[Draft Proposal for WFA] Cross-Media Measurement System for Reach and Frequency CONFIDENTIAL NDA

Executive Summary

As part of the WFA's cross-media measurement working group, this document outlines an initial proposed technical blueprint that local industry bodies could implement to deliver cross-media measurement in markets around the world. This blueprint is designed to include both digital and linear channels, with further input needed from measurement experts on the buy and sell side, advertisers, research vendors, agencies, and other media (e.g., TV broadcasters). The measurement technology described here:

- Lays out a common technical infrastructure that can be built at scale to support distinct measurement systems in local markets.
- Can support various panels, methodologies, data science models etc, to accommodate for local market preferences and needs.
- Works for digital and linear data inputs (*with further input and review required from industry experts).
- Prioritizes reach and frequency use cases, with a path to outcomes measurement (e.g., Multi-Touch Attribution) over time.
- Focuses on meeting advertiser requirements (referred to in this document as the Advertiser "North Star"), as defined by the WFA working group.

We expect the proposed technical blueprint to be an ongoing collaboration with industry partners and will update this document on an ongoing basis as testing and other research is completed.

Before outlining the blueprint, the technical working group would like to highlight a few key points below, on: the specific advertiser use cases this system is aiming to deliver; global vs. local market considerations; system inputs and outputs; call for feedback; and timelines & next steps. These topics are also covered in more detail throughout the document where applicable.

Advertiser Needs & Industry Requirements

There are many potential applications for the cross-media measurement system. This system should focus on prioritized, specific needs and advertiser supported Industry Requirements (as referenced as the Advertiser 'North Star' in the *WFA Industry Framework For Establishing A New Approach To Cross-Media Measurement*):

Advertiser Needs

- Comprehensive: Enable cross media reporting across all media formats
    - Build for full cross-channel measurement, including linear TV and all digital formats.
        - Inclusive of all eligible digital and traditional media platforms and publishers.
        - Enable measurement of all relevant publishers/channels regardless of size, relationship to users, technical expertise, etc…
    - Enable advertisers to measure their entire campaign not just specific media formats.
    - Solution should be designed such that it could work for all major global markets, clearly identifying areas that require market-specific innovation.
- Continuous: Tagless, always-on data capture
    - Provide tagless always-on data capture and on-demand availability for eligible marketers. *Specifically, Advertisers who have opted-in and met other requirements (e.g., subscriptions, cleared legal and data sharing aspects, etc.) should be able to access measurement on an ongoing basis, rather than campaign-by-campaign, limited-duration, fee-based campaign tracking.*
- Full Life Cycle Measurement: Enable all phases of measurement
    - (1) pre-campaign audience planning
    - (2) intra-campaign audience and frequency management and optimization
    - (3) post-campaign audience evaluation
- Full Funnel: Reach, Frequency Management and ultimately Outcomes
    - De-duplicated (1) reach and (2) frequency of media campaigns to start
        - (*Advertisers recognize that this would not be individual-level information and would adhere to all privacy rules/regulations, etc.)*
    - Clear path to integration of outcomes measurement as the program evolves, to enable media audience and related analytics, such as attribution modeling, media mix modeling and brand/sales lift studies leading to better media performance and outcomes
- Comparability across channels
    - Comparing, counting and de-duplicating ad exposures is one of the necessary foundations of cross-media audience measurement. This system should enable comparability across channels.

Advertiser Supported Industry Requirements

- Privacy Centric: Ensures respect for the consumer and that user privacy is safeguarded
    - Built to the highest standards in terms of privacy
    - Ensures technical measures are in place to address risks, such as ensuring no entity knows more identifying information about individual users than the entity did prior to participating in this system, and that re-identification of users is prevented.
    - Built to be future proof, meeting the privacy requirements of today with development roadmaps focused on the future
- Fair & Objective: Enables fair cross-media comparisons that provide apples-to-apples comparison across TV and digital advertising through technology that solves for cross-media, open standards, and a neutral governance model
- Advertising & Content: Measurement priority should be focused on media exposure that is either advertising specific or capable of carrying advertising, but the system should be capable of supporting both ads measurement and content measurement.

- Global Trust & Transparency: The technical design and any implemented version should be sufficiently transparent to build trust in the measurement service. This could mean open-sourced technical approaches as well as opportunities for audits and verification.

This list is not comprehensive, and additional requirements and details are outlined in the WFA's Framework, which includes the Advertiser 'North Star' .

Global vs. Local Market Considerations

This proposed technical blueprint outlines the technical infrastructure needed for privacy-focused deduplication of reach and frequency across media. The infrastructure can scale globally to speed up implementations, reduce inefficiencies and support some degree of global comparability. However, local markets will still need to make critical decisions about the measurement system before implementing this design. For example, reach modelling requires a number of inputs and model assumptions. These can and should be determined at a local level to reflect each market. This will ensure that the solution meets market-specific needs, and that advertisers have flexibility and choice. A non-exhaustive list of high level areas for local markets can be found in the detailed technical blueprint below.

