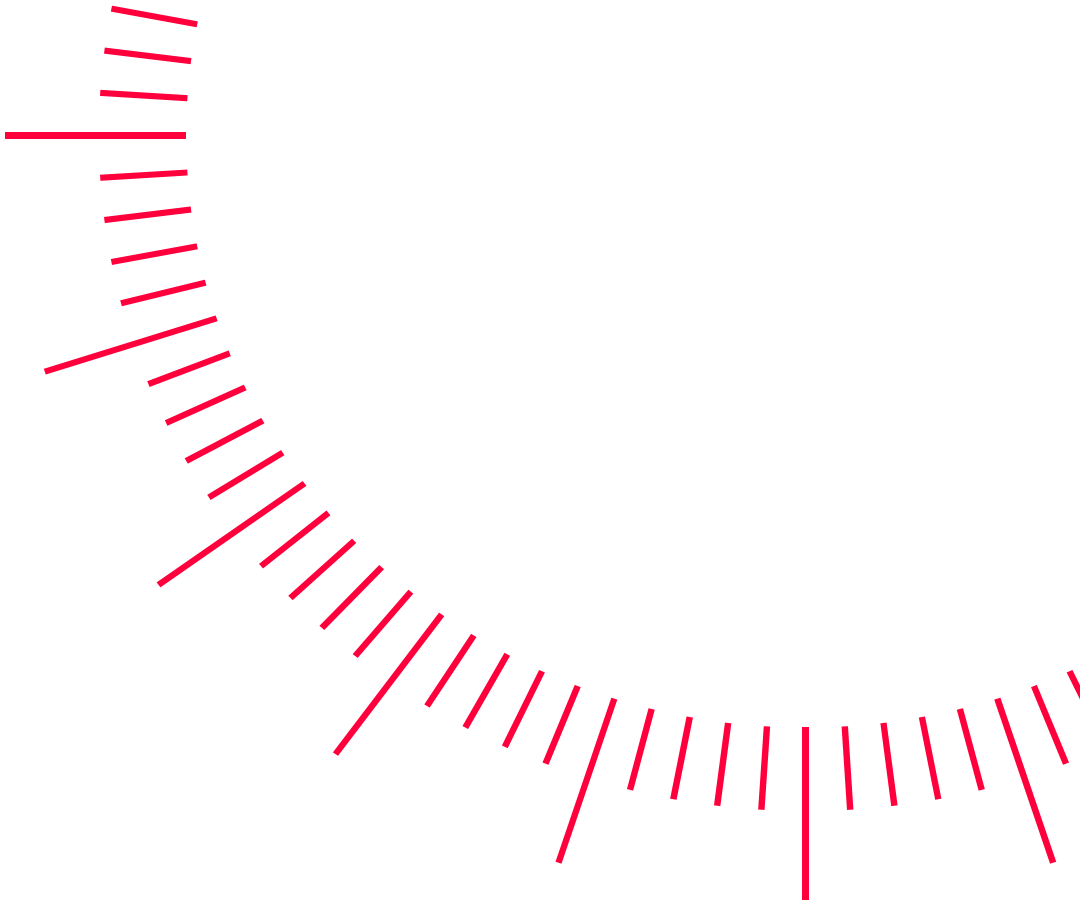


# Halo Technology Overview

19th April 2023





# Objective

**Understand Halo's role in a Local measurement Service that upholds the WFA North Star for reach and frequency measurement.**





# Agenda

- Review of WFA North Star
- Halo Overview
- Technology Overview
- Conclusions





# WFA Northstar

# WFA Northstar



# 2020 Requirements

	Requirement		Description
Foundational Feature Set	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 lookahead; 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.

Available Today

2023/4 Roadmap Item

\* In September 2022 the UI was moved up in priority.





