TUTORIAL:

Virtual IDs and Private Sketches for Data-Protected ID Resolution
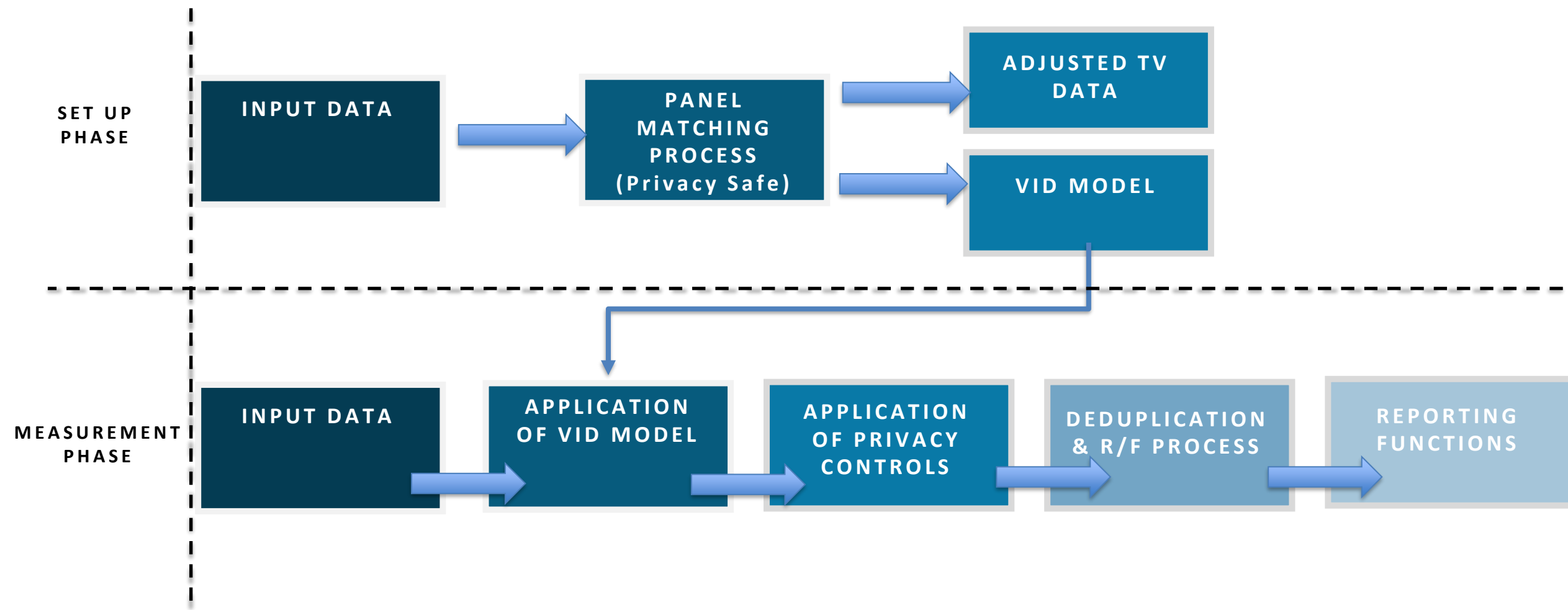
Ron Pinelli, Media Rating Council
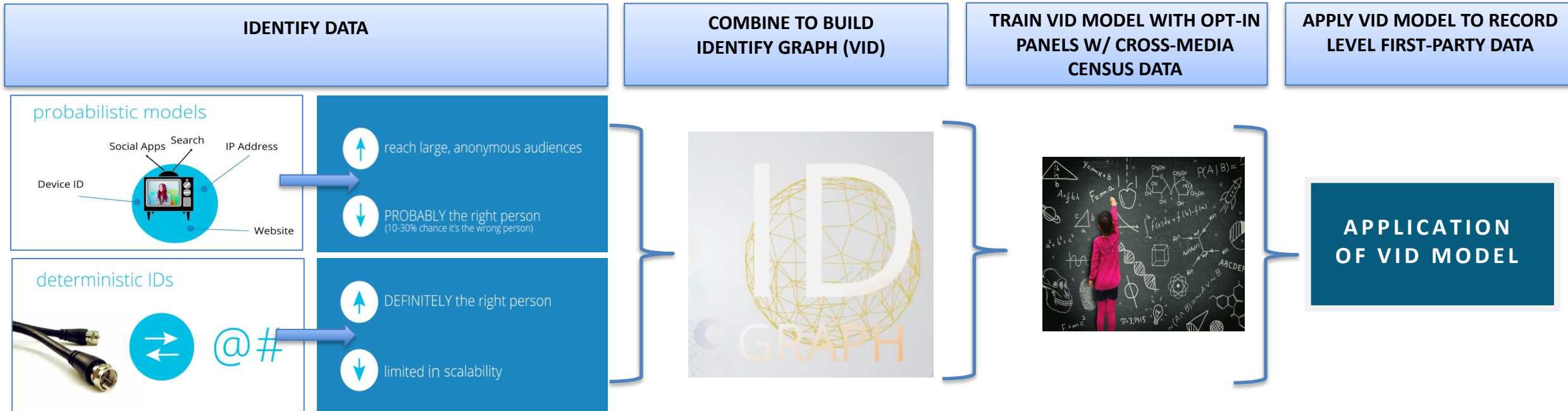
cimm
coalition for innovative
media measurement

