

# The Halo Cross-Media Measurement Framework



# Foreword

For as long as advertising has been around, an important question for marketers has been to understand how it is performing. It's never been an easy problem to solve but in today's complex media ecosystem, with rapidly changing consumer behaviours as well as with the proliferation of media channels, it takes on many new dimensions. A key one of those is Cross Media Measurement (CMM).

Advertisers seek a better understanding of the effectiveness of their media investments across a number of traditional and new media forms. We want to avoid audiences being excessively exposed to our ads while our consumers demand increased transparency and control over their data. In short, CMM is an extraordinarily important utility.

Yet, our ability to conduct this has been rather limited, being rooted in the bygone era of a much simpler media landscape. It's our view that the panel approach, which has traditionally underpinned audience measurement, alone, is not sufficient for purpose. Likewise, techniques which are over-reliant on identifiers, like cookies and device IDs, are not in-step with the evolution of our ecosystem towards 'privacy-first' and comprehensiveness. Measurement systems had to evolve.

So we've sparked an extraordinary effort which, although not complete yet, is surely one of the industry's most successful collaborations. From the development of North Star principles to the identification of a technological concept, to today - where we are now code-complete on the first full release of a new cross-media measurement system – we are driving change. And in this new system we are advocating for a hybrid approach, combining panels and census data, in privacy-safe and scalable ways.

It's an incredible journey, and a significant achievement by anyone's yardstick. The core concepts we've developed provide advertisers with the means to plan and optimise their media investments with more ease and accuracy, paving ways to revolutionise the way media is measured.

So it is very exciting to see this system now being deployed by two pioneering advertiser associations, ANA and ISBA/Origin, in two of the prominent markets, the US and the UK, respectively. While these two local pilots have yet to reach maturity, they are both showing considerable promise.

As a design principle, and as ANA and ISBA/Origin have shown it working in practice, it is up to local market groups to decide if and how they want to follow the routes available under the Halo CMM system, towards their own local deployment. Markets will need to go through similar processes: building governance groups, potentially standing up new panel assets and deploying the Halo code to local environments. Not necessarily a straight-forward exercise, let's admit. But

building on the back of the experience acquired from these two initial deployments we are looking forward to what can be scaled to other markets.

And beyond Reach and Frequency, we're excited to see how we can further develop the core measurement technologies, potentially establishing the means to measure other outputs, and even outcomes.

This journey is just getting started and there's much more to come. As ever, we encourage all actors in the industry to join us on the mission towards enabling a modern and holistic approach to cross-media measurement, that will serve not only the advertisers or the media owners, but most importantly, our consumers!

**Atin Kulkarni, VP Global Media & Commercial Capabilities, PepsiCo**



# Executive Summary

Supported by global brands and national advertiser associations, including ANA and ISBA, WFA has been facilitating a powerful programme of work ('Halo') designed to expedite the implementation of a new wave of cross-media measurement solution, globally. The programme has been participated in by partners from across the ecosystem.

The Halo Framework incorporates an innovative set of technologies and software components that empower local markets to use data assets like panels, exposure events, non-census measurement data, and other inputs to create measurement reports that are available via a set of Halo APIs. The resulting measurement is intended to meet the needs of the most ambitious advertiser use cases, and the demands of media owners for content measurement.

The initial objective is the secure computation of reach and frequency measurement reports, but with a roadmap that includes outcomes measurement.

The Halo Framework is built around two technological pillars: (1) the Virtual People Framework; and (2) the Private Reach and Frequency Estimator (PRFE). Halo's Virtual People Framework provides a distributed method for calibrating and mapping raw measurement data, including digital events and non-census measurement data alike, onto a unified census-level representation of the population under measurement.

Three core data assets are required to achieve this: (1) a population Enumeration Survey of the population under measurement; (2) panel(s) for modelling and debiasing; and (3) when available, raw exposure event data, first party identifiers, and demographic profiles for the panellists (used for modelling) and for all users (used for measurement) from as many sources as possible. Additional non-census data, such as from set-top boxes or smart TVs, may also be used if desired, and access to a single-source panel is preferred.

The Private Reach and Frequency Estimator accepts the output of the Virtual People Framework and produces privacy preserving estimates of cross-media/publisher reach and frequency. This is done with a secure multiparty computation (MPC) protocol, which allows for the computation of desired cross-party outputs while guaranteeing the secrecy of all inputs and intermediate values through advanced cryptographic methods.

The Halo Framework is not a global centralised measurement service. Rather the Halo Framework and its constituent software components, which are available under the Apache 2.0 Licence, are intended to be deployed under the auspices of a local market as a Local Measurement Service. The open source software code is a critical foundation of the service but significant local effort will be required to address the issues of governance, commercials, contracts and auditing, not to mention the technical and operational customisation of the Halo Framework, given local data assets and market requirements.

At first reading, the technical infrastructure that surrounds the core hybrid measurement methodology may appear overly complex. However, the system is the first of its kind to deliver privacy-first media measurement while enabling the secure and fair use of census data combined with first party identifiers and demographics. If available, such data, corrected by the panel assets at the heart of the Halo Framework, help overcome the limits of those very same panels in the face of fragmented and personalised media. As such, the Halo Framework passes the parsimony test of Ockham's razor.

This document outlines what is included in the Halo Framework and proposes how local markets can get started to evaluate and deploy their Local Measurement Service, if they decide to use it.

<b>Executive Summary.....</b>	<b>4</b>
<b>What is Halo?.....</b>	<b>9</b>
Industry Consortium.....	9
Framework.....	9
<b>WFA Industry Principles.....</b>	<b>12</b>
Advertiser North Star Requirements.....	12
Advertiser Supported Industry Requirements.....	12
Privacy Principles.....	13
Data Security.....	14
Prioritisation of Capabilities.....	14
<b>Technology Pillars.....</b>	<b>16</b>
Virtual People Framework.....	16
Population Modelling.....	17
Virtual People Modeling.....	18
Virtual People Labelling.....	19
Non-Census Data Processing and Fusion.....	20
VID Pilot Results.....	21
Private Reach and Frequency Estimator.....	21
PRFE Inputs.....	22
PRFE Output.....	23
Privacy Budgeting.....	23
PRFE Results.....	24
Balancing Accuracy, Privacy, and Performance.....	24
Privacy Principles in Action.....	24
<b>Design Overview.....</b>	<b>26</b>
System Architecture & Roles.....	26
System Flows.....	30
Setup Phase.....	30
System Setup.....	30
Enablement.....	31
Metadata Distribution.....	31
Training Phase.....	32
Panellist Data Exchange.....	32
VID Model Training and Distribution.....	33
Measurement Phase.....	34
VID Labelling.....	34
Measurement.....	35
Putting It All Together.....	36
The EDP Aggregator.....	37

<b>Local Measurement Service Deployment.....</b>	<b>39</b>
Governance.....	39
Commercial.....	40
Technical.....	40
Legal.....	41
Audit.....	42
Timeline.....	43
<b>Conclusions &amp; Next Steps.....</b>	<b>44</b>
Setting the goal.....	44
Delivering on the goal.....	44
Next Steps for Halo.....	45
Next Steps for Local Markets.....	45
Future Directions of Halo Framework.....	45
<b>References.....</b>	<b>47</b>

# What *is* Halo?

Throughout this document ‘Halo’ is principally used to describe the technologies underpinning the new Framework. But Halo is also used to refer to the industry consortium that has built this schematic, initially in service to the two markets leading local pilots of the tech.

## Industry Consortium

Companies involved in the development process to date include advertisers, advertiser associations, digital platforms, measurement companies and others.

ABinBev	Mars
Accenture	Mastercard
ACA (Canada)	Meta
Amazon	Media Rating Council (MRC)
ANA (US)	Nestle
The Coca-Cola Company	OWM (Germany)
Comscore	PepsiCo
Deutsche Telekom AG	P&G
General Motors	TikTok
Google	Unilever
IAB Tech Lab	Videoamp
ISBA (UK) / Origin	WFA
Kantar	

After two years of development work the Halo group is pleased to announce that the first release of the open-source [Halo Cross-Media Measurement System](#) (CMMS) is available under the [Apache 2.0 software licence](#).

## Framework

The Halo Framework employs a new set of technologies – primarily a Virtual People Framework and a Private Reach and Frequency Estimator (PRFE) – to enable comprehensive, always-on, privacy-preserving cross-media measurement. These technologies allow the system to deduplicate reach and frequency across multiple data providers’ (e.g. broadcasters and publishers) census data and non-census data alike, while using a panel for calibration and correction. Outputs are computed securely and in a way that preserves user privacy, without compromising accuracy.

However, the Halo Framework is not just a reference architecture and set of technology recommendations. Today it provides a set of documented ‘common components’ that include both software libraries and fully deployable systems that implement these technologies.



Collectively, Halo's deployable components are called the Halo Cross-Media Measurement System (CMMS).

The CMMS orchestrates the interactions between participants that provide inputs like panels and event exposures, and others who desire measurement reports. To achieve this the CMMS provides various mechanisms for collecting the inputs and transforming them into the desired outputs. This broader framework that is composed of the CMMS and the participants with their inputs and desired outputs is called the Halo Cross-Media Measurement Framework (Halo Framework). To simplify integrations, the Halo Framework also provides a set of software libraries, documentation, and support channels.

The inputs to the CMMS and the outputs it produces allow for a range of configurations that can be determined by each market's Local Measurement Service. A Local Measurement Service is operated by a local market under its own governance, commercial, auditing and contractual terms, keeping a view of global best practices in mind. These relationships are summarised in the following diagram.

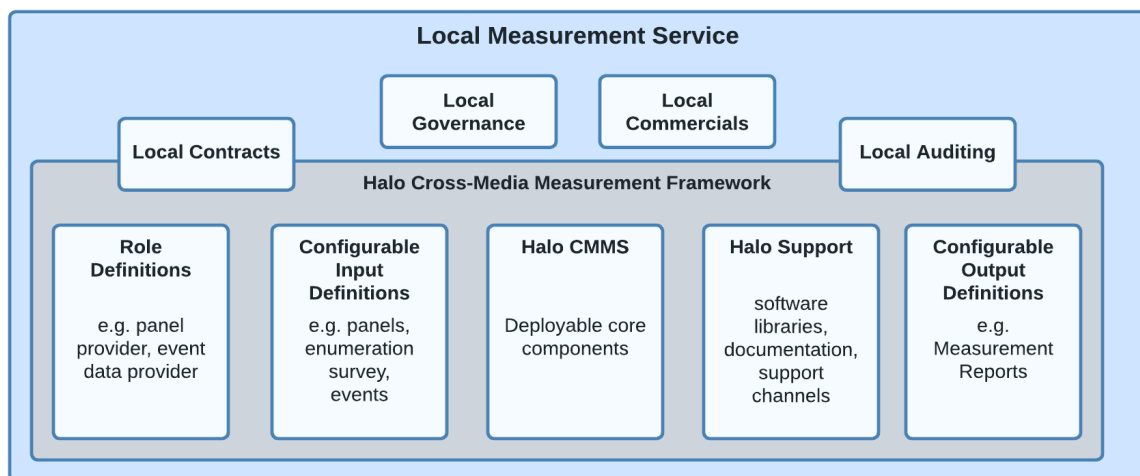


Figure: The Halo Framework situated within a broader Local Measurement Service

This document provides a high-level view of the Halo Framework and its CMMS, which includes both technological and non-technological aspects. The Halo Framework is designed to be deployed in multiple markets, so local markets need to understand what decisions are required to correctly deploy the Halo CMMS.