System Inputs and Outputs

This technical infrastructure is designed to be data agnostic - it will support various sets of data inputs and can be built to produce many outputs, subject to privacy requirements and reasonable computational considerations. A detailed global data inputs spec will be required before this system can be implemented in local markets, including: 1) an outline of raw data that aggregators & publishers must collect, process and permit audits of; 2) inputs the measurement system must process; and 3) outputs the measurement system will produce for default reports and APIs. This will help ensure consistency and viability across participating data providers at a global level. The MRC, in collaboration with WFA, is developing an initial recommendation on these global data inputs which, once finalized, can support implementation work at the local level.

Additionally, based on feedback from the agency community, the output of this system should produce a "transformational dataset in familiar environments", that is: (1) compatible with existing tools and processes used in the industry; and (2) available to multiple vendors, agencies, advertisers, publishers etc. However, to accommodate industry stakeholders who may not wish to use an API, the system also enables local markets to produce standardized reports and simple dashboards as part of the service. More detailed input will be required from industry experts to ensure the API and standardized reports are useful.

Timeframes and Open Questions

While this document is a meaningful step forward, there are still a number of open questions that the technical working group will need to resolve over time. A non-exhaustive list includes:

- Completing testing and final selection of private reach and frequency estimators to enable deduplication (estimated timing: mid-May).
- Development of a standardized campaign taxonomy to help implement access controls for the measurement system so that advertisers only access to their own campaigns. At a minimum,

advertiser/brand-level standardization is required to operate the measurement system. (estimated timing: workstream to start in May-June)
    - In addition to advertiser/brand-level standardization of taxonomy, standardization at campaign/placement level can significantly enhance the actionability of the measurement system to aid with advertisers' ask of continuous data processing and always-on measurement. (*Note: This would be a separate workstream, but wouldn't be a dependency for 1st version of the solution*).
- Development of a technical infrastructure blueprint for Secure Universal ID (SUID) to improve the quality of R&F and enable additional use cases, including MTA and Sales Lift (estimated timing: workstream to start May/June, with timing dependent on research results).
    - *Note: This SUID technical blueprint should be a complementary effort to the MRC's outcomes standards workstream.*

The WFA hopes to publish a version of this blueprint with cross-industry support by July 2020. At that point, national advertiser bodies should lead the implementation of the design in their local markets. Specific timing will depend on market stakeholders and requirements.

Call for Feedback

This proposed blueprint is the first step to defining a solution that can accurately and fairly measure advertiser campaigns across channels. It can only be successful with additional feedback and engagement from across the industry to improve it. The technical working group considers this to be a draft, and will continue to evolve the design as feedback is received. We encourage all interested parties to engage in the WFA's peer review process to help improve and iterate on this design.

Detailed Technical Blueprint

Motivation

As advertisers seek to better understand the efficiency and effectiveness of their media investments across traditional and digital channels, measurement systems have to evolve.

Today, most existing measurement systems rely primarily on traditional panel-based approaches designed for linear TV. However, these approaches cannot measure and deduplicate a significant portion of online and Over-The-Top (OTT) user activity without additional census data. For example, existing measurement panels cannot measure or deduplicate in-app activity, despite this being a major channel for consumers. Even when publisher logs are provided for panelists, panel sample sizes are often too small to provide the fidelity, confidence and granularity of measurement desired by advertisers. Even when aggregate census data is available, extrapolating the panel to the census totals significantly reduces the effective sample size of the panel. On the other hand, measurement using only census data would be subjected to biases in those datasets.

Despite the adoption of new media over the last couple of decades, most cross-media measurement solutions have not sufficiently progressed beyond panel-based approaches. Existing and emerging technologies are and will continue to be impacted by evolving privacy regulations and increased user demand for transparency and control over data. To that end, we believe census data must continue to play a role in the richness and completeness of future measurement solutions and metrics, while single-source panels will be required to ensure the accuracy and fairness of measurement systems. An approach that combines census and panel data, which uses a panel (or panels) as the source of truth to calibrate (i.e. remove bias) census data will best serve advertiser needs.

We propose a new set of technologies, primarily a census identification framework and a private reach and frequency estimator, to deduplicate reach and frequency across multiple data providers' census data, using a panel for calibration. Computing these outputs securely and in a way that preserves user privacy, without compromising accuracy, is a key design goal.

Principles and Requirements

In order to inform design choices in this proposal, the WFA's technical working group focused on:

1. Delivering the advertiser's "North Star" over time; and
2. Meeting the set of advertiser supported Industry requirements based on the WFA's Industry Framework (*referenced above)*.