2021 SUMMIT

High Level View of ANA Blueprint

SET UP PHASE

INPUT DATA → PANEL MATCHING PROCESS (Privacy Safe) → ADJUSTED TV DATA

PANEL MATCHING PROCESS (Privacy Safe) → VID MODEL

MEASUREMENT PHASE

INPUT DATA → APPLICATION OF VID MODEL → APPLICATION OF PRIVACY CONTROLS → DEDUPLICATION & R/F PROCESS → REPORTING FUNCTIONS
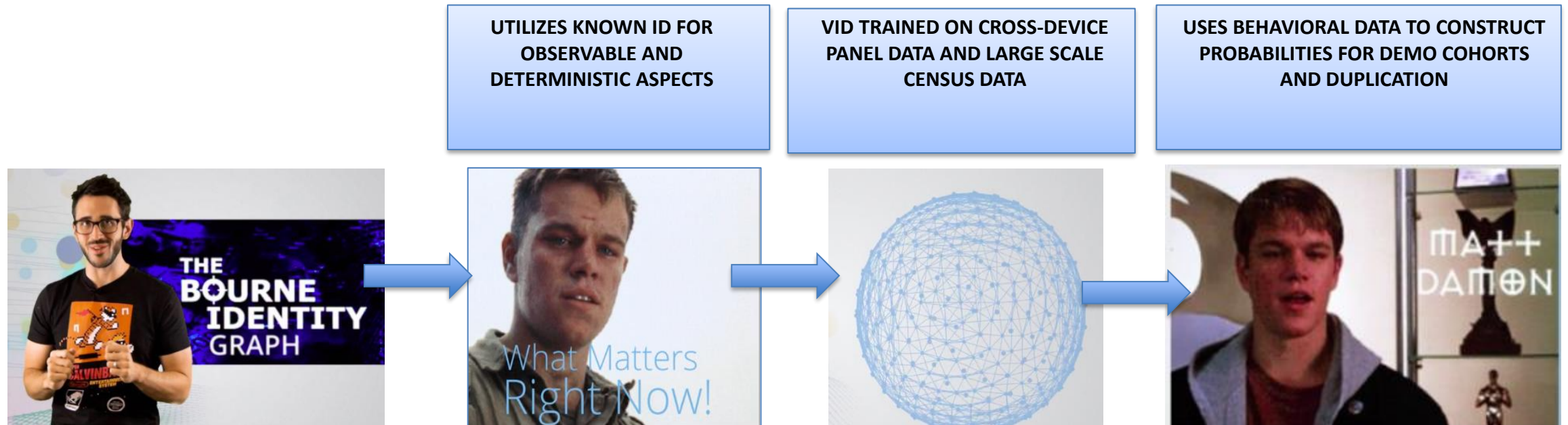
How VID Works

- Virtual ID or VID is an approach to assign record level data instances to demographic or behavioral groups
- At a high level VID is an identity graph used for assigning records to demo or behavioral groups

