



# Combining Set Top Box and Smart TV ACR Data

January 2021

# Table of Contents





















- [Study Background/Goals](#).....3
- [Summary of Findings](#).....11
- [Phase 1](#).....16
  - [Set Top Box/Smart ACR Use Cases and Ecosystem](#).....19
  - [Viewing Sources Overview](#).....21
  - [Reporting Eligibility](#).....23
  - [Audience Measurement and Metadata](#).....29
  - [Matching & Data User Rights](#).....37
  - [Reported Viewing Metrics](#).....42
- [Phase 2](#).....49
  - [Set Top Box/Smart TV ACR Integration Recommendation](#).....50

# Study Background/Goals

# 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 TV transactions.
- Demand for analytics that allow advertisers and agencies to precisely plan digital video and CTV on top of linear is rapidly accelerating.
- There is a growing industry need to improve TV measurement systems that supports the exploration and discovery of best practices to commingle these two complementary data sources.

## 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		

# CIMM Project Goal

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: collect vital marketplace feedback on existing data sources and best practices for commingling

- Phase 1 – **Review of Smart TV ACR and Set Top Box providers** - collection of vital landscape statistics including sample size, data captured and reported, data processing rules 18 Interviews

Deliverable: Phase 1 preliminary report

- Phase 2 – **Review of existing methods used to integrate Smart TV ACR and Set Top Box providers** - 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 9 Interviews

Deliverable: Phase 1 and 2 combined report; best practices for combining Smart TV ACR and Set Top Box data

# Phase 1 Data Gathering

## Interview Themes

### Underlying Viewing Data Sources

- Source(s) of viewing data
- TV access types
- Data collection cadence
- Average monthly churn

### Reporting Eligibility

- % of total universe represented
- %not eligible for reporting
- Process for determining homes/sets used in reporting

### Sample Used for Reporting

- Current reporting sample size
- Reporting cadence
- Set Top Box: Coverage of homes with DVRs
- Set Top Box: Coverage of homes with VOD
- Set Top Box: boxes per home
- Smart TV ACR: Is data set aggregated to HH level?

### Metadata - Ad Occurrence

- Source of ad occurrence data
- Granularity of ad occurrence data matching to program/content

# Phase 1 Data Gathering: Interview Themes (cont'd)

## Metadata - Programs & Schedules

- Source of program data; fingerprinting, syndicated
- Source of program schedule data

## Metrics Reported

- Universe estimate base for reporting
- Common metrics provided
- Time-shifted reporting
- Program rating calculation
- Process of validating channel mapping
- Time zone adjustments; east/west coast feeds
- Reporting clock synchronization
- Editing rules
- Smart TV ACR: Local mkt assignment

## Data Licensing Rights and Usage

- Restricted elements of data set
- Primary data use cases
- Match rights

## Data Matching & HH ID & Sample Representation

- Method for matching HH viewing data with demo comp
- Provider of HH demos
- Match rate
- Process for appending attributes to un-matched homes
- Smart TV ACR: process for determining IP address for match when multiples IPs



# Feedback gathered from 18 entities

## MVPDs



Major MVPD



## OEMs/ Smart TV ACR Providers



## Third-party integrators



# 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.

# Summary of Findings

# Summary of Phase 1 Findings

- **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 to assess results, 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.
- 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
- 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

# Summary of Phase 1 Findings

- **Metadata overview**

- Nielsen Gracenote (programming and scheduling) and Experian (demo/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 will gain access to this data.

- **TV data processing, while not standardized across data providers and third-party firms, is far more systematized 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 Smart TV ACR and Set Top Box 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, 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 piqued curiosity regarding CIMM study results.

# Summary of Phase 2 Findings

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

## **Stage One - Data set selection - best practices**

- Utilize Set Top Box data sets that span multiple traditional/virtual MVPDs and Smart TV ACR data providers
  - To ensure representativeness of viewer footprint and amplification of complementary measurement properties of both data collection techniques
- Recognize diversity of household TV access on tuning behaviors that reflect changing landscape of TV viewing
  - Apply consistent definition and sample inclusion of Over-The-Air, Pay TV, and Broadband-Only homes
  - Leverage 2021 ARF Media Universe Estimates study

## **Stage Two - Establish match and commingling design - best practices**

- Use tuning data from homes with Set Top Box-to-Smart TV ACR device matches to inform calibration of combined data set estimates, 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

## **Stage Three - Match execution - best practices**

- Deploy high quality matching agent, able to match on postal and IP address
- Leverage HH device graph to ensure representation of Over-The-Air, Pay TV and Broadband-Only homes

# Summary of Phase 2 Findings

## Stage Three - Match execution - best practices (cont'd)

- Validate 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.
- Ask IP match provider questions regarding quality of data records
  - Recency, churn rate, deterministic vs. probabilistic, life span, etc.

## Stage Four - Calibration & Weighting - best practices

- Key calibrations made to data sets
  - **Set Top Box adjustments to Smart TV ACR** - # of sets in home, DVR/VOD, backfill reference for 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
  - 1. U.S. demographics, TV access universe, tuning metrics, geographics

## Stage Five – Validation – best practices

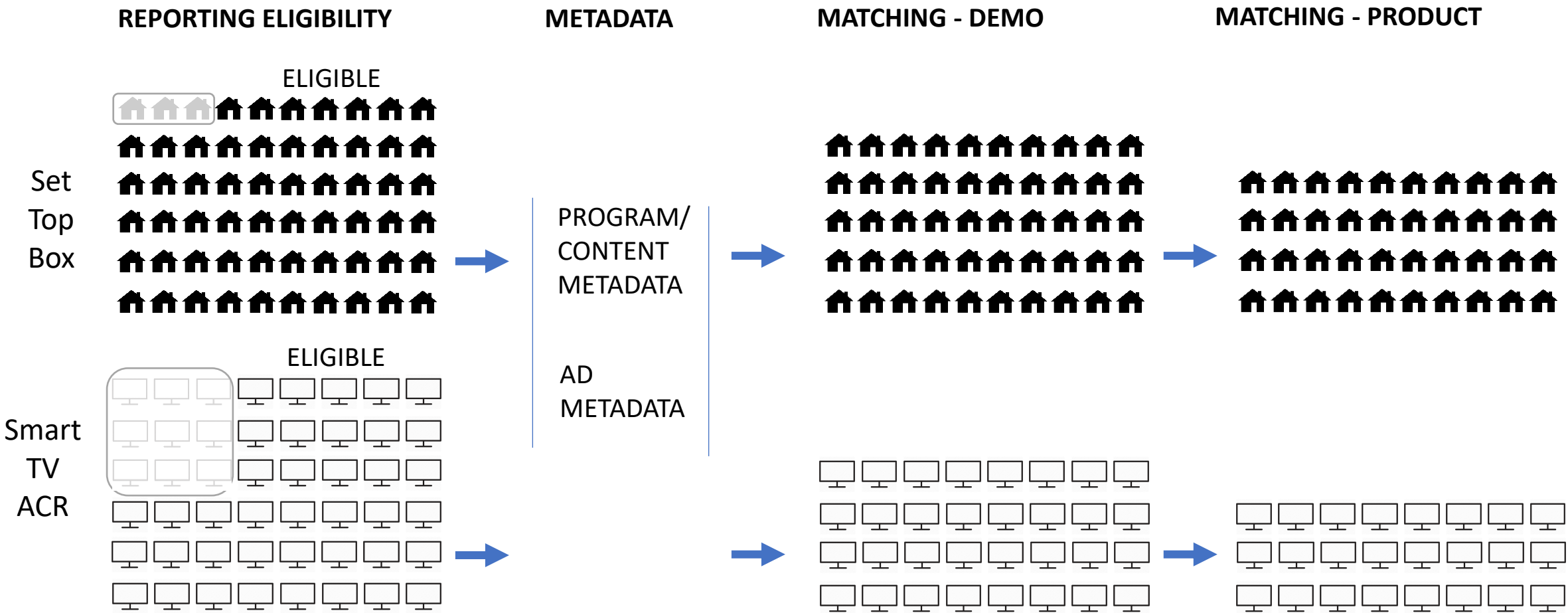
- Validate universe and tuning estimates