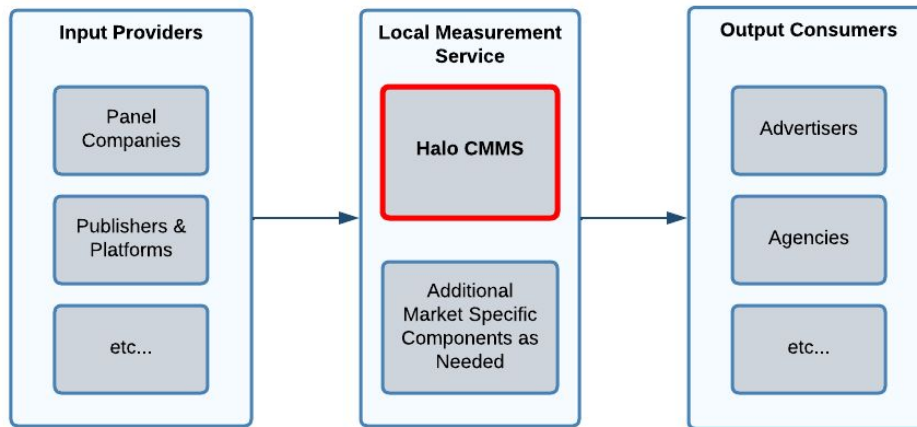
# Halo Overview

# Halo is a Software Framework

Halo provides **configurable core modeling and measurement components** that local bodies can deploy as **part of a privacy-centric Local Measurement Service**

Collectively, these core components are known as the **Halo Cross-Media Measurement System (CMMS)**

To simplify integrations, Halo also provides a set of software libraries, documentation, and support channels.

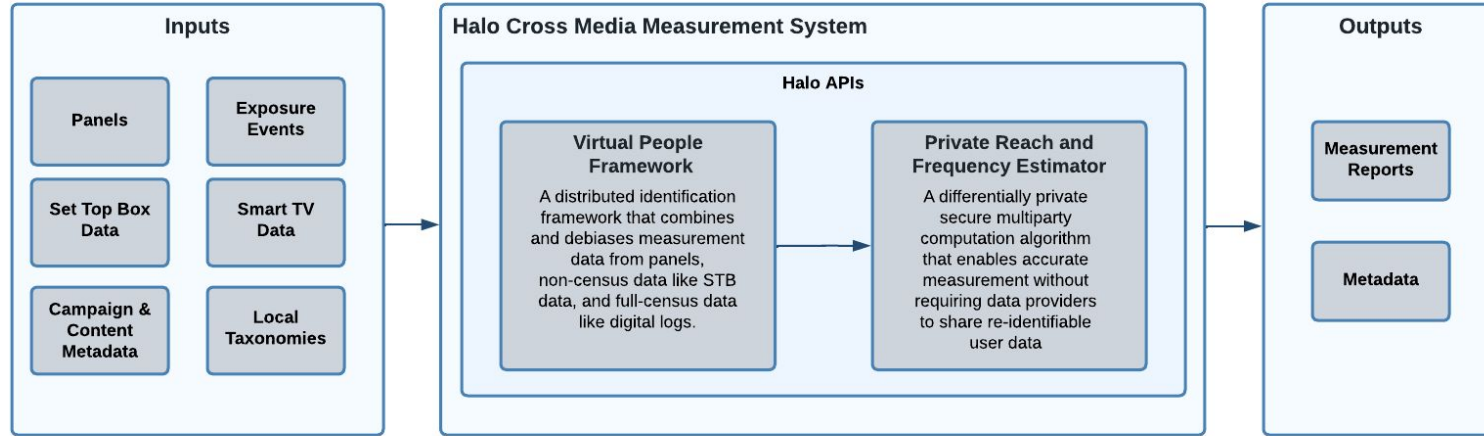


Halo components are open-source software available under the Apache 2.0 License and can be found on the [WFA GitHub site](#)





# Halo Cross-Media Measurement System (CMMS)



Supported by **two technological pillars** - the Virtual People Framework; and the Private Reach and Frequency Estimator

The Halo CMMS **collects and transforms local inputs** and configuration to **produce outputs that meet local measurement guidelines** via a suite of APIs



