synthetic data: a very real path to cross-media measurement

iab. australia

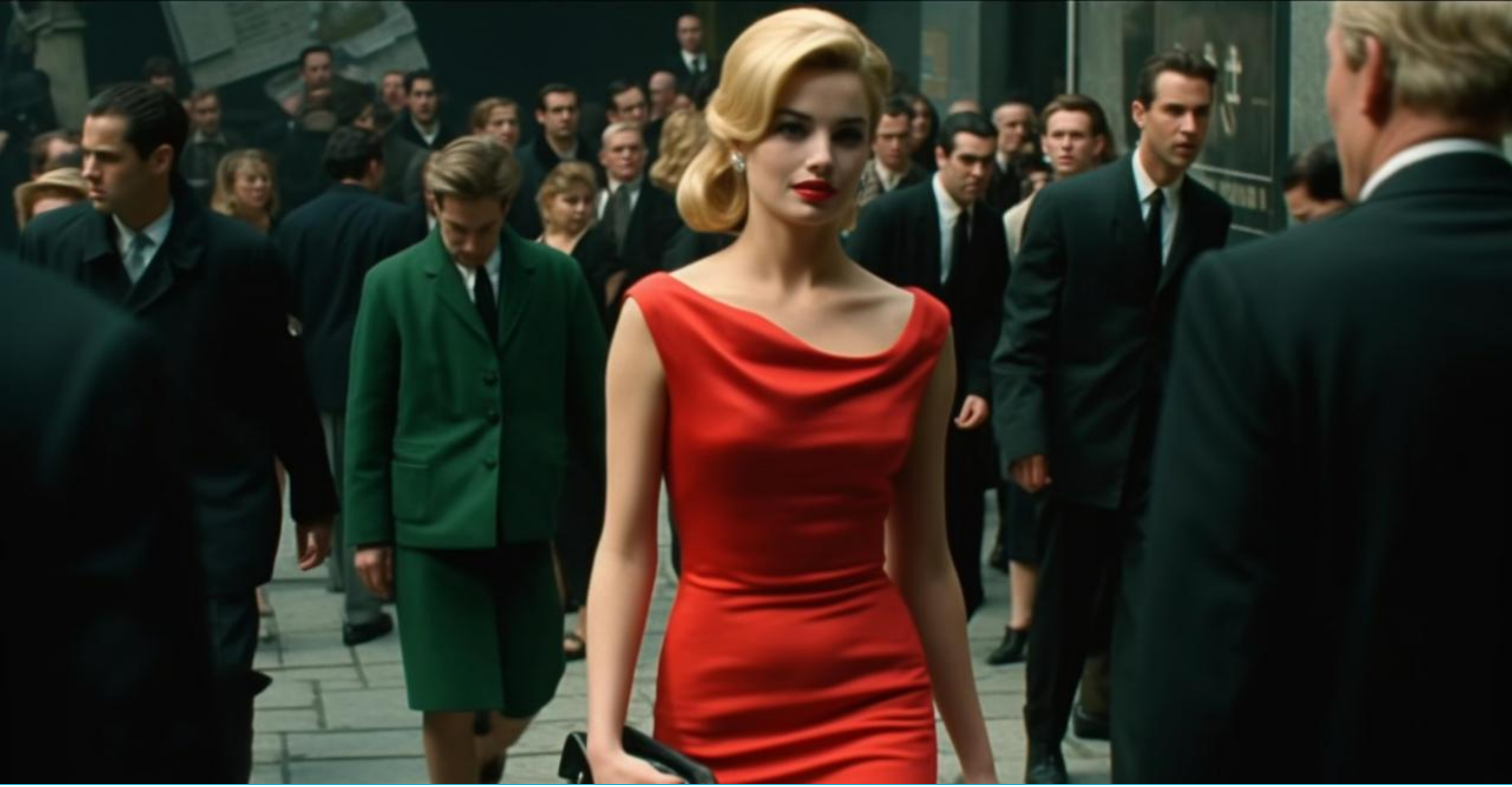
ian garland

managing director | milton data

[Milton Data]

Synthetic Data: a very real path to cross-media measurement

11 September 2024



Synthetic Data Basics

- Synthetic Data most common currently in medical fields
  - Privacy focus
  - Alternative scenarios
  - Capture distribution/structure of real data

Synthetic Data Basics

- Synthetic Data most common currently in medical fields
  - Privacy focus
  - Alternative scenarios
  - Capture distribution/structure of real data

- Military & Security applications
  - Image classification
  - Cyber-security
  - Autonomous system development

Synthetic Data Basics

- Synthetic data most common currently in medical fields
  - Privacy focus
  - Alternative scenarios
  - Capture distribution/structure of real data

