

# UPDATE ON CIMM'S DATA MATCHING STUDY

## Status Report



**Jim Spaeth, Sequent Partners**



**2022 SUMMIT**

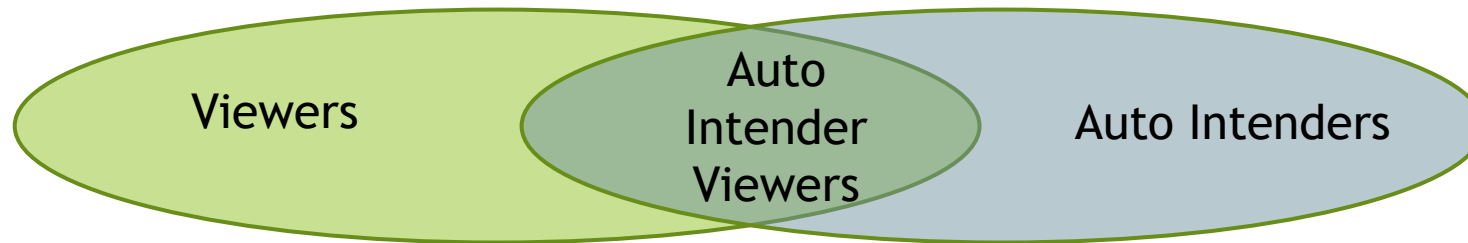
# **The CIMM Study of Identity Resolution and Its Impact on Matched Datasets**

Janus Strategy & Insights, Pre-Meditated Media and Sequent Partners

February 17, 2022

# Why?

- ▶ Advanced advertising and attribution depend upon matching TV viewing data to audience characteristics and outcomes data




- ▶ Data records are lost in the matching process
  - ▶ Do we still have enough data?
  - ▶ Has the viewing or outcome data been biased in the process?
  - ▶ Are there best-practices for maintaining data integrity?

# Who?




Thanks to those who are making this study possible ...

## IDR PROVIDERS

- ▶  EPSILON®
- ▶  experian™
- ▶  CADENT
- ▶  THROTTLE
- ▶  infutor
- ▶  neustar®  
A TransUnion® Company
- ▶  TransUnion™
- ▶  LiveRamp  
an axxiom company



## TEST DATA PROVIDERS

- ▶ Viewing data providers:  
 SAMBA TV  TiVo
- ▶ Outcome data provider:  
PlaceIQ (QSR visits)
- ▶ Audience characteristics data:  
 TBD (Household demos)

# How?

- ▶ Compare original test data to post-match test data
  - ▶ Match Rate? Data retention?
  - ▶ How did test data aggregate profiles change post-match?
- ▶ Eight matches:
  1. STB viewing data to IDR provider's Identity Graph
  2. Retail visit data to IDR providers Identity Graph
  3. STB viewing data to retail visit data
  4. Demo data to IDR providers Identity Graph
  5. STB viewing data to demo data
  6. ACR viewing data to IDR provider's Identity Graph
  7. ACR viewing data to retail visit data
  8. ACR viewing data to demo data

# Pre-Phase: What Have We Learned

- ▶ The path IDR providers take to develop ID Graphs is varied
  - ▶ Some start with the household: postal address, Zip11
    - ▶ HH devices associated (e.g. CTVs)
    - ▶ Persons (or personas) associated
    - ▶ Personal devices associated (e.g. mobile phones)
  - ▶ Others start with devices or email addresses associated to persons and build-up to households
    - ▶ Depends on their foundational data sources
  - ▶ Some maintain separate HH and persons graphs and link them
  - ▶ Mostly deterministic, with some probabilistic associations

# Pre-Phase: What Have We Learned

- ▶ IDR providers assemble as many match keys as possible, filling-in any gaps in their device graphs
  - ▶ HH: name, address, Zip11, lat/long, IP address, phone number, CTV IFR, user agent strings
  - ▶ Persons: hashed email, Maid, phone number
  - ▶ For this study, match keys: STB= address, CTV=IP address, QSR Visits=MAIDs, Demos=Zip11
- ▶ Quality assurance
  - ▶ IDR providers consult multiple data sources to check the likelihood of each match key association
- ▶ Privacy is the top priority
  - ▶ Firewalls, clean-rooms, personification, pseudonymization, anonymization, ...

# Pre-Phase: What Have We Learned

- ▶ IP addresses are not fixed – main causes: purposely cycled, router/internet interruptions
  - ▶ Freshness of data is critical
- ▶ Street address files may not be 100% accurate due to move rates
  - Some issue with single vs. multi-unit residences
- ▶ Majority of television STB data matches go through Experian
  - ▶ Legacy source for name-and-address records
  - ▶ Requiring two-step matching with outcomes and audience characteristics



# Where Are We?

- ▶ Great efforts are made to ensure the accuracy, coverage and completeness of identity graphs leveraging tons of data
- ▶ But we still don't know what happens to viewing and outcome data as it is transformed through the matching process
- ▶ And what could be done to improve the process
- ▶ But we will know soon!
  - ▶ Pre-phase Report waiting for final veracity checks
  - ▶ Still chasing a few legal agreements.
  - ▶ Working out technical specs and timing for data delivery
  - ▶ Shooting for final delivery by June