# Phase 1



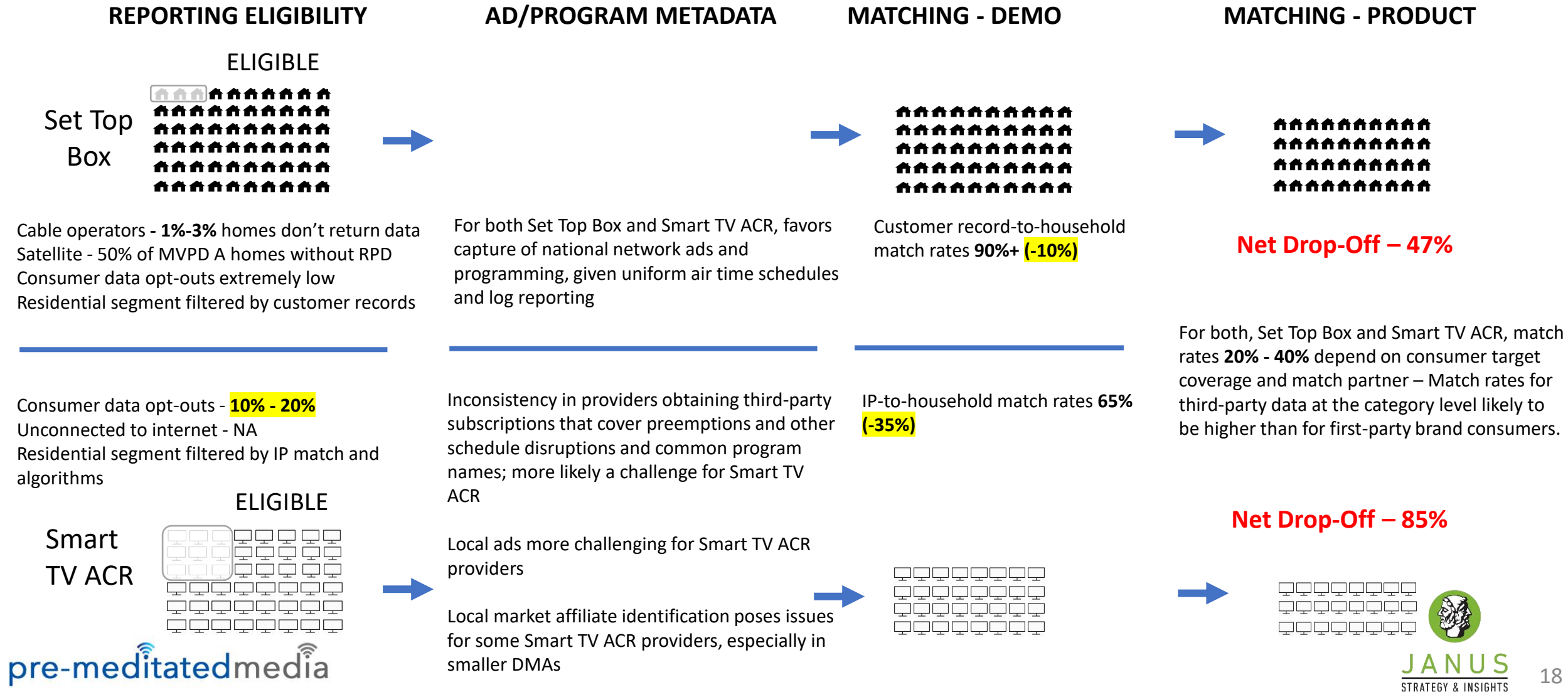
# Data Integration Pathway

A key goal of any data integration is to produce a reported output that reflects the original population in terms of media and product consumption



# 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.



# Set Top Box/Smart TV ACR Primary Use Cases and Provider Ecosystem

# Use cases for more granular, integrated data sets

- TV audience ratings
  - Programming analytics for MVPDs and national networks
- Campaign Planning and Optimization
  - Interconnect, national for local cable inventory, national for network inventory
  - Age/sex and advanced audiences
- Addressable TV campaign activation and measurement
  - Advanced audiences
  - Incremental reach
- Deduplicated reach/frequency with digital video and CTV
  - Campaign Planning
  - Activation- Impression placement to support integration of linear with digital video, addressable, CTV
  - In-flight and post campaign measurement
- Attribution/Outcome measurement

# Viewing Sources Overview

# Viewing data sources

## Underlying Sources

### Set Top Box

MVPDs tap customer event-level tuning data from set top boxes, many of which have been updated during the past several years. Some issues with non-coverage

## Collection Cadence

Data for boxes is generally transmitted on a time-staggered basis, usually every hour or once or twice per day. Transmission latency for some boxes may be due to turned-off boxes, or devices that are programmed to send data during a certain point during the day.

## Churn

Cord cutting has been the primary catalyst for driving MVPD pay subscription churn in recent years. Rule-of-thumb turnover is pegged at a 5%-6% annual clip for most MVPDs

### Smart TV ACR

OEMs collect data from Smart TV ACR signals, reflecting content and advertising appearing on the TV screen.

Virtually all Smart TV ACR data is amassed on a continuous basis, then aggregated to reporting specifications.

Churn is driven by the add/replace cycle of Smart TVs versus changes in TV access status that encompasses pay, broadband-only and pure over-the-air models. Some Smart TV ACR providers cited 2% monthly mark. Others tied churn to the dynamic rotation of IP addresses that makes it challenging to gauge turnover of specific devices.

# Reporting Eligibility

# Reporting eligibility – Homes and Devices

## Device 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 (e.g., one MVPD excludes partially-active homes where some boxes aren't returning data, about 30% of footprint)

## Consumer opt-outs

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

## Non-residential Homes

Providers have ability to filter non-residential homes from data sets. MVPDs utilize subscriber files and different codes for residence and business.

### Smart TV ACR

Most Smart TV ACR providers work from a base of smart TV sets that are no older than five years, must be internet enabled and with no data transmission issues.

For Smart TV ACR, about 10% - 20% of TV set buyers decline to have their viewing data collected. Most data providers lack intel on how much the demographic and TV viewing profile of decliners differs from the rest of the customer base.

Smart TV ACR providers use a combination of matching and multiple IP addresses to identify likelihood that device location is a bar, restaurant, retail store, etc. versus a household.