IDENTIFY DATA	COMBINE TO BUILD IDENTIFY GRAPH (VID)	TRAIN VID MODEL WITH OPT-IN PANELS W/ CROSS-MEDIA CENSUS DATA	APPLY VID MODEL TO RECORD LEVEL FIRST-PARTY DATA

probabilistic models

Social Apps Search IP Address

Device ID

Website

↑ reach large, anonymous audiences

↓ PROBABLY the right person
(10-30% chance it's the wrong person)

deterministic IDs

@#

↑ DEFINITELY the right person

↓ limited in scalability

ID GRAPH

APPLICATION OF VID MODEL

High-level: VID is an Identity Graph

- VID starts by determining if a record belongs to a known ID (deterministic); if so characteristics are maintained

- "Known" records could include matched users, previously observed cookies, IDs or logged in users

- If the record was not previously observed, it is assigned characteristics (probabilistic) based on learning data

UTILIZES KNOWN ID FOR OBSERVABLE AND DETERMINISTIC ASPECTS	VID TRAINED ON CROSS-DEVICE PANEL DATA AND LARGE SCALE CENSUS DATA	USES BEHAVIORAL DATA TO CONSTRUCT PROBABILITIES FOR DEMO COHORTS AND DUPLICATION

VID Requires a Training Dataset = Market Level Panel

- VID must be trained on a learning data set where census data across media is matched to opted-in panel data:

    – Panel *is not* intended to be the measurement of record

    – Panel *is* used to match to census data to adjust TV data, estimate overlap and inform the VID identity graph

    – Panel(s) *must-be cross-media*, but aspects can be from separate providers if able to be connected accurately

    – Panels *must be probability based* with good *cross-media coverage*

    – Panel *sample size can be smaller for training* than what's required as the source of measurement records, but still *must be of high quality* (and may serve as the measurement of record for instances of low quality or missing input data such as OTA)

    – ANA design *requires census data to be double-blind matched to panel data* to inform the VID process

    – *Models must be periodically updated* and ANA is working on setting requirements based on testing

VID is NOT a Unique ID

- VID **only** determines what demo or behavioral group a record belongs to and labels it consistently for R&F processing
- VID does not create persistent IDs that can be matched across datasets
- VID does not enable re-identification for targeting or outcomes measurement
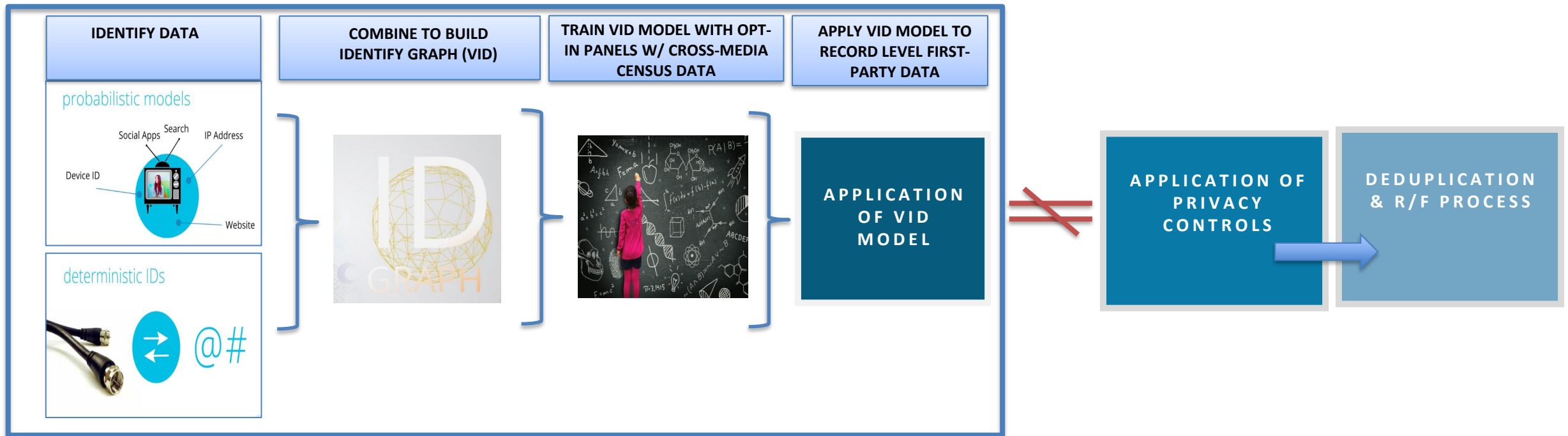
VID is NOT a Privacy Approach