# Inputs

# Inputs: Adaptable and Configurable

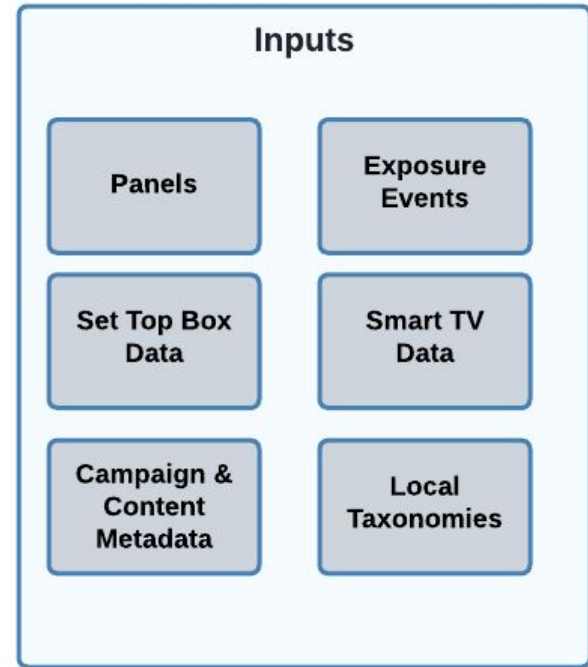
The Halo CMMS **supports a range of census and non-census data assets** like panels, set top box data, and digital exposure logs that the local market is expected to supply

Halo can be **configured to support local taxonomies** for content and campaign meta

A flexible Event Template allows **customization of data cuts**

**Local measurement standards** can also be enforced (e.g. impression qualification)

Tools for enforcing data retention policies are also provided

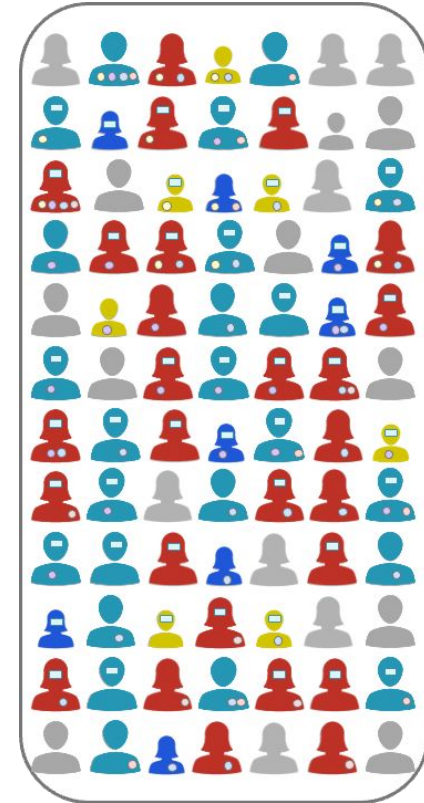




# Virtual People Framework

# Virtual Population

- Based on census data and enumeration survey of the country / market
- Consists of Virtual Persons (VIDs)
  - One VID per person in the universe under measurement
    - Each VID has demographic attributes, e.g., gender, age-group, geographic area, ...
    - Attributes are in the right proportions matching the enumeration data
  - For example in Italy the Virtual Population might consist of 50 million Virtual Persons
- All activity being measured is assigned to Virtual Persons
  - Campaign exposure (and content consumption) at different publishers/broadcasters/platforms
  - Mobile, desktop, CTV, TV, Print, Radio, Out-of-home Advertising
- Is the basis for all reporting



# Assignment of Activity to Virtual Persons

Two main ways to do it:

1. Using VID Models
2. Data Fusion



# Assignment Using VID Models

## Demographic Correction

- Publisher-provided demographic attributes per digital ID/event provided through the panelist data exchange are corrected using a correction method trained using the panel as the ground truth
- If the publisher doesn't have (complete) demographics, missing information is filled in using information from the panel