Therefore Halo is working on the next level of documentation that is split into two parts:

- **Halo Core Specification:** describes the roles and responsibilities of participants in any Local Measurement Service and the decisions that must be made to successfully deploy the CMMS locally.

- **Halo Local Specification:** provides a template that allows a local market to consolidate all the local decisions and configurations in one place.

These documents are introduced here so that the reader can appreciate in the following sections how local market deployments require local decisions, and allow for local market variations.

This document continues with a review of the requirements that drove the creation of the Halo Framework. We then move on to an overview of the technologies that power it. This is followed by a more detailed dive into the Halo Framework's architecture. We conclude with a discussion of local market deployment considerations and proposed next steps.

# WFA Industry Principles

In the [WFA Industry Framework, Establishing Principles For A New Approach To Cross-Media Measurement](#) articulates a set of *North Star* advertiser and advertiser supported industry requirements for cross-media measurement. These are summarised below.

## Advertiser North Star Requirements

**Full Life Cycle Measurement:** enable all phases of measurement

- Pre-campaign audience planning
- Intra-campaign audience and frequency management and optimization
- Post-campaign audience evaluation

**Continuous:** always-on data capture, no buy-side tagging required

- Advertisers who have opted-in and met other requirements as specified by the local market (e.g., subscriptions, cleared legal and data sharing aspects, etc.) can access measurement on an ongoing basis, rather than campaign-by-campaign, limited-duration, fee-based campaign tracking.

**Comprehensive:** cross-media reporting across all media formats

- Full comparable cross-channel measurement, including linear TV and all digital formats inclusive of all digital and traditional media platforms and publishers regardless of size, relationship to users, technical expertise, etc.
- Measurement of an entire campaign without regard to media format
- Applicability to all major global markets with the ability to adapt to market-specific innovations

**Full Funnel:** reach, frequency management and outcomes

- Deduplicated reach and frequency of media campaigns
- Integration of outcomes measurement to enable media audience and related analytics, such as attribution modelling, media mix modelling, brand lift, and sales lift

## Advertiser Supported Industry Requirements

**Privacy Centric:** respect the consumer and safeguard user privacy by incorporating the highest level of technical privacy standards and guarantees in order to meet the privacy requirements of today and to provide a development roadmap for meeting them in the future

**Fair & Objective:** provide apples-to-apples comparison across TV and digital advertising through technology that solves for cross-media, open standards, and a neutral governance model

**Global Trust & Transparency:** technical design and implementation are sufficiently transparent to build trust in the measurement service via open-source implementations, audits and verification, etc.

**Advertising & Content:** capable of supporting both ads and content measurement with priority given to ads measurement

## Privacy Principles

Privacy principles have a deep influence on technical design choices and use of specific privacy-centric technologies is required to meet current and future regulatory requirements, and evolving user expectations. The Halo Framework was designed to:

- minimise the risk of re-identifying consumers;
- allow users to control the collection and use of their data;
- protect panellist identity;
- allow for compliance with applicable global and local privacy laws and regulations.

As such, the Halo Framework does not rely on identity graphs, browser fingerprinting, or third party cookies, although it is compatible with all of these.

### Re-identification

[Re-identification](#) is the process by which records in a de-identified data set can be linked to specific individuals by combining them with records from another dataset, and is a risk that many industries face today. While it is impossible to fully eliminate the possibility of re-identification, to minimise its likelihood, data providers must have both strong contractual protections and quantifiable technical guarantees that guard against re-identification. As such, the Halo Framework incorporates technical measures like differential privacy and secure multiparty computation to help ensure that no participating entity learns more identifying information about individual users than the entity had before participating in this system, and that the ability to re-identify users is minimised.

### User Control

Data providers should be able to provide their users with transparency and control over the collection and use of their data as it pertains to its availability in the measurement system. The Halo Framework has allowed for this through the adoption of measurement technologies that allow user data to remain on systems that are either controlled by the data provider itself or when that is not feasible, a trusted delegate. The system also employs technical guarantees that prevent the data from being used for anything other than verified measurement use cases.

### Special Consideration for Panellists

Panellists are those individuals and households who have provided consent for data providers to share their data for measurement. As such, to ensure fair and objective measurement, panellist identities must not be divulged to any data provider, save for those that are responsible for the panel itself and, when necessary, for the construction of reach models by a model provider. As with re-identification, the Halo Framework assumes a set of strong contractual

protections and quantifiable technical guarantees to ensure this, and further assumes that any controls available to non-panelist users are available to panellists as well. Finally, as the technology progresses, we look forward to model training techniques that would preclude even the model provider from learning panellist identities.

### Privacy Laws and Regulations

While the Halo Framework has been constructed with an eye toward complying with applicable global and local privacy laws and regulations, whether the solution actually achieves this is left to local markets to determine as their particular circumstances warrant.

## Data Security

As built, data from publishers, including campaign metadata, event data, and any derived aggregate, may only be used for the purposes of enabling specific cross-media advertising measurement use cases for individual advertisers, agencies, or other authorised users as determined by local markets. Similarly, once computed, outputs of the system are only made available to those authorised users (or their delegates) who requested them. Achieving this currently requires that each advertiser or agency opt-in to using a Local Measurement Service deployment of the Halo CMMS with each participating data provider. Additional mechanisms are being explored to make this process less onerous.

Once opted-in, a cryptographic consent signalling system, the details of which are beyond the scope of this document, ensures that data can only be accessed and decrypted by authorised parties.

## Prioritisation of Capabilities

As part of the construction of the initial Halo [2020 Blueprint](#), the below requirements were consolidated and divided into three categories: foundational features, reach and frequency use cases, and an advanced feature set to include outcomes. The following stack rank was the result of this exercise:

	Requirement		Description
<b>Foundational Feature Set</b>	1	Always-on	Tagless implementation; begins from advertiser opt-in; forecasting based on available data
	2	Data Granularity (e.g. for viewability)	Sufficiently granular data to comply with relevant industry standards, report across all relevant metrics etc. while respecting privacy principles, and applicable laws and regulations
	3	Basic Segments	Creative, geo, device, placement, campaign, age, gender, format, household income

	4	Output Metrics	Deduplicated reach, frequency, impressions, watch time
Delivery 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	12 month forecast; can be based on historical data to start (inventory data from publishers added as available); includes daily scenario planning across channels to optimise R/F (this is the mechanism to manage frequency) * note: frequency management is not a centralised targeting approach
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	First Party (1P) Data	Custom segments based on advertiser data (reporting is top priority, forecasting secondary)
	9	Outcomes	All effectiveness measures (brand/sales lift are top priority; conversions, MMM, MTA secondary)
	10	Advanced segments	Psychographic, behavioural audiences based on common third party (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.

We are pleased to report that the progress on this list has been substantial and, as of April 2023, the Halo CMMS supports appropriately granular, always-on reach and frequency metrics for the set of basic segments, while an additional component supports the R/F reporting capability. APIs are provided for all functions. We will return to this list in the last section of this document, when considering Halo's future directions.

# Technology Pillars

The Halo Framework is built around two technological pillars: (1) the Virtual People Framework; and (2) the Private Reach and Frequency Estimator (PRFE). Together these technologies allow the system to offer accurate measurement of reach, frequency and other key metrics, while simultaneously preserving consumer privacy. The next two sections explore each of these pillars in more detail.

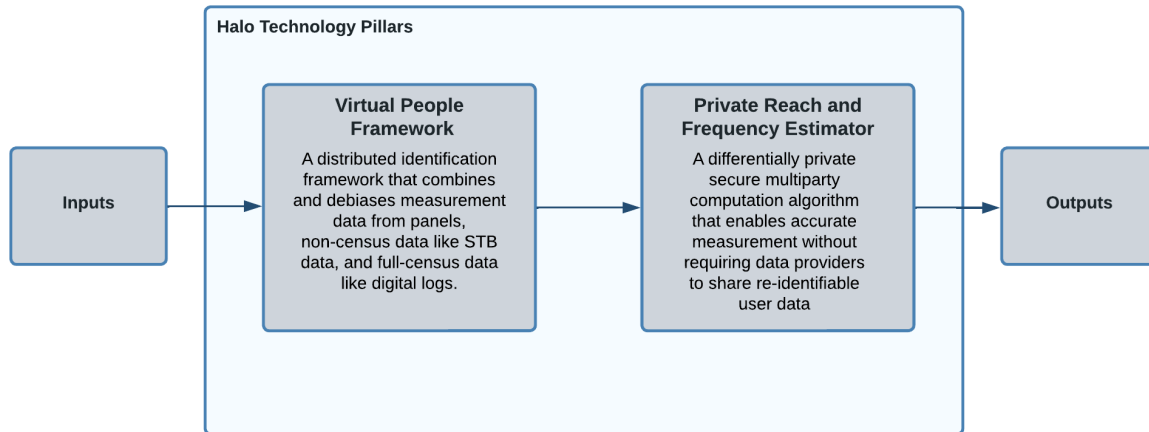


Figure: Halo Technology Pillars

## Virtual People Framework

A common method for measuring reach in the target population is desired to compute deduplicated reach across data providers and media types. This method must allow data providers (also known as Event Data Providers or EDPs) to combine their exposure data at the census-level, and this process must be consistent across all channels. The method must also be capable of accommodating *any* raw measurement data including single-source panels, measurement panels, non-census sources like set top box (STB) data, and full-census sources like digital event logs. Achieving this while providing a high quality solution that adheres to the above privacy principles is a considerable challenge.

Halo's Virtual People Framework (VID Framework) consists of several processes, the overall job of which is to calibrate and map raw measurement data, including digital events and non-census measurement data, onto a unified census-level representation of the population under measurement. Three core data assets are required to achieve this: (1) a population enumeration survey of the population under measurement; (2) one or more panels for modelling and debiasing; and (3) when available, raw exposure event data, first party identifiers, and demographic profiles for the panellists (used for modelling) and for all users (used for measurement) from as many sources as possible. Additional non-census data may also be

used if desired. Note that use of a single-source panel is preferred. The diagram below shows the overall flow, which is expanded upon in the next several sections.

