Understanding the AGF Census Files

On a daily basis, 7 distinct Census files are delivered. Generally, all Census files are delivered in a highly aggregated format to reduce the enormous data volume. The exception is the EventLevel file for livestream usage, which is aggregated at the Stream ID level and therefore contains all livestream sessions individually.

- agf_S2S/SDK_ContentByAssetIDCensus_PUBLISHER_daily_YYYYMMDD_F5_V0.csv
- agf_S2S/SDK_EventLevel_PUBLISHER_daily_YYYYMMDD_F3_V0.csv
- agf_S2S/SDK_summarycensus_PUBLISHER_daily_YYYYMMDD_F5_V0.csv
- agf_S2S/SDK_AdsByAssetIDCensus_PUBLISHER_daily_YYYYMMDD_F4_V0.csv
- agf_S2S/SDK_ContentByAssetIDCensusOutlier_PUBLISHER_daily_YYYYMMDD_F3_V0.csv
- agf_S2S/SDK_AdsByAssetIDCensusOutlier_PUBLISHER_daily_YYYYMMDD_F3_V0.csv
- agf_S2S/SDK_summaryoutliercensus_PUBLISHER_daily_YYYYMMDD_F3_V0.csv

Versioning Logic

- **Format Version (F):** The number following the "F" (e.g., F3) indicates the format version. If there are structural changes to the table (e.g., adding new columns), this version number increases.
- **Delivery Version (V):** The number following the "V" (e.g., V0) represents the delivery version. If an error occurs that requires a corrected redelivery for the same data day, this version number will be incremented (e.g., to V1).

Aggregation & AssetID Handling

The **AssetID** serves as the basis for Nielsen's census data aggregation. It is an identifier from the publisher's CMS which must be unique per video file when combined with the Client ID (CID) and Channel ID (VCID). For advertising, the provider-specific unique ID of an ad video file (Ad-ID) is passed as the AssetID.

To ensure absolute uniqueness, Nielsen creates a Content_ID by combining these elements. The final aggregation is based on this ID (Content_ID = CID_VCID_AssetID, or for ads, AdID = AssetID).

Note:
Due to the aggregation by Ad-ID and the freezing of metadata and c-variables, views and duration in the AdsByAssetID files cannot be simply summed up for an ID contained in the c-variables (e.g., you cannot simply aggregate to calculate the total ad duration spent on a specific Video-ID).

Metadata Freezing

During census data aggregation, metadata and custom variables (c-variables) for each Content_ID are locked in once per data day (3:00 AM to 3:00 AM). They are extracted from the very first video view of the day and frozen for the remainder of that day. Any adjustments or corrections made to the metadata or c-variables of an AssetID later in the day will not be reflected in the daily data. So if the metadata file for an AssetID is missing for the hour in which the video was first watched that day, all views for that AssetID on that day are reported with UNWN metadata and therefore cannot be included by AGF for reach reporting.

Metrics & Device Classification

The measured views (stream_views) and duration (total_time_spent) for streaming are broken down per Content_ID by platform, device type, OS group, usage type, and VOD/Live status. Device classification is performed by **DeviceAtlas** based on the transmitted User Agent:

- **platform**: Differentiates between Desktop (DSK), Mobile (MBL), and OTT devices.
- **device_type**: Provides a more granular breakdown into device groups such as smartphones (PHN), tablets (TAB), Smart TVs (STV), and Set-Top Boxes (STB).
- **os_group**: Indicates the operating system of the device.

Importance of Accurate User Agents:

Accurate device information is crucial for precise measurement. The more correct and detailed the transmitted User Agent is, the better DeviceAtlas can map usage to the specific device types - ensuring an accurate differentiation between DSK, MBL, and big screen versus small screen consumption.

For industry best practices on formatting User Agents, particularly for Connected TV and OTT devices, please refer to the guidelines provided by the IAB Tech Lab:

IAB Tech Lab: OTT / CTV User Agent Guidelines (PDF)

Geographic Breakdown

Traffic originating from Germany is marked with "DE" in the global_or_country column. Please note that these German views are **already included** within the overall global traffic counts marked with "G".

Important Note:

Running a simple SUM() across the entire table without filtering global_or_country will cause German traffic to be counted twice.

- To query total **worldwide traffic** (including DE), filter by global_or_country = 'G'.
- To query **Germany-only traffic**, filter by global_or_country = 'DE'.