



JANUS
STRATEGY & INSIGHTS

pre-meditatedmedia

Combining Set Top Box and Smart TV ACR Data

A study in industry best practices

January 2021

Contents

- Executive Summary** 3
 - Phase 1 Results Summary 3
 - Phase 2 Results Summary 5
 - Background - The Case for Commingling 6
 - Project Scope and Definitions 7
 - CIMM Project Goals 8
- Phase 1** 9
 - Participating companies – Phase 1 9
 - Key criteria/use cases for integrated panel utility 9
 - The data integration pathway 10
 - Data Source Overview 11
 - Reporting Eligibility 12
 - Household Count Overview 13
 - Metadata – Ad Occurrences & Content 15
 - Matching Viewing and Demographics 17
 - Matching and Data Usage Rights 18
 - Reported Viewing Metrics – Gaps and Overlaps 20
- Phase 2** 23
 - Set Top Box/Smart TV ACR Process Integration overview 26
 - Stage 1 - Select data sets 27
 - Stage 2- Establish Match and Commingling Design 30
 - Stage 3- Execute Match 30
 - Stage 4- Calibration and Weighting 35
 - Stage 5- Validation 37



Executive Summary

Study Set Up

Recognizing the need for TV audience measurement to keep pace with the ongoing shifts in consumer viewing across TV platforms and more granular use cases for these data sets like Data Driven Linear (DDL), CIMM commissioned a study to evaluate how best to combine big data Set Top Box and Smart TV ACR sources. The two sources possess complementary measurement properties that make a compelling case for combining them. The research entailed two phases:

Phase 1 – Collection of data provider vitals - vital landscape statistics were collected for pre-commingled data sets including sample size, data captured and reported and data processing rules. Companies that provide Set Top Box and Smart TV ACR data sets, as well as those that aggregate those data sets, were interviewed.

Phase 2 – Develop best practices for commingling and Set Top Box and Smart TV ACR data - collection of detailed account of methods for integrating Smart TV ACR and Set Top Box data, covering matching methods at the device and household level as well as the co-mingled processing of viewing data. Companies that are currently commingling these data sets were interviewed, and those interviews were used as inputs in developing a recommendation for best practices.

Phase 1 Results Summary

General application beyond TV currency

Virtually all applications of Set Top Box and Smart TV ACR are geared for attribution, measurement, optimization and campaign management versus the creation of new audience measurement currency.

Sample size versus representativeness

Some providers and processors make data available from the matched portion of their data sets without modeling the remaining consumers. Rationale: matched sample sizes are sufficiently large and representative, but assumes that demographics account for viewing differences between matched sample and unmatched sample.

One third-party processor with a moderate sample size, produces a highly-curated data set that maps closely to the U.S. population. Value proposition: representativeness and accuracy.

There exists very limited transparency into the depth of weighting utilized by each provider.

At least one provider makes unweighted data available to end users, allowing them to utilize their own weighting schemes.

Weighting schemes are based on consumer data such as provided by Experian; no attempt made to correct for any potential biases in underlying Experian data.

End users call shots on defining metrics

Trend is towards providing end users flexible data access to define their own metrics such as ad exposure time minimums, time-shifted viewing windows and particular geographies that will meet their brand-specific needs.

Metadata overview

Nielsen Gracenote (programming and scheduling) and Experian (household demography/product matching) were cited most often as primary partners. The use of common metadata may make for cleaner comparisons of early-stage data set outputs. On the programming and scheduling front, however, some providers may not deploy the full depth of metadata, for example, actual program airings to capture pre-emptions and cancellations and sports telecast overruns or exact program naming conventions. This omission may be due to the fact that third-party metadata vendors are unable to gain access to certain types of data and are therefore unable to offer it to customers. Programmers have historically provided ad log and program name data to Nielsen and by default Nielsen Gracenote, for example. Ideally, other companies such as Red Bee may gain access to this data.

TV data processing, while not standardized across data providers and third-party firms, is far more systematized and advanced within individual organizations versus 3-5 years ago, for example.

Procedures for data ingestion, integration and formatting are in place as well as editing rules.

Algorithmic rules for filling data gaps, e.g., Smart TV ACR distinguishing DVR and VOD, modeling room in house and using in weighting scheme, residence vs. non-residence is still a work in progress.

Fundamental difference in processing of Set Top Box and Smart TV ACR data

Smart TV ACR- Fingerprint detected on screen matched to fingerprint library based on monitoring, actual time viewing assigned based on when fingerprint was detected, matched to network/program name from external provider.

Set Top Box – Set Top Box records actual time of viewing event (channel set top box is tuned to), matched to station/network from cable lineup, matched to program name from external provider.

Appetite for experimentation

Most participants were highly engaged in the discussion, many now working with multiple data sets either within or across Set Top Box and Smart TV ACR. All expressed curiosity regarding CIMM study results.

Phase 2 Results Summary

Five stages of best practices for commingling Set Top Box and Smart TV ACR data were identified:

Stage One - Data set selection - best practices

- **Representation** - Ideally, utilize a wide diversity of Set Top Box and Smart TV ACR data sets to best represent viewing behavior and footprint while amplifying complementary measurement properties of both data collection techniques.
- **Use cases** - Data set selection will depend upon use cases which could include TV estimates for national TV networks, specific MVPD footprints, cross-platform and data-driven linear reporting, for example.
- **Standardization** - Establish one data set for metadata – i.e., nomenclature for TV networks, stations, programs, channel numbers, etc.

Stage Two - Establish match and commingling design - best practices

- Use tuning data from matched homes with both Set Top Box and Smart TV ACR device matches to inform calibration of combined data set, including un-matched homes. Three core cells emerge: **1.** Set Top Box only, **2.** Set Top Box & Smart TV ACR and **3.** Smart TV ACR only.
- Account for diversity of household TV access on tuning behaviors that reflect changing landscape of TV viewing within different homes: Pay TV, Broadband-only and Over-The-Air. Ensure proper identification and inclusion of these household types.

Stage Three - Match execution - best practices

- **Deploy high quality matching agent** - Use a partner with the ability to match on postal and IP address.
- **Leverage household device graph** - to ensure representation of Pay TV, Broadband-Only and Over-The-Air homes.
- **Benchmark the match process**
 - Smart TV ACR tuning matched versus unmatched homes
 - Set Top Box tuning, matched versus unmatched homes
 - Demographics of matched homes to total U.S.
- **IP address quality** - device IP addresses may change over time, leading to incorrect assignments to homes. Ask IP match provider questions regarding quality of IP data records
 - Recency, churn rate, deterministic vs. probabilistic, life span, etc.

Stage Four - Calibration & Weighting - best practices

Calibrations should be made between Set Top Box and Smart TV ACR data to address the measurement strengths and weaknesses of both data sources. This process is followed by weighting.

- Key calibrations made to data sets
 - **Set Top Box adjustments to Smart TV ACR** - # of sets in home, DVR/VOD, backfill reference for Smart TV ACR signature library.
 - **Smart TV ACR adjustments to Set Top Box** – CTV access/tuning, set-on/set off, on-screen ad exposure.
- Apply weights to four benchmark cells
 - U.S. demographics, TV access universe, tuning metrics and geographics.

Stage Five – Validation – best practices

- Validate universe and tuning estimates to industry benchmarks
 - Overall TV usage, share of TV network viewing, share of linear, on-demand and CTV viewing.

Background - The Case for Commingling

Set Top Box and Smart TV ACR data sets have quickly gained an influential marketplace position as metric sources for planning, scheduling, stewardship and post evaluation of linear TV transactions. Demand for analytics that allow media companies, advertisers and agencies to precisely plan digital video and CTV on top of linear is rapidly accelerating.

Both Set Top Box and Smart TV ACR data sources satisfy the need for more granular data where panel measurement falls short. But at this point in time neither one can check the boxes on all aspects of an ideal audience measurement system. Rather, the two technologies possess a set of complementary measurement pros and cons. For example, there are more Set Top Boxes per home providing data than captured by any one Smart TV ACR provider. Access to name-and-address customer records also tip the advantage to Set Top Box providers in terms of match rates and the ability to append demographic attributes. Set Top Box providers also hold an edge in more accurately distinguishing DVR and VOD activity than Smart TV ACR providers, although many providers of Set Top Box data do not include VOD data at a user/session level. On the Smart TV ACR provider side, Smart TV ACR technology can measure and report some degree of connected TV (CTV) viewing which cannot be detected by Set Top Boxes and more accurately capture content and advertising that actually appears on screen. Smart TV ACR reporting is also speedier. Neither of the two technologies is able to collect data on persons' viewing so TV activity is reported at the individual set and household level, using household level consumer descriptors.

Bringing together Set Top Box and Smart TV ACR audience measurement capabilities holds the promise of developing future measurement systems that are not only more granular by virtue of sample size but that more accurately reflect the rapidly changing consumption of TV, the most obvious benefit being CTV audience measurement. Another benefit is the need for greater sample sizes and greater data granularity to support data driven linear TV advertising. And while IP delivery of TV content is expected to eventually become the dominant mode of TV signal transmission, there will be years ahead of us where Smart TV ACR data will coexist with

Set Top Box data as both providers seek to leverage their data assets. Since we'll be living in a two-source TV measurement world for years to come, there is a growing need to improve measurement systems by supporting the exploration and discovery of best practices to commingle these two complementary data sets.