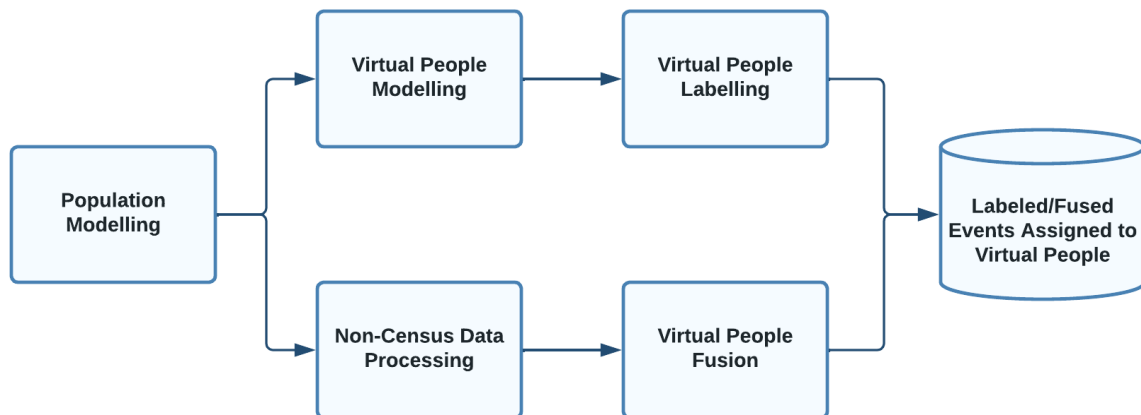


Figure: The Virtual People Framework

## Population Modelling

The Virtual People Framework uses the Universe Estimates from the Enumeration Survey to generate a model of the population under measurement. In this process an identifier, which is called a Virtual Person ID or VID, is generated for each individual in the target population. The VID is then assigned demographic attributes and other metadata, for example a geographic location so that the distribution of the characteristics of VIDs match those of the population under measurement. Note however, that the population of VIDs is synthetic, which means that while all individuals in the population are *represented* by the set of VIDs, *there is no correspondence between any particular VID and a real person*. The output of this step is a Virtual Population that represents the population under measurement.



Figure: Population Modelling



## Virtual People Modeling

A Virtual People Model (VID Model) is created by combining the Virtual Population, the panel(s), and the panellists' raw exposure event data from all available sources, along with any other desired measurement data. All panellists should be associated with the participating data providers' ad and/or content exposure event information, including user device identifiers, and other contextual and demographic information useful for developing measurement models. To the extent that multiple panels are available (e.g., a single-source panel and other single-media panels), [they may be combined](#).

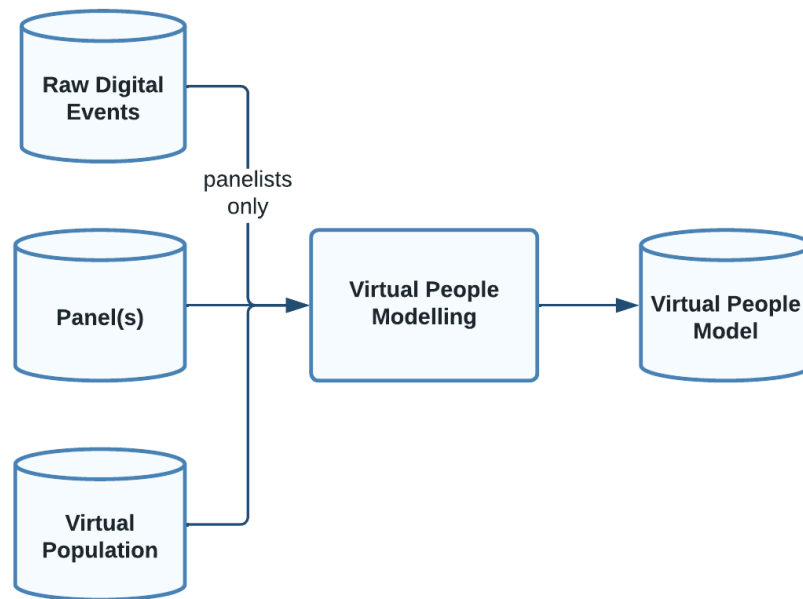


Figure: Virtual People Modelling

To ensure completeness of exposure event information, the Halo CMMS requires that all panellists consent to having their digital exposure event logs shared with the Panel Provider. Digital exposure event logs for all panellists can then be provided by participating Census Event Data Providers (an Event Data Provider that provides census data to the CMMS), whose logs are queried by the Panel Provider via a double-blind cryptographic protocol. The protocol ensures that Census EDPs do not learn the identity of any panellist, while third-party audits ensure that Panel Providers do not query for logs to which they are not entitled. Throughout this process only data that belongs to consenting panellists leaves the Census EDP's environment. Additional details are provided in the Design Overview below.

When combined, the panel data and the census logs allow the modeller to learn the relationship between email addresses, device IDs and other identifiers for panellists across all media properties, the details of which are used to construct a VID Model that faithfully reproduces the

reach curve for each media property. As part of this process, demographic data associated with the panel serves as ground truth to debias the demographic signals present in the digital event data. During this process the relationship to non-census measurement data may also be considered.

In summary, common steps for building VID Models are:

- de-biasing census data and data providers' demographic information
- determining the relationship between user IDs and people across sources, and
- determining how media consumption varies across different devices and media.

## Virtual People Labelling

Virtual People Labelling uses the Virtual People Model to label raw digital event data with VIDs. During this process, each event record and its accompanying exposure metadata is labelled with one or more VIDs, where multiple VIDs may be produced to accommodate co-viewing or account sharing scenarios. The result is that each record is labelled with one or more VIDs, which can be easily counted and deduplicated.

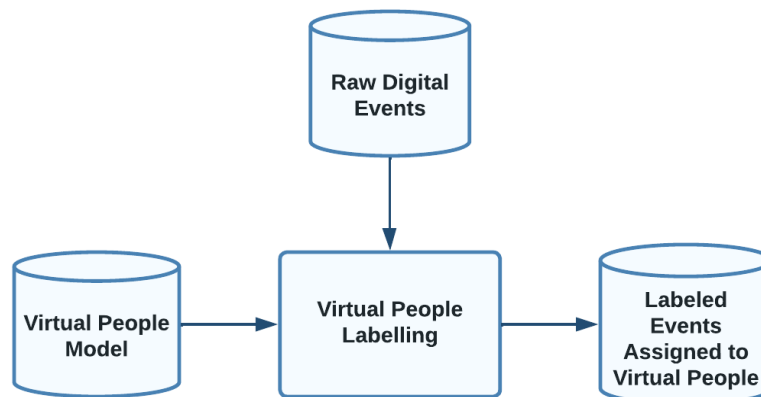


Figure: Virtual People Labelling

A key property of Virtual People Labelling is that multiple parties can apply it independently. In fact, the Virtual People Model is actually a set of models, one for each Census EDP, which means that distribution of these models to the Census EDP for which they are intended does not allow that Census EDP to learn anything about any of the other Census EDPs participating in the overall system. An important feature of this distributed labelling framework is that Census EDPs can ensure that their event-level data (with their associated first-party identifiers and demographic profiles) stay on systems under their control, while third-party audits are used to ensure Census EDPs faithfully label their events.

The labelling process has the following characteristics:

- All exposures are assigned one or more VIDs, meaning the entire census can be used for measurement (assuming that the necessary user consent has been obtained)
- Labels, while probabilistic in nature, are applied deterministically, meaning that for any specific input the output is always the same.
- An input identifier, while often mapped to the same VID, may be mapped to different VIDs depending upon the other metadata associated with a particular exposure.
- Similarly, different input identifiers can be mapped to the same VID.
- While a set of VIDs can be used to compute reach estimates, any particular VID has limited usefulness. For example, individual VIDs can not be used effectively for targeting, explicit audience definition or attribution.

## Non-Census Data Processing and Fusion

Non-census data can take many forms. Two of these are Linear TV records and measurement panels, where Linear TV records include STB data, automated content recognition (ACR) data, and similar. The Virtual People Framework expects that all non-census data is obtained through the usual channels. Once obtained, this data is processed, appropriately weighted, and personified. The result is a set of weighted exposure events.

The next step is Virtual People Fusion, which combines the weighted exposure events with the Virtual Population. The result of this is a data set where each weighted exposure event record is labelled with a number of VIDs according to the weight of the record. Thus non-census and digital census exposure data are mapped onto a common set of VIDs that can then be easily counted and deduplicated to arrive at total impression counts, reach and frequency.

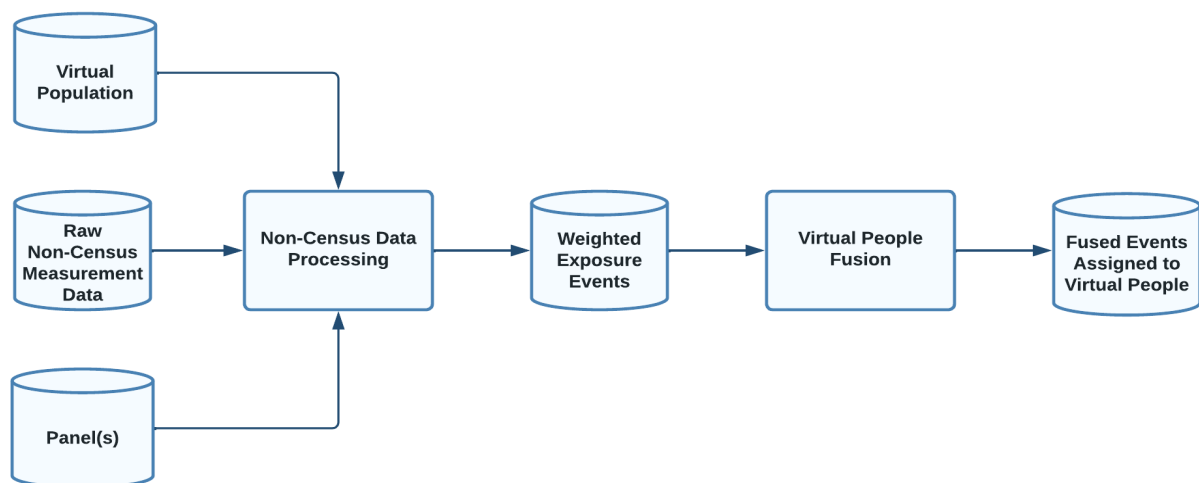


Figure: Non-Census Data Processing and Fusion

## VID Pilot Results

Since the Cross-Media Measurement Working Group published the tech blueprint in 2020 several VID pilot studies have been conducted. The first was sponsored by ISBA/Origin and [conducted in the UK](#). The UK study deployed the VID Framework to [model online and linear TV campaign reach and frequency](#). Results from the study found the VID Model to be an elegant solution, [capable of closely replicating panel results and correlations whilst preserving the TV currency](#). Origin in the UK, supported by Kantar and Accenture, will be conducting live Alpha and Beta tests of the Halo Framework with a number of data providers and advertisers across 2023.

