



IAB Tech Lab
116 East 27th Street, 7th Floor
New York, New York 10016
United States of America

6 December 2020

Feedback on IAB Podcast Measurement Guidelines v2.1

I'm grateful for the circulation of this document to allow those of us involved in the industry to give you feedback; and have pleasure in giving you feedback as requested.

Overleaf, you'll find two recommendations for two extra sentences as part of these guidelines:

- A recommendation that all libraries used should have an identifiable user agent
- A recommendation that identifiable user agents are used when accessing RSS feeds

If you've any questions, I can be reached by email at james@podnews.net or by mobile phone (at a suitable time for my local timezone) at +61 447 692743.

Thanks so much for reading, and for circulating this document.

Yours,

James Cridland
President, Podnews LLC

The user agent structure section on page 20, which sets best practice for the user agent structure, is welcome. I have been working for a while on an open user agent list, <https://github.com/opawg/user-agents>, which lists all agents visible out there. Already in production at many podcast hosts, I'd be keen for more adoption of it from IAB members, and would welcome collaboration on this open resource.

From this, I would like to make two recommendations:

A recommendation that **all libraries used** should be identifiable

Libraries, like AppleCoreMedia or Dalvik, are difficult to attribute to a specific podcast agent. It would be helpful if this document highlighted that all libraries used should ideally identify the agent (app name, device type, OS) wherever possible. My experience is that many apps only identify themselves upon file download, not on play-on-demand progressive downloads. Apple, particularly, does not allow a different user agent for the AppleCoreMedia library: it would be helpful were the IAB to request this, within this document.

In my tests, only 47% of "AppleCoreMedia" is directly attributable to Apple Podcasts, as one example.

"If any third-party libraries are being used, such as Android Download Manager, Dalvik, wget, AppleCoreMedia or others, the IAB recommend that the useragents are modified to be identifiable wherever possible, and requests that they are made modifiable within the host OS."

A recommendation that identifiable user agents are set when **accessing RSS feeds**

User-agents are also in use when accessing podcast RSS feeds - either by an app directly or, more usually, by a server I keep an open list of these here:

<https://github.com/opawg/podcast-rss-useragents> - it would be very useful if this document highlighted that an identifying useragents should be used when accessing the RSS feed.

If you generate a podcast RSS feed dynamically, you can add a querystring to the audio that is downloaded that reflects the RSS user agent. As an example, if an RSS feed is downloaded by a user agent of "**Tentacles**", that user agent is actually used by Castro, an iOS podcast app. You could then stamp the audio files with that user agent - to make the enclosure read https://example.com/audio/podcast.mp3?_from=castro.fm

This then means that you have certainty that this audio filename, with this query string, has been downloaded as a result of being available in Castro's directory - irrespective of whether Castro has set a correct useragent, or if it using the AppleCoreMedia library to progressively download the audio.

Even if Castro has a website with a web-based player, the RSS user agent allows us to stamp the audio file regardless. So the audio used by a web-based play can still be stamped with **?_from=castro.fm** and therefore this play can be correctly attributed to Castro, and not just to a web browser.

In my tests of a sample of 3,880 total podcast plays:

- Only 2,193 downloads provided an identifiable device user agent (56%)
- 3,492 downloads could be identified from the RSS user agent (90%)
- For the 179 AppleCoreMedia records (5%), only 85 (47%) were from Apple Podcasts.
- 457 records (12%) were from web browsers and could be correctly identified with this method.

You can see a view of data used by monitoring both RSS and audio user agents in a live environment here: <https://podnews.net/pages/podcast-analytics/parse-analytics.php> ...at the time of writing, only 3% of downloads are from an unattributed source: significantly lower than the usual numbers from audio useragents alone.

My recommendation would be that user agents should be set both when downloading audio but also when downloading RSS feeds.

"A identifiable useragent should always be used when accessing a podcast's RSS feed."