Complementary Audience Measurement

	Set Top Box	Smart TV ACR
Demographic representation	●	●
Contiguous U.S. representation	●	●
Sample Size	●	●
People Measurement	●	●
Cross Device Measurement	●	●
Delayed/non-live viewing DVR, VOD	●	●
Household-level match rate	●	●
Reporting speed	●	●
CTV data collection	○	●
Multi-set data capture	●	●

Project Scope and Definitions

Scope of study

- Focus on Smart TV ACR and Set Top Box datasets in the U.S.
- TV viewing estimates on Smart Phones, tablets, laptops and PCs are not included, although processes for matching device IDs to accommodate cross-platform integrations is covered

Definitions

- Smart TV ACR - datasets referenced in study are sourced from Smart TVs versus other sources such as audio captured by mobile devices in proximity to TV sets, for example
- Pay TV homes - households that subscribe to MVPD pay services from cable, satellite firms as well as vMVPDs
- Broadband Only homes - households with broadband connection and no vMVPD subscription
- Over-The-Air homes - households that obtain TV signal primarily through antenna
- Metadata - descriptive information appended to tuning data, including, program/content names, program schedules, ad occurrences, ad creative, broadcast and cable networks, channel designations, time stamps, etc.

CIMM Project Goals

CIMM elected to execute an industry study to pursue the advancement of measurement and reporting of commingled Set Top Box and Smart TV ACR data. The project goal was as follows:

Assess the strengths and weaknesses of Smart TV ACR and Set Top Box data to inform best practices for combining them at the household level to create granular nationally representative data sets for linear TV programming and advertising use cases.

Two-phase study

CIMM retained two advisory firms, Janus Strategy & Insights and Pre-Meditated Media to inform design and implement execution of the study. The research was conducted in two phases:

- **Phase 1 – Set Top Box and Smart TV ACR provider review** - the purpose of this initial stage was to lay groundwork for Phase 2 discussions by obtaining vital data collection and processing statistics from 18 providers including MVPDs, OEMs and third-party data integrators. This feedback was key for shaping Phase 2 expert discussions as it provided perspective on the similarities and differences between Set Top Box and Smart TV ACR measurement and methodology approaches. Phase 1 data points included:
 - Current reporting sample size
 - Percent of installed home or device base providing data for reporting
 - Demographic characteristics of homes/devices not reporting vs. reporting
 - Metadata sources for program content and ad occurrences
 - Assignment of program viewing, assignment of ad viewing
 - Editing rules – e.g., current Set Top Box adjustments for TV on/screen off
 - Time-shifted reporting via DVR, VOD, etc.
 - Data processing rules - determining active sets for ratings calculation
- **Phase 2 – Review of methods for integrating Set Top Box and Smart TV ACR data, use to create recommendation for ideal process for integration** - this stage of the study sought to identify best practice methodological approaches for commingling Set Top Box and Smart TV ACR viewing data.
 - Defining a primary data set
 - Approach for creating a household data set that reflects all types of TV access, i.e., Pay TV, Broadband Only and Over-The-Air
 - Data processing and matching - aligning event level tuning data between sources
 - Metric calculations
 - Matching and weighting

Phase 1

Participating companies – Phase 1

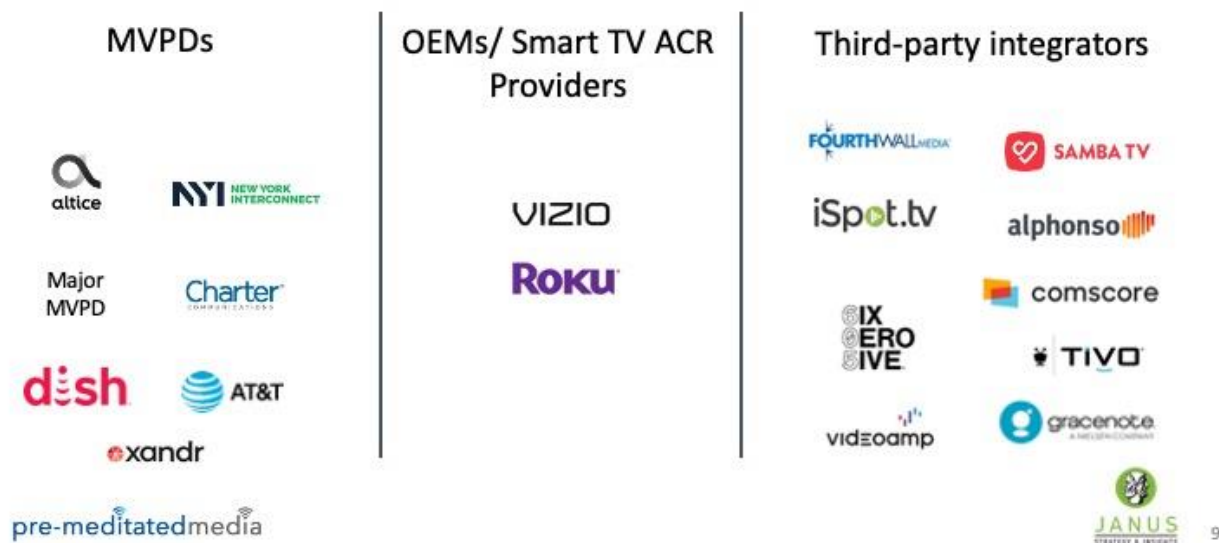
During the months of May through October 2020, 18 interviews were conducted which represented a cross section of data suppliers from three different categories:

MVPDs (7) - Major cable operating and satellite companies provided details on the licensing and processing of their Set Top Box data.

TV OEMs (2) – Smart TV manufacturers and measurement enablers supplied details on their licensing and processing of Smart ACR TV audience data.

Third-party integrators (9) – These companies license data from MVPDs and OEMs and process either Set Top Box or Smart TV ACR data or both. Some of these firms license data from the major MVPDs that participated in the study while others partner with as many as 20 smaller cable operating companies.

Feedback gathered from 18 entities



Key criteria/use cases for integrated panel utility

The ultimate purpose of enhancing TV data sets via commingling Set Top Box and Smart TV ACR data is to support better TV measurement and investment decision making. To that end, below is a roster of important applications or use cases of such data sets to guide TV ad spending.

TV audience ratings/views - provide relative benchmarks for media strategists to compare advanced target audience size composition across dayparts, time periods, programs and networks at scale.

Deduplicated reach/frequency with digital video and CTV - enable cross platform measurement in a holistic manner, i.e., across TV network groups, CTV, VOD and DVR and digital video for better ad frequency management.

Campaign management - monitor audience delivery of advanced targets on timely basis, i.e., days or weeks versus months

Attribution/outcome measurement - determine which combination of creative and ad placements work best to amplify TV/video ad effectiveness

Linear TV integration with addressable TV- as more TV ad inventory is sourced for addressable placement, measuring and reporting both linear and addressable the same way becomes critical

Our Phase 1 discussions indicated that virtually all providers claimed to have capability or at least interest in pursuing nearly all of these use cases with the exception of establishing a new ratings currency. Only one firm expressed formal interest in developing a transactional currency.

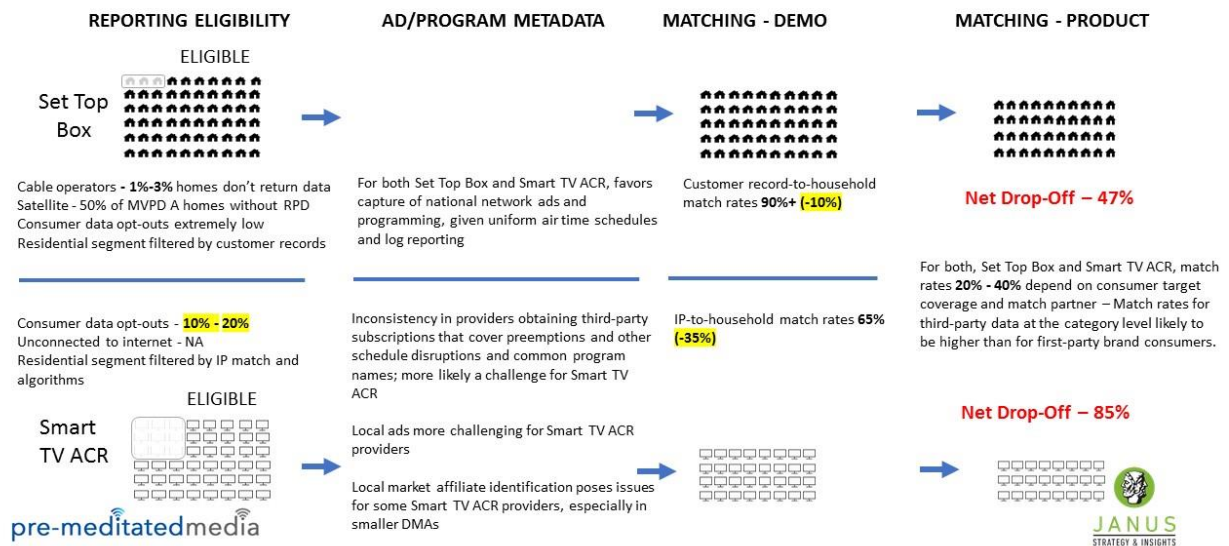
The data integration pathway

An ideal TV and consumer data collection system is one with a representative, large sample size that captures the different modes of TV viewing, e.g., live, DVR, VOD, CTV, etc. in addition to online and offline transactions. Today's mega-sample data sets aspire to move closer to fulfilling this vision of perfection in terms of scale and granularity, however, gaps often exist in the completeness and representative properties of the output data. If we visualize individual data sets as ingredients within a larger integration recipe, these shortfalls are understandable, given the variation in ingredients that can include geographic footprint, device eligibility for reporting, device capability for data capture and third-party coverage of consumer transactions. Layered on top of this is the potential to lose data during the matching process across homes, devices, demographics and product or service consumption. This framework is useful for helping to understand how best to combine Set Top Box and Smart TV ACR data, given the gaps encountered along the data integration pathway.