- Military & Security applications
  - Image classification
  - Cyber-security
  - Autonomous system development

- Training AI Systems
  - Alpha Go most famous example
  - Rendering engines as inputs for autonomous vehicles
  - Amazon training Alexa language system on synthetic data
  - American Express & synthetic financial data for fraud detection

Applications in Media and Marketing

- Early days but showing potential

- Projects around the world
  - Iris (Australia/UK)
  - VOZ (Australia)
  - Dovetail (UK)
  - WFA Halo/“North Star” (USA/UK)

- Core elements in each
  - Audience measurement applications
  - Privacy preserving approaches
  - Integration of different data sources
  - Emphasis on cross-media/cross-platform measures

VOZ
TOTAL TELEVISION MEASUREMENT

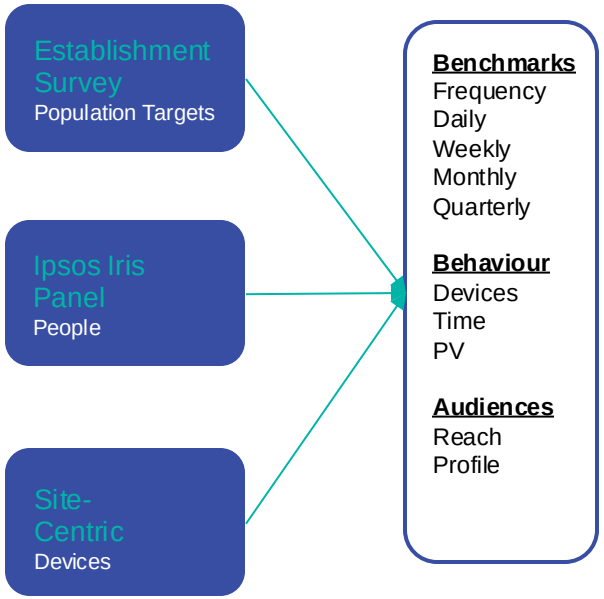
Making Synthetic Data – General Version

- "Classic" approaches
  - Generative Adversarial Networks (GAN)
  - Variational Auto-encoders (VAE)

- Problems for audience measurement applications
  - LLMs revealing training data
  - No control on representativeness

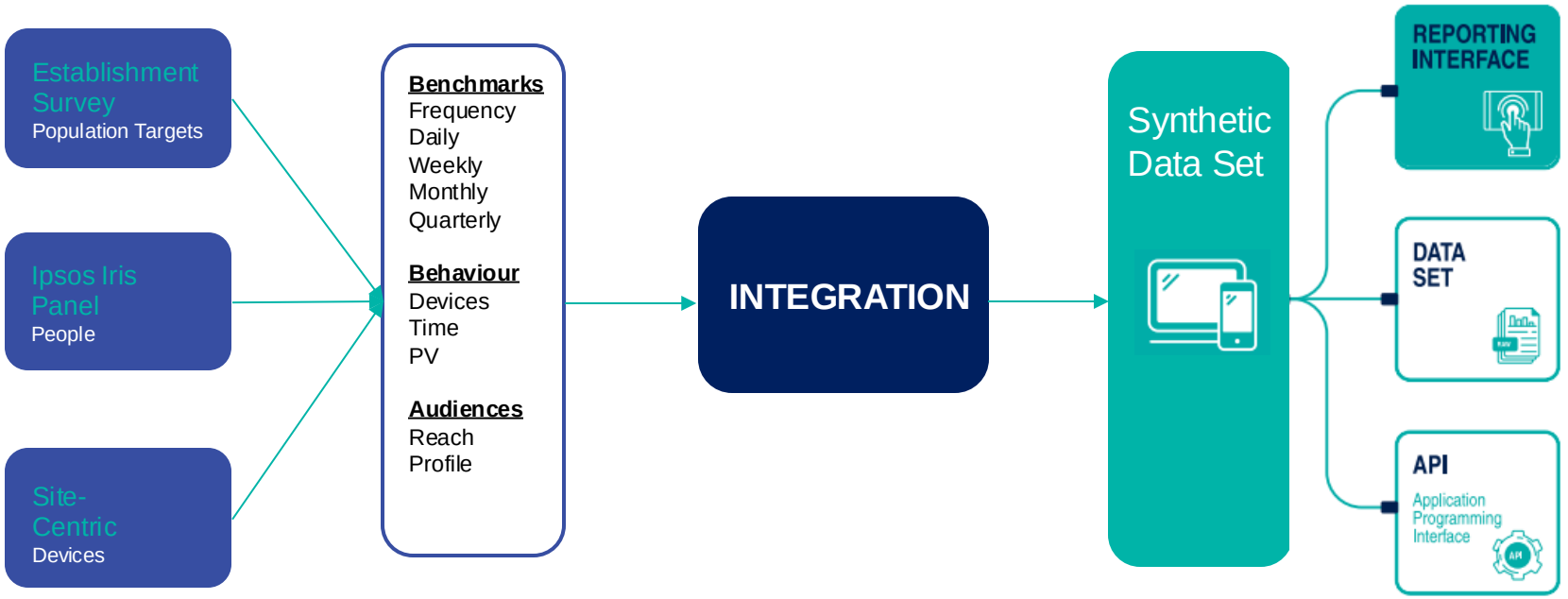
Making Synthetic Data – Media/Marketing Version