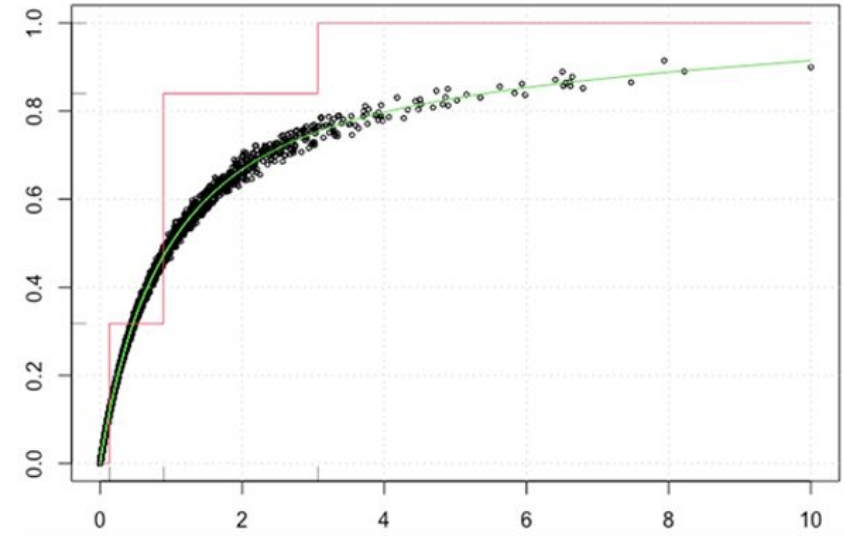
	M 18-24	M 25-34	M 35-44	M 45-54	M 55-64	M 65+	F 18-24	F 25-34	F 35-44	F 45-54	F 55-64	F 65+	18-24	25-34	35-44	45-54	55-64	65+	M	F	Unknown
M 18-24	0,539	0,026	0,001	0,000	0,000	0,000	0,093	0,000	0,000	0,000	0,000	0,000	0,322	0,015	0,000	0,000	0,000	0,000	0,121	0,015	0,064
M 25-34	0,236	0,545	0,014	0,000	0,000	0,000	0,000	0,111	0,000	0,000	0,000	0,000	0,107	0,312	0,008	0,000	0,000	0,000	0,177	0,023	0,097
M 35-44	0,105	0,237	0,578	0,016	0,001	0,000	0,000	0,000	0,135	0,000	0,000	0,000	0,040	0,105	0,322	0,007	0,000	0,000	0,169	0,021	0,090
M 45-54	0,050	0,112	0,280	0,711	0,029	0,000	0,000	0,000	0,000	0,160	0,000	0,000	0,018	0,045	0,120	0,365	0,012	0,000	0,147	0,019	0,080
M 55-64	0,013	0,030	0,074	0,191	0,762	0,004	0,000	0,000	0,000	0,000	0,139	0,000	0,006	0,016	0,043	0,110	0,430	0,002	0,175	0,022	0,094
M 65+	0,003	0,006	0,015	0,039	0,147	0,903	0,000	0,000	0,000	0,000	0,000	0,137	0,001	0,003	0,009	0,023	0,079	0,492	0,157	0,020	0,083
F 18-24	0,055	0,000	0,000	0,000	0,000	0,000	0,577	0,032	0,001	0,000	0,000	0,000	0,329	0,015	0,000	0,000	0,000	0,000	0,007	0,109	0,062
F 25-34	0,000	0,044	0,000	0,000	0,000	0,000	0,212	0,544	0,016	0,000	0,000	0,000	0,114	0,320	0,008	0,000	0,000	0,000	0,011	0,184	0,103
F 35-44	0,000	0,000	0,038	0,000	0,000	0,000	0,074	0,193	0,522	0,014	0,000	0,000	0,040	0,103	0,319	0,007	0,000	0,000	0,009	0,157	0,088
F 45-54	0,000	0,000	0,000	0,043	0,000	0,000	0,030	0,083	0,223	0,567	0,021	0,000	0,018	0,047	0,125	0,374	0,012	0,000	0,008	0,140	0,078
F 55-64	0,000	0,000	0,000	0,000	0,060	0,000	0,011	0,029	0,082	0,206	0,667	0,003	0,005	0,014	0,037	0,091	0,383	0,002	0,009	0,139	0,078
F 65+	0,000	0,000	0,000	0,000	0,000	0,092	0,003	0,007	0,021	0,052	0,174	0,860	0,001	0,004	0,009	0,024	0,083	0,504	0,009	0,150	0,084



# Assignment Using VID Models

## Reach Modelling

- We train a model based on the digital ID-generating behavior of panelists
- Assignment of digital IDs and their associated events (campaign exposure and content consumption) to VID models reproduces the reach curve (#IDs → #Virtual Persons)