What follows is a summary chart example of based on participant feedback which shows that data loss was found to be more pronounced for Smart TV ACR integrations than Set Top Box, due mainly to a smaller report-eligible set universe versus MVPD subscriber base and device-to-home match keys. Net match rates, beginning with eligible homes and TV sets and culminating in TV viewer product integrations showed a hypothetical drop-off of 47% for Set Top Boxes and nearly double for Smart TV ACR data (85%).

Data Integration Pathway – Mapping Data Loss Example

As each integration layer is added, the number of viewing records enriched with ad and content, demographic and consumer data diminishes, with the original population potentially misrepresented.



Data Source Overview

This section provides a general description of the basic Set Top Box and Smart TV ACR data set properties in regard to impact on processing and reporting.

Underlying Sources

Set Top Box – Traditional MVPDs and virtual MVPDs tap customer event-level tuning data from Set Top Boxes; many of these devices have been upgraded during the past several years. For the most part, return path data are available from all cable operator Set Top Box devices within a typical subscriber home. For satellite pay TV services, not all homes and devices are returning data because of the lack of a pathway to return data. In cases where satellite pay TV homes are not returning data the missing data can be modeled. For example, Comscore uses the reporting homes to model the viewing of non-reporting homes/devices.

Smart TV ACR – TV OEMs and measurement enablers collect data from Smart TV ACRs, reflecting content and advertising appearing on the TV screen.

Data Collection Cadence

Set Top Box – Set Top Box data are generally transmitted on a time-staggered basis, usually every hour or once or twice per day. Transmission latency for some devices may be due to turned-off boxes, where the data transfer resumes after the Set Top Box is turned back on. There are also some boxes that are programmed to send data during certain points during the day. One MVPD described a process where tuning activity is collected in batches from the Set Top Box collecting agent to a central repository on an hourly basis, then released for processing every 24 hours.

Smart TV ACR - Virtually all Smart TV ACR data are returning data on a continuous basis, then aggregated to reporting specifications.

Churn

Set Top Box – Cord cutting has been the primary catalyst for driving MVPD pay subscription churn in recent years. Some subscriber churn is being stemmed as exiting customers swap their pay subscriptions for broadband-only access. Once the switchover to broadband-only is complete, subscribers can continue to watch TV content through SVOD and AVOD offerings whose streaming activity is detectable and reportable through Smart TV ACR technology.

Smart TV ACR - Churn is driven by the add/replace cycle of Smart TVs versus changes in TV access status that encompasses Pay, Broadband and pure Over-The-Air viewing models. Some ACR processors reported a 2% monthly turnover mark. Others tied churn to the dynamic rotation of IP addresses that make it challenging to gauge the turnover of specific devices and track homes over time.

Reporting Eligibility

This section describes the first-stage criteria for how households and TV sets come to be included or filtered in data sets.

Device Inclusion Criteria

Set Top Box - All MVPD boxes are considered eligible for reporting except in edge cases where the device(s) malfunction or aren't return-path capable. Satellite MVPDs, for example, may exclude partially-active homes where some boxes aren't returning data due to legacy uplink technology.

Smart TV ACR - Most Smart TV ACR providers and measurement enablers work from a base of Smart TV sets that are no older than five years and must be internet enabled.

Consumer Opt-outs

Set Top Box - Virtually all MVPD customers grant permission for data usage as included in subscriber agreements, which also carry terms for privacy and confidentiality.

Smart TV ACR - For Smart TV ACR, about 10% - 20% of TV set buyers decline to have their viewing data collected. Most Smart TV ACR data providers lack intelligence on the magnitude of difference between the demographic profile and TV-viewing activities of the decliners versus the rest of the TV set owner customer base.

Non-residential homes

Set Top Box - MVPDs hold the advantage of utilizing subscriber files to distinguish residences from business establishments. Both customer segments are coded accordingly, providing the ability to filter non-residential homes from data sets, if desired.

Smart TV ACR – Unlike MVPDs, Smart TV ACR providers cannot tap customer files to separate residential from non-residential TV sets. Instead, they use a combination of IP-to-household matching and evaluation of tuning activity to identify the likelihood that device location is a bar, restaurant, retail store, etc. One example might be a waiting room in a doctor’s office where the set is turned on every day at 8:30 am and shut off at 6pm, reflecting visiting hours.

Household Count Overview

Gathering reported household counts for individual Set Top Box and Smart TV ACR providers is a somewhat challenging task. There is no public validated source of truth, and there are multiple ways to report the available database sizes.

These include:

- US versus global figures
- 30-day active viewing counts versus the number of total records in the datasets
- Sets versus households, accounting for instances where there are more than 1 set per home from the data provider
- Whether counts are pre-match or post-match

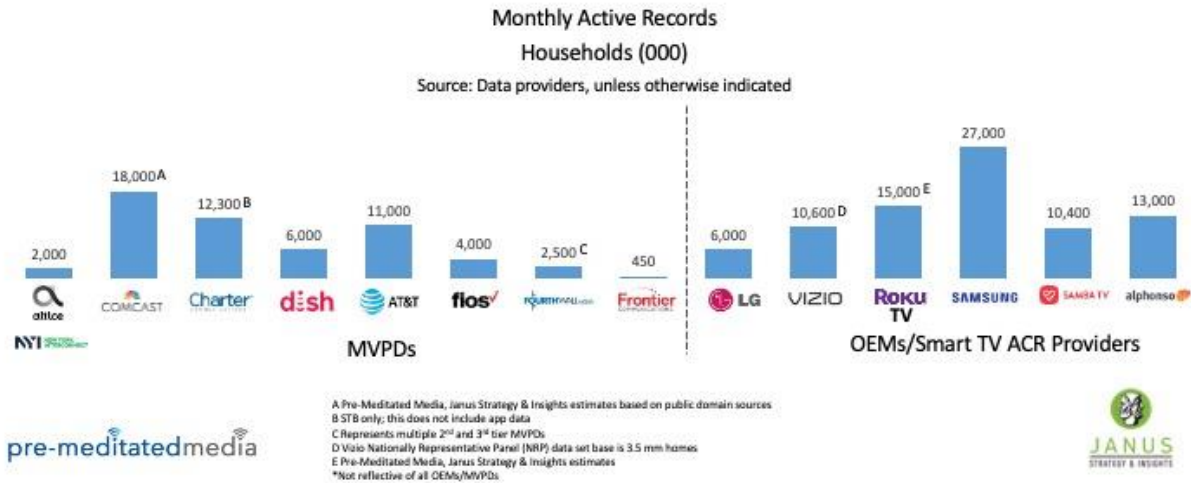
In addition, there are other challenges when interpreting sample size figures supplied by companies that are bringing co-mingled databases to market. A major issue is whether the reported base size reflects gross sample size from all the contributing data sources versus a net count of homes after matching and co-mingling is done.

The counts were sourced from the MVPDs, OEMs and third-party aggregators during study interviews, unless otherwise specified. The counts have not been verified by an independent entity, e.g., third-party auditor.

The chart below provides a monthly viewer count for firms that were contacted for the CIMM study. Note that numbers for some small- to medium-sized TV OEMs may be nested within one or more partner third-party integrators. For example, several partner TV manufacturer household counts are combined within the 14 million monthly actives reported by Samba and included with the 13 million actives for Alphonso/LG. There are also many cases where multiple third-party data integrators license one or more of the same data sources. VIZIO (Inscope) data, for example, are included in more than three MVPD and data aggregator sources.

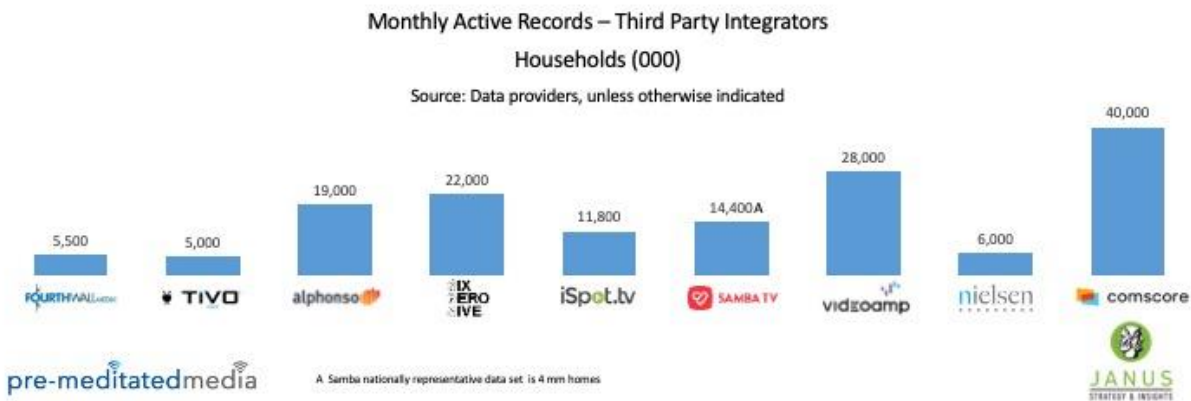
Reporting household count overview* MVPDs & OEMs

These household counts represent the number of homes from **each individual source** with active viewing during a 30-day period



Reporting household count overview – Third-party integrators

These household counts represent the gross number of homes from **multiple sources**, prior to commingling Set Top Box and Smart TV ACR data, with active viewing during a 30-day period.