- Synthetic data in media more about integration
  - Survey data (establishment survey)
  - Panel data (consumer behaviours)
  - Census data (site tagging data)



Making Synthetic Data – Media/Marketing Version

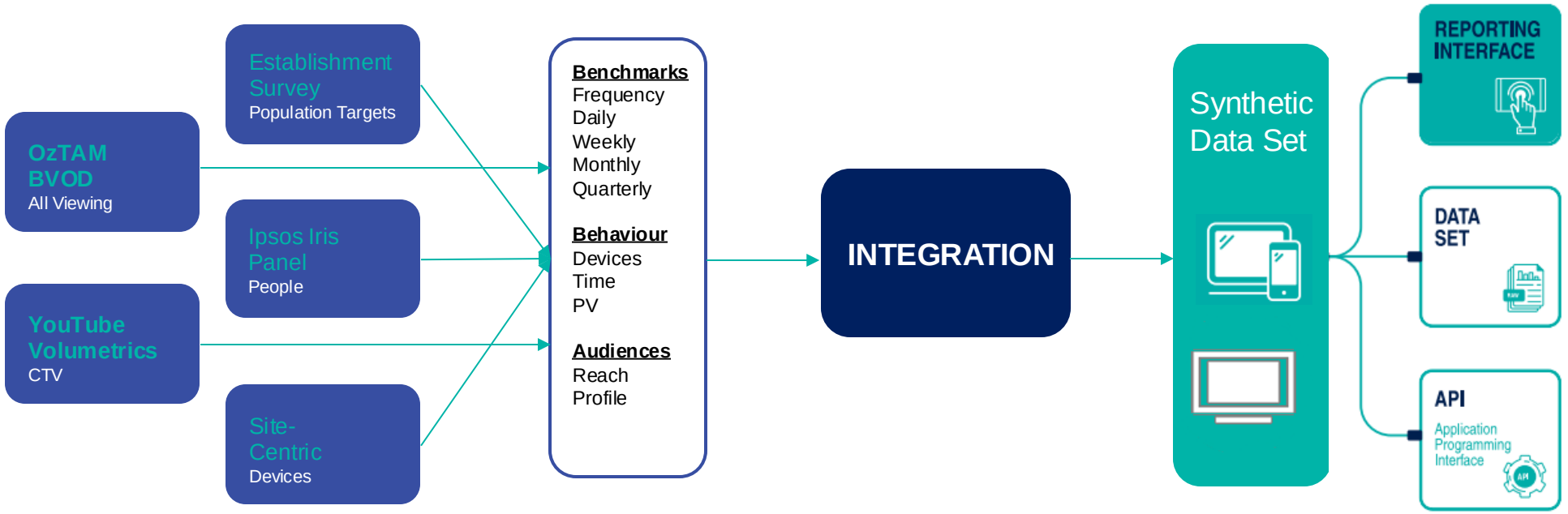
- Synthetic data in media more about integration
  - Survey data (establishment survey)
  - Panel data (consumer behaviours)
  - Census data (site tagging data)



Making Synthetic Data – Media/Marketing Version

- Synthetic data in media more about integration
  - Survey data (establishment survey)
  - Panel data (consumer behaviours)
  - Census data (site tagging data)