After VID model is trained and applied at the first-party record level, the data is not yet privacy safe
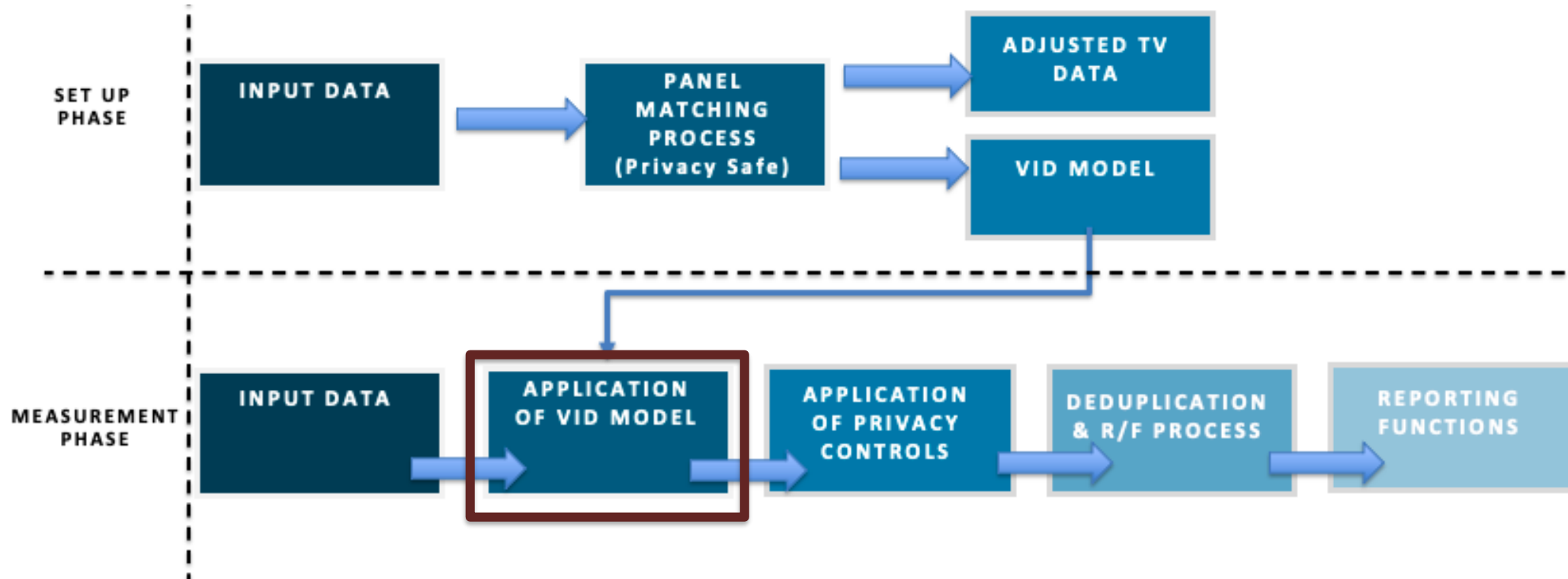
Privacy Algorithms and Cardinality Estimator

Design them separately but they must work together

- Before VID applied records can be shared externally for processing, privacy preserving algorithms must be applied

- Privacy algorithms add noise and aggregate the data in a manner that allows them to be processed together for R&F in the aggregate on a privacy compliant basis, but does not allow for record-level identification or analysis (privacy sketches)

- Privacy approaches have been suggested at the WFA level and must be subject to peer review as well as tested at the market level

The Difference between VID vs SUMID

VID –
Virtual ID

– VID is more accurate when more deterministic data is present

– Deprecation of cookies, IDFA and other identifiers will weaken VID as it will be more reliant on probabilistic assignment

– VID also does not allow for targeting or outcomes measurement

SUMID –
Secure Universal Measurement IDs

– Enable more persistent ID and matching

– SUMID works with VID and does not replace it

– SUMID proposals are present and must be tested, but also considered when designing and testing VID