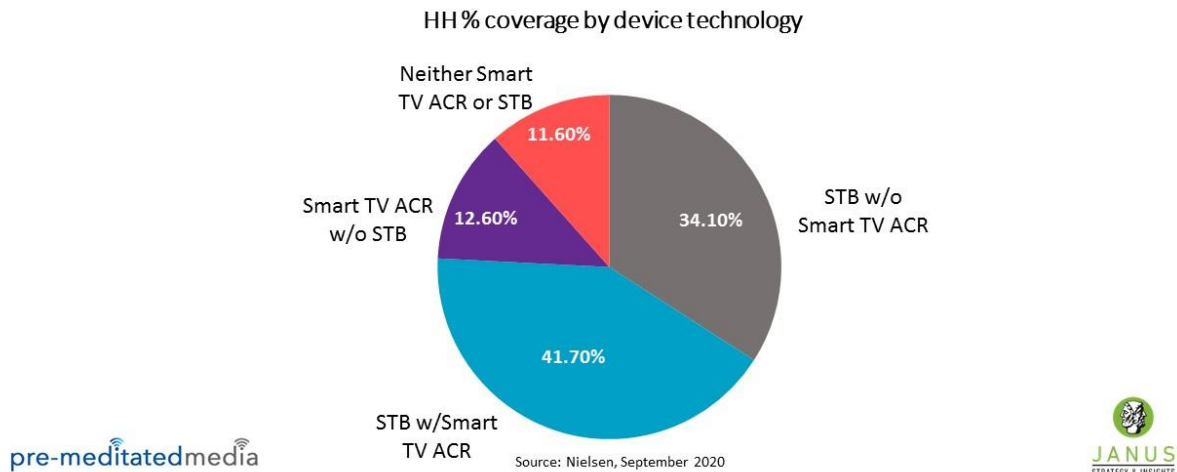


Looking at the entire viewing universe using Nielsen data, we find that 54.3% of U.S. homes have at least one Smart TV that is connected to the internet, a requirement for collecting and reporting viewing data. Homes that are both Smart TV internet-enabled and where Set Top Box technology is present represent nearly 42% of the U.S. It's important to note that this number represents a net mix of different Smart TV OEMs across homes versus the same, multiple set

top boxes in each MVPD subscriber household. Smart TV presence is generally 1.1 TV sets per home for the average OEM while there are 2.4 set top boxes per Pay TV home.

The universe for commingling Set Top Box and Smart TV ACR

- Nearly 42% of US homes are equipped with both Set Top Boxes and internet-enabled Smart TVs
- Note: Smart TV ACR universe reflects the net sum of all OEM devices



Metadata – Ad Occurrences & Content

Metadata, including ad occurrences, content, channel-mapping and clock times, plays a critical role in producing and reporting finished products. For example, the accuracy, precision and timing of second-by-second integration of metadata grows increasingly important for properly crediting viewing of commercial ad units, which are getting shorter in duration.

AD OCCURRENCE METADATA

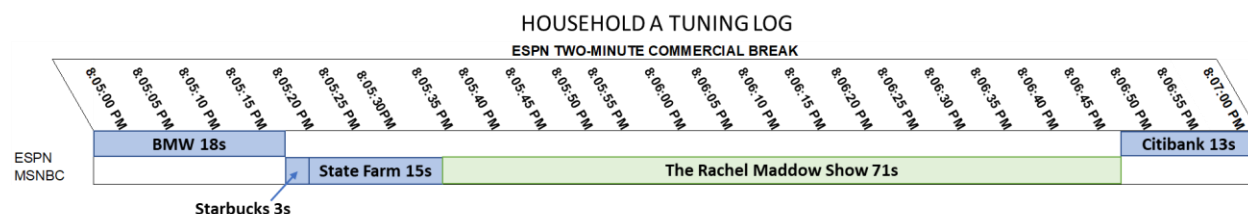
Ad occurrence data are sourced both internally and from third-party firms. For both Set Top Box and Smart TV ACR, time-stamped ad occurrences are mapped to program tuning on a second-by-second event level basis to render ad exposures estimates.

Set Top Box - MVPDs may source ad occurrences from their own local ad breaks in cases where they're supporting a sales platform, for example. The bulk of network ad activity is provided by third-party firms such as Kantar, Kinetic and third-party integrators (one MVPD uses iSpot.tv, for example)

Smart TV ACR – providers such as iSpot.tv, Alphonso and Samba.tv create ad reference libraries by collecting their own ad data via monitoring advertising activity in 35 or more local markets. In most cases they are able to pick up network TV feeds but have limited line of sight to local MVPD ads. Other data providers tap third-party firms such as Kantar and Kinetic, etc. as a core source for ad occurrence data.

Partial ad exposures will occur during channel switching as some viewers explore other content options, particularly during commercial breaks. During all viewing sessions, event-level viewing data must be precisely aligned with ad log file start and end times so that partial commercial exposures are properly credited. In the 2-minute commercial break example below, only one of four ads that Household A tuned to had an opportunity for 100% of the ad to be viewed. As shorter ad lengths proliferate, capturing these partial views becomes more critical.

	length	% viewed
BMW	30	60%
Starbucks	15	20%
State Farm	15	100%
Citibank	30	43%



CONTENT/PROGRAM METADATA

Program data are sourced from third-party firms. There may be inconsistencies across providers licensing both planned and actual air schedules since not all providers incorporate final “as-run” schedules in their reporting. For example, stop times for sporting events that run beyond their scheduled time slots and that preempt the start of shows that follow may not be captured. Additionally, program naming conventions made available by Nielsen and other providers are sometimes used inconsistently.

METADATA – MAPPING THE CHANNELS

Channel mapping

Set Top Box - For MVPDs, tying system channels to cable networks is simple, utilizing channel guide information sourced from third-party firms such as TiVo and Gracenote and their internal channel lineup information. Actual programming may be looked at for locality indicators such as news or sports coverage.

Smart TV ACR - the channel is not used for direct TV viewing assignment since the ACR signal captures only on-screen activity, independent of the channel tuned. The program signature captured on the set is matched to a content library. Within top DMAs, once the program is identified, it is compared to monitored content and program schedule metadata for channel assignment.

It can be challenging for Smart TV ACR providers to know which DMA station is associated with when two channels in an MVPD system are affiliated with one network. For example, ABC World News Tonight airing on the Altice MVPD in Fairfield county Connecticut would run on both channel 7 (WABC, New York) and Channel 8 (WTNH, New Haven). The Smart TV ACR provider can make a probabilistic guess on assigning station to appropriate market, based on

programs watched, i.e., local news, sports, etc. and based on local ads. Proper identification becomes more challenging in smaller DMAs, where local station monitoring may not exist. In these instances, a source like Gracenote can be used to supplement in-house monitoring.

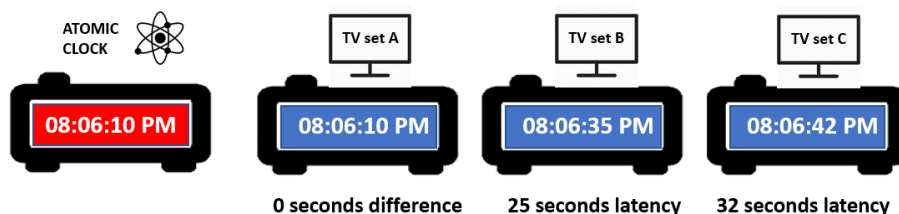
METADATA – SYNCING THE TIME

Syncing clock times - variations in signal transmission speed across broadcast distribution points, means of distribution, head ends, and set top boxes can offset the time when programs or commercials render on different TV sets. If not calibrated to one source of time, discrepancies in second-by-second viewing reporting would surface. Most providers calibrate their device clocks with the U.S. Atomic clock to create uniform times that are reflected in event-level reporting.

Set Top Box - MVPD providers claim that calibrating clock and signal latency is a much smaller issue than it was several years ago, however Set Top Box data sets do not know when a frame of content actually hits a specific TV screen.

- Clock drift is addressed by frequent syncing with MVPD control center time.
- Signal latency is handled by tying all scheduled viewing to one schedule time source, e.g., program logs.

Smart TV ACR - is highly accurate in reporting consumption based on minute of program, however Smart TV ACR cannot distinguish between clock drift and time shifted viewing.



- **Syncing time zones** - some networks run dual feed programming, with delayed, or split air times across different time zones. Most providers make time-zone adjustments in their reporting either based on their channel line ups or based on their monitoring, rolling up audience metrics across the split feeds.

Matching Viewing and Demographics

Set Top Box - MVPDs household match rates for basic demographics are high (90+%) due to use of customer name and address files; Experian is primary match partner for most participants. Churned customers are pulled from subscriber data bases.

Smart TV ACR – Data providers use device IP address as the primary match key, claiming a 60%-70% household demo match rate when combining with name-and-address files. This match rate is lower than for Set Top Boxes, where customer name and address is a primary match key; there's also the penchant for IP address churn and the chance that some households will be

incorrectly matched. Expected match rates would be lower when combining demo and product usage data, depending on data partners.

Non-response bias potential - In virtually all Set Top Box and Smart TV ACR measurement, there is an inability to evaluate potential non-response bias created from homes that don't meet reporting eligibility, don't opt in for data sharing or fail to achieve a match.

Variations in weighting - Weighting tended to be executed for general demographic profiles within the geographic footprints represented by the data set. In many cases end users can apply their own weighting schema. There appears to be an assumption by some providers that weighting to the US population corrects for errors not only in demographic profiles but for viewing and product usage activity as well. From a TV viewing perspective, two providers indicated that they use volume of TV usage and broadcast/cable network shares to calibrate final viewing estimates. Virtually all providers recognized the importance of including different types of TV access homes in their data sets, i.e., pay, broadband-only and Over-The-Air, however, no weights were applied to reflect TV access incidence in the population. (See Reported Viewing Metrics for more detail) There was also no indication that weighting was applied to homes in accordance with product or service usage data.

Data providers indicated no efforts to correct for underlying data issues with companies like Experian e.g., gender and ethnicity and household composition. Additionally, unmatched viewing records are not modeled for demographics and product usage when matched sample size is sufficiently large; the resulting match may not represent true universe.