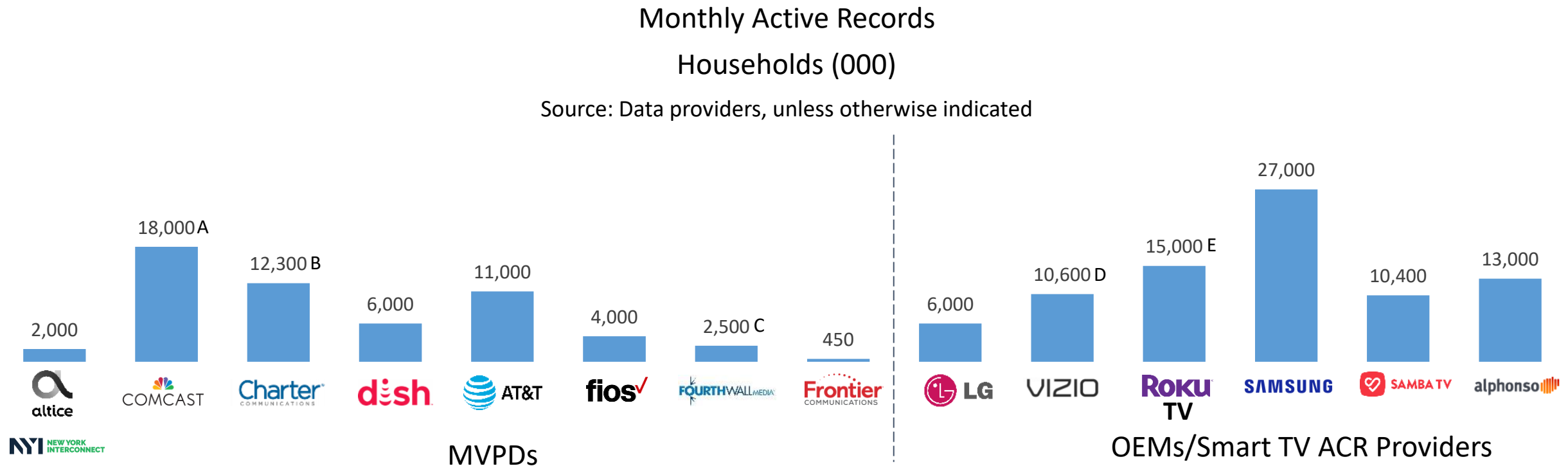


# Reporting household count overview - Preface

- Household counts represent the **gross** total of U.S. homes reporting viewing during a 30-day period, gathered from all data sources used **prior** to the commingling process.
- After commingling procedures are executed, household counts reporting viewing will be lower and **not included in this report**. The depth of this reduction may be impacted by the following:
  - Match rate levels between and across the TV datasets
  - Deduplicating the Set Top Box and Smart TV homes yields a **net count** of homes which is **lower than a gross count**
  - Exclusion of households that do not register minimum reportable levels of viewing during a 30-day period
  - Potential technical Set Top Box or TV set device reporting failures that occur on a sporadic basis
- The counts were sourced from MVPDs, OEMs and third-party aggregators during the study interviews, unless otherwise specified
- The counts have not been verified by an independent entity, e.g., third-party auditor

# 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



A Pre-Meditated Media, Janus Strategy & Insights estimates based on public domain sources

B STB only; this does not include app data

C Represents multiple 2<sup>nd</sup> and 3<sup>rd</sup> tier MVPDs

D Vizio Nationally Representative Panel (NRP) data set base is 3.5 mm homes

E Pre-Meditated Media, Janus Strategy & Insights estimates

\*Not reflective of all OEMs/MVPDs

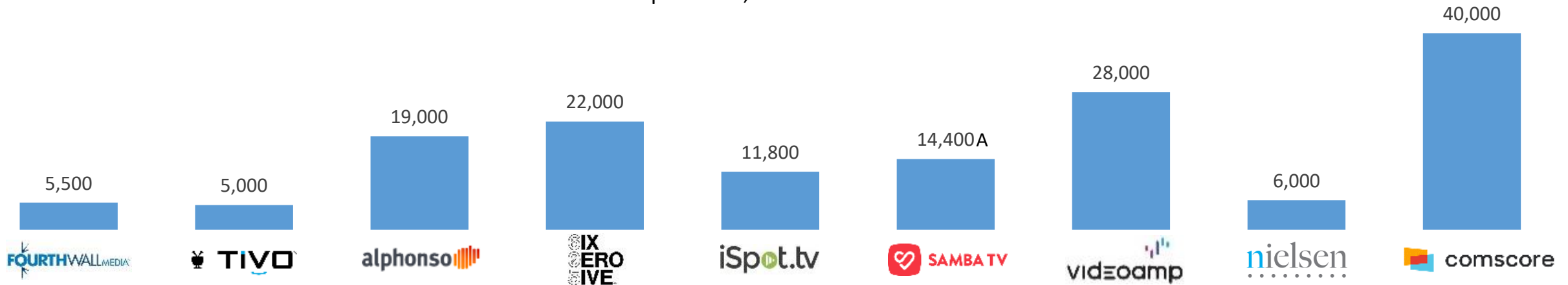
# Reporting household count overview – Third-party integrators

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.

## Monthly Active Records – Third Party Integrators

Households (000)

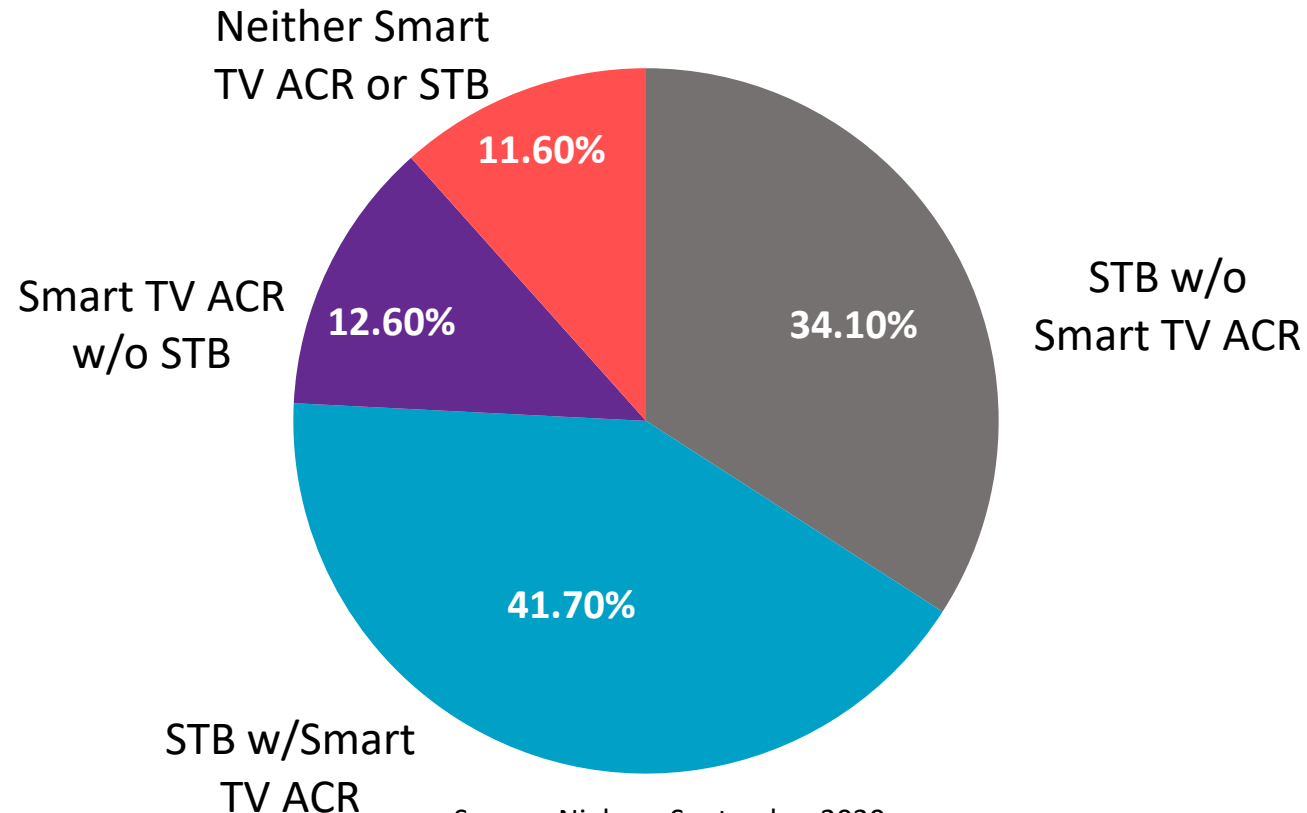
Source: Data providers, unless otherwise indicated



