



Credibility In Media

Audio On Demand Methodology

Version 2025



Table

1) Introduction	3
2) Scope.....	4
2.1) General scope	4
2.2) Server-side radio measurement	5
3) Methodology.....	6
3.1) Data collection	6
3.2) Data processing.....	8
3.2.1) Data processing steps.....	8
3.2.2) Challenges.....	10
4) Output	11
4.1) Metrics.....	11
4.2) Publication.....	13
4.2.1) Dashboard.....	13
4.2.2) Top Ranking	13
4.2.3) Ranking Site.....	15



1) Introduction

CIM is proud to introduce a new component to its suite of audio measurement studies: Audio On Demand (AOD). This addition complements our existing studies—CIM Radio Audience Measurement (RAM), and CIM Radio Stream Monitor (RSM)—by capturing data on previously unmeasured audio on demand content (radio replays, native podcasts).

By expanding into these areas, CIM is aligning with market trends and reinforcing its commitment to providing comprehensive insights. This enhanced measurement approach enables CIM and its members to track the growing segments of podcasts and on-demand audio, ensuring more accurate and relevant audience data.



2) Scope

2.1) General scope

Currently, the measurement covers audio-on-demand (AOD) content from participating broadcasters (i.e. subscribers to the CIM Radio study and providing their server log files), including radio stations and news brands.

It includes:

- Streams and downloads of on demand content
 - o Streaming or progressive download: Listening to content, either live or on demand, without the ability to download, as data is transmitted progressively.
 - o Direct download: A listener's direct request to download content from the platform, allowing offline listening, for example.
- both native podcast content (audio in podcast format made available online without prior broadcast on live radio), and on-demand streaming of previously aired radio programs, whether as full episodes or fragments.
- content streamed through both owned platforms and third-party services such as Spotify or Apple, via pass-through streams, or aggregators.
- Audio on demand streamed through connected speakers like Sonos, ...

The above allows a comprehensive view of listening habits across various streaming services and methods.



2.2) Server-side radio measurement

A similar measurement method is used between the measurement of live radio (see our RSM-, or Radio Streaming Monitor study) and AOD. Broadcasters deliver logfiles in what is called 'server side radio measurement'. It involves an exhaustive, census, measurement of all traffic that was distributed from servers to devices, which is logged in log files and then anonymized and processed to be usable by the research institute NeuroMedia to produce all necessary metrics and outputs.

This method differs from client-side tracking, which relies on data collected directly from users' devices (like web browsers or apps). Server-side measurement, on the other hand, gathers data directly from the server hosting the content. The server-side approach directly uses the logfiles generated by the streaming servers. In these logfiles, all events (stream requests by/and sent to a 'web service' through IP) are stored by the streaming server.

- The biggest advantage of this server-side measurement is that it allows the measurement of 100% of the delivered streams, independently of the receiving platform.
- The disadvantage is that the method cannot identify the exact user interaction with the delivered stream (play-pause, sound on-off, listening of downloads...).

A client-side approach is based on the scripting/tagging of the participating players. All subscribing stations are required to implement a small JavaScript tag into the source code of their players (on website or in app). Each time a stream is requested by a player, a call is sent to one of the measurement servers of the research institute.



- The advantage of the client-side measurement is that it allows the precise measurement of all actions generated by the user with the players and to objectify what has really been listened to.
- The disadvantage of the client side is that the scope of measurement is limited to the scripted players. In other words, it means that streams that are listened outside the proprietary scripted players are not measured (e.g. radio aggregators, wifi radio, ...)

3) Methodology

3.1) Data collection

The log files are text-based records that capture every interaction between a user and the audio content. Each event, such as a request to stream or download a podcast episode, is stored as a statement in the logfile. These statements contain key pieces of information that help measure and analyze listening behavior.

For example:

User IP Address	Timestamp	Content Identifier	HTTP status + bytes transferred
123.123.12.123	-- [02/Oct/2024:11:16:17 -0600]	"GET/ podcasts/programA/episode1.mp3"	200 213036
"Referrer"	"GooglePodcasts/1.0.0 (Linux; Android 13)"	103	
Referrer	User Agent	Duration	

The text string contains several useful parts (labelled in red):

- The User IP address:
 - This is the IP address of the user accessing the podcast. It helps identify unique listeners, though it does not directly reveal personal information. IP addresses can also provide approximate geographic location data.
- Connection time stamp
 - This records the exact date when the request was made.