Based upon the advertiser needs ("North Star") and industry requirements the group considered the design and each subcomponent to assess the ability to meet the full set of requirements. Advertisers provided a stack ranking of these requirements, to guide solution development (listed below). This list was used to establish a prioritisation of the Advertiser 'North Star' and a phasing for proposed technical work. For a detailed list of cross-media measurement principles, please refer to the [*link to WFA Framework*].

Figure 1. Advertiser Use Cases & Features - Stack ranked and phased

		Requirement	Description
Foundational Feature Set	**1**	Always-on	Tagless implementation; begins from advertiser opt-in; forecasting based on available data
	2	Data Granularity (ex. For viewability)	Sufficiently granular data to comply with relevant industry standards, report across all relevant metrics etc while respecting privacy principles
	3	Basic Segments	Creative, geo, device, placement, campaign, age, gender, format, HHI
	4	Output metrics	Deduplicated reach, frequency, impressions, watch time (absolute, %, distribution)
Phase 1: Deliver R/F Use Cases	**5**	R/F Reporting	Reports updated at least weekly; 12 month lookback; exportable data (via API & CSV)
	6	R/F Forecasting & Mgmt.	12 months forecast; can be based on historical data to start (inventory data from publishers added as available); includes **daily** scenario planning across channels to optimize r/f (this is the mechanism to manage frequency) **note: freq mgmt is not a centralized targeting approach*
Phase 2: Deliver Outcomes Use Cases & Advanced Feature Set	**7**	System integration	Data API that connects to existing systems (i.e. agency tools, other tools like Datorama)
	8	1P Data	Custom segments based on advertiser data (reporting is top priority, forecasting secondary) - Priority 1: Reporting; Priority 2: Forecasting

	9	Outcomes	All Effectiveness measures (brand/sales lift are top priority; conversions, MMM,MTA secondary)
	10	Segments - advanced	Psychographic, behavioral audiences based on common 3P definitions
	11	On-demand analytics	Ability to query database (considerations: data availability, latency, UI) for advanced analytics
	12	Global view	Single UI to access multiple markets R/F reports, forecasts etc

***Note:** Use cases and advanced features in phase 2 may be made available at different times. Development will be prioritized according to the ranking provided here by advertisers.

Privacy Principles in Depth

Privacy principles have a particular bearing on technical design choices, as ensuring data can be deduplicated while meeting regulatory requirements and user expectation requires specific, privacy-centric technologies. For the purposes of this blueprint, we defined the following privacy principles:

Consumer Privacy

- **Re-identification:** Data providers should be able to ensure that their users' data is not re-identifiable (i.e., an observer seeing any data in this system cannot tell if a particular user's information was used in the computation), unless data belongs to consenting panelists
    - Why? Re-identification is a risk that many industries face today. If data is not properly protected, encrypted data can be combined with other datasets to reveal sensitive user information.
    - In addition to strong contractual protections, data providers will also need technical guarantees against re-identification of users, through measures such as differential privacy[1]. All methods must provide a quantifiable measure of protection against re-identification.
- **User control:** Data providers should be able to provide their users with transparency and control over the collection and use of their data
- **Panelists:**
    - Data providers may request explicit consent to share panelist data
    - Data providers should not learn the identity of any panelist
    - Panelists should have the same user controls per above

We will meet these principles by designing this system such that data providers' inputs are either combined cryptographically and/or are differentially private, and that the system's outputs are compliant with privacy principles (e.g., differentially private). In short, no party to the system should be able to learn anything beyond the agreed upon outputs of the system.

Data Provider Data Security

By default, data from publishers could only be used for the purposes of enabling cross-media advertising measurement for individual advertisers or agencies. Measurement service operators must get data

providers' explicit consent to use their advertising measurement data for any additional use cases (e.g., reporting overall trends in digital advertising).

Advertiser Data Security

While we acknowledge the need to have continuous data collection to enable always-on cross-media measurement and on-demand analytics, only data from campaigns of advertisers (or agencies representing advertisers) that have opted into such a solution may be provided to the measurement system.

Additionally, the Reporting and Planning APIs that the system offers should provide enough access control features and transparency to advertisers to allow them to limit and understand the exposure of their campaign data.

Open source principle

The technical working group also developed an "open source" principle, to ensure transparency, as it is an important part of building trust in the technical solution. We will provide a high quality reference implementation for this technical blueprint that includes technical papers describing the methods, test results on the accuracy of the methods, and open source software that implements the methods. The intent is to demonstrate that privacy-centric cross-media measurement is technically feasible and to encourage locally-implemented solutions to be equally transparent and open. In addition, we hope that making a reference implementation available will improve quality and lower the cost of building and deploying local solutions.

Design Overview

A future-resilient cross-media measurement solution should consist of four key components to offer highly accurate reach and frequency measurement while still preserving consumer privacy. Those components include: (1) Panels; (2) Data Provider Census Data (3) Private Reach and Frequency Estimation; and (4) Output Data. These components meet our guiding principles and advertiser requirements per above and are described in further detail below.