VideoAmp and Comscore carried out two more studies for the ANA with [positive initial results](#). The VideoAmp results showed that the VID Framework is capable of measuring linear TV in the US. Comscore found that the VID Framework is a robust approach for content audience measurement in the digital environment without third-party identifiers. Comscore, on behalf of the ANA, have also run the first of a number of live-data pilots for both linear TV and digital campaigns that tests the entire Halo Framework, including the VID Framework. We look forward to sharing the results as they become available, which will be posted, along with other research papers, on the WFA's [GitHub project site](#).

Overall, the VID Framework provides a distributed architecture that allows EDPs to independently assign VIDs to their exposure data. Thus, outside of the training steps outlined above, an EDP need not share any exposure data outside of their systems. The VID Framework itself, while stochastic in nature, is far from random, and is capable of providing high quality reach and frequency estimates that are consistent across all dimensions including demographic segments, device types, geographic location, campaigns, and time. What's more is that the VID Framework is capable of consuming any type of identifier as input, and is therefore likely to be compatible with technologies like the OpenAP OpenID framework. A description of the mathematics underlying the Virtual People Framework can be found in "[Virtual People: Actionable Reach Modeling](#)."

## Private Reach and Frequency Estimator

The second pillar of the Halo CMMS is the Private Reach and Frequency Estimator (PRFE). And in spite of its name, the PRFE can produce reach and frequency measurements as well as other metrics like impression counts and watch duration. The purpose of the PRFE is to accept labelled exposure events from multiple sources, deduplicate them, and produce privacy preserving estimates of cross-media/publisher reach and frequency. The PRFE achieves this by cryptographically protecting each EDP's input, thereby ensuring that no other party to the system learns it. The EDP inputs are then combined without ever being decrypted and a privacy preserving estimate of the cross-EDP deduplicated reach and frequency is produced. This is done using secure multiparty computation (MPC), which is an advanced cryptographic technique that allows multiple parties to contribute inputs to compute a mutually desired output

without having to divulge the values of their inputs. This process is shown graphically in the figure below. The following sections will describe each component in more detail.

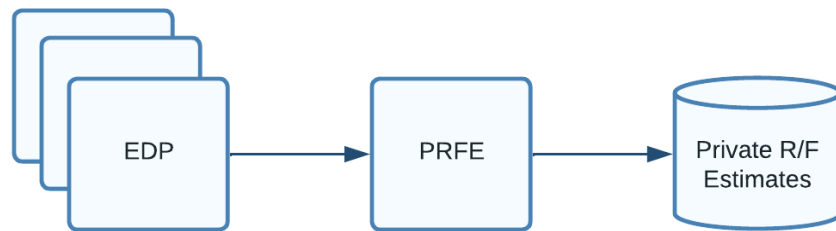


Figure: Private Reach and Frequency Estimator

## PRFE Inputs

EDP exposure data for any given measurement can be extremely large, consisting of hundreds of millions of events or more. At this scale it is simply not feasible to deduplicate at the level of each individual event, and several measures have to be taken to make deduplication tractable. The following diagram, which is explained below, shows the process an EDP goes through to provide an input to the PRFE.

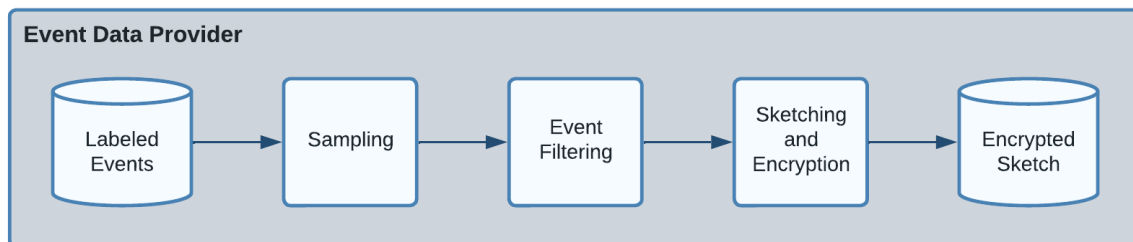


Figure: Event Data Provider Input Generation

First, the EDP samples its set of events, which significantly reduces the number of VIDs that must be sent to the PRFE. The sampling rate must be chosen to properly balance scaling concerns with the ability to deliver an accurate measurement. Sampling is also used to conserve privacy budget, which will be discussed in further detail below.

Next, EDPs filter the sampled events and export only those events that are required for a particular computation. For example, an EDP may be requested to filter its events by a desired age range. Moreover, once filtering is applied, an EDP only has to send the values of the VIDs

for those events that passed the filter. Thus, filtering trims the number of events that must be sent and allows an EDP to avoid sending any event metadata.

Finally, the filtered set of VID's are compressed into a memory efficient data structure called a sketch. This VID sketch is then encrypted and sent to the PRFE for further processing.

## PRFE Output

After each EDP creates its input and sends it to the PRFE, the PRFE begins its job of computing the deduplicated reach and frequency of those inputs. As previously mentioned, this is done with an MPC protocol, which means that the PRFE is actually a collection of independently operated systems that work together to compute the desired output. The Design Overview will describe these systems in more detail.

An essential part of computing the output is ensuring that it adheres to our previously stated privacy principles. In practice, this means that any output produced by the PRFE is differentially private. Simply stated, this means that, within a threshold defined by a privacy parameter, the presence of any particular user's data cannot be inferred from the output. Differential privacy is achieved by adding noise to the output, which the PRFE implements as an intrinsic part of the MPC protocol. This means that even the PRFE does not learn the denoised value of the output. For those that wish to learn more about differential privacy, "[Why Differential Privacy is Awesome](#)" provides a great non-technical introduction, while a mathematical introduction is provided by [The Algorithmic Foundations of Differential Privacy](#).

## Privacy Budgeting

For differential privacy to be practical, the amount of noise applied to any given output must be balanced with the error incurred from doing so. This balance is called the privacy-utility trade-off. Utility is measured using typical measures of error, while privacy-loss is measured using one or more privacy parameters, which are real-valued numbers that bound the probability that a particular user's presence can be detected in a given output.

Privacy loss, while applicable to individual metrics, also applies to the entire data set used to compute those metrics, and accumulates as the data set is used to compute additional metrics. Specifically, the privacy-loss associated with each output metric is deducted from an overall privacy budget associated with the underlying input data set, such that when the privacy budget for that data set has been exhausted no further outputs can be calculated using it.

A fundamental question about privacy budgeting is the unit of account, which could be defined in terms of a range of time, a particular campaign or program, an advertiser, or similar kinds of groups. Currently, the recommended unit for privacy budget accounting as it relates to ads measurement is an advertiser campaign day. This means that for each advertiser there would be a unique budget for each day of each campaign. We believe that this unit strikes a

reasonable balance between privacy loss and utility, and research has shown that tens of thousands of reasonably accurate queries can be performed using this unit of budgeting.

## PRFE Results

The algorithms used by the PRFE were arrived at after a substantial amount of research. The first of these was the sketching algorithm, which was selected in late 2020 based upon the [results](#) of the WFA's [Cardinality Estimation Evaluation Framework](#) ([summary deck of results](#)). In early 2021, a reference implementation of the MPC protocol was provided and a detailed description of it [was published](#) later in year in the Proceedings of the Privacy Enhancing Technology Symposium. Research in this area continues and we expect significant improvements to the Halo MPC technology over the coming months and years.

On behalf of the ANA, Comscore have run the first of a number of live-data pilots for both linear TV and digital campaigns that tests the entire Halo Framework, including the PRFE. Origin in the UK is planning to run a similar testing of the entire Halo Framework in 2023.

## Balancing Accuracy, Privacy and Performance

Through the use of sampling, sketching, and differential privacy noise, the Halo PRFE seeks to strike a balance between accuracy of measurement, user privacy and cost. Several parameters can be tuned to find the right balance for each deployment. Their values will depend on both the local market's thresholds for accuracy and cost, as well as the size of the population being measured. Finding the right balance can be tricky, and we anticipate providing utilities for making these decisions.

The Local Specification will outline the configuration decisions to be made by the local market. Learnings from the proof of concepts in the US and UK will also be shared to inform these choices.

## Privacy Principles in Action

The above technology pillars provide the core capabilities for achieving Halo's privacy principles. Specifically:

- The Virtual People Framework allows EDP census logs to remain on systems that the EDP controls with a notable exception for consented panellists whose logs are supplied to Panel Providers.
- Since VID sketches can reveal information about users, the MPC protocol ensures that even the sketches cannot be read individually. Thus an EDP's input is known only to itself.
- The MPC protocol ensures that all outputs have had the requisite amount of noise applied so that the likelihood of inferring the presence of any particular user's data in the output is mathematically bound.

In short, the distributed nature of the Virtual People framework combined with MPC and differential privacy form the core of Halo's privacy-centric approach.



# Design Overview

This section describes how the technology pillars described above are realised in a model implementation of the Halo Framework. We start with an overview of the overall architecture and operational roles, which are performed by participating companies that have been selected by or registered with the Local Measurement Service. This section also shows how the Halo CMMS components and software libraries are oriented within the overall Halo Framework. We then proceed to discuss several key flows.

## System Architecture & Roles

The diagram below shows the two major components of the Halo CMMS and the roles that participate in our model instance of the Halo Framework. As we progress, this diagram will be updated to illustrate the relationship between the various components and roles across several flows.