Comparing Set Top Box and/or Smart TV ACR data to nationally representative panel data, for example, may be useful for better understanding differences in demographic profiles and viewing behavior of individual data sets. Results can potentially be used to inform weighting for national representativeness as well as validating demographics of panel homes versus those reported by third parties such as Experian. Another application would be to guide Set Top Box-Smart TV ACR integrations in terms of set-on/set off editing rules and/or modeling persons estimates.

Matching and Data Usage Rights

Every provider ensures that subscribers and TV set owners have granted data usage permission.

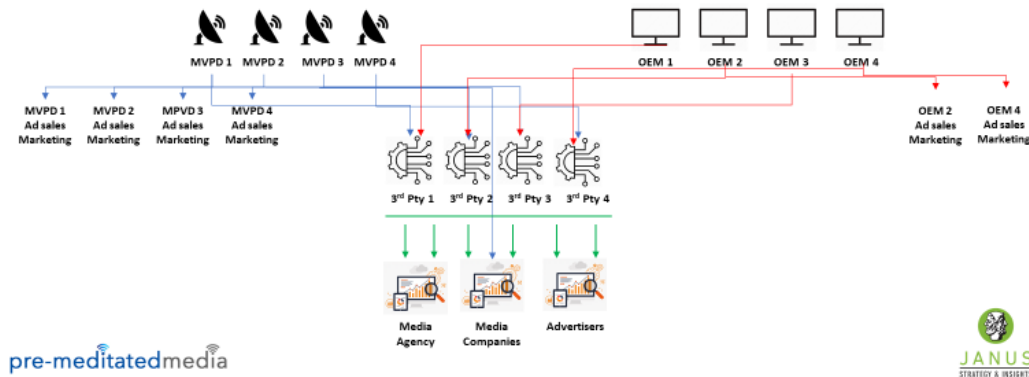
Set Top Box – MVPDs tend to access and confirm subscriber data usage permissions on a periodic basis. For example, one data aggregator stated that MVPD data partners sent subscriber notifications and conducted audits every two years to monitor permission status. MVPD data usage restrictions tend to be competitive in nature, including prohibition of sharing of viewing data with competing MVPDs and use for program carriage valuations and negotiations.

Smart TV ACR – No remarkable data match and use restrictions were mentioned, except where one OEM controls for usage when viewing metrics are combined with select health data.

All data providers are free to engage with multiple match partners, usually with one matching firm designated as a safe haven (e.g., Experian) for matching to other data bases for data-driven linear and addressable.

Data licensing and distribution channels are complex

- MVPDs' first data applications were used for marketing, then internal ad sales prior to licensing to third party processors
- OEMs joined the fray, with some opting to focus on developing ad operations and TV set marketing



Commingling of data is made possible when at least one Set Top Box and one Smart TV ACR data set has been licensed for use. In general, data owners prioritize use for internal functions such as marketing and ad sales, followed by third-party licensing at their own discretion, often entering into exclusive relationships.

With the exception of Frontier, major MVPDs have focused on processing data for internal purposes, e.g., ad sales, while also entering into limited data usage rights agreements with aggregators such as Comscore. One large MVPD is expected to become a Comscore partner in 2021. Altice may potentially pursue licensing.

VIZIO is perhaps the most prolific free-flow Smart TV ACR data provider, focusing on delivery of flexible data sets to end users. Measurement for many middle-tier TV OEMs has been enabled by measurement enablers like Samba and Alphonso, who provide the Smart TV ACR measurement technology and data capture and reporting. Roku and Samsung will likely continue to concentrate on supporting of their ad platforms.

Data licensing:

The model for releasing and licensing data varies from unlimited access “free flow” to curated partnerships with third-party processors

MVPDs

With the exception of Frontier, major MVPDs have focused on processing data for internal purposes, e.g., ad sales, while also entering into limited data usage rights agreements with third-party integrators such as Comscore. Comcast, for example, will become an audience measurement Comscore partner in 2021. Altice may potentially pursue licensing.

Free flow	
Internal/partner exclusive	   
On horizon	 

OEMs

Vizio is perhaps the most prolific free-flow Smart TV ACR data provider, focusing on delivery of flexible data sets to end users. Many middle-tier OEMs have been packaged together in partner relationships with various Smart TV ACR aggregators. Roku and Samsung will likely continue to concentrate on supporting their ad platforms.

Free flow	
Internal/partner exclusive	          
Internal ad sales exclusive	  

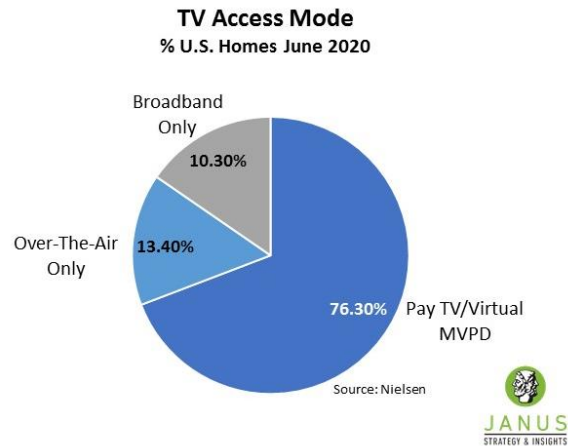
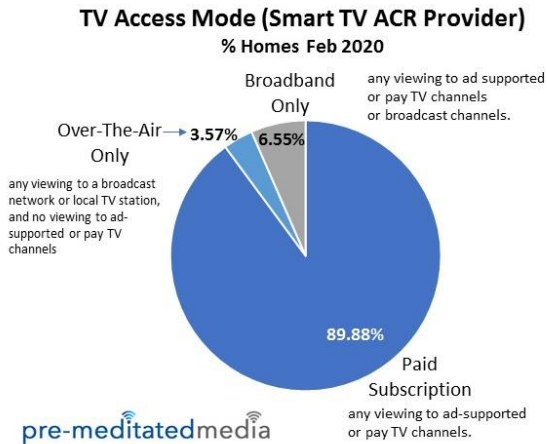
pre-meditatedmedia



Reported Viewing Metrics – Gaps and Overlaps

Representation of all modes of TV access across Set Top Box and Smart TV ACR is challenging, with broadband-only and Over-The-Air homes now representing 23.4% of the TV universe. MVPDs can report viewing to broadcast and cable channels but do not have access to streaming OTT viewing data from their broadband-only customers. Smart TV ACR providers capture data across both linear and OTT, however, DVR and VOD reporting processes are nuanced. Smart TV ACR providers’ reporting is generally limited to time spent viewing within AVOD/SVOD apps, without using ACR to establish the content viewed. This is based on their agreements with AVOD/SVOD services.

As will be discussed in Phase 2, because of the need to accurately size and measure all viewership segments, it will be important to normalize Smart TV ACR data sets based on viewership segment. Data sourced from one Smart TV ACR provider revealed a heavy skew towards Pay TV homes.



Viewing metrics reported - Users define viewing thresholds

Emphasis on measurement and analytics: catalyst for user-defined metrics - With the exception of ratings products, most data suppliers provide users the option to define reporting metrics that may align with their own brand-specific ad impact thresholds. Examples:

- Ad exposure – 1+, 2+ 3+ seconds or minimum of 50% ad viewed.
- Program audience - 20% + of program viewed.
- Campaign reach - exposure to any ad for 3+ seconds.

Time Shifted viewing - data suppliers offer capability to report viewing of recorded and on-demand programming well beyond C3 and C7 conventions. Users can define the telescoping period, however:

- Set Top Box data is limited to DVR playback with VOD reporting supplemented from other sources, e.g., comScore, and not available on a viewer/session level; also blind to all OTT consumption- i.e. watching network app on Smart TV.
- Smart TV ACR can recognize non-live viewing but providers and processors must create rules for distinguishing whether it is DVR or VOD. For example, a distinction could be inferred by referencing ad load (greater for DVR vs VOD), specific ads (DVR reflecting ads in original broadcast) and air dates (DVR playback closer to original air date than VOD).

Viewing Metrics Reported - general gaps

Reporting of app activity – Smart TV ACR providers generally cannot report on viewing within AVOD/SVOD apps but can provide general measures, e.g., time spent viewing reported for Netflix and Hulu. Viewing on connected devices such as Apple TV, Roku, Firestick, etc. can also be reported with more granularity.

- Rule of thumb: any ad-supported linear TV content can be reported where content fingerprint exists and viewing is done through a connected device.

Reporting of digital consumption - while MVPDs have access to digital consumption for their own business purposes, e.g. cross-device ad sales, this information is not available for 3rd party uses

Third-party supplemented - For Smart TV ACR, third-party companies can provide ad occurrence and content library data (e.g., for specific ad campaigns) that can be deployed to produce ACR viewing reports. Inability to identify AVOD ad exposure to ads that do not exist in fingerprint library (i.e. ACR can report everything that hits the screen for which there is a library fingerprint, however, channels or networks can't always be identified)

Demographic appends - Both Smart TV ACR and Set Top Box data providers are reliant on household data appends from companies like Experian. These data sources are used for weighting and profiling. There is limited effort taken to compensate for accuracy issues with those data sources.

Phase 2

Phase 2 involved interviewing various companies who are currently commingling Smart TV ACR data and Set Top Box data about their current approach and rationale behind the approach they are taking, and with that learning, creating our recommendation for the optimal approach for building a commingled data set.

We are very grateful for the cooperation from the companies that we interviewed and thank them for their time and contributions.

Below highlights the companies that we interviewed and the discussion themes.