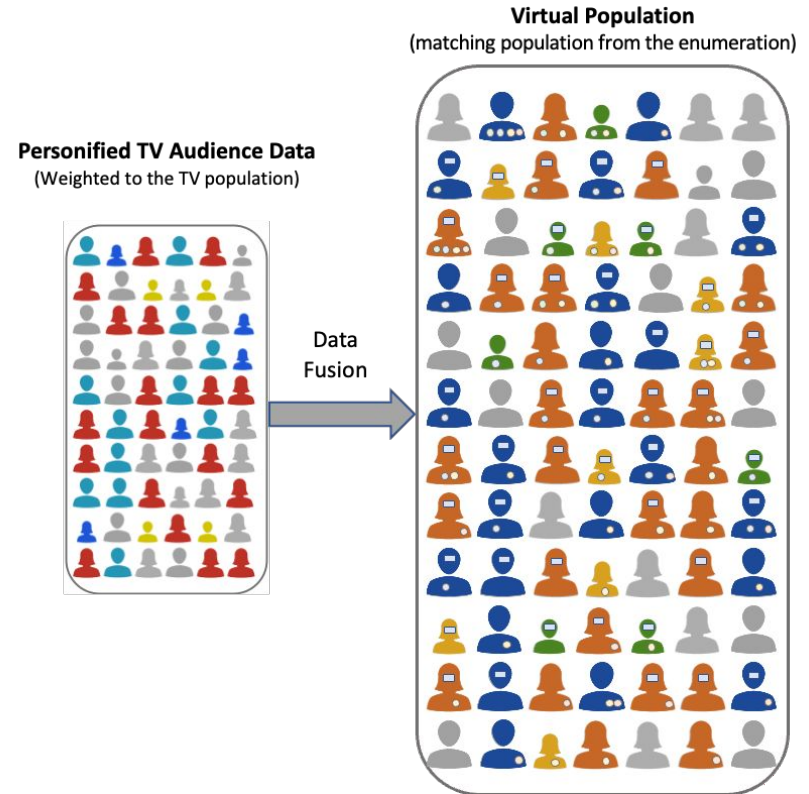



$$r = \sum_i \alpha_i (1 - e^{-\delta_i c})$$



# Assignment using Data Fusion (Example)

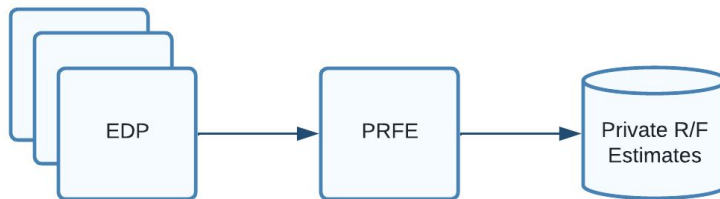
- Household-level viewing from set-top-box households
- Personified viewing using household demographic information and surveys
- Large sample weighted to the Virtual Population
- Viewing of people in the weighted sample fused with VIDs in the Virtual Population respecting demo-/geographic attributes and weights and other fusion clues.
- Similar method can be applied to any weighted respondent-level dataset, e.g., print, radio, and out-of-home advertising data





# Private Reach & Frequency Estimator (PRFE)

# PRFE at a Glance



- **Basic idea**

- Query for all VIDs associated with a campaign's exposures
- Compute reach by counting the number of distinct VIDs
- Compute frequency with a VID histogram

- **Constraints:**

- Inputs must only be used to compute agreed upon outputs
- Any Event Data Provider's (EDP's) input may not be learned by any other party
- The output must bound the probability of detecting the presence of any particular user

# Thought Experiment:

## Use Incorruptible Third Party to Compute R/F

The third party agrees to

- keep all input data secure; and
- only reveal agreed upon outputs

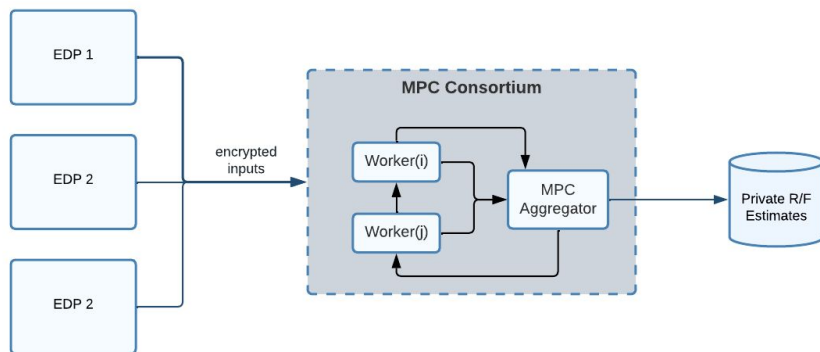
There's a problem with this...