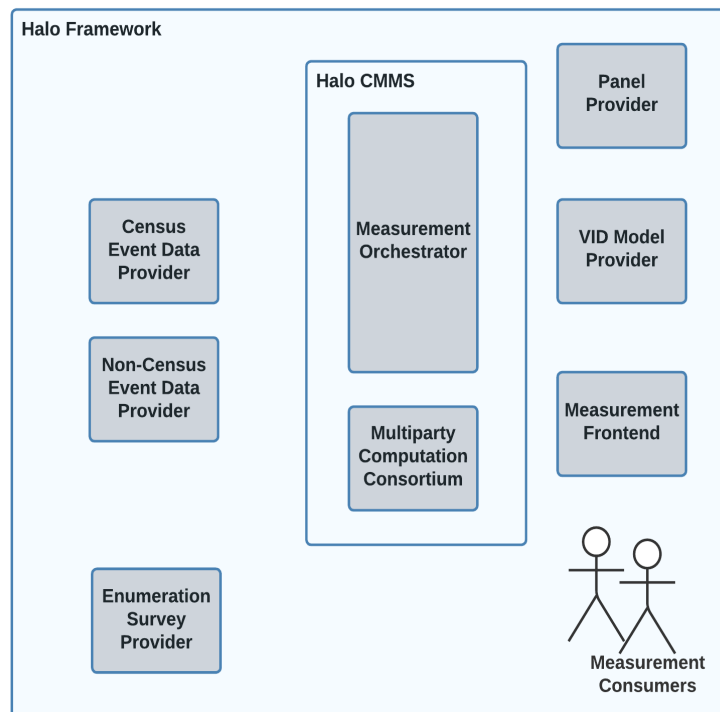


Figure: Components and roles of a Local Measurement Service

The Halo Framework's software components are as follows:

- **Measurement Frontend.** This is a set of APIs and UIs that provide an easy to use interface for the Measurement Consumer (see definition below) to access the

Measurement Orchestrator. For any given deployment one or more measurement front ends may exist. However, without loss of generality, we assume the existence of just a single frontend in this document. Halo provides reference implementations of several frontend services, which are housed in the Halo Reporting Server. It includes basic Reporting and Metadata APIs; however, at this time it does not provide a UI.

- **Measurement Orchestrator (MO).** This is the hub of the Halo CMMS. Its job is to mediate the relationships between the rest of the components and roles and to coordinate interactions between them. It is also sometimes called the Measurement Coordinator or the Kingdom.<sup>1</sup> The Measurement Orchestrator is deployed in a single Kubernetes cluster and consists of a collection of microservices and other jobs. It is expected to be operated by the local measurement body.
- **Multiparty Computation (MPC) Consortium.** This is actually a collection of systems, which together provide the computation engine for the Halo CMMS. Each component of the MPC Consortium is called an MPC Node or a Duchy. MPC nodes can be thought of as subsidiaries of the Measurement Orchestrator. Each MPC Node is deployed in its own Kubernetes cluster and consists of a collection of microservices and other jobs. MPC Nodes can be of two types, Aggregator or Worker, and there is exactly one Aggregator and at least one Worker in the overall Consortium.<sup>2</sup> Anyone can operate an MPC Node, but the Aggregator must *not* be operated by an Event Data Provider.

The Halo Framework roles are:

- **Panel Provider.** One or more entities that provide panel(s) suitable for training a VID Model and/or providing media measurement.
- **VID Model Provider.** An entity whose responsibility it is to train a VID Model. Typically this entity is referred to as just the Model Provider.
- **Event Data Provider (EDP).** An umbrella term that encompasses both Census EDP and Non-Census EDP. These are the entities that provide exposure data to the system. There is no limit to the number of EDPs that can integrate with the system.
- **Census Event Data Provider (Census EDP).** An entity that provides census-level exposure data to the system. Sometimes a Census EDP is also a publisher, but this need not be the case as some publishers could choose to rely on a third party (see EDP Aggregator below) to provide their events to the system. Additional examples of non-publisher Census EDPs are demand-side platforms and supply-side platforms.
- **Non-Census Event Data Provider (Non-Census EDP).** A Non-Census EDP is responsible for processing non-census measurement data, like measurement panels or STB data, and providing an event level representation of that data to the CMMS.
- **Measurement Consumer (MC).** An entity that requests and receives the output of the system and includes, but is not necessarily limited to, advertisers and agencies. There is no limit to the number of MCs that can integrate with the system.

---

<sup>1</sup> Note that the closer you get to the code, the more that the term Kingdom is used.

<sup>2</sup> The closer you get to the code the more the term Duchy is used and the distinction between Aggregator and Worker is less necessary.

- **Measurement Orchestrator Operator.** The entity that operates the Measurement Orchestrator.
- **MPC Aggregator Operator.** The entity that operates an Aggregator in the MPC Consortium.
- **MPC Worker Operator.** An entity that operates a Worker in the MPC consortium.
- **Measurement Frontend Operator.** The entity that operates the Measurement Frontend. This may include use of the Reporting Server, which is described in the components and libraries table below.
- **Enumeration Survey Provider.** The entity that provides the enumeration survey of the population under measurement. The Enumeration Survey Provider is not used in any of the flows presented below, but availability of its data is assumed where it is required, for example during VID Model creation.

The following table shows the set of roles discussed above alongside Halo provided software libraries and/or components to support deployment or integration for that role. Roles that are not listed are not provided with any client libraries at this time, however API access to deployed components is simple enough not to require any special support.

Role	Halo provided Components/Libraries
Panel Provider(s)	Panellist Exchange Client Libraries
VID Model Provider	Model Training Toolkit
Census EDP	Panellist Exchange Client Libraries VID Labelling Library Privacy Budget Management Library Sketching Library
Non-Census EDP	Sketching Library
Measurement Orchestrator Operator	Kubernetes Kingdom Deployment Measurement CLI
MPC Worker Operator	Kubernetes Duchy Deployment
MPC Aggregator Operator	Kubernetes Duchy Deployment
Measurement Frontend Operator	Kubernetes Reporting Server Deployment Reporting CLI (Measurement Frontend)

All libraries are compatible with the JVM, except for the Virtual People Labeler and Sketching Library, which have both C++ and JVM compatible implementations. The CMMS components for the Measurement Orchestrator, MPC Nodes, and Measurement Frontend are available as complete Kubernetes deployments. The APIs exposed by the Kubernetes deployments are available as [gRPC](#) services and can be accessed natively in most modern programming

languages. The complete set of libraries and components mentioned above are described in the following table.

Component/Library	Description
Panellist Exchange Client Libraries	Libraries used by the Model Provider and Census EDP pairs to execute a panellist exchange
Model Training Toolkit	Libraries that can be used to help train VID Models
VID Labelling Library	A library that can be imported by Census EDPs to label events given a trained VID Model
Privacy Budget Management Library	A reference library with database hooks used by EDPs to maintain Privacy Budgets for Measurement Consumers
Sketching Library	A library used by EDPs to generate and encrypt the inputs for reach computations that are carried out by the MPC Consortium
Kubernetes Kingdom Deployment	A complete Kubernetes deployment of a set of gRPC services and daemons that coordinate measurement requests and panellist data exchanges. This is the software component deployed by the Measurement Orchestrator Operator.
Measurement/Reporting CLIs	A set of command line interfaces used to interact with the Measurement Orchestrator and Reporting Server
Kubernetes Duchy Deployment	A complete Kubernetes deployment of a set of gRPC services and daemons that participate in the secure multiparty computation that computes reach and frequency estimates. The deployment can be configured to act as an MPC Worker or an MPC Aggregator.
Reporting Server	A complete Kubernetes deployment of a set of services that provide higher level reporting and metadata search abstractions on top of the Measurement Orchestrator. These components also support simplified authorization, measurement caching, and privacy budget optimization. A Demo UI will be provided soon. This component can be used as a constituent part of a complete Measurement Frontend.

Halo is exploring the concept of a multi-tenant Measurement Orchestrator that could serve more than one local market and thus save costs and reduce time to market. Were this to be developed, it would still remain the choice of the local market whether to use this or to deploy their own Measurement Orchestrator.

## System Flows

This section describes how the framework components interact to accomplish several important flows. Each of these flows fits into one of three phases: (1) the Setup Phase, (2) the Training Phase and (3) the Measurement Phase.

Within the Setup Phase, there are three steps: an overall setup of the system, enablement of measurement for Measurement Consumers, and metadata distribution. The Training Phase is concerned with preparing the system for measurement and includes the panellist data exchange, and training and distribution of the VID Models. Finally, the Measurement Phase consists of applying the VID Labeler to raw measurement data and the actual computation of measurements.

As we describe each flow, the diagrams will show only the components and interactions that are active in that flow. We will conclude with an overall system diagram that superimposes the entire set of interactions. Flows that are specific to Non-Census EDPs are omitted as we anticipate a degree of variability in these cases.

## Setup Phase

### System Setup

System Setup entails several steps:

1. Deploying
  - the Measurement Orchestrator,
  - the set of MPC Nodes that compose the MPC Consortium;
  - and a Measurement Frontend.
2. Creating Measurement Orchestrator accounts for each participant and exchanging credentials for authentication.
  - Measurement Consumers will also get an account ID and API key.
  - The system uses [mTLS](#) for all system-to-system authentication, which is the same technology that allows web browsers to securely communicate across the Internet, and as such [X.509](#) certificates must be exchanged by all parties.
3. The Panel Provider, VID Model Provider, and EDP(s) must also integrate with a suite of APIs, and Measurement Consumers may integrate with frontend APIs if desired.

### Enablement

Once the system has been set up, Measurement Consumers provide each EDP with their account ID. This ensures that EDPs know how to associate accounts on their side with the Measurement Consumer account in the Halo CMMS. Currently, this step can be done either manually via customer service representatives or automatically via an EDP's own UIs. The exact details are dependent upon the EDP in question, and for linear TV, enablement may not be required at all.

## Metadata Distribution

In order to measure a set of events, metadata that describes the groupings of those events is necessary. Specifically, for all enabled Measurement Consumers, EDPs must share metadata that describes the set of campaigns or content that a Measurement Consumer can report on. We refer to these units of metadata as Event Groups, which for the sake of simplicity can be thought of as being associated one-to-one with a campaign, though in practice Event Groups can be associated with content or more granular details like creatives or ad placements.

Event Groups perform a function similar to that of the ad or content dictionaries and may need to be aligned with existing local market dictionaries to enable cross media reporting. Event Groups are expected to provide rich descriptors that allow a Measurement Consumer to search for and identify those Event Groups they wish to report on. The next diagram shows how Event Groups enter into and flow through the system.

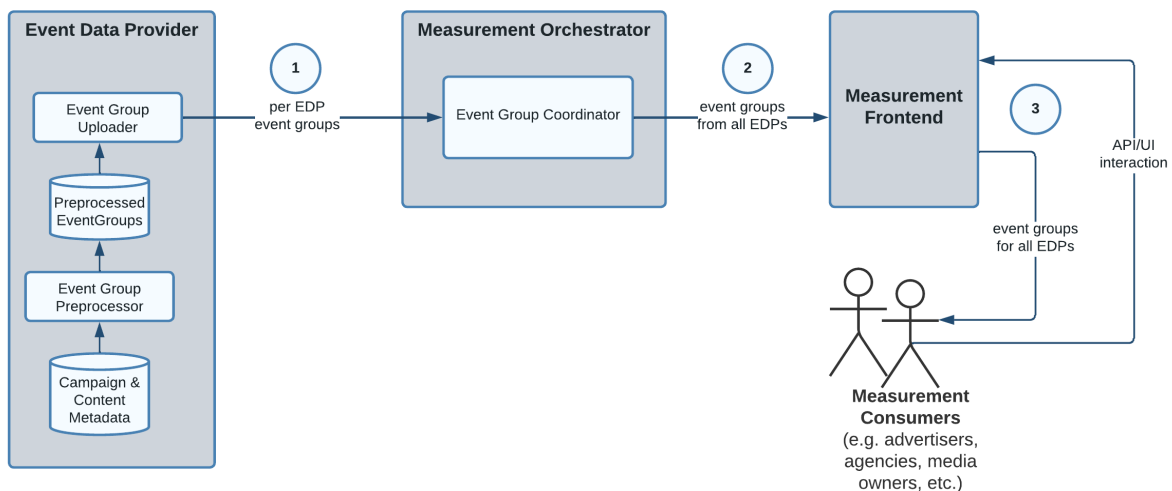


Figure: Metadata Distribution

1. Upon enablement, EDPs proactively upload the Measurement Consumer's campaign and/or content metadata to the Measurement Orchestrator via its Event Group Coordinator.
2. The Measurement Frontend accesses the Event Groups via the Measurement Orchestrator's Event Group Coordinator.
3. Measurement Consumers request Event Groups from all EDPs via the Measurement Frontend's Event Group Search API, which are then returned.