1. Panels as source of truth

In order to capture all relevant media consumption data, data providers will need to contribute census data in a privacy centric way. However, to ensure the system is fair and objective, the technical blueprint must enable independent verification and transparency. The main mechanism for this is a single-source panel (or multiple separate panels) to act as the arbiter of truth, providing benchmarks for the use and overlap of media consumption and correcting for bias in census data logs.

The single-source panel will be used to calibrate and adjust census data collected directly from data providers where appropriate as well as to provide inputs into reach, frequency and deduplication models.[2, 3] The single-source panel may also be used to measure exposure to media where census data is

unavailable. For example, smaller publishers without census data can still be measured by the panel. This will ensure a broad coverage of publishers and reduce barriers to entry for those without census data.

If a single source panel is not available in a local market, other data sources will be required. If separate digital and TV measurement data is available, they could be used in conjunction with relationships measured between digital and TV from countries that have a single source panel, or other data sets or assumptions available in the market.

Panel Requirements

Panel data must contain the participating data providers' ad exposure event information, associated with user device identifiers, and additional contextual and demographic information that can be useful for developing measurement models.

The panel measurement technology must provide mechanisms to prevent participating data providers from learning the identity of the panelists. Technologies can be used to perform double-blind joins of census log data with panelist sessions. This will protect panelists' privacy, and among other benefits, ensure no data provider can influence panelists inappropriately.

In order to uphold our privacy principles of user control and transparency, panelists must provide verifiable consent for their participating data controller event data to be collected and shared with the panel operator.

Digital and TV media need to be handled somewhat differently (*Note: there is significant local expertise in handling TV census data, which should be relied upon in developing a solution for measuring TV. The following description is one possible approach to incorporating TV census data into the Census Identification Framework*). For both digital and TV media, census data will be calibrated with the single-source panel. However, TV census data, including Set-Top-Box, Smart TV, and OTT data, could be considered household-level data (depending on the device). This means that it has higher co-viewing rates than other devices. Therefore, this data will have to be converted to individual-level data, using the single-source panel, before applying the Census Identification Framework, as described below. To enable this, TV census data must be accompanied by the number of people in each household and their associated contextual and demographic information. To the extent that the TV census data is a subset of the full census, viewership will need to be scaled up using the single-source panel and Census Identification Framework. In addition, there are data quality challenges specific to TV census data that will need to be addressed using the single-source panel.

To the extent that multiple panels are available (e.g, a single-source panel and other single-media panels), there are methods for combining these panels in an optimal way for reach and frequency estimation.[4]

2. Census Data

Quality Controls

Each data provider that has the means to collect census level impression data can choose to contribute this data to the system through the Census Identification Framework and Private Reach and Frequency Estimation.

Data providers are responsible for implementing the necessary mechanisms to filter invalid traffic, appropriate quality controls, and auditing mechanisms required by local markets for both content and ad impressions, including potential impression-level signals such as contextual and demographic information. The local body governing this system, should establish an appropriate process for verification and audits of these mechanisms.

Census Identification Framework

In order to deduplicate reach across data providers and media types, a common method for identifying users is required. An identification framework will allow individual data providers (e.g., publishers and broadcasters) to identify users at a census-level, and make sure this process is consistent across all channels.

There are at least two challenges with this. The first one is designing an identification framework that upholds our privacy principle of non re-identification. The second one is ensuring the quality and coverage of the existing data.

The technical working group has identified two identification frameworks that address both of these challenges. We propose:

1. Virtual People IDs (VIDs) as a way to solve this problem for reach and frequency immediately[5];
2. A cross-industry effort to explore more advanced methods like the creation of a Secure Universal ID (SUID) that could both power metrics for sales lift and multi-touch attribution, and enhance the quality of reach and frequency measurement

We believe these two approaches complement each other and recommend investing in both in parallel. However, additional work is required to determine if a SUID can meet our requirements and principles. The technical working group therefore recommends an ongoing WFA workstream focused on relevant research to develop a SUID over time.

Virtual People IDs (VID)

Virtual People IDs allow us to establish a common virtual identity for every impression, injecting the statistical information of user activity measured by the Panel into the Data Providers' census data.

The main value of VIDs comes from the fact that:

- Every impression from every Data Provider is assigned a VID.
- The reach and frequency distribution of VIDs over a particular segment of census impressions is a high quality estimator of the true people reach and frequency distribution of that segment.

- VIDs can be assigned independently by each Data Provider, using a common Panel trained model. No census data needs to be shared between Data Providers in order to assign VIDs.

It is worth noting that VIDs are statistical in nature, and there is no guarantee that a single person will be assigned the same VID across multiple participants. This does not impact the accuracy and overall quality of the measurement outputs for advertisers.[5] Because of this, VIDs should not be used for targeting purposes, since they do not correspond to real people individually.