# 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

HH % coverage by device technology



Source: Nielsen, September 2020

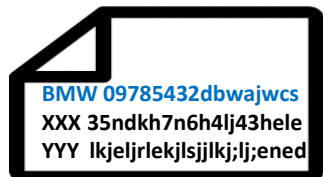
# Audience Measurement and Metadata

Ad Occurrences and Program data

# Audience measurement and metadata – Set Top Box vs Smart TV ACR

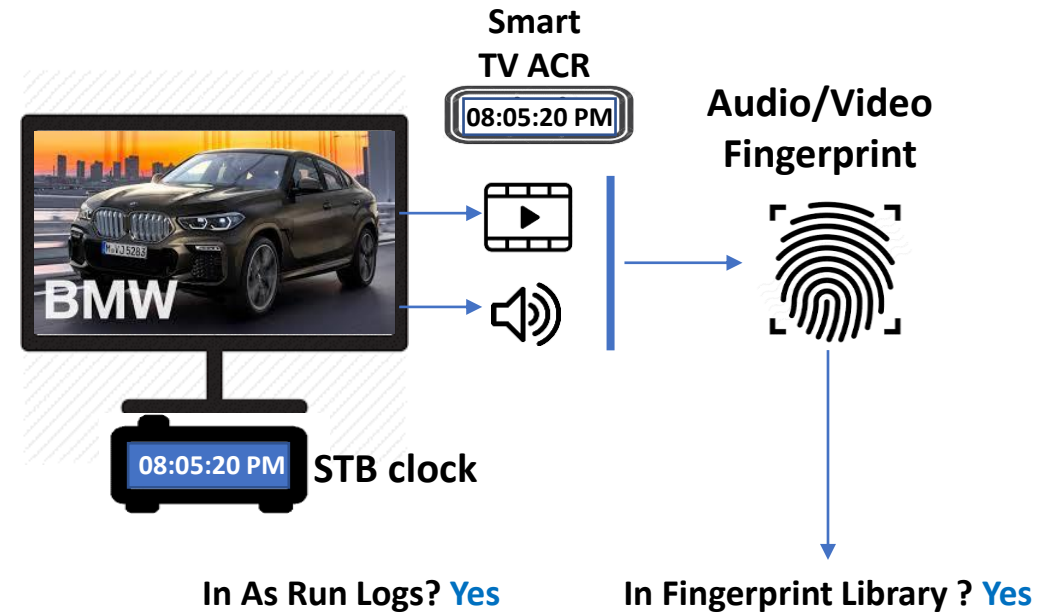
## National ESPN/BMW ad

- Primary distinction in viewing data collection – Set Top Box **tuning** records versus Smart TV ACR **fingerprinting**
- As run ad logs from third-party sources and fingerprinting libraries provide ad occurrence data



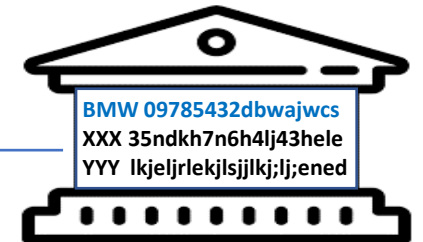
### Sources

Hive, Kantar  
Kinetic, Nielsen Ad Intel



### Sources

Hive, Kantar  
Kinetic, Nielsen Ad Intel



# Metadata - Ad occurrences

- Ad occurrence data are sourced both internally and from third-party firms

## Ad Data Sources

Set Top Box

Internally produced - from log files of local MVPD ad breaks

Externally sourced – Hive, Kantar, Kinetic, Nielsen Ad Intel  
Smart TV ACR providers (one MVPD uses iSpot.TV)

Smart TV ACR

Internally produced - 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. Limited line of sight to local MVPD ads.

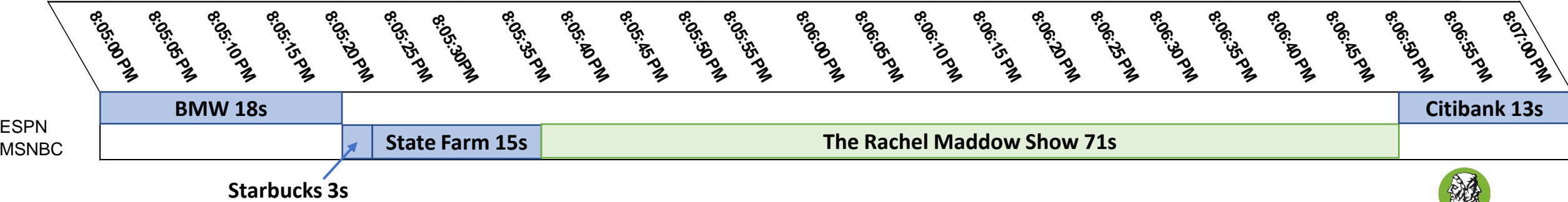
Externally sourced - other data providers tap third-party firms such as Hive, Kantar, Kinetic, Nielsen Ad Intel, iSpot, etc. for ad occurrence data

- Second-by-second mapping of ad occurrences to viewing is critical for crediting shorter unit and partial ad views