## Training Phase

### Panellist Data Exchange

As discussed above, VID Model training can take advantage of census logs provided by Census EDPs. The process by which the logs are retrieved from the Census EDP's by the Panel Provider is described in the figure below.

The steps are as follows:

1. The Panel Provider initiates an exchange via the Panel Exchange Coordinator, which is itself a collection of subcomponents residing in the Measurement Orchestrator.
2. The Census EDP acknowledges this signal.
3. The Panel Provider encrypts its panellist identifiers and sends them to a shared storage bucket.
4. The Census EDP retrieves the encrypted identifiers and uses them directly to query its logs.
5. The Census EDP outputs an encrypted set of log lines for each panellist identifier to the shared storage.
6. The Panel Provider retrieves the encrypted events and decrypts them.

Status information about the exchange continues to be communicated via the Panel Exchange Coordinator for the life of the exchange, the end of which is a status that indicates whether the exchange was successful or not. It is important to note that this entire process, including the query itself, happens under encryption and the Census EDP does not learn the unencrypted values of either the inputs or the outputs.

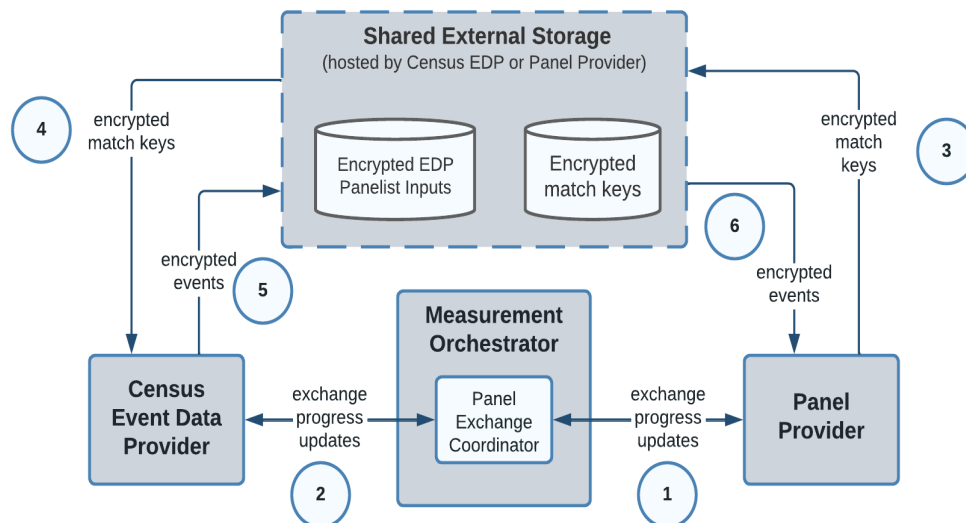


Figure: Panellist Data Exchange

## VID Model Training and Distribution

Once the panellist data exchange has been completed, the next step is to train a VID Model. This is described in the following diagram.

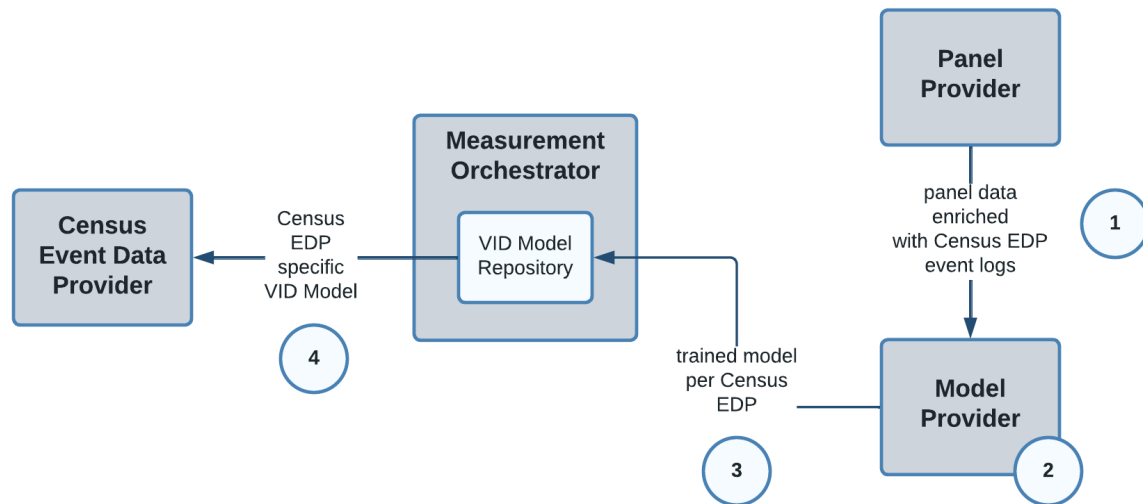


Figure: VID Model Training and Distribution

The steps are as follows:

1. The Panel Provider enriches the panellist data with the Census EDP's census logs and transfers the combined data set to the Model Provider. This is done via standard data transfer mechanisms and may require that the Panel Provider obfuscate the identities of the panellists. The Panel Provider and the Model Provider may also be the same entity, or in the case of multiple panels, one of the Panel Providers may also serve as the Model Provider.
2. The Model Provider receives the enriched panellist data, and trains a VID Model for each Census EDP as described in the Technology Overview.
3. The Model Provider encrypts and uploads each Census EDP's model to the VID Model Repository, which is itself a component of the Measurement Orchestrator.
4. Each Census EDP downloads, decrypts, and installs its VID Model.

Model rollout schedules, which ensure that all Census EDPs are using the appropriate model versions, are transmitted in a similar fashion from the Model Provider to all of the Census EDPs.

Model Providers may also collaborate with Non-Census EDPs so that their data is fused in a manner compatible with the VID Model. Furthermore, it may be the case that the Model Provider also plays the role of a Non-Census EDP. These relationships are not shown in the diagram.



## Measurement Phase

### VID Labelling

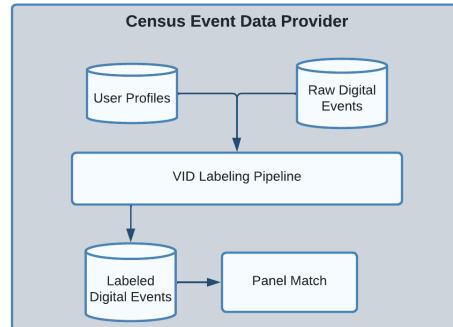


Figure: VID Labelling

As described in the Technology Overview, Census EDPs perform VID Labelling. This consists of applying the VID Model to exposure event data. The previous diagram is similar to those presented in the Technology Overview, but shows that labelled events become available to the panellist match process. This allows the Model Provider to access the VID labels for panellists, for example to validate whether Census EDP's are applying labels correctly.

### Measurement

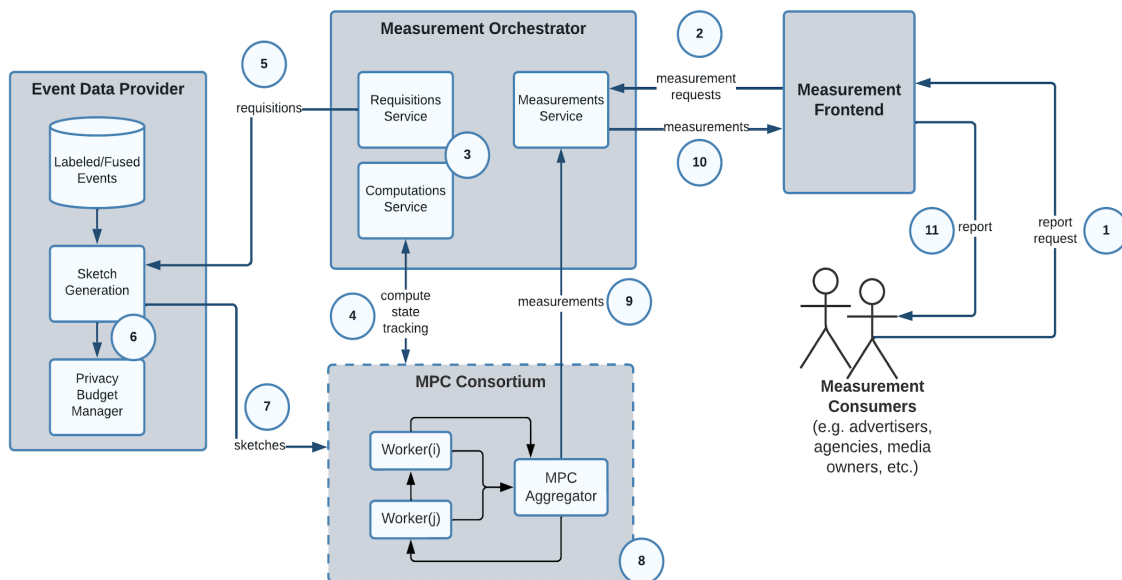


Figure: Measurement Flow

The above diagram shows the Halo CMMS measurement flow, which consists of the following steps:

1. The Measurement Consumer directs the Measurement Frontend to compute a report.
2. The Measurement Frontend decomposes the Report Request into one or more Measurement Requests which it then sends to the Measurement Orchestrator.
3. The Measurement Orchestrator splits each Measurement into a set of Requisitions, one per EDP involved in the measurement. Requisitions describe to the EDP what it must produce in order for the Measurement to be computed.
4. The Measurement Orchestrator notifies the MPC Consortium to prepare to compute the Measurement, and status information continues to be shared between the Measurement Orchestrator and the MPC Consortium during the lifetime of the computation.
5. The EDP receives its Requisitions.
6. The EDP determines whether there is adequate privacy budget remaining and if so deducts the requisite amount before generating its input to the computation, which, as previously described, is an encrypted sketch.
7. The sketch is then sent to an MPC Node of the EDP's own choosing.
8. After all participating EDPs have sent their inputs to an MPC Node, the multiparty computation begins. This consists of the MPC Nodes communicating amongst themselves to compute the requested deduplicated reach and frequency.
9. Once computed, the MPC Aggregator sends the output Measurement to the Measurement Orchestrator.
10. The Measurement Frontend retrieves the Measurements from the Measurement Orchestrator and assembles them into a Report.
11. Finally, the Measurement Consumer can retrieve and review the Report.