The VID framework implementation can be broken into two phases, described in detail below. Additionally, to ensure that the system upholds our privacy requirements, VIDs need to be combined with methods that can deduplicate counts of VIDs across Data Providers in a privacy centric way. We discuss these private reach and frequency estimation methods below.

Secure Universal Identity

Given the current and upcoming changes to how third-party cookies behave on browsers and potential future changes to persistent mobile identifiers, more data providers (both large and small) are relying on signed-in experiences to establish a direct relationship with their users. This enables data providers to leverage email addresses or other data as common identifiers for the purposes of cross-media measurement.

It should also be possible to add multi-key identifiers to the system, which will improve quality and coverage over just a single identifier such as email. However, for this to be viable, the multi-key identifier has to be both secure and private, which requires significant technical work and research.

The use of common identifiers for measurement also necessitates industry-wide legal and policy discussions (and consensus) that may take time to develop. On the other hand, leveraging such frameworks will enable more advertiser use cases, such as sales lift and multi-touch attribution. To that end, the technical working group recommends an ongoing WFA-led workstream to continue development of a privacy centric, industry accepted SUID.

Figure 2. Comparison of VID and SUID frameworks

Considerations	Virtual People ID (VID)	Secure Universal ID (SUID)
Meets Privacy Principles	Yes	TBD. Subject to additional research.
Produces high quality measurement outputs	Yes, for R/F	Yes, for all use cases
Enables key use cases	R/F	R/F (improved when used with VID) Sales outcomes (lift, MTA)
Scales (i.e. accessible to all types of data providers)	Yes	No (excludes data providers without 1P relationship with users)

Does not preclude other ID framework (i.e. allows us to use both VID and SUID)	Yes	Yes
Recommendation:	Over time the system should support both VID and SUID. In the short term, we recommend using VID to deliver reach and frequency measurement without delay, while we continue the WFA-led workstream to develop SUID.	

3. Private Reach and Frequency Estimation

Once a common method for identifying users has been established, it becomes possible to count impressions across data providers and to compute deduplicated reach and frequency. Due to its consistent approach to modeling, the virtual people id model makes this a simple matter of counting unique VIDs across data providers. To accomplish this, data providers would provide VIDs by campaign and user segment to an independently operated measurement service that can combine and deduplicate VIDs to produce an overall estimate of deduplicated reach and frequency.

The WFA technical working group is currently evaluating several candidate solutions for securely and privately computing deduplicated reach and frequency in this way. Each of these solutions ensures that no information beyond the agreed upon outputs of the system is learned by any party and that the presence (or absence) of any single individual's data does not appreciably change the output of the system. In short, regardless of which method is chosen, data providers should feel confident that their data is being used only for the agreed upon calculations, and that the privacy of users will be maintained.

We are currently in the process of defining a comprehensive evaluation framework with input from the MRC. A description, as well as a link to source code, will be included in this document when it is available. In addition, once the evaluation is complete, results will also be made available for review.

[Hold for section on test framework & results]

4. Output Data

This system is designed to produce both standard metrics (i.e. default reporting) and a data API that can be accessed broadly. The API can be made available to multiple organizations, including advertisers, agencies, research vendors, publishers to support new measurement products and/or existing tools. In addition to the API, local markets may also choose to develop a standard user interface that supports default metrics for users who cannot leverage the API. A list of output metrics that the API and these default reports produce will be determined at the local level.

Key Technical Roles

This design will require the support and participation of many industry stakeholders. We've listed some of the roles and responsibilities below. These roles are not mutually exclusive - i.e. a 'User Impression Data Provider' may also be a 'Private Estimation Node Operator' as appropriate. The specific stakeholders selected for each of these roles should be decided at a local market level. Local markets may also require additional roles to support specific governance structures.

User Impression Data Providers

Participating user impression data controllers integrate these data sets with the Panel(s), in coordination with Panel Operators, for the purposes of training VID models in the Setup Phase, and using those models to label user impression data to produce measurement outputs in the Live Measurement Phase *(Note: detailed explanation of VID models and phases below).*

Panel Data Provider(s)

Licenses or provides the panel data, coordinating with different User Impression Data Providers, and providing the output panelist data to the VID Model Operator for the purposes of training a VID model.

TV RPD Providers

Provides RPD TV (STB, smartTV, etc) data to the VID model operator that can be used to enrich TV panel data sets to create more accurate VID models for TV, and more comprehensive TV assets.

Private Estimation Node Operators

Private estimation methods require running a decentralized (multi-party compute) system where the security is guaranteed by the decentralized nature of the system. In other words, multiple operators are necessary to run the system in a secure way. Participating User Data Providers must trust at least one of the Private Estimation Node Operators in order to trust the security of the system.