VID Must Be Tested

- VID approaches have been proposed, but it is important that they are thoroughly vetted
  - VID must be tested with digital and linear data to ensure persons level and HH or device level data can be accounted for and work together
  - There have been theoretical reviews of the VID approach that have highlighted concerns or questions (some complimentary and some contrarian)
  - These concerns and questions are contemplated in the design and are key aspects of planned testing by WFA, ANA and ISBA
  - It will be important to use real data to evaluate VID at the local market level and adjust it accordingly or explore alternative approaches
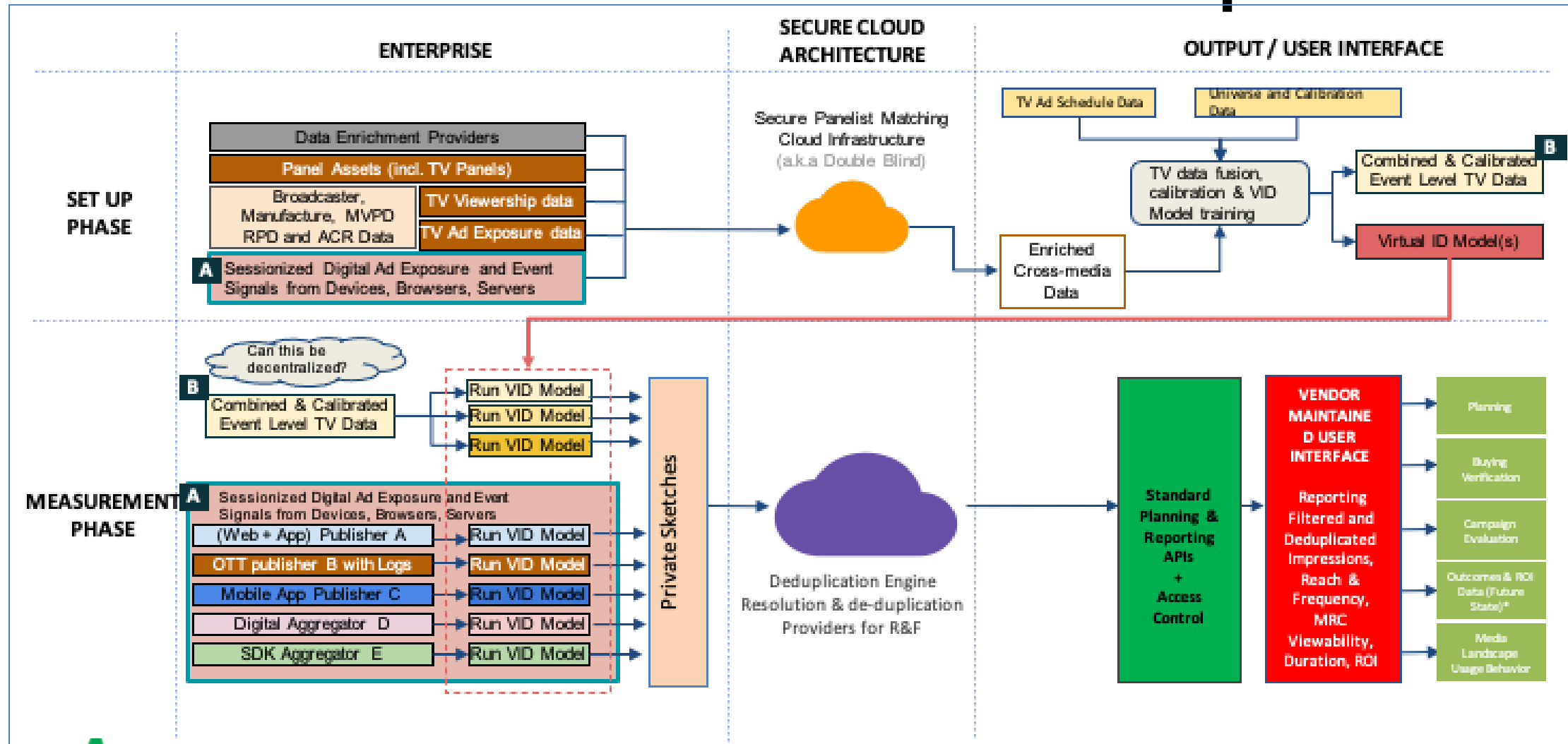
Marketer-Led Cross-Media Measurement

Mission: To create a marketer-led system of unified cross-media campaign measurement restoring essential measures of deduplicated reach & frequency. This will address four key areas …