Phase 2 research discussion themes

Best practices

- Defining the primary data set
 - What are the key considerations in choosing?
- Approach for creating a representative household data set
- Data processing and matching
 - Aligning event-level tuning data between the sources
 - Using Set Top Box data to create a model for Smart TV ACR homes that only have one TV set reporting
 - Using Smart TV ACR data to improve edit rules for SOSO (set on set off) in Set Top Box homes
 - Need for calibration panel to assign persons tuning and co-tuning
 - Set Top Box/Smart TV ACR discrepancies - what rule-based processing situations would the default be to Set Top Box/ Smart TV ACR?
 - Integrated processing time for reporting
 - Metrics calculations - e.g., ad exposure, rating, reach, etc.
 - Matching and weighting

Other considerations

- Cost/benefit
 - How much additional accuracy is achieved when combining Set Top Box and Smart TV ACR?
 - Premiums paid for licensing of multiple data sets and for staffing
 - Anticipated growth rate of attribution, analytics, targeting, campaign optimization, etc.
- End user marketplace - types of clients most likely to use integrated Set Top Box/ Smart TV ACR data
 - Advertiser brands or media companies with younger consumer skews?
 - DTC brands intensely focused on cross platform linear, CTV and digital video?
- Scaling Set Top Box/ measurement system
 - Vision for workable industry system, including governance
 - Likelihood to license data/participate

Phase 2 research interviews



pre-meditatedmedia

JANUS
STRATEGY & INSIGHTS

Because of the large scale of homes with traditional cable services and the growth of TVs that are connected to the internet, the reality is that a commingled database could potentially include direct measurement of the great majority of US households.

Based on Nielsen National People Meter data from September 2020, only 12% of US homes are not covered by either the presence of cable or the existence of an internet enabled Smart TV. So, the challenge from the industry is then one of cooperation from the various data providers.

For owners of Pay TV Set Top Box data, we've seen greater willingness to make data sets available of late. Comcast has licensed data to both Nielsen and Comscore. With Comcast's move, all the major cable providers have made their data available into the research market.

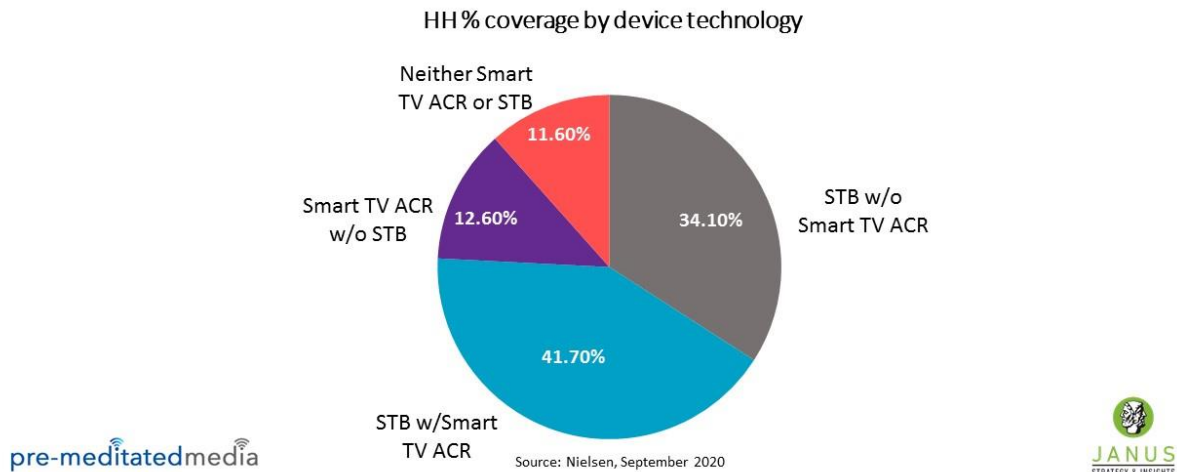
For the owners of Smart TV ACR data, the outlook is somewhat mixed. Certain TV manufacturers like VIZIO have made their data widely available, while other major

manufacturers have decided to use their data solely to fuel consumer experiences and their own advertising monetization efforts.

As you will learn later in when we present our recommendation, the availability of a matched data set, where Set Top Box and Smart TV ACR data is combined at a household level, is an important component of building a commingled data set. That data set can be used to correct for certain limitations known to exist in each data set - e.g., Smart TV ACR data generally reflecting viewing from only 1 home; Set Top Box data needing a Set On/Set Off model to adjust for homes that leave their Set Top Box on when the TV is off. Nielsen data shows that the potential exists for very robust matched data sets.

The universe for commingling Set Top Box and Smart TV ACR

- Nearly 42% of US homes are equipped with both Set Top Boxes and internet-enabled Smart TVs
- Note: Smart TV ACR universe reflects the net sum of all OEM devices



While Set Top Box and Smart TV ACR data provide the ability to work with sample sizes far greater than what panel measurement offers, it is very important to understand that neither data set properly reflects the US population, so any commingling will require a robust weighting scheme to correct for these imbalances, if the use case is to project TV ratings to a broader population. In addition, there are certain viewership segments such as broadcast-only homes without an enabled Smart TV that cannot be represented in these databases, and may have unique skews, such as older Hispanic and Black households, for example. All efforts must be made not only to properly weight for these skews, but also to be able to estimate how viewing levels may differ. This is a role that the matched Set Top Box/Smart TV ACR data set may be able to help with.

Areas of Concern about Integrated Data Set

There are unique demographic skews to the contributing data sets, and to the segments that are not captured in contributing data sets

Viewership Segment Profiling Analysis

Segment % Total US - Index to Total US

	STB	ACR	Calibration	Calibration		ACR	Not Captured		ACR	Not Captured
	Pay TV	Enabled Smart TV	Pay TV and Enabled Smart TV	Pay TV w/o Enabled Smart TV	Over-The-Air	Over-The-Air and Enabled Smart TV	Over-The-Air Only w/o Enabled Smart TV	Broadband Only	Broadband Only and Enabled Smart TV	Broadband Only w/o Enabled Smart TV
HOH Age Range < 35	55	102	70	37	95	122	71	216	218	211
HOH Age Range = 35-54	96	121	117	69	110	130	91	126	128	122
HOH Age Range = 55+	120	84	98	146	94	70	117	39	36	43
HOH Origin = Hispanic	87	113	102	68	162	187	140	96	105	76
HOH Race = Black	95	96	95	95	130	113	146	82	84	80
HOH Race = White	106	102	104	108	93	94	91	101	99	105
Asian Household	63	91	78	45	73	87	60	136	147	114
Language Class = Spanish Dominant	72	91	76	68	210	210	210	56	62	45
HHLD Income = <\$75,000	92	79	72	117	128	105	149	94	87	107
HHLD Income = \$75,000+	109	125	133	80	67	94	42	107	115	92



Source: Nielsen 9/28/20



Beyond the need to weight and calibrate to adjust for unique skews and coverage of Set Top Box and Smart TV ACR databases, there are some other concerns about the contributing databases that must be transparent to end users. There are three general areas:

- **Adequate household coverage** - Beyond the 11% of homes that are not available for measurement, there are other areas where a commingled database will have issues with overall coverage or complete measurement within the home.
- **Weighting schema** - Any weighting scheme requires use of a demographic append company, and assumes that the household variables that are used in weighting are accurate. The industry needs greater transparency on Set Top Box homes and Smart TV ACR homes that don't match to that demographic append company, and whether weighting sufficiently adjusts for any viewing differences.
- **Lack of standardization** - Clear, consistent definitions of Pay TV, Over-The-Air and Broadband Only, and OTT enabled homes. This is important to support reach and frequency planning applications.

Areas of Concern about Integrated Data Set

- **Adequate HH coverage** - Fully integrated database may have incomplete coverage:
 - Some Smart TV ACR and owners of Set Top Box providers will not license their data
 - Older TV sets without Smart TV ACR
 - People who have opted out of measurement
 - People with Set Top Box w/o return path data
- **Weighting schema**
 - **Demographics**
 - No direct capture of household demographics from Set Top Box or Smart TV ACR homes, accuracy of weighting scheme dependent on accuracy of appended demographics
 - **TV access mode** - need to ensure representation and weighting of Pay TV, Broadband-Only and Over-The-Air homes to reflect different viewing behaviors
 - **Unmatched homes** - Process needed to address whether weighting appropriately corrects for un-matched homes biases in TV and product consumption
- **Lack of standardization** - Proper integration of Set Top Box and Smart TV ACR would require common definitions and universe estimates that can be applied to both databases
 - Cable homes with internet enabled TVs
 - Smart TV ACR homes with cable TV access; Smart TV ACR homes who are Broadband-Only
 - Some general segmentation of viewing to OTT versus viewing to linear TV

pre-meditatedmedia



Set Top Box/Smart TV ACR Process Integration overview

Based on the learnings from Phase 1, what follows is our recommendation for the optimal process for creating a co-mingled Set Top Box and Smart TV ACR database, one that capitalizes on the complementary benefits of each contributor, leverages the direct learnings from a single-source matched database.

It's important to note that these are general guidelines and that the optimal process for any given end-user scenario, whether it is an owner of Set Top Box data looking to enhance their data set for ad sales execution purposes, or a research company looking to create a nationally representable database to forecast linear TV ratings, may differ.

It's also important to note that the recommendation may differ based on the applications that will be enabled by the commingled data set. For the purposes of our recommendation, we're envisioning an integrated data set that will report a national estimate of program and commercial ratings for broadcast and cable, and be used as a planning database to guide the planning of digital video and connected TV in coordination with linear TV.

There are 5 stages to the integration, as detailed on the following slide; each stage will subsequently be discussed in depth.