VID Model Operator

The VID Model Operator has the responsibility of training and releasing VID models using the Panel and TV data assets. It is responsible for collecting other sources of data and the necessary parameters to train the model, as well as timely updates to the model to meet the local market requirements.

Measurement API Service Operator

The measurement API operator owns the reporting layer on top of the Measurement System, and is responsible for enforcing the necessary access control, as well as additional functionality built on top of the outputs of the measurement system, including potentially standard reports published to the market participants with the required licenses (also enforced by the Measurement API Service Operator).

Design Details

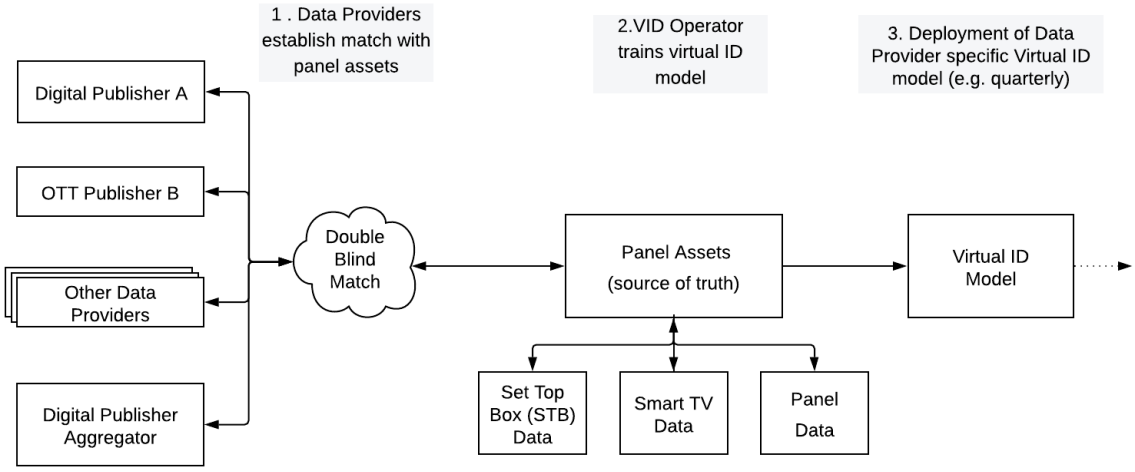
System Overview and Lifecycle

Advertisers require that the cross-media measurement system be always-on and not require a resource-intensive set-up process. While a common campaign taxonomy may still be needed to address this (i.e. how to name campaigns, creatives, etc., consistently across data providers), the technical design should support always-on, tagless measurement and not place an onerous technical burden on participating data providers.

The system is built to work in two phases: a set-up phase, and an ongoing measurement phase.

Set-Up Phase

Figure 3. Overview of measurement system set-up phase



The main objective of the set up phase is to train a market-specific virtual people identity model, which will provide the foundation for cross-media deduplication. This set-up process must be completed at the country/market level and refreshed a few times per year. The specific cadence of updates will depend on a number of factors, including the country's rate of population growth, rate of internet penetration etc. For example, a developed market with a relatively stable population and higher internet penetration (e.g., US, CA) may only require annual updates. On the other hand, developing countries with rapid population growth and evolving internet penetration (e.g., IN, ID) may benefit from a quarterly cadence. This should be left up to local governing bodies to decide based on relevant market factors.

Step 1. Data Providers establish match with panel assets

In the first step of the set up phase, the panel operator and the data participants establish a mechanism to measure ad event data for the participating panelists. This is accomplished through metering technologies

as part of the panel, and through an offline double-blind data exchange with data providers. Double-blind data exchanges allow data providers to share consented panelist data without knowing who panelists are, and without revealing sensitive information.

For the TV data providers, the Set-Top-Box (STB) data, Smart TV data, and Over-The-Top (OTT) data will be used to establish the match with a Return Path Data (RPD) Panel and provide feature data for the matched panelists.

This data obtained through direct panel measurement or through the double-blind match process will enable the collection of the necessary event- and user-level signals required to train the VID model.