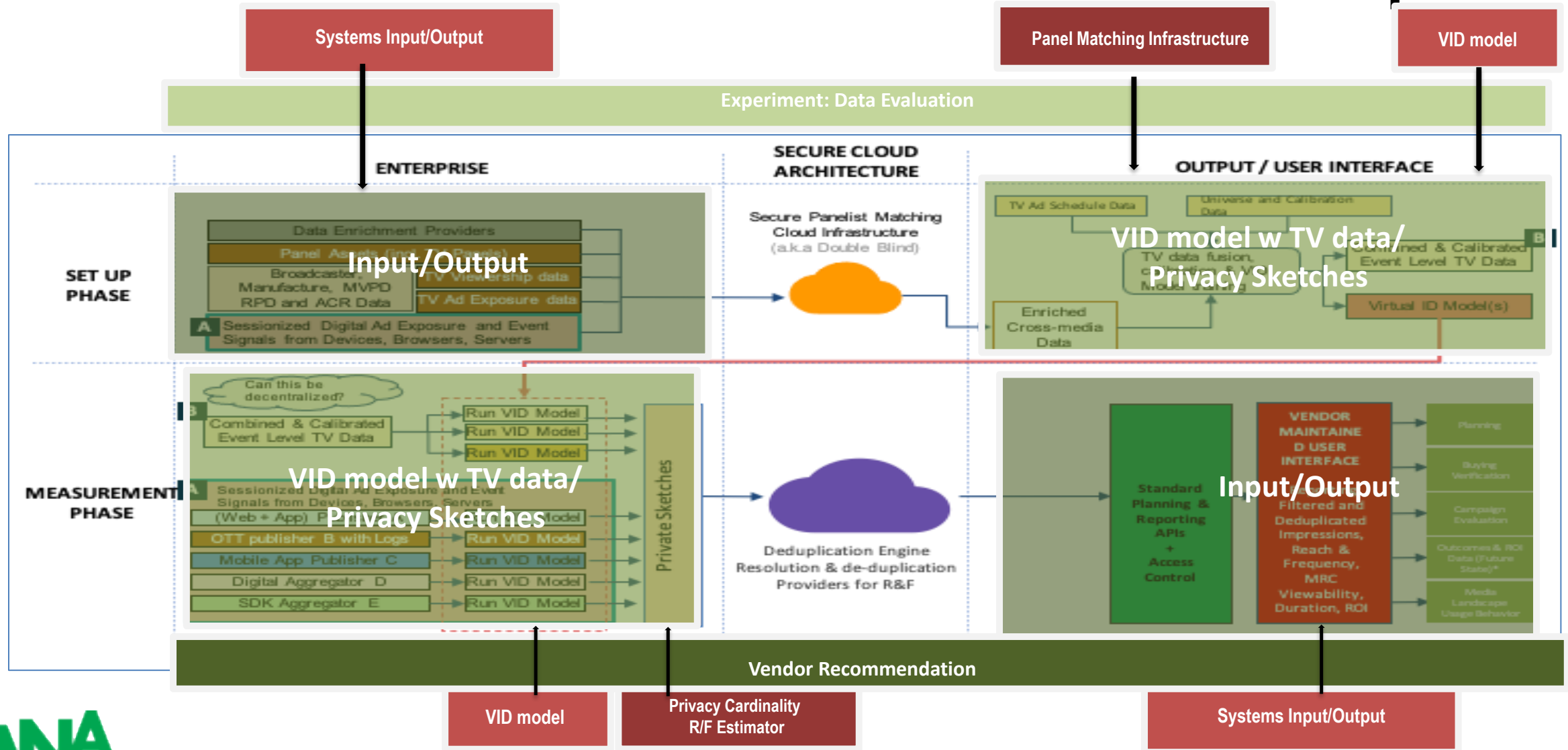|	
CONSUMER BEHAVIOR and MEDIA TRANSPARENCY	Restoring consumer centric, objective, and complete transparent view of ad exposure across all media.
PLANNING	Enable planning and optimization for de-duplicated reach and frequency across all media channels to improve efficiency and consumer experiences.
CAMPAIGN EVALUATION	Ability to evaluate the complete and cumulative effects of a campaign delivery across all media channels and devices, including value of specific media channels.
OUTCOME and ROI	Improve precision of decision making leading to greater ROI and value across all effectiveness measures.

ANA Cross-Media Measurement Blueprint

13

ANA SBPs and Common Connection to Blueprint

© 2021 Confidential ANA (2-4-21)