- Identifier: 'Get request' for a podcast episode
 - This is the request made by the user's device to access a specific podcast episode. It shows what content was accessed and helps track the popularity of different episodes and programs.
- HTTP status code (200)
 - This shows whether the request was successful. A 200 OK response means the file was successfully delivered. Other status codes, such as 206 Partial Content, indicate that only part of the file was transferred—common in progressive downloads where the audio is played while downloading in chunks. Streaming is also possible.
- Number of bytes transferred
 - This represents the size of the data sent in response to the request. It helps determine whether the full episode was downloaded or only a part of it.
- The referrer
 - This field indicates where the request came from—whether from a website, an app, or another source. It helps track how listeners find the content.
- User Agent
 - This provides details about the software or application used to access the content. It helps understand which platforms or devices are most commonly used for listening.
- Duration
 - Duration is the time used to download the content.

3.2) Data processing

Once collected, logfile data undergoes processing to ensure accurate measurement of podcast and audio-on-demand (AOD) consumption.

3.2.1) Data processing steps

Metadata and content identification

The logs only contain file names or episode IDs.

Each request in the logs must therefore be matched with relevant content identifiers:

Each AOD-program and -episode is assigned a unique identifier (programID and episodeID). When log files are processed, each download is associated with a unique piece of audio content, which is identified according to a matching rule specific to the technology used by each broadcaster. The detected content is then reconciled with the metadata supplied by each broadcaster. This metadata contains, for example, the unique episode ID, the unique program ID, the program name, the episode title and the associated radio brand. NeuroMedia maintains its own metadata database. This is fed by files sent to NeuroMedia in advance by broadcasters, or by the provision of an API which is called up when NeuroMedia does not yet know the content.





In some cases, a new program or episode may be queried but not found in the metadata database. When this happens, it is assigned to "UNKNOWN" program and/or episode.

Filtering of non-human traffic

Based on the player identification, traffic coming from specific players (robots, link, ripper) is systematically excluded. A filtering on excluded IP addresses is also applied (blacklisted IP addresses, broadcaster IP addresses, ...).

Sessions bridging

When people download or stream online content, interruptions can sometimes occur. These interruptions can lead to measurement errors by mistakenly counting one session as two separate ones. This issue is even more common with podcasts that use delivery methods like progressive download, HLS, or Dash. These methods break an audio file into smaller parts, which are sent to the player one by one and then reassembled during playback. For streamed content, these fragments are sent as the listener or viewer progresses.

Some measurement methods try to solve this by counting a session only once if the same IP address and user agent access the content on the same day. However, NeuroMedia's session bridging technology ensures that two separate downloads, especially in streaming, are correctly identified as a single session when appropriate. It does this by reconstructing the session based on matching fragments, allowing for a gap of up to 60 seconds for standard downloads or progressive downloads, and up to 180 seconds for HLS or Dash.

This method also works when downloads are split across two different log files, such as when logs are saved by hour or day. By applying this approach, we can provide additional insights beyond the standard international metrics (see Metrics section below), giving a more complete picture of how on-demand audio content is actually consumed.



Deduplication of unique downloads/streams

Finally, gross downloads or streams are deduplicated (within a day) based on IP address and user agent (e.g., browser, operating system) to determine the net downloads or streams. For instance:

On 1/1/25, I downloaded podcast A, episode 12, twice.

- The gross download count is 2.
- The net download count is 1, as the duplicate download from the same IP and user agent is deduplicated.

This is further explained in the chapter on 'Metrics'.

3.2.2) Challenges

Log files can sometimes be incomplete, contain errors, or be missing for certain dates, leading to gaps in the data. These issues can arise from server problems or interruptions in data collection and affect data accuracy. Additionally, missing or inaccurate metadata may result in temporary misattributions, such as 'UNKNOWN' labels for programs or episodes. Data can be redelivered and reprocessed up to a certain period to solve these issues.

Guidelines for data exportation and reprocessing:

- 1) Monthly ranking production is based on an export taken eight days after the last day of the month. This allows sufficient time to correct any technical errors. The final day of a month, therefore, has a seven-day correction buffer.
- 2) Data available in the internal dashboard for participating broadcasters is more flexible and can be redelivered up to one month. In exceptional cases, CIM may agree on processing redeliveries beyond this period on a case-by-case basis.