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

### HOUSEHOLD A TUNING LOG

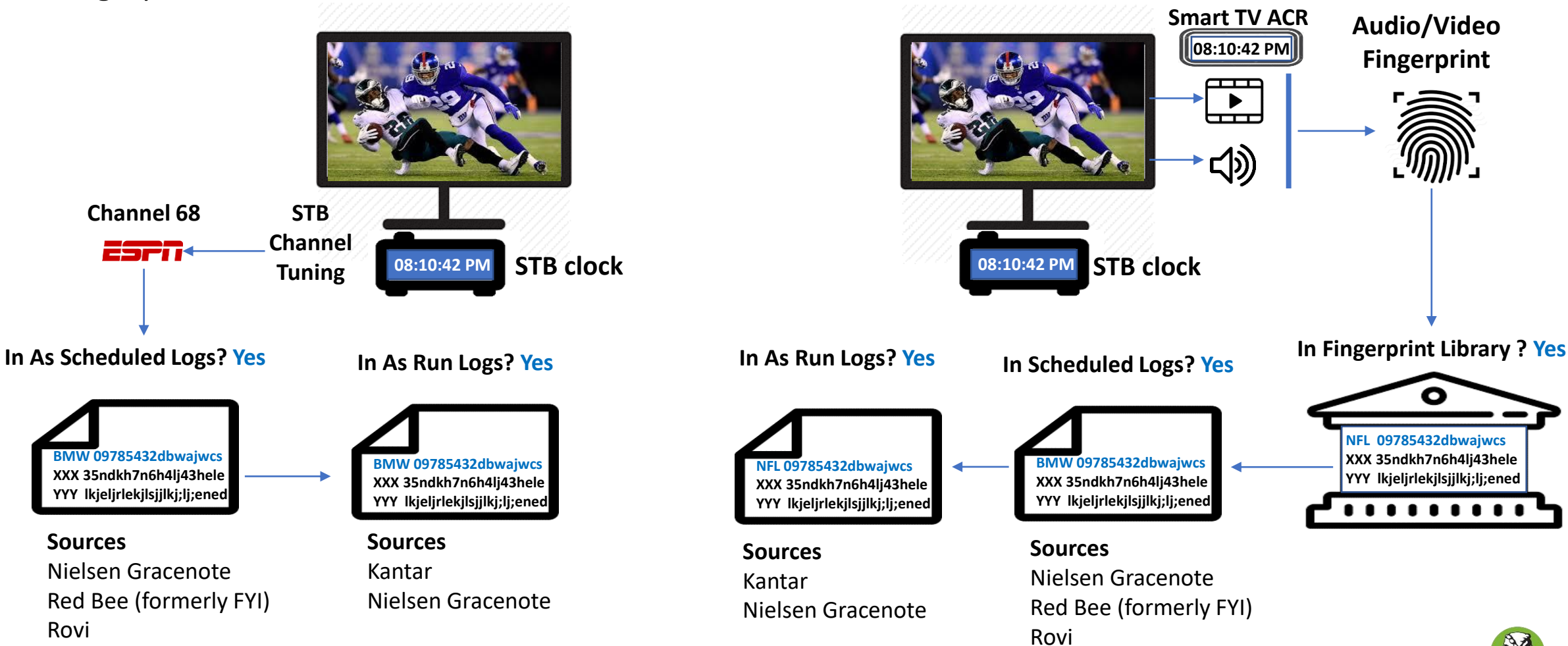
#### ESPN TWO-MINUTE COMMERCIAL BREAK



# Audience measurement and metadata – Set Top Box vs Smart TV ACR

## National ESPN NFL program example

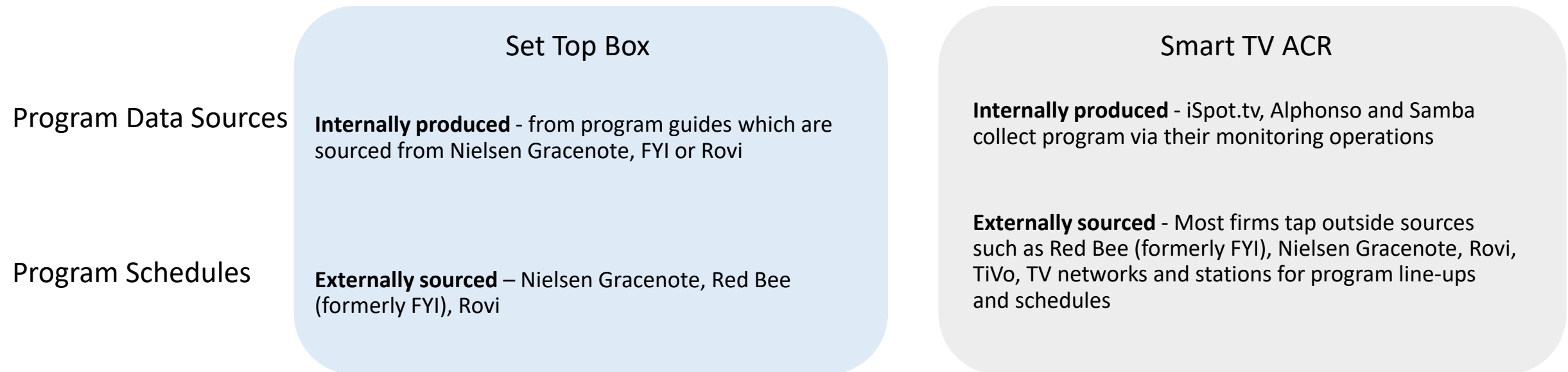
- Program content and schedule data are sourced largely from third party sources. Smart TV ACR fingerprint libraries





# Metadata – Programs and schedules

- Program data are sourced from third-party firms
- Inconsistency in providers licensing both planned and actual air schedules
  - Not all providers incorporate final “as-run” schedules in their reporting. For example, they may incorrectly capture sporting events that run beyond their scheduled time slots and that preempt the start of shows that follow. Additionally, program naming conventions made available by Nielsen\* and other providers are sometimes used inconsistently.










# Metadata – Source Summary

## Ad Occurrences

- National and local data
- Number of networks and stations will vary
- Local coverage is stronger for broadcast stations, limited for local cable





### Sources

	Screen-based monitoring of video content Modeled based on local airings for 7 monitored markets
	Automated monitoring in all local markets
	Kinetiq's global monitoring network
	Automated monitoring in all local markets
	Smart TV ACR in 30+ DMAs
	Alphonso's proprietary Ads catalog
	Not Provided

## Program content/schedule data

- National and local data
- Number of networks and stations will vary

### Sources

	Gracenote, Kantar Editorial Work Station (see Gracenote)
	Gracenote TV networks/stations provide schedules directly to Gracenote
	TV networks/stations provide schedules directly to Red Bee
	Not Provided



# Metadata - Mapping the channels

- **Channel mapping**

- For MVPDs, tying system channels to cable networks is simple, utilizing channel guide information
  - Look at actual programming for locality indicators
  - TiVo and Gracenote metadata sometimes used
- For Smart TV ACR providers, channel is not used to directly assign viewing
  - Look at program signature captured on set relative to library
  - Assign to channel based on monitoring and metadata

- **Local market stations** - 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

- Smart TV ACR provider can make 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 affiliation changes are more frequent and monitoring may not exist
- Other sources are used: Nielsen Gracenote



WABC  
New York

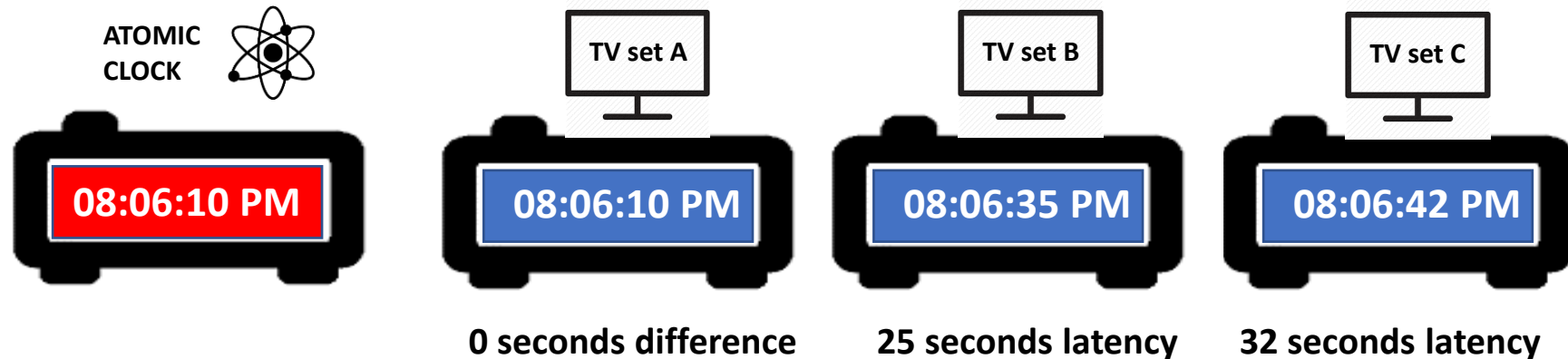


WTNH  
New Haven



# 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 sec x sec 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.
  - MVPD providers claim that calibrating clock and signal latency is a much smaller issue than several years ago
    - 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 , rolling up audience metrics across the split feeds

# Matching & Data User Rights

# Matching viewing and demographics

- **High MVPD demographic household match rates (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.
- **Lower Smart TV ACR demographic household claimed match rates (70%-80%)** - providers use IP address as primary match key when combining with name-and-address files.
  - Match rate appears to be high, given the device-to-household match key and bears further investigation. Expected match rates would be lower when combining demo and product usage data, depending on data partners
- **Inability to evaluate potential non-response bias** - in virtually all Set Top Box and Smart TV ACR cases:
  - Homes without data collection, homes who don't opt in
  - Homes that don't match with match partner

Assumption that weighting to US population corrects for errors

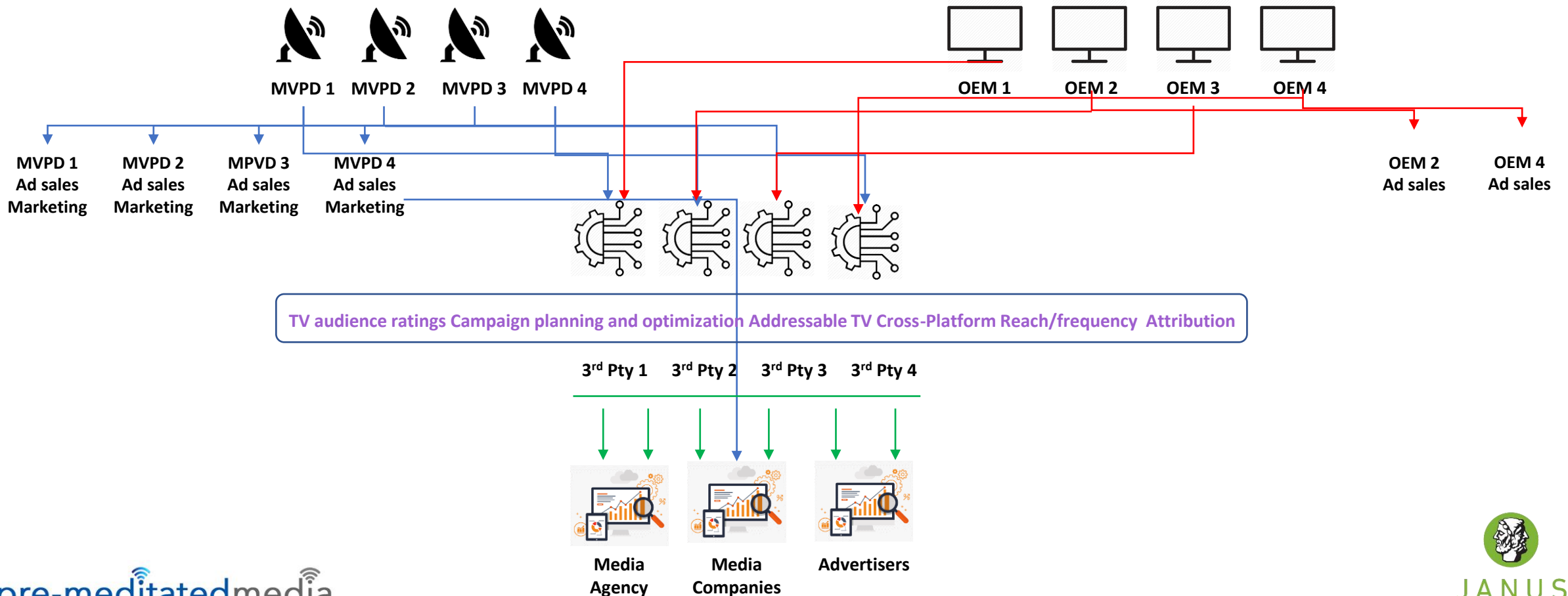
- **No/low reconciliation of underlying data issues** - e.g., Experian, accuracy of gender and ethnicity classifications. For Smart TV ACR, unmatched viewing records are not modeled for demographics and product usage when matched sample size is sufficiently large; resulting match may not represent true universe
- **End users have option to apply their own weighting** - many raw data sets are sent to end users who can select and apply weighting variables at their discretion
- **Need for persons panel to inform estimates** - Comparing Set Top Box and/or Smart TV ACR data to nationally representative panel data 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 soso (set-on/set off) editing rules and/or modeling persons estimates.