Step 2. Measurement Service trains virtual ID model

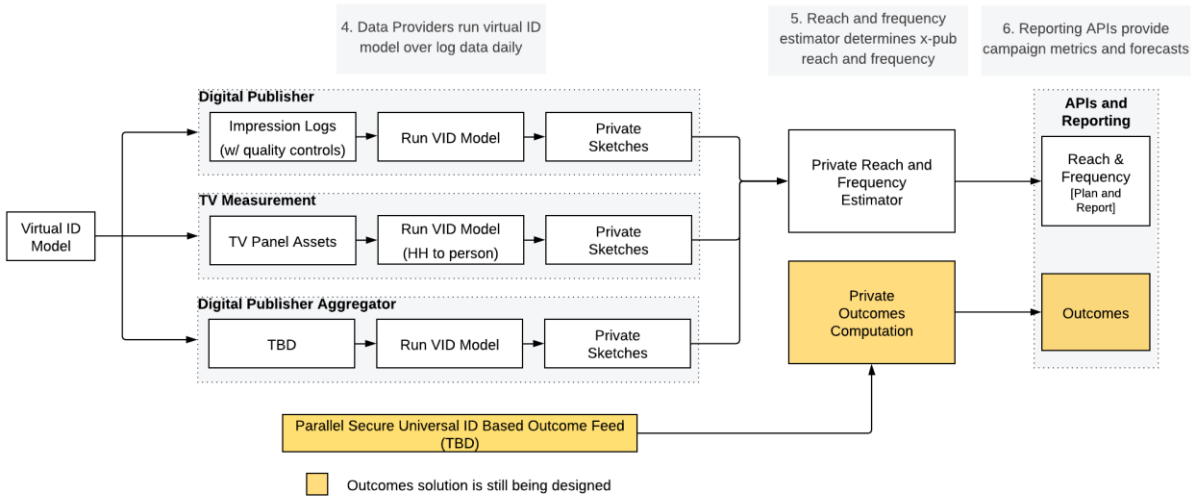
After the local measurement service receives all the data for the matched panelists, a VID model will be trained. This model encapsulates the relationship of user activity across demographic groups and data providers, measured and calibrated by the panel. This step is performed by the VID Operator, using local market insights and benchmarks (e.g., country specific population data), as well as the panel assets, in order to calibrate the output to the target population.

Step 3. Data-provider-specific virtual ID model

In the last step of the set-up phase, the operator of the measurement service will send the trained data-provider-specific VID model to each participating data provider.

Live Campaign Measurement Phase

Figure 4. Overview of live campaign measurement phase



The second phase of the cross-media measurement is the live campaign measurement phase. This phase happens on a daily basis for the duration of each campaign. In this phase, data providers will label their campaign impression data with VIDs, and then transform the labelled impression data to a private format to be shared with the private reach and frequency estimator. This measurement service will then combine the private reach data from multiple data providers and supply standard reach and frequency metrics through an API(s) and/or default reports.

Step 4. Data Providers run virtual ID model on campaign impression data

After a campaign begins, digital and TV publishers will collect campaign impression data on a daily basis. This impression data will become the input of their data-provider-specific VID models. When data providers run the VID model, they will be able to assign a VID and reporting attributes (e.g., age and gender) to the data-provider-specific identifier (e.g., email address, cookie, device ids, etc.) associated with their campaign impression data. Since all the VIDs are drawn from the same virtual identity population, they become the foundation for deduplication across multiple data providers and media.

Once the raw campaign-level impression data is processed by the VID labeller, each impression will have been assigned a VID and a user segment. Next, the data provider will generate an aggregate representation of the VIDs associated with each combination of campaign and user segment. This aggregate representation is called a sketch, which in its raw form is not privacy preserving. Next, the sketch will be transformed into a private sketch in order to help preserve user privacy. This transformation, depending upon the details of the algorithm used to estimate reach and frequency, can entail either the addition of differentially private noise or encryption. Finally, once the private sketch has been built, the data provider can transmit it to the measurement service where it can be combined with sketches from other data providers.

It is worth noting that both the running of the VID model and the construction of sketches are performed by data providers within their own computing environments and that raw impression data will never leave their respective environments. Moreover, this process is straightforward and should not create any barriers to adoption for data providers.

Step 5. Reach and frequency estimator determines cross-publisher reach and frequency

Upon receiving private sketches from multiple data providers, the measurement service will combine the sketches for each campaign and segment and determine the deduplicated reach and frequency across data providers.

[Additional details will be added to this section once the exact method for computing reach and frequency has been established]

Step 6. Reporting APIs provide campaign metrics and forecasts

The next step of the live campaign measurement phase is for the measurement service to provide the standard reach and frequency reporting and APIs, which can be accessed by any licensed user subject to data governance and usage rules established by local governing bodies.

Additional details will be provided in a subsequent version as our user research track continues to make progress.

Global vs. Local Market Considerations

While some of the technical infrastructure can be built once and scale globally, local markets will still need to make critical decisions about the measurement system before implementing any design. This will ensure that the solution meets market-specific needs and that advertisers have flexibility and choice. We have captured high-level areas for local markets to review below. This list is not exhaustive. The WFA technical working group will continue to update this list on an ongoing basis, as needed.

Global forums, with local input, should determine:

1. Privacy principles to guide technical design choices.
2. Input metrics to the cross-media system that enables multiple comparison metrics for local markets.
3. Scalable technical infrastructure and required data inputs to produce desired outputs.
    a. Census Identification Framework *(note: specific VID model inputs will be determined at the local level)*
    b. Private Reach & Frequency estimator
4. Consistent campaign taxonomy to make sure campaigns, creatives etc., are globally identifiable across channels.
5. Advanced outcomes measurement approaches. For example, multi-touch attribution for offline sales measurement will require additional technical infrastructure. This can and should be designed in a global forum to make sure the solution will scale properly.