Set Top Box/Smart TV ACR Process Stage Overview

- Stage 1. Select data sets - situation analysis, key considerations, specific use cases
- Stage 2. Establish commingling match cell design, example:
 - 1. matched Set Top Box/Smart TV ACR, 2. unmatched Set Top Box, 3. unmatched Smart TV ACR
- Stage 3. Execute matches
 - Household (demographics)
 - Device match (Set Top Box to Smart TV ACR)
 - Match validation
- Stage 4. Calibration and Weighting
 - Calibrations – Set Top Box to Smart TV ACR data; Smart TV ACR to Set Top Box data
 - Weighting - Demographics, TV access universe, tuning metrics to all match cells
 - Scaling- Scaling Smart TV ACR to network reach levels from Set Top Box data
 - Consolidation - combine all match cells to form one reporting data set
- Stage 5. Validation
 - Universe estimates
 - Core tuning metrics to industry benchmark

pre-meditatedmedia



Stage 1 - Select data sets

As stated earlier, the decision on what data sets to use should be guided by the intended use cases for the commingled data sets. Our recommendation is based on creating a database to report national ratings for broadcast and cable, and creating a database to serve as the basis for cross-screen planning.

Stage 1 – Select data sets

Data set selection will vary based on use cases

- Use case examples
 1. Report national ratings for all broadcast and cable programming
 2. Provide integrated database to enable cross-platform planning- linear TV, digital video, CTV
 3. MVPD enhancing accuracy of Set Top Box data with Smart TV ACR
 4. Provide infrastructure for data driven linear TV reporting
 5. Provide infrastructure of addressable TV planning and measurement, also allow for reporting of underaddressable where network doesn't fill entire unit with addressable
- Our best practices recommendations focus on #1 and #2, leveraging Set Top Box data with Smart TV ACR data:
 - National viewing estimates and ratings for broadcast and cable networks
 - Data output guides planning of digital video and connected TV in coordination with linear TV and measurement of reach and frequency of cross-platform campaigns

pre-meditatedmedia



Ideally, a database integrator should select multiple data sets that reflect a diversity of viewing behaviors. A Comcast subscriber with an Xfinity X1 device, for example, may view TV differently than a DirecTV customer who signs up for NFL Sunday Ticket. Also, viewing differences may be driven by contrasting geographic footprints between MVPDs; West coast TV time spent viewing, for example, is much lower than in the South. Similarly, and at the household level, Smart TV ACR providers may have different skews in terms of TV set locations in the home. Having multiple providers of Set Top Box and Smart TV ACR data will help reduce bias in an individual provider's database.

An important role that Smart TV ACR data plays in a commingled dataset is reflecting the viewership of Over-The-Air and broadband-only homes. Smart TV ACR data sets cannot directly identify whether the viewing source is from a Pay TV household, an Over-The-Air household, or Broadband Only household. Our recommendation is to use viewing data to classify homes into these segments.¹ Another important role that Smart TV ACR data sets can provide is the identification of homes that are OTT-accessible, either based on viewing behaviors (i.e. time spent with ad supported VOD apps) or based on matching a Smart TV ACR database to a DSP with access to CTV inventory. This is important from a cross-platform planning perspective.

The final important data set is an accepted source of household demographics from one of the various data append companies. This data will be important at a later stage when the matched data sets are weighted and projected to their intended universe.

An important industry development that promises to greatly improve accuracy of TV viewer universe estimates is the ARF Media Usage Study. This research, to be conducted in 2021, will capture how consumers access video content within a representative cross section of the U.S. population to better inform counts of Over-The-Air, Pay TV and Broadband-Only homes. Overall, this new data could enhance the weighting/projection schemes utilized for commingled data sets.

¹ example chart on page 20

Stage 1 – Select data sets

Best practices

- Utilize Set Top Box data set that spans multiple traditional/virtual MVPDs
- Utilize Smart TV ACR data set that spans multiple Smart TV ACR providers
 - Apply consistent definition of Over-The-Air, Pay TV, Broadband Only homes based on tuning behavior
 - Over-The-Air = household that watches broadcast stations and no pay networks
 - Pay TV = households that watch any pay networks
 - Broadband Only = homes that watch neither broadcast stations nor pay networks
 - Ensure Smart TV ACR provider is able to provide Over-The-Air viewership through traditional and digital antenna
 - Include identification of OTT- accessible homes, based on agreed upon definition that can be implemented Smart TV Across all contributing Smart TV ACR providers
- Accepted source of household demographics
 - Include universe estimates for Over-The-Air, Pay TV and Broadband Only, OTT enabled homes, scored onto matching agent's database and used in weighting
 - Leverage planned ARF Media Universe Estimates study



Prior to executing the actual data matching, it's important to ensure that both data sets have been appropriately cleansed, so that the actual matching reflects current MVPD/vMVPD subscribers and active Smart TV ACR sets in the United States. Other issues that might impact the quality of the match, like unstable IP addresses, should also be addressed prior to beginning the matching process.

Stage 1 – Ensure that both data sets have been cleansed

Set Top Box - confirm/verify valid subscriber records

Smart TV ACR - confirm/verify valid 30 day/90 day active sets

Remove unstable device IDs/unstable IP addresses

Duplicate TV_ID Removal - Based on the TV Manufacturer's error. Raw Smart TV datasets frequently contain TV ID errors due to manufacturing inconsistencies. If not accounted for, the attribution of viewership will not be accurate.

International ID Removal - Smart TV datasets may contain viewing records from TV sets outside the U.S. which must be removed.



Stage 2- Establish Match and Commingling Design

Stage 2 involves establishing the proper matching design for the commingled database, which includes:

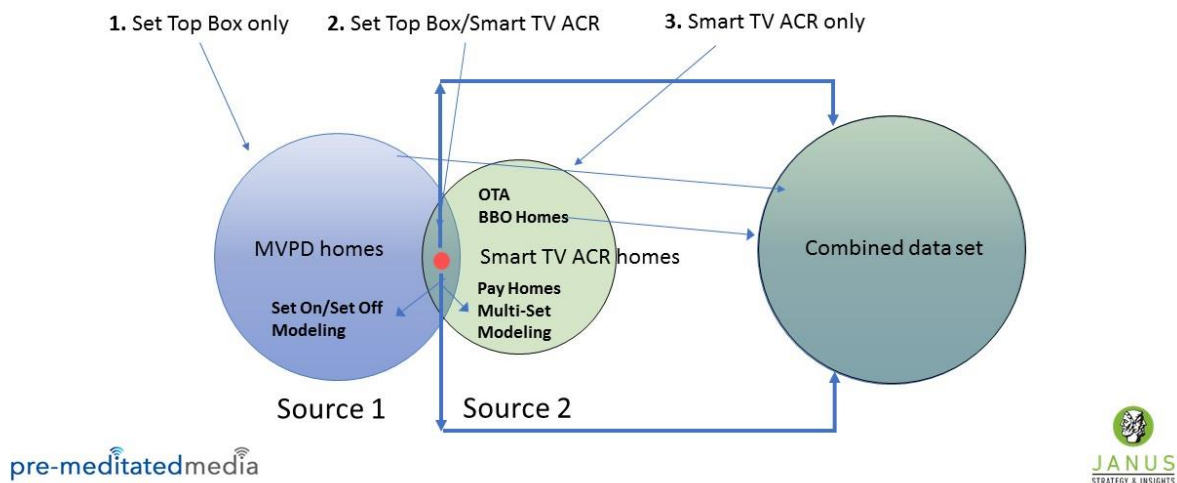
- A sample of MVPD homes providing set top box data
- A sample of Smart TV ACR homes
- A matched sample of homes that exist in both the Set Top Box data and Smart TV ACR data sets

The proper integration design includes using the matched data set to model specific attributes onto both the Set Top Box data set and Smart TV ACR data set, which will be detailed in the next stage. The modeled data sets will then be projected to the U.S., eliminating duplicated viewing sessions from the matched sample.

Stage 2 - Establish Match and Commingling Design

- Most commingled approaches use tuning data from two or more sources where device-to-device integration provides insights for calibrating the combined data set

Two-source example: Three data sets underlie the combined data set:



Stage 3- Execute Match

Stage 3 involves executing the specific match and executing specific validation to identify and minimize introduction of bias and to begin preliminary assessment of weighting and calibration requirements of the commingled data set. During this stage, Set Top Box data sets, because they originate from billing files from the various MVPDs, can easily be matched using name and postal address. Smart TV ACR data sets lack this information, but do capture device ID and IP address from homes that have enabled measurement. Many matching companies maintain databases that connect IP address to postal address. Utilizing this process provides for appending household demographics onto Smart TV ACR data, and allowing the creation of the Set Top Box/Smart TV ACR overlap sample.

For the Set Top Box providers, the commingled data set would be greatly enhanced by including flags for broadband access in the matching file. One challenge that the commingled data set will encounter is representation of linear TV viewing of Over-The-Air homes without broadband access, which are not represented in ACR. As previously reported, there are certain ethnic and socio-economic segments that over-index those homes. Receiving the Set Top Box dataset with this broadband access flag might provide signals for estimating the viewing of that segment.

As mentioned previously, the Smart TV ACR data set should be separated into the 3 segments of Pay TV, Over-The-Air, and Broadband Only homes.

It is commonly accepted that match rates for Smart TV ACR data sets are at best in the 60% range, meaning that at least 40% of data sets are not able to be used for any advanced analytic purpose. The industry should evaluate whether there is a modeling approach, using viewing data and demographic data from the matched data set that can be used to ascribe household level demographics to the unmatched sets.

The launch of Blockgraph, and its access to IP addresses from the participating MVPDs, could potentially help improve match rates for Smart TV ACR data sets.

It's important to execute the match in a highly transparent, auditable process. While the standard process is to use weighting to adjust for any demographic differences between the matched data sets and Total US population, it is important to also study the actual viewing behavior of the matched homes and the total contributing data set. Based on our discussions with data integrators, there appears to be a de facto assumption that TV viewing behavior of matched homes is similar/the same as within un-matched households, which is likely not the case.