# Data user rights

- Every provider ensures that subscribers and TV set owners have granted data usage permission
  - Third-party processor example: MVPD subscriber notifications are sent periodically and audits conducted every two years
- Very few restrictions on data usage and match rights. Exception examples:
  - MVPD prohibits sharing of viewing data with competing MVPDs
  - MVPD prohibits TV network use for program carriage valuations and negotiations but can use for ratings and ad hoc analyses when data is aggregated with data from other MVPDs.
  - One OEM restricts usage when viewing metrics 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












# 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	  

# Reported Viewing Metrics

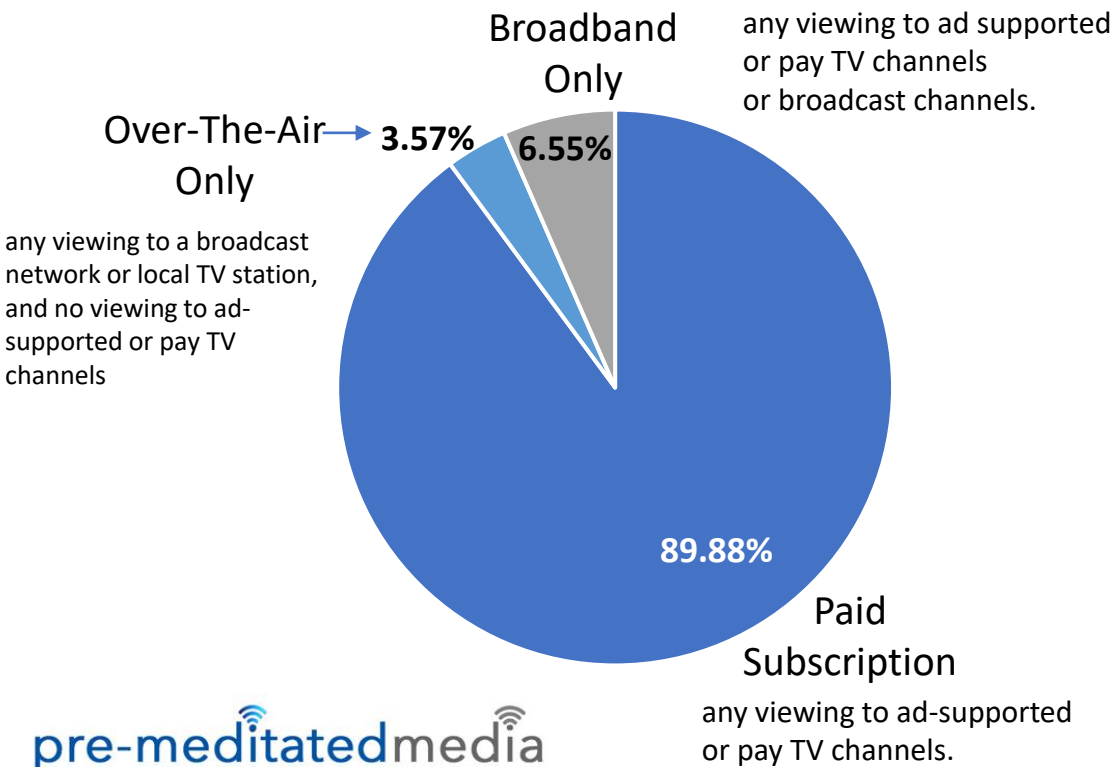
Gaps and overlaps (across sources)

# Reported Viewing Metrics - Universe representation

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 pay TV but for the most part do not have access to streaming OTT viewing data from Broadband-Only customers. Smart TV ACR providers capture data across both linear and OTT, however, DVR and VOD reporting processes are nuanced. Viewing data sourced from one Smart TV ACR provider revealed a heavy skew towards pay TV homes.

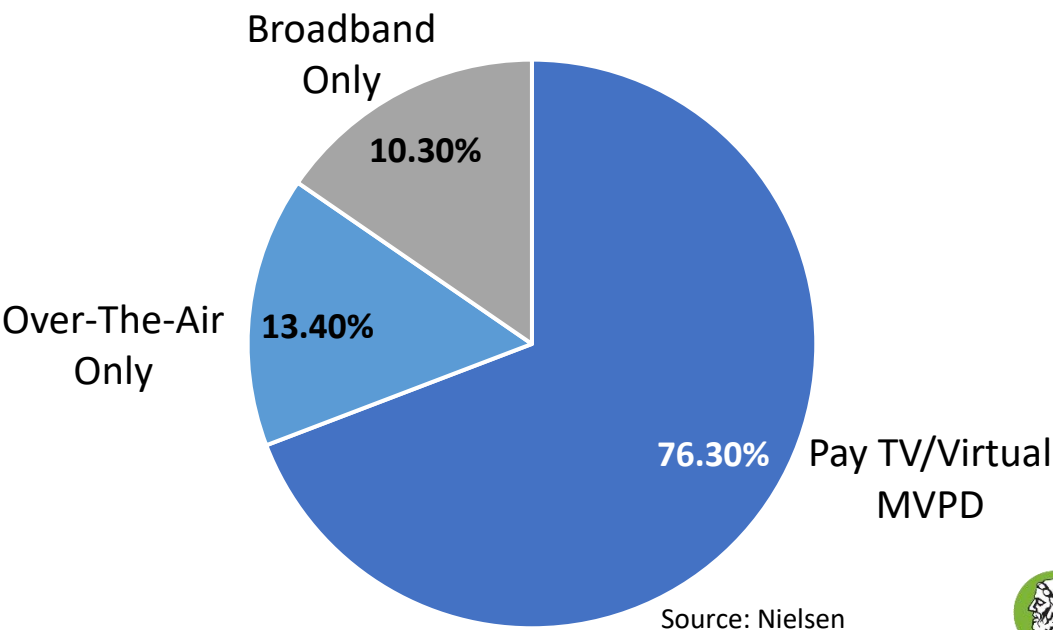
**TV Access Mode (Smart TV ACR Provider)**

**% Homes Feb 2020**



**TV Access Mode**

**% U.S. Homes June 2020**



Source: Nielsen

# 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
- **Non-live 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 Boxes are limited to DVR playback with VOD reporting supplemented from other sources, e.g., comScore; 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 DVR or VOD source. For example, a distinction could be inferred by referencing ad load (greater for DVR vs VOD) 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 OEM 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.
- **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 3<sup>rd</sup> 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 Smart TV ACR viewing reports. Inability to identify OTT ad exposure to ads that do not exist in fingerprint library (i.e. Smart TV ACR can report everything that hits the screen for which there is a library fingerprint, however, channels or networks can't always be ID'd)
- **Demographic appends** - Both Smart TV ACR and Set Top Box 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.