In addition to these design decisions, advertisers in local markets will have to define:

1. A common definition of an impression for comparisons across channels (e.g.,MRC cross-media video standard)
2. Key reach and frequency deduplication model inputs and parameters. For example, which panel should be used in a given market? How should reach models define total reachable population? Deciding and sourcing parameters that govern the training of the VID model, and other census identification techniques.
3. Governance & funding models, including data governance & usage. For example, who will deploy the VID model? Who will have access to the system's data APIs? Under what conditions?

Glossary

Census Identification Framework - A system for applying identifiers to census data in either a deterministic or statistical matter so that accurate measurement can be enabled.

Cryptography - A family of techniques that allows data to be encoded in such a way that only those parties who have the correct key can decode it. Without this key the encoded data is indistinguishable from random bits. One cryptographic technique, called homomorphic encryption, allows data (e.g.

numbers) to be combined (e.g. by summing) while encrypted such that the combined encoded data can be decoded while preserving the properties of the function that did the combination (e.g. the sum).

Data Provider - An entity that provides census data for reach and frequency estimation. This term is used in place of publisher and/or broadcaster and is meant to abstract away the difference between providers of various media types.

Differential Privacy - A privacy framework that provides a mathematical guarantee of whether anything can be learned about any individual whose data may (or may not) have been used as an input to an reach/frequency estimate.[1]

Double Blind Match - A process whereby multiple entities share data via a cryptographic protocol that allows comparison of data without divulging the actual data itself.

Measurement Phase - The part of the system that processes publisher census and produces reach and frequency measurements.

Re-identification - The process by which seemingly anonymous data is joined with other data sources in order to reveal the identities of individuals in the first "anonymized" data set. For example, poorly anonymized New York City taxi cab records revealed the comings and goings of the city's taxi cab drivers over the course of about 173 million trips.

Secure Universal ID (SUID) - A deterministic cross media user identifier. The SUID system uses multiple common user identifiers to securely and privately match users across different data providers.

Setup Phase - The part of the system that is concerned with collecting panelist data and training a Virtual ID model for deployment.

Virtual People ID (VID) - An identifier that is assigned to census data based on a panel-calibrated model. The VID model also assigns segment information to the census data.

F.A.Q.

Who is managing this? WFA is facilitating the production of this proposed Reach & Frequency Technical Blueprint as part of an initiative seeking to introduce a global Industry Framework for cross-media measurement. The Framework sets advertiser-centric but industry aligned principles for measurement which seeks to expedite the local implementation of a new wave of cross-media measurement solutions. Technological Infrastructure is a key component of the Framework and WFA has asked the industry to develop proposals which meet advertiser 'North Star' needs.

Who is involved? The WFA's process has been cross-industry and many companies have been involved. As a follow-up to the development of the WFA Industry Framework, companies participating in the

process were asked to collaborate with the WFA on a technical blueprint as a starting point for industry discussion.

Is this a Digital only solution? No. From the outset the group's intention was to include as many channels as possible for true cross-media measurement. To date, we have created a proposed technical blueprint which includes digital publisher data and TV data, and we will continue refining the design to include more detail on data from other channels.

What are the next steps? The proposed Reach & Frequency Technical blueprint will be made available to industry stakeholders (including advertisers, agencies, broadcasters and others) for peer review, with the express intent of developing a high-level industry-endorsed blueprint, which could form the basis for more detailed local market consideration and ultimately implementation. In addition, ongoing workstreams, which include those needed to support outcomes measurement, are required.

The WFA and its members welcome this work and our request to all stakeholders is to engage in the peer review process, and to help further develop this draft design. Only through broad industry collaboration can we successfully serve the needs of the entire ads ecosystem.

Is this a global solution? The purpose of the Reach & Frequency Technical blueprint is to provide a proposed framework for local cross-media measurement solutions. In and of itself it is not a solution and requires considerable industry engagement (global and local) and adaptation before it could become such. But the intention is to create as much global consistency as possible to address advertiser needs. That's why WFA has been facilitating this process alongside key advertiser associations and global advertisers.

What is the origin of this? The proliferation of number of players in ecosystem, technologies, and increasingly regulations are causing significant change. As such we recognize the need to balance the consumer needs better while maintaining a healthy ad ecosystem with good value exchange / accountability etc.

For additional FAQs please link here

Link to other docs

- WFA Cross Media Measurement Framework draft

References

1. The Algorithmic Foundations of Differential Privacy (2014)
2. A Method for Measuring Online Audiences (2013)
3. Measuring Cross-Device Online Audiences (2016)

4. Data enrichment for incremental reach estimation (2014)
5. Virtual People: Actionable Reach Modeling (2019)