4) Output

The study's output includes:

- An internal dashboard, accessible only to participating broadcasters, offering in-depth data exports on the metrics listed below.
- The Top Ranking, which is publicly available on the CIM website ([CIM Audio On Demand](#)) or via [CIM AOD Dashboard](#).

For more details, see the 'Publication' chapter below.

4.1) Metrics

The following metrics are available to the participating broadcasters in the AOD internal dashboard:

- AOD Downloads (Downloads On Demand or DOD)
 - Total downloaded audio files with a minimum of 60 seconds worth of bytes downloaded in case of 'progressive downloads' (206 HTTP status code) or total file size in case of "downloads" (200 HTTP status code). Two downloads from the same listener will be counted twice.
- AOD Streams (Streams On Demand or SOD)
 - Number of streamed audio files with a minimum of 60 seconds duration with a bridging of max 180 seconds. A listener accessing two times during the same grouping will be counted twice.

The sum of the above is aggregated into the D+S OD metric

- Unique AOD Downloads (UDOD)
 - Count of unique* downloads with at least 60 seconds' worth of bytes** downloaded for "progressive downloads" (HTTP status code 206) or the total file size downloaded for "direct downloads" (HTTP status code 200).
- Unique AOD Streams (USOD)



- Count of unique* on-demand streams with a minimum duration of 60 seconds**, allowing for a maximum interruption (bridging) of 180 seconds.

** unique listener = same combination of IP address and user agent properties.*

*** calculated based on two of the following variables: audio file size, audio bit rate, audio duration. When not available, the default parameter of 1000000 bytes is used.*

Both metrics are aggregated into the UD+SOD metric, which is used in the production of the 'Top Ranking' (see below).

Important: Deduplication of unique downloads and streams occurs only at the lowest level (per episode). When aggregating data at higher levels—such as per broadcaster or per radio station for the Top Ranking—unique downloads and streams per episode are summed without further deduplication.

For example:

- If I listen to episode A in both January and February, I am counted as a unique download or stream in each month.
- If I listen to episode A twice on the same day in January, I am counted as one unique download or stream.
- If I listen to two different episodes (A and B) (even on the same day) from the same broadcaster, I am counted twice for that broadcaster.



4.2) Publication

4.2.1) Dashboard

Currently, participating broadcasters can access the data through an internal dashboard, which provides the metrics outlined above. This dashboard is not publicly available.

4.2.2) Top Ranking

Scope & Timing

The data available in the dashboard is used to generate a public Top Ranking each month. As mentioned earlier, this ranking is based on a full month's data, representing the 'monthly' ranking.

The data snapshot is taken eight days after the last day of the month, allowing a minimum seven-day buffer for error corrections. CIM aims to publish the final Top Ranking 15 days after the end of the month, ensuring sufficient time for technical validation and processing.

Content

This ranking is based on unique downloads and streams, which determine the order of the rankings. Additionally, the number of unique episodes and programs (counted by distinct episode- or program IDs) is tracked.

It is important to note that the ranking reflects content listened to during a given month, not content published that month. As a result, old episodes with small listening volumes still contribute to the listening figures.



The Top Ranking consists of three levels, with separate rankings for the Dutch-speaking (NL) and French-speaking (FR) content:

1. Broadcaster Ranking
2. Radio Brand Ranking
3. Program Ranking (top 100)

Broadcaster & Radio Brand Rankings

These rankings are relatively straightforward, as they involve summing unique downloads and streams per broadcaster or radio brand, as well as counting the number of different programs and episode IDs per brand or broadcaster.

Program Ranking

The program ranking is more complex due to several factors:

- Aggregation by Program ID
- Aggregation by Program Title
- Metadata Cleaning:
 - Programs with 'UNKNOWN' titles are removed.
 - Programs with incorrect metadata or 'generic' titles (e.g: 'sport', ...) are excluded from the top 100.

By following these steps, the ranking ensures that programs are correctly aggregated, and only valid metadata is included in the final list.



4.2.3) Ranking Site

After generating the rankings for broadcasters, radio brands, and programs in both Dutch (NL) and French (FR), they are published and publicly available on

- the ranking website: cimaod.neuromedia.io
- and on the CIM website: <https://www.cim.be/nl/audioondemand>

The website allows users to filter data by month, broadcaster, brand, and language.

Users can also export the data.

The data is sorted and ranked by the unique streams & downloads volume.