# Areas of Concern

- **Adequate HH coverage** - Fully integrated database will have incomplete coverage:
  - Some Smart TV ACR 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

# Drilling down on the concerns

- Commingled databases will still have incomplete coverage:
  - Name/address files or IP address files without matches to third-party append companies
    - Cannot contribute to weighting
    - Cannot be included in data-driven linear (DDL) or be targeted for addressable or CTV
- Because there is no direct capture of household demographics from Set Top Box or Smart TV ACR homes, accuracy of weighting scheme is only as good as accuracy of appended demographics
  - Impact on weighting scheme difficult to determine
- Proper integration of Set Top Box and Smart TV ACR would require common second-by-second metric definitions and universe estimates that can be applied to both databases
  - Weighting scheme and integration will be incomplete until this is accomplished

# Areas of Opportunity

Phase 1 feedback bodes well for future development of granular Set Top Box/Smart TV ACR data sets that support advanced targeting and placement optimization on linear TV

- Integration processes are maturing, enabling more flexibility and potential for standardization
  - Most firms have mastered advanced ETL processes for prepping and providing data sets
  - Data is made available in flexible format so that end users can define metric parameters, e.g., thresholds for ad exposure and time-shifted viewing, with second-by-second being the lowest common denominator
  - End users looking to standardize metric parameters across multiple sources will have an easier time of it versus three years ago, for example
- Set Top Box and Smart TV ACR data formats possess some similarities
  - Both can be processed at the second-by-second event level
- Early metadata consistency - Some companies are being widely tapped for metadata, thereby offering some level of reporting consistency that can support cleaner comparisons of different data set outputs during these early developmental days
  - Nielsen Gracenote was most prominently mentioned as a source for program, channel and scheduling data. The networks provide “as run” logs to Nielsen, which is why the data are more accurate; ideally this data will be made available to other providers to improve accuracy of “as run” reporting of program names.
  - Experian was cited as the primary demographic matching partner for most of the providers and processors, however, other partners were mentioned for cross platform and product usage including Epsilon, Acxiom, Oracle, Neustar, LiveRamp and Merkle
- Experimentation is accelerating
  - Many providers are now working with several data sources either within or across MVPD and OEM universes
  - Most interviewed were enthusiastic about providing insights for their data and are highly interested in CIMM’s study results



# Phase 2

# Set Top Box/Smart TV ACR Integration Recommendation

# 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

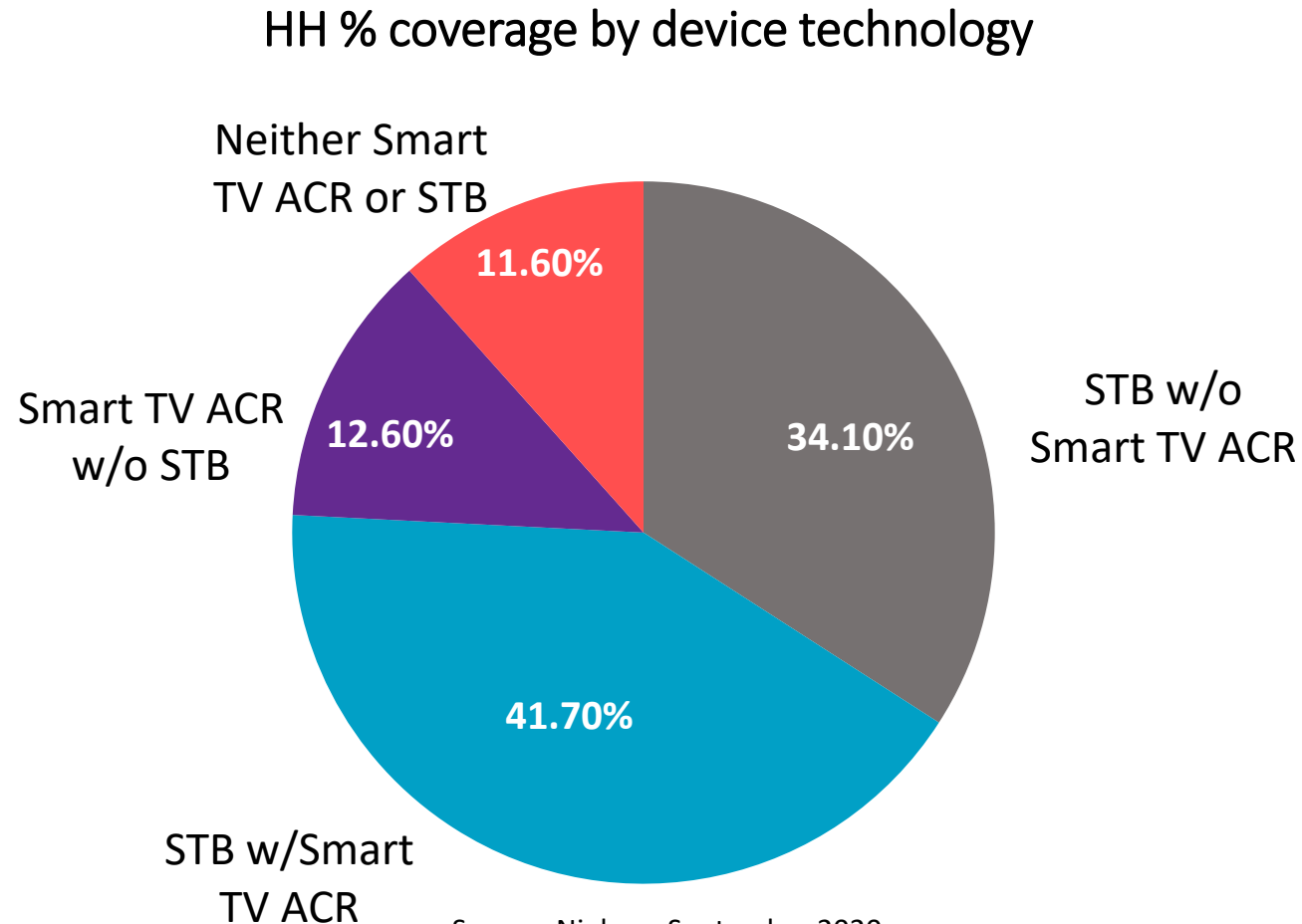
- Cost/benefit
  - How much additional accuracy is achieved when combining Set Top Box and Smart TV ACR?
  - Premiums paid for licensing of multiple data sets and for staffing
  - Anticipated growth rate of attribution, analytics, targeting, campaign optimization, etc.
- End user marketplace – types 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



# 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



Source: Nielsen, September 2020

# 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
HHLI Income = <\$75,000	92	79	72	117	128	105	149	94	87	107
HHLI Income = \$75,000+	109	125	133	80	67	94	42	107	115	92