Another scenario, which is not shown in the diagram, occurs when only a single EDP is involved in a computation. This occurs when the reach and frequency of just one EDP is requested, or when the impression count or watch time from any number of EDPs is requested. When this happens the EDP can compute the result and the output can be sent directly back to the Measurement Orchestrator, thus bypassing the MPC Consortium. These are called Direct Measurements.

## Putting It All Together

Having discussed the major flows of the Halo CMMS, we now zoom out and consider the system as a whole. The diagram below is an overlay of all of the diagrams discussed above, and shows all of the major interactions between components and roles as well as the internals of a hypothetical Census EDP. The Non-Census EDP, which is shown behind the Census EDP, will share some similarities with the Census EDP as discussed above, especially related to sketch creation, but depending on the data it is aggregating, each Non-Census EDP could differ substantially.

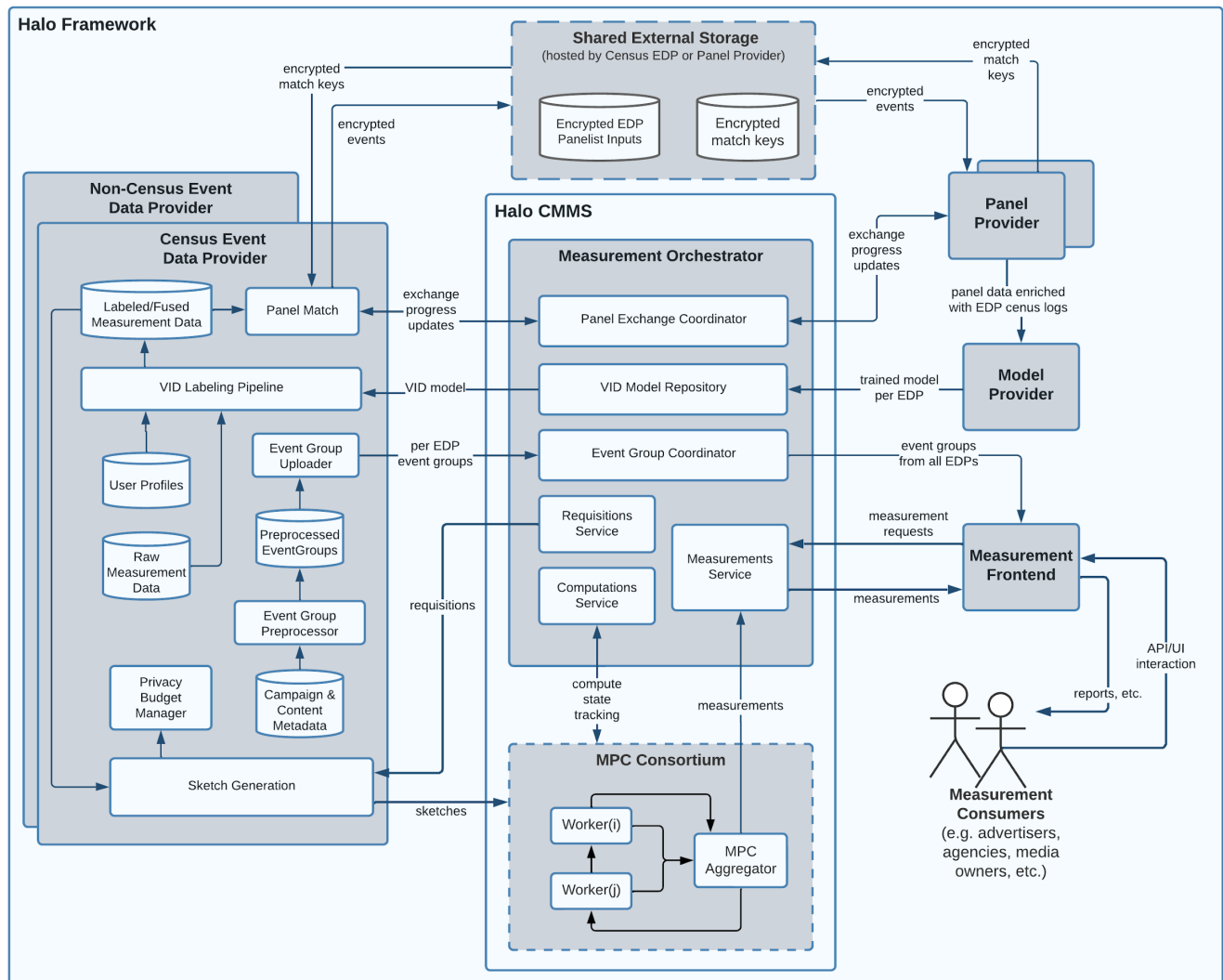


Figure: Complete Framework Diagram

## The EDP Aggregator

So far we have introduced two types of EDPs, the Census EDP and the Non-Census EDP, which have been distinguished on the basis of whether they provide census or non-census events to the Halo CMMS.

An EDP can also be distinguished by how many media properties it integrates on behalf of. For example a digital platform might serve as an EDP for only its owned and operated media properties. Another possibility is where an EDP Aggregator integrates on behalf of multiple media properties, which would allow media owners that do not want to take on a full Halo integration a simpler path to having their properties measured. An EDP Aggregator could be

either a Census EDP, a Non-Census EDP, or serve simultaneously as both depending upon the details of the media properties it was integrating on behalf of.

EDP aggregation services may be offered in a local market to perform the obligations of an EDP on behalf of one or more media media owners. Aggregators would perform roles including the collection of events, panellist data exchange, application of the VID Labeler, and fulfilling sketch requisitions as part of integrating with the Halo CMMS.

A consequence of this approach is that the media owner's raw measurement data would need to be shared with the EDP Aggregator, perhaps via a cleanroom or similar technology.

Comscore has successfully acted as an EDP Aggregator as part of the ANA proof of concept.

# Local Measurement Service Deployment

Halo exists to expedite the implementation of a best-in-class cross-media measurement system. The collaboration's sole focus is on enabling local stakeholders to use the Halo Framework and its supporting libraries in a deployment of a Local Measurement Service under the auspices of a local body.

But while the open-source code is a key asset – perhaps even the foundation – of what local markets need, there is much to do for local groups.

Using the experience obtained from the initial local pilots in the UK and US, the following section explores some of the choices, processes and considerations that local groups may need to reflect upon, as they decide whether the Halo Framework is appropriate for their market.

It is worth noting that the success of a Local Measurement Service will depend upon widespread adoption by advertisers and media owners, and that the specific combination of the governance, commercial, technical, legal and auditing choices will determine the attractiveness of each local service.

Development and testing of the Halo Framework has involved certain advertisers and media owners, who are generally very supportive of this measurement initiative. But it's worth pointing out that the selection of the Halo Framework as the measurement methodology in a particular local market does not necessarily guarantee the participation of any individual advertiser or media owner in the local programme.

## Governance

Local groups considering a Halo CMMS deployment will need to think carefully about how they govern the solution.

The following is a non-exhaustive set of potential considerations for governance of a solution:

- It is recommended that local market groups seeking to deploy the Halo Framework explore early on whether they can enlist the support (and participation in governance) from both digital and TV sell-side parties as well as advertisers.
- As a governance group designed to make executive decisions on the measurement it is over-seeing, agreeing on a decision-making process among the various stakeholders (e.g. voting, consensus, etc) is recommended
- Sub-committees to oversee specific areas (e.g. technical, business and others) may also be needed.
- The Halo CMMS is Apache 2.0 licensed open-source software that has been developed through a global collaboration. The Halo team will continue to administrate, develop and

release new open source code. The local governing group should consider whether it is comfortable being part of such a global programme, and using open-source software.

It is noteworthy that the approaches to governance chosen by the ANA and ISBA, the two markets leading the first pilots of the Halo Framework, have clear differences.

## Commercial

Local markets must determine how to (1) fund the initial build of the system; and (2) pay for its ongoing maintenance and operations.

Whether under Joint Industry Committee or Media Owner Contract models, funding for single-media measurement solutions has traditionally been born predominantly by the sell-side, either through membership fees or commercialisation through third-party measurement vendors.

Industry-wide measurement systems are expensive to build. For example, recruiting a single-source panel will certainly be a significant local cost. There are also costs associated with the deployment and development of the Halo open-source code. These include costs of deploying the Halo CMMS to a cloud environment and other participant integration costs, which must be undertaken before the system can be used. This process may also take an extended amount of time, as we demonstrate below.

Local markets need to identify how this pre-commercial ‘build’ phase will be funded. Some non-exclusive options include:

- Contracting with one or more vendors that are familiar with the Halo CMMS to build and/or operate components of it.
- Working with other markets and the Halo development team to prioritise a multi-tenancy feature, which would allow multiple markets to share the same Halo CMMS deployment, while also allowing for custom configuration on a per market basis.
- Working with the broad body of industry stakeholders to make contributions to develop and launch the service.

Beyond the build phase of the Halo CMMS, local markets need to establish how they will commercialise their measurement system in the longer-term. Some options include, but are not limited to:

- subscription fees;
- percentage paid by advertisers on ad spend (‘fractional contribution’);
- membership fees; and
- advanced reporting fees.

It’s noted that the ambition for ISBA’s Origin programme is to transition the funding model from majority sell-side funding to majority buy-side. In the latest ‘Industry Consultation on Origin Funding Model’, it is reported that the organisation “[expects buy-side funding to be near 60% of the total revenue.](#)”

## Technical

Technically, a Local Measurement Service consists of the necessary inputs, the Halo CMMS, and the resulting outputs, with all participants clear on who is playing which role(s) with their related obligations.