- Incorruptible third parties do not actually exist
- Consider the ramifications of breaching the third party



# Better Idea: Simulate an Incorruptible Third Party Using Advanced Cryptography!

- Trust is distributed across a consortium of independently operated secure computation workers
- They accept encrypted inputs from each Event Data Provider
- Operate on inputs under encryption
- Only reveal agreed upon outputs - (i.e. private estimates of reach and frequency)



This technique is known as secure multiparty computation or just “MPC”



# | Outputs

# Privacy Preserving Outputs

- Outputs of the PRFE are *differentially private*
- This means that:
  - Within a threshold defined by a privacy parameter, the presence/absence of any particular user's data can not be inferred from the output
- This is achieved by adding noise to the outputs

The amount of noise applied to any given output must be balanced with the error incurred from doing so

Local markets can balance this by applying appropriate configuration





# Output Availability

- Halo provides a suite of basic Reporting and Metadata APIs
- APIs enable
  - Construction of UIs
  - Integrations with existing reporting tools
- Halo APIs continue to evolve and enhancements are under development
- A Halo Demo UI is in early stages of development





# Auditing

- In order for the system to be trusted and reliable, EDP inputs must be collected, processed and contributed to the system in a consistent manner
  - Event Groups, Events specs and use of common component systems aids consistency across local markets
- Local market implementation of HALO components must also be deployed per global configurations and adhere to local requirements
  - Panel and VID training
  - VID deployment and EDP event labeling/sketch generation
  - PRFE computation and reporting
- Auditing of inputs, processes, involved vendors and outputs stipulated by local market governing bodies
  - There is opportunity for leveraging of audits of global aspects of local market implementations, if agreed to and accepted by local governing bodies (in discussion)





# Conclusions

# Available Components

- 15+ engineers from 5+ organizations
- 2.5 years of development
- Over 250k lines of Kotlin and C++ code
- 95% test coverage

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.

# WFA North Star Report Card

<b>Full Cycle</b> - report only	<b>Privacy Safe</b> - use of cutting edge cryptographic methods and differential privacy
<b>Continuous</b> - tagless, always-on	<b>Fair and Objective Metrics</b> - local markets are empowered to achieve this
<b>Comprehensive</b> - TV & digital demonstrated; tech supports more	<b>Trust and Transparency</b> - open source code-base with clear break down of operational roles to support auditing
<b>Full Funnel</b> - reach, frequency, impressions, duration	<b>Advertising and Content</b> - designed to support ads and content

**Today, Halo supports the North Star for reach and frequency reporting!**

Learn more at our GitHub project-site:

<https://github.com/world-federation-of-advertisers/project-site>



# Pilot Results

- Project Origin
  - RSMB provide an analysis of the Virtual People Framework showing that it was suitable for cross-media measurement
- ANA
  - VideoAmp showed the Virtual People Framework could be used to measure TV in the US
  - Comscore has shown that the entire Halo Framework can be deployed to successfully generate cross-media measurement (Digital + TV) for both advertising and content



# Summary

- Today, the Halo Platform provides configurable open-source core software components to enable privacy centric media measurement
- Local measurement bodies use Halo components as part of deploying their own measurement service
- Technology
  - The Virtual People Framework provides a distributed modeling methodology that incorporates both census and non-census data
  - Secure multiparty computation ensures that only agreed upon outputs are computed and revealed
  - Noisy outputs mathematically bound the likelihood of inferring the presence of any particular user's data in the output
- Additional information about Halo's system design, which was not covered here, can be found in An Introduction to the Halo Cross-Media Measurement Framework