# 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

# 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

# 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 1 – Select data sets

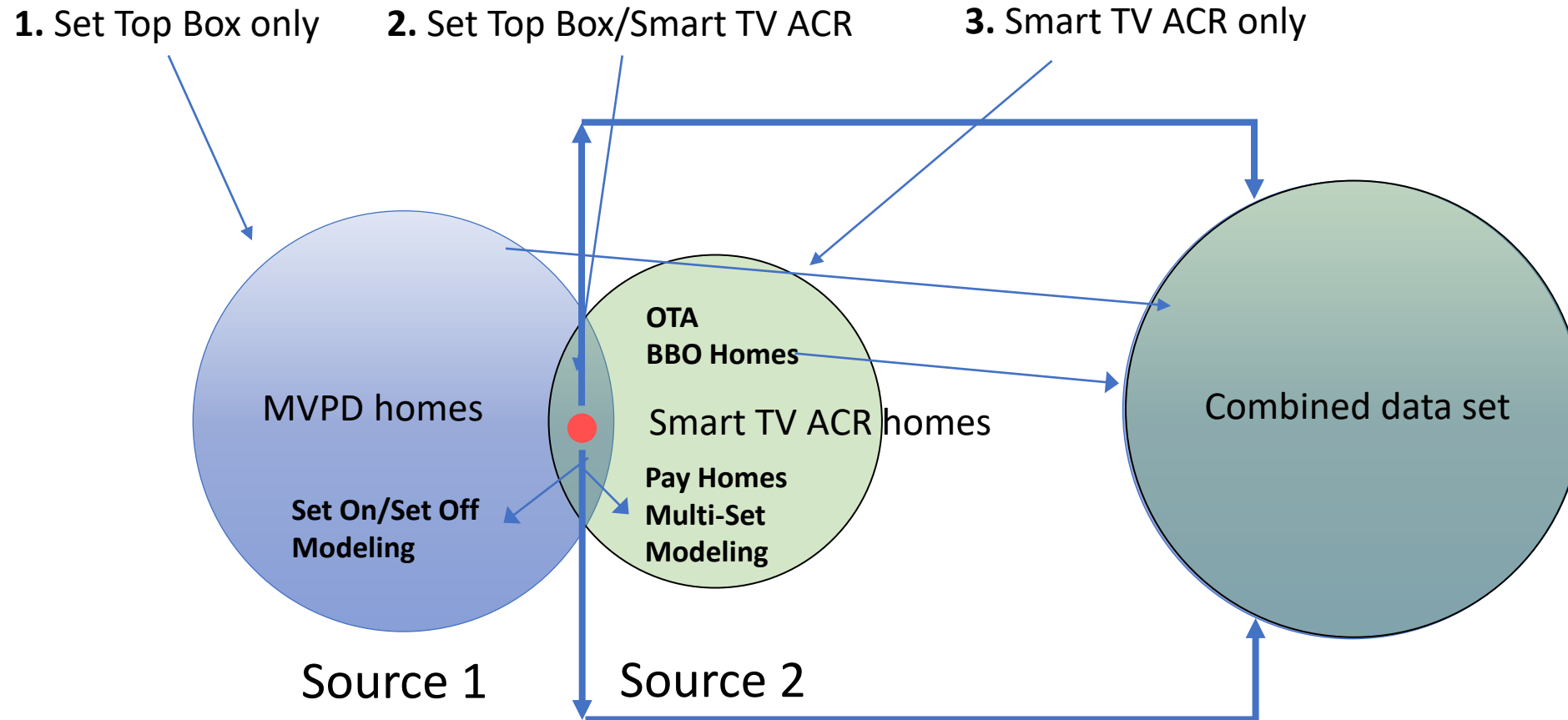
## *Best practices*

- 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

# 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:



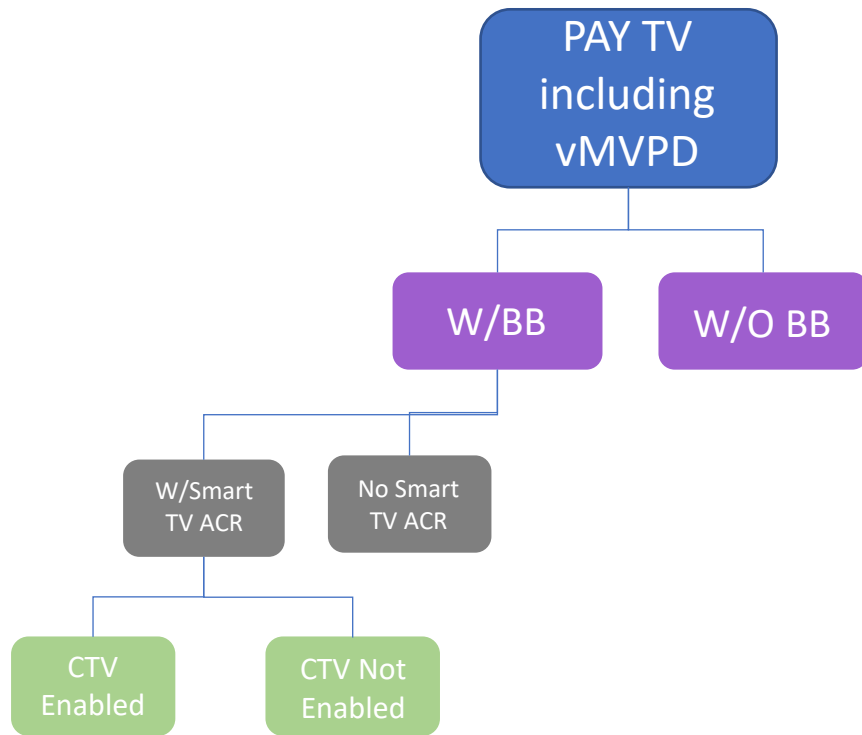
# Stage 3 – Execute matches

## *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

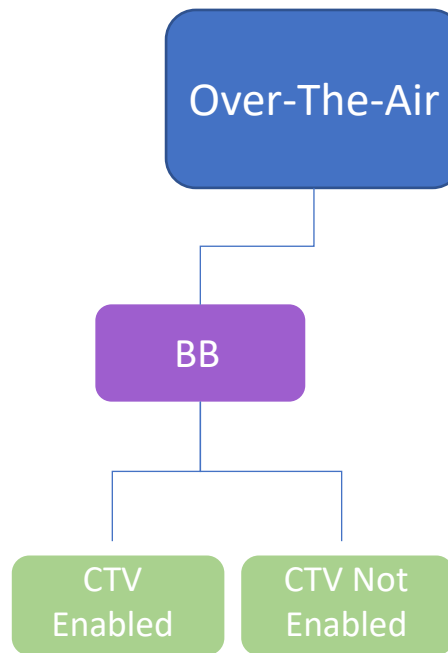
# Stage 3 – Execute matches

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



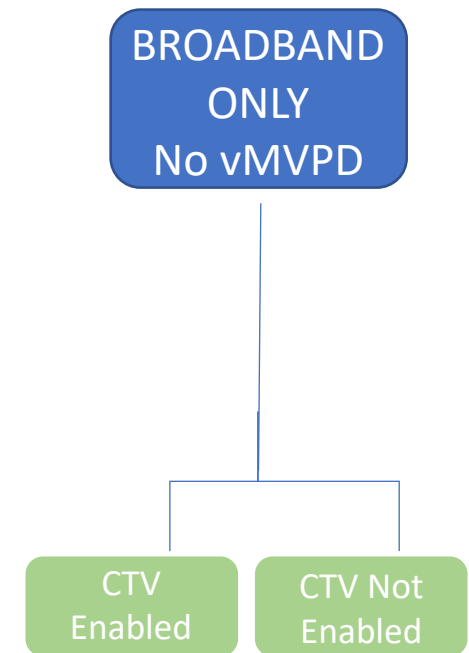
Heavy Linear

Moderate CTV



Moderate Linear

Moderate CTV



Light Linear

Heavy CTV

# 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

# 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			% ?

# 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.

# 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:
  - Using Smart TV ACR data to adjust Set Top Box data:
    - Nationalize Set Top Box data sets
    - Model set-on/set-off
    - Calibrate for time of actual ad exposure based on time ad hit screen
    - Differences in viewing- homes with enabled ACR smart TV measurement versus those without
    - Model CTV accessible homes
  - Use Set Top Box data to adjust Smart TV ACR data:
    - Account for limitation in number of sets per home, use to build model for primary/second set identification
    - Adjust for linear TV tuning level differences - enabled smart TVs households compared to households without enabled smart TVs
      - Adjust by ethnic segment, including Spanish Dominant Hispanics
    - Model DVR/VOD breakout for time shifted tuning
    - Bridging rules to account for tuning to networks during times without signature libraries-i.e. local cable breaks
    - Use Set Top Box to attempt to fill in unidentified tuning sessions that may be ad supported TV



# 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	

# Stage 4 – Weighting

## *Best practices for weighting*

- Apply weights to three cells for national estimate



### Demographics

Age  
Gender  
Children  
Ethnicity  
Income  
etc.



### Tuning Metrics

Tuning volume and shares by:  
Broadcast network  
Pay network  
etc.

Scale Smart TV ACR to Set Top Box levels



### TV Access Universe

Pay Subscribers  
Broadband Only  
Over-The-Air



### Geographics

Applies to MVPDs when nationalizing estimates

# Stage 5 – Validation

## *Best practices for validation*



### Universe Estimates

Ensure estimates reflect tuning population

TV access mode: Pay TV, Broadband Only, Over-The-Air

CTV-Accessible Homes



### Core Tuning Estimates

Track and compare tuning estimates to industry benchmarks

Overall TV usage

Share of broadcast and pay networks

Share of linear, on-demand and CTV consumption

Top 100 rated programs