Complementary to the Core Specification, the Local Specification will make available a detailed framework that outlines the key technical decisions that a local market needs to make. At a high level these are:

- Decide who performs which roles in the Halo Framework (see list of roles in Design Overview section above) and by what mechanism of selection or registration.
- Define the necessary panel assets ensuring that:
  - The sample frame covers the universe of behaviour to be measured
  - The panellists have given clear consent for data collected by the properties under measurement to be joined to their panel data
  - The panel has the necessary match keys for such data linkages
  - The panel has the required size given the desired error margins for reach of properties under measurement, and deciding if a single-source panel can be augmented at a lower cost to reach the required size.
- Set the universe estimates with an enumeration survey
- Determine what configurations of the CMMS provide the best balance of accuracy, privacy, cost, and performance
- Agree on a security model, such as defining which parties can perform which roles, or how many MPC nodes are needed (must be more than two and at least one must not be run by an EDP)
- Map how Event Groups will align with existing or new cross-media ad or content taxonomies
- Agree on reporting metrics, including demographic breaks or other slicing criteria
- Define service level objectives (e.g. data completeness, response time, uptime)
- Define audit processes (see following section)
- Define change request processes
- Define software update processes

While the engineering effort behind Halo has been extensive, it is not a full ‘production’ solution. Much work is still required to deploy the software. Measurement vendors or partners may need to be identified to help with this process in addition to operating parts of the solution.

The separation of the Core Specification from the Local Specification not only enables a common development roadmap for Halo and easier technical deployments in new markets, but also allows for the option to use a common underlying legal framework for operational deployment, as discussed in the next section.

## Legal

The contracts for each Local Measurement Service are developed and signed under the local governance structure and according to local law.

A local market may wish to:

- Decide whether to separate the contract that covers the formation of the governing entity and its commercial model from the contracts that cover the operational deployment of the Local Measurement Service. (The governance and commercial contracts are expected to be highly specific to the local market, while there is scope for some commonality between markets in the operational contracts.)
- Given the multiparty nature of the Halo CMMS, consider how the legal framework will scale, avoiding the need for  $n \times n$  contracts between each and every participant in a local market.
- Decide whether the local market should create its own glossary of terminology and related technical documentation or use a common glossary that aligns with Halo technical documentation.
- Examine ways to ease the privacy and security approval processes of the participants, such as by making it clear that it is deploying a standard Halo solution with defined customisation to suit the local needs.
- Consider if the tendering process can lead more seamlessly into the contracting process, reducing legal costs and time to market where possible.

Given the issues raised above, there is a Halo workstream looking to develop a suggested Halo legal framework that, subject to customisations to adapt to local law and circumstances, could provide a consistent and transparent contractual framework, including a common glossary, that describes the data exchanges and related obligations between the participants according to the roles that they are performing in the Local Measurement Service with the aim of making available a multilateral contracting option.

In short, a common legal framework provides an alternative to the default position of locally defined roles and responsibilities with custom documentation and a patchwork of contracts.

*It is of course up to a local market if or how it wishes to leverage such a legal framework. At a minimum, it serves as an example for a local market, as it decides on its approach.*

## Audit

In order for the overall system to maintain trust, data within a Local Measurement Service should adhere to a locally defined audit framework with criteria and design specified by a locally appointed independent audit organisation.

A local market audit may consider how to ensure data quality, data integrity, data completeness, and data privacy in the following areas:



- General Audit Guidelines
- Validation of the Halo CMMS Component Implementation
- IT General Controls
- Data Collection
- Data Normalisation and Processing
- Data Filtration
- VID Model Creation
- VID Labelling
- Sketch Creation
- Computation of Estimates
- Reporting and Disclosures

Given that the Halo Framework is designed to be deployed in multiple markets, and that there are advertisers and media owners that may operate in more than one market, a local audit framework could choose to accept existing separate or global audits for publishers and platforms conducted in relation to a separate Halo CMMS deployment.

To facilitate the acceptance of “equivalency” where appropriate between audit bodies, an Audit workstream has been formed within Halo to outline the high-level audit responsibilities for each role in the Halo Framework. The ANA is working with the MRC who have issued their [Cross-Media Measurement Standards](#) that provide one possible standard to audit against.

## Timeline

The timeline below provides an estimation of the time to launch a Local Measurement Service. While there is some scope to run activities in parallel, such as finalising the governance structure while running tender process, it is a broadly sequential process. To the extent that a local market does not have the required data assets, and especially an appropriate panel, the timeline can be extended significantly.

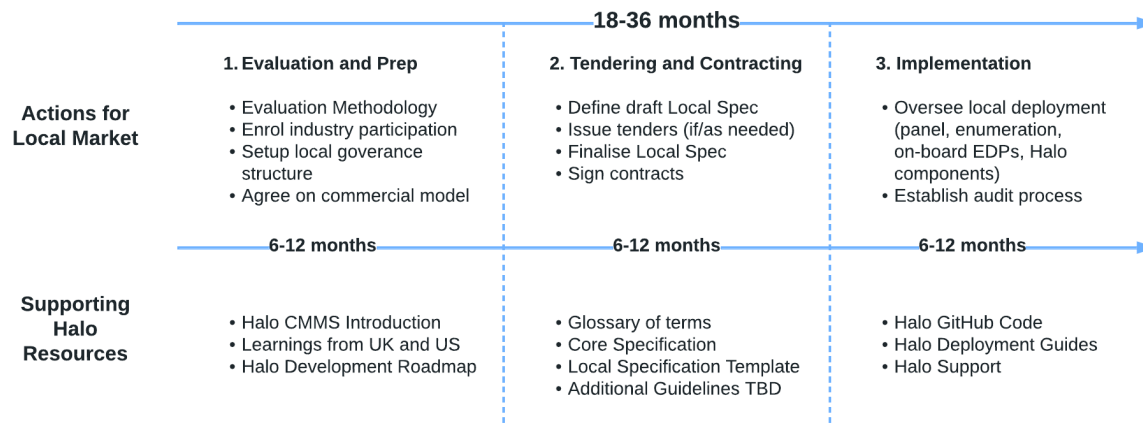


Figure: Local Market Launch Timeline

# Conclusions & Next Steps

## Setting the goal

ANA, ISBA, WFA and its advertisers set up the Cross-Media Measurement initiative to break down the siloed nature of media measurement, both within digital and between media channels, but especially between digital and TV. They wanted it to address the demands for increased privacy for consumers while also delivering the granularity required in a fragmented and personalised media landscape. They also sought to strike a balance between empowering local markets and developing a scalable methodology that could address the complexity of the challenge.

## Delivering on the goal

The Halo Framework delivers on all of these requirements: comparability of media, consistency of methodologies, enhanced privacy, improved granularity, and scalability with the ability to be customised by local markets.

For the first time, a measurement system securely and fairly enables the large-scale use of census data with first party identifiers and demographics with such granularity that panels alone cannot deliver. The open-source and collaborative nature of the Halo initiative ensures a transparent approach that is responsive to the demands of its varied stakeholders, especially as more local markets join the Halo programme.

Critically, the Halo Framework and its code components have been under a rigorous process of testing and validation for both the Virtual People Framework and the Private Reach & Frequency Estimator.

The Virtual People Framework has been shown to address the basic measurement requirements for census events by resolving the relationship between digital identifiers and real people, correcting or assigning demographics, and accounting for overlaps between media properties, platforms and channels. The Virtual Population at its core, representative in size and composition to the population of the country under measurement, has had both census events and the non-census measurements, such as from TV panels or large scale set-top box datasets, assigned to it. This has worked for both content and ads, and has supported distributed processing so that first party data can be more readily included in the system.

The Private Reach & Frequency Estimator has been deployed in the US with secure multiparty computation for the first time in audience measurement, which has allowed the distributed processing of digital properties, digital platforms and linear TV to be integrated into insightful measurement reports for advertisers and agencies, as well as media owners.

The silos are no more. Advertisers and agencies are beginning to get the consumer-centric view of media that they have long sought.

## Next Steps for Halo

For Halo, there is still work to be done to conclude the two technical proofs of concept underway in the US and the UK. For example, the fine tuning of the system to balance the demands of accuracy, privacy and performance is still underway and will determine key variables in the system costs. The consolidated learnings from those tests will be shared with other markets and inform the guidance to complete the Local Specification.

Beyond executing on its 2022 engineering roadmap, Halo's next priority is the release of further documentation in Q2/Q3 2023 that is necessary for local markets to conduct their detailed planning. These include:

- Halo Core Specification,
- Halo Local Specification template,
- Glossary of Terms, and
- proof of concept findings (updated during the year as tests continue).

As there is progress in the legal and audit workstreams, further documents will also be published.

## Next Steps for Local Markets

For local markets or companies considering adopting or participating in the Halo Framework, please send any questions about this document or the overall system to [halo@wfanet.org](mailto:halo@wfanet.org). Halo intends to issue an FAQ in response in Q2 2023.

For local markets, on reviewing the considerations laid out in the Local Measurement Service Deployment section, please prepare your expected local market timelines to share with Halo, so that Halo and its participants can consolidate interest and estimate resource requirements.

Please also consider joining Halo; as an open-source industry collaboration, it is dependent on the funding and time contributions of its participants. The best way to ensure Halo delivers what you want is to be part of it.

## Future Directions of Halo Framework

At the conclusion of our September 2022 Halo Summit, a revised set of requirements were agreed upon. Specifically, it was agreed that the Halo CMMS fulfils requirements 1-5 in the originally requested feature set presented in the first section of this document. The revised advertiser priorities are below. Note that detailed requirements for most of these items are yet to be defined.

1. System Integration
2. R/F Reporting UI (visualisation)
3. R/F Forecasting
4. Adding additional media channels
5. 1P Data
6. Outcomes measurement
7. Advanced segments
8. On-demand analytics

These items were detailed in the first section of this document.

In addition to the above features, the Halo team plans to consider several other non-functional enhancements. These include:

- Improved MPC performance
- Improved privacy budgeting algorithms
- Privacy budgeting transparency and usability improvements
- Energy-footprint transparency
- A comprehensive at scale end to end testing and deployment strategy
- Multi-tenant Measurement Orchestrator
- An enhanced Reporting Server
- An EDP integration toolkit.

We also continue to polish and refine the existing set of APIs and functionality, and to tune the system in response to pilot results.

## References

A comprehensive set of references can be found on our project site at <https://github.com/world-federation-of-advertisers/project-site>.

## WFA Competition Law Compliance Policy



The purpose of the WFA is to represent the interests of advertisers and to act as a forum for legitimate contacts between members of the advertising industry. It is obviously the policy of the WFA that it will not be used by any company to further any anti-competitive or collusive conduct, or to engage in other activities that could violate any antitrust or competition law, regulation, rule or directives of any country or otherwise impair full and fair competition. The WFA carries out regular checks to make sure that this policy is being strictly adhered to.

As a condition of membership, members of the WFA acknowledge that their membership of the WFA is subject to the competition law rules and they agree to comply fully with those laws. Members agree that they will not use the WFA, directly or indirectly, (a) to reach or attempt to reach agreements or understandings with one or more of their competitors, (b) to obtain or attempt to obtain, or exchange or attempt to exchange, confidential or proprietary information regarding any other company other than in the context of a bona fide business or (c) to further any anti-competitive or collusive conduct, or to engage in other activities that could violate any antitrust or competition law, regulation, rule or directives of any country or otherwise impair full and fair competition.