# Questions?

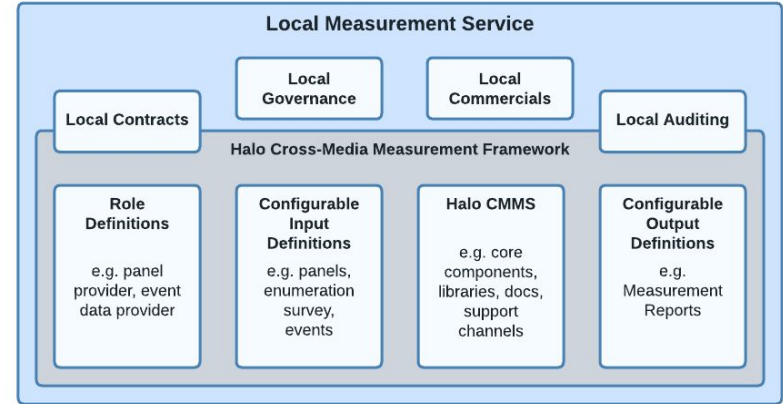
Contact [halo@wfanet.org](mailto:halo@wfanet.org)  
or see our GitHub project site  
<https://github.com/world-federation-of-advertisers/project-site>  
for more information

# | Appendix



# The Halo Framework

- A platform that enables local bodies to deploy their own Measurement Service
- The Halo Cross-Media Measurement System (CMMS) is a set of core software components
- The CMMS orchestrates interactions between participants playing particular roles:
  - Panel Provider
  - Event Data Provider
  - Advertiser
  - Agency
  - etc.
- Who:
  - Provide inputs; and/or
  - Consume outputs



To simplify integrations, the Halo Framework also provides a set of software libraries, documentation, and support channels.



# Halo Framework Components and Roles

**Measurement Frontend** APIs and UIs that provide outputs to Measurement Consumers (i.e. advertisers, agencies, etc.).

**Measurement Orchestrator (MO)** - Mediates the relationships between the rest of the components and roles and to coordinate interactions between them.

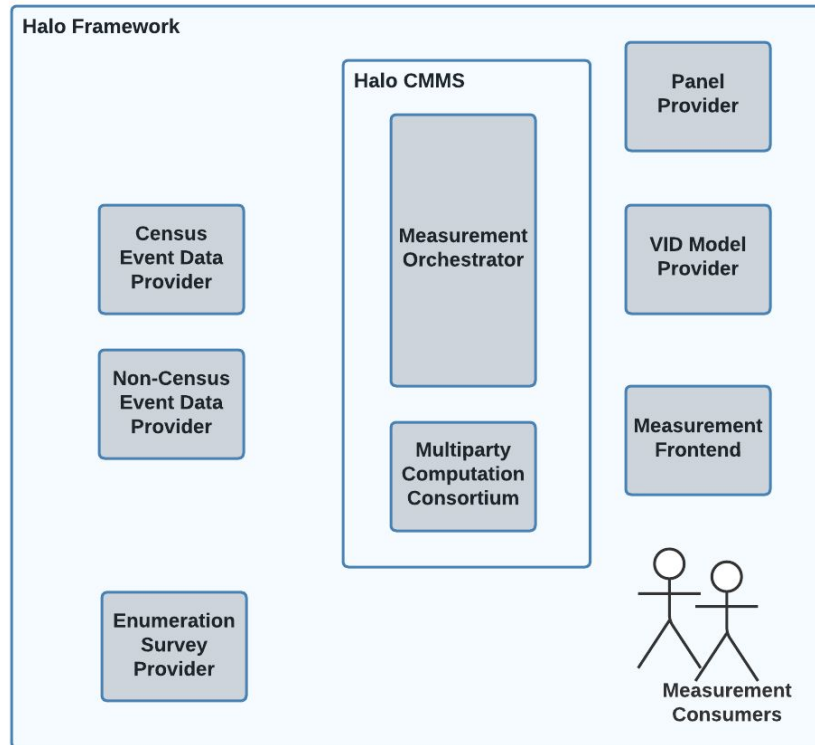
**Multiparty Computation (MPC) Consortium** - This is actually a collection of systems, which together provide the private computation engine for the Halo CMMS.

**Panel Provider** and **VID Model Provider** - Provide panel(s) and a VID Model to the system.

**Event Data Provider (EDP)** - Provide census and non-census exposure data to the system.

**Measurement Consumer (MC)** - Consumers of outputs (e.g., advertisers)

**Enumeration Survey Provider** - Provides estimates of population under measurement.



# Halo Framework in Detail