The matching process will identify both Set Top Box and Smart TV ACR homes that cannot be matched to the selected data append company. It is estimated that approximately 10% of Set Top Box homes and up to 40% of Smart TV ACR homes will not be matched, and because of the lack of match cannot be used in the commingled data set. Our view is that the industry needs to carefully study the impact of the match, comparing both the demographics of the matched data set to the Total US population or the geography the Set Top Box database covers, and the viewing levels of matched homes and unmatched homes. Greater transparency with this stage will greatly increase the market's acceptance of this commingled data set.

Stage 3 – Execute matches

Best practices

- Segregate data into following cells:
 - 1.HHs with Set Top Box only, separated into HHs with broadband access, HHs without broadband access
 - 2.HHs with Set Top Box and Smart TV ACR, separated into HHs with broadband access, HHs without broadband access
 - 3.HHs with Smart TV ACR only, separated into the following cells:
 - Over-The-Air
 - Traditional Pay/Virtual Pay
 - Broadband Only
- Leverage HH device graph to assist in matching process, and identify potential holes in overlap data
- Metadata selection - for all post-integration reporting, utilize one dataset's metadata that is most accurate and complete. E.g., the most extensive, valid list of broadcast and cable networks
- Validate match process, comparing:
 - Smart TV ACR tuning, among matched versus unmatched homes
 - Set Top Box tuning, among matched versus unmatched homes
 - Demographics of matched homes to Total US
- Note: industry should investigate whether a methodology can be created to model HH demos onto unmatched Set Top Box and Smart TV ACR homes, using matched homes as seed set

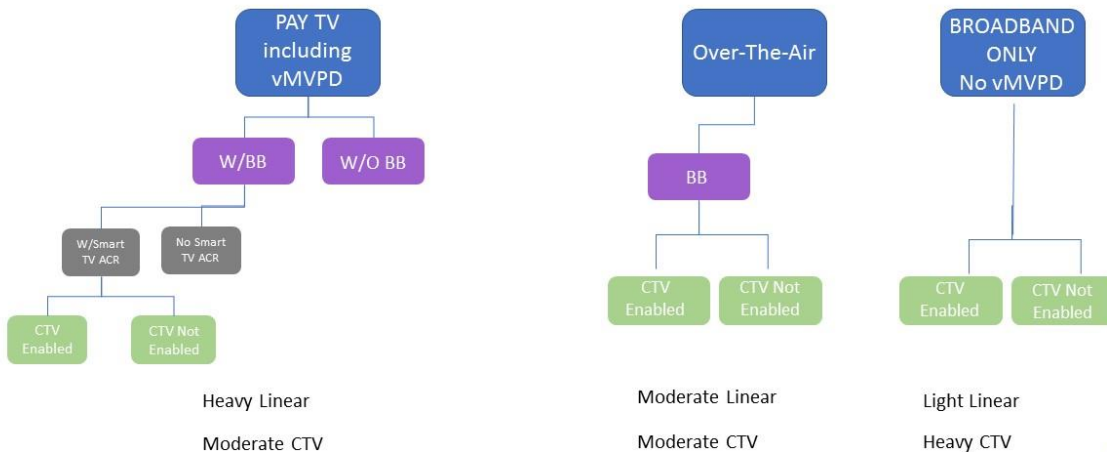
pre-meditatedmedia



Once the match is completed, the match provider will then be able to segregate the resulting databases to reflect how consumers access video content, which will be important for reach and frequency planning applications.

Stage 3 – Execute matches

Segregating matches by Pay TV, Broadband-Only and Over-The-Air recognizes the way consumers access TV, impacts the diversity of available viewing options and what they choose to watch



pre-meditatedmedia



As mentioned previously, our belief is that the matching process needs to occur in a highly transparent, auditable fashion. There are many available companies who can execute matches on both name/address and IP address, it's important to use the matching agent who delivers the highest quality match - combination of high match rate and minimization of bias from the match. Blockgraph may play a growing role in this process, because of the quality of their IP information, being based on MVPD broadband information. It may also be necessary to consider multiple match partners, based on the strength of both their matching methods and data they can append post match.

Stage 3 – Execute matches

Best practice

- Audit data matching process, especially for Smart TV ACR data with lower matching rates
 - Transparency on the demographic composition of the matched universe compared to the total US
 - Process to adjust Set Top Box and Smart TV ACR data to account for any viewership level differences between matched and un-matched HHs
- Deploy high quality matching agent, able to match based on postal and IP address
 - Evaluate if MVPDs can provide IP addresses for MVPD broadband subs to use in resolving matching issues
- Establish identity resolution technique
 - Household subscriber base (preferred)
 - Wifi/device centric
 - Decentralized (e.g., Blockgraph) vs. Centralized (e.g., LiveRamp, Experian)
- Consider utilizing multiple match partners, based on individual provider's strength with matching elements

pre-meditatedmedia



One challenge that exists with matching Smart TV ACR data is that IP addresses may churn, presenting a challenge in mapping IP addresses to postal addresses. Another challenge is that IP addresses may migrate from one household to another.

It is important when choosing a matching provider to understand the steps they go through to cleanse their IP addresses and insure highest possible rates of matching.

Stage 3 – Execute matches









Assigning IP addresses to the correct homes

- IP address churn may present challenges for correctly matching TV devices to homes
- It is possible for an IP address associated with one home to migrate to another household which could lead to misassignment

Best practice

- Ask match provider questions about data curation and process
- Consider weekly/monthly match

Example: One IP Address 255.255.255.0

	Match Provider A	Match Provider B	% of cases
Both Assign IP Correctly			% ?
One Assigns IP Correctly			% ?
One Assigns IP Correctly			% ?
Both Assign IP Incorrectly			% ?

pre-meditatedmedia



To address this issue, we have provided a list of questions to ask potential match providers to better understand their matching process, how they deal with challenges like IP addresses being associated to multiple homes.

Stage 3 – Execute matches

Questions to ask IP match provider

Best practice

- Please describe the sources for IP addresses in your database.
- What portion of your IP addresses are associated with deterministic household records?
- What portion of your IP addresses are associated with probabilistic household records?
- What is the estimated household coverage of IP addresses in your database?
- What is the range of ages for IP addresses used for matching?
- What is the average monthly churn rate of IP addresses used in the matching process?
- How do you validate that an IP address is matched correctly to a household?
- Please provide the time window for introducing new IP addresses and sunseting older ones.

pre-meditatedmedia



Stage 4- Calibration and Weighting

Stage 4 includes two important steps. The first is calibrating Smart TV ACR data sets and Set Top Box data sets based on analyses of the overlap Set Top Box/Smart TV ACR sample.

Smart TV ACR data sets can be utilized to build a better model of set on/set off and capping long viewing sessions, can be used to add precision to Set Top Box's identification of the actual time that an ad hit a TV screen, and be used to nationalize a data set and adjust for differences in viewing behaviors in each Set Top Box providers' subscribers.

Set Top Box data sets can be utilized to both adjust Smart TV ACR data sets to correct for the fact that ACR tends to reflect just over 1 set per home, also can improve models that estimate set location. In addition, Set Top Box data sets can be used to adjust Smart TV ACR data sets to account for differences between internet enabled sets and non-internet enabled sets.

Stage 4 – Calibration

Best practices for calibration

- In-depth process to leverage matched Set Top Box and Smart TV ACR panel to adjust for known limitations of each data set:

Calibration item	Smart TV ACR Adjustments to Set Top Box data	Set Top Box Adjustments to Smart TV ACR data
TV sets per home/Set Location		Model # of sets in home Use Set Top Box data to develop more refined model of set location, use in weighting Smart TV ACR only data
Over-The-Air Viewing	Use Smart TV ACR data to model OTA household viewing	
DVR/VOD playback		Model DVR/VOD breakout for time shifted tuning
Set on/Set off	Model set on/set off	
Unidentified Tuning Sessions		Attempt to fill in unidentified tuning sessions that may be ad supported TV
Bridging Rules		Account for viewing to times w/o signature library



Stage 4 – Calibration

Best practices for calibration

- In-depth process to leverage matched Set Top Box and Smart TV ACR panel to adjust for known limitations of each data set:

Calibration item	Smart TV ACR Adjustments to Set Top Box data	Set Top Box Adjustments to Smart TV ACR data
CTV Accessible Homes	Model CTV accessibility	
Actual Ad Exposure	Calibrate time of actual ad exposure based on time ad hit screen	
Viewing differences- Broadband homes versus non-broadband homes		Use Set Top Box data to adjust for any potential viewing biases in Smart TV ACR data due to limitation of representing only broadband homes
Total U.S. Coverage	Nationalize Set Top Box data sets	

Once both data sets are calibrated, the final stages are to weight each data set to the Total US population, or whatever geographic universe the commingled data set is attempting to represent. A detailed analysis of what variables and combinations of variables best explain differences in viewing for both the Set Top Box and Smart TV ACR data set prior to applying weighting. Should universe estimates by TV Access exist, from a study like the one planned by the ARF, the weighting should be done separately for Pay TV, Over-The-Air and Broadband Only homes. Monthly reach levels for all broadcast and cable networks from Smart TV ACR data should be scaled to Set Top Box levels, to account for the likelihood that the impact of Smart TV ACR data sets' limitation of one set per home on reported viewing would vary by TV network, dayparts and programs.

The individual data sets should be weighted, duplicate households removed, and then combined for reporting.

Stage 4 – Weighting

Best practices for weighting

- Apply weights to three cells for national estimate



Stage 5- Validation

The final stage is to execute a validation of the combined data set. The universe estimates from the combined data set can be compared to universe estimates from the matching provider, overall and by TV Access method if those are used.

If available, the viewing estimates from the combined data set should be validated against industry benchmarks. Both the overall levels and the relative ranking between network and between programs can be a component of the evaluation.

Stage 5 – Validation

Best practices for validation

