysed curated broadcast data from 73 radio stations, including 14 public broadcasters affiliated with the European Broadcasting Union. These public stations have explicit mandates to promote cultural diversity—a useful benchmark for assessing algorithmic performance.

The central question was whether streaming exposes users to a broader musical spectrum than radio. While commercial radio often sticks to tight rotations of high-performing hits, public broadcasters are expected to feature more varied content.

To quantify this, researchers applied the **Gini index**, a standard measure of distribution inequality (where 0 represents perfect equality and 1 total concentration). Among streaming users, the median Gini index was approximately **0.6**, suggesting moderate inequality in what people listen to. Radio playlists, by contrast, showed higher levels of homogeneity, with Gini scores approaching **0.8**, especially among commercial stations. Public broadcasters fared somewhat better, though still showed tendencies toward repetition.

One interviewee summed up the appeal of streaming this way:

“I think it's a nice way to be introduced to new music (...) also that it's not just the radio hits, which feels like the alternative, where it's the same 20 songs, so it's nice to have more new songs that you can also discover.” (Respondent #36)

The implication is clear: while streaming platforms have greater potential for diversity, that promise is not always fulfilled, and may be compromised by platform design, opaque algorithms, and commercial logics.

Metadata Matters: The Hidden Barrier to Fairness

One of the most persistent obstacles identified by the study is the absence of standardised, complete, and accessible metadata. Without reliable information about an artist's nationality, gender, genre, or rights ownership, it becomes exceedingly difficult to measure whether digital ecosystems are inclusive or equitable.

Moreover, the workings of recommendation algorithms remain largely proprietary and unaccountable. This opacity prevents both users and policymakers from understanding how music



is selected, ranked, or promoted. Researchers stressed that **auditability and transparency** are essential prerequisites for any serious fairness reform. Without them, efforts to improve algorithmic diversity will remain aspirational.

Moving Forward: Micro-Level Analyses and Policy Innovation

While this phase of Fair MusE's work provides a macro-level overview, the next step is to zoom in. Upcoming research will delve into user-level listening paths, comparing behavioural trends across different demographic groups and attempting to enrich the dataset with external metadata sources wherever possible.

Researchers also plan to examine how specific genres or cultural expressions fare across platforms and how algorithms respond to varied inputs over time. Investigating royalty distribution, content ownership, and platform-side decision-making will be key to fleshing out the full fairness picture.

Ultimately, these findings will inform the development of **fairness indicators**, **algorithm audit tools**, and **transparency guidelines**, all aimed at supporting a more equitable European music landscape.

Discoverability, Not Availability, Is the New Bottleneck

In today's digital era, access to music is no longer the core challenge. What matters now is **discoverability**—what gets surfaced, recommended, and heard. The Fair MusE project reminds us that the future of music diversity will not be determined by how much content exists online, but by how platforms curate and present that content to listeners. Ensuring that visibility is not dictated solely by market logic or algorithmic opacity is one of the defining fairness challenges of our time.