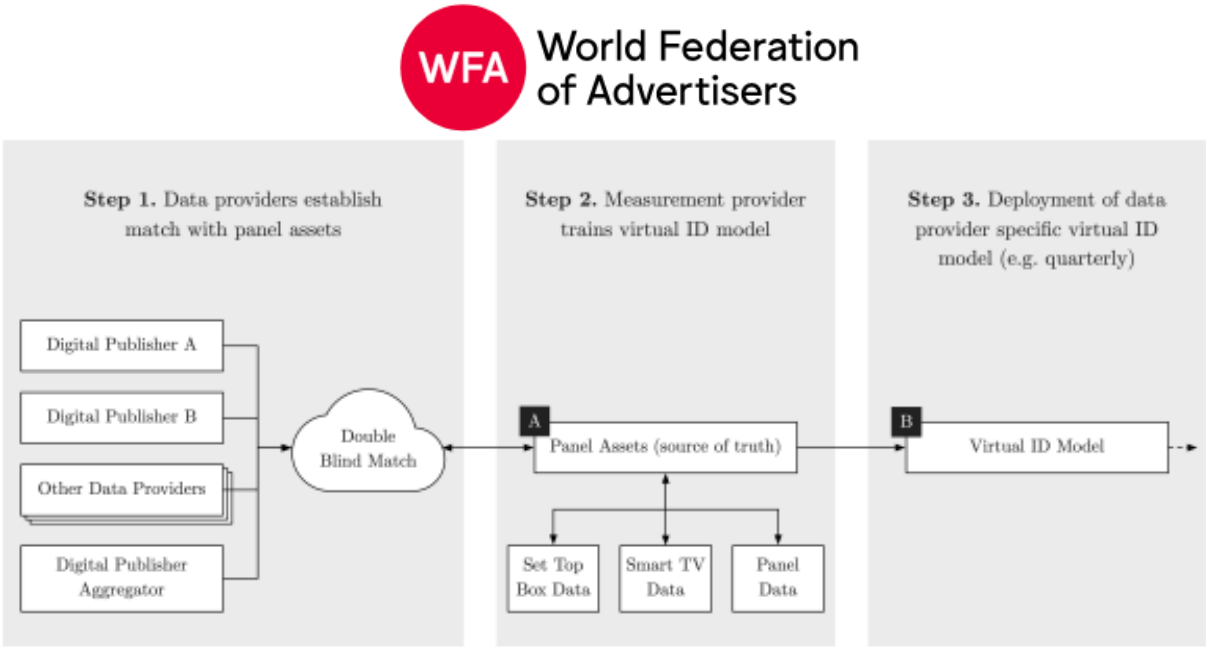
Can be extended to include additional data sets



Making Synthetic Data – Media/Marketing Version

- Synthetic data in media more about integration
  - Survey data (establishment survey)
  - Panel data (consumer behaviours)
  - Census data (site tagging data)

- **Alternate approach**
  - Project Origin/WFA Halo (North Star)
  - OzTAM VOZ

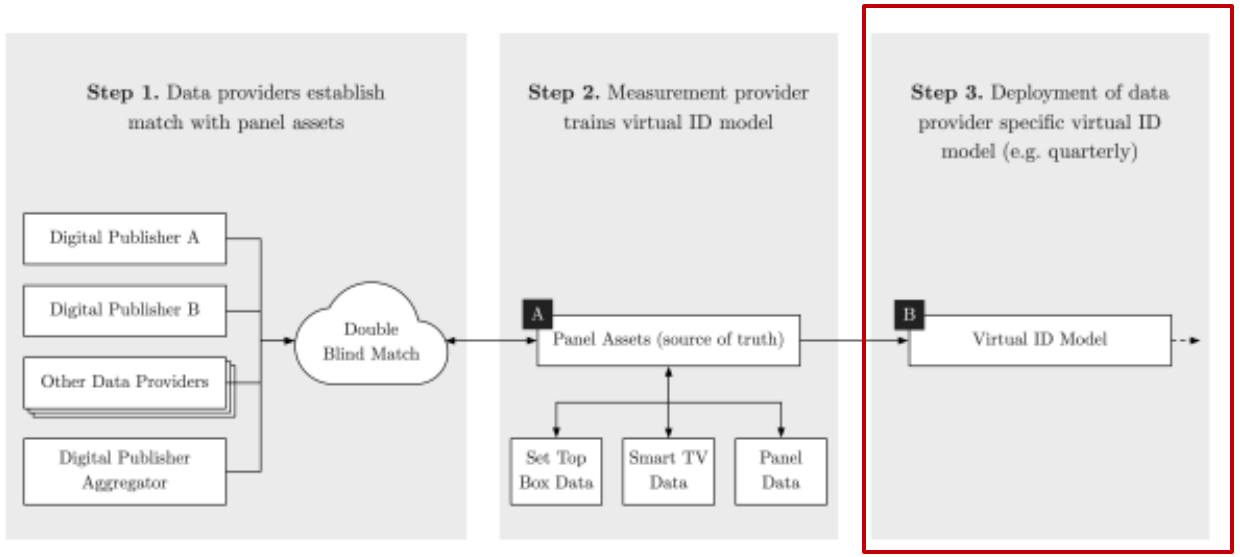


Making Synthetic Data – Media/Marketing Version

- Synthetic data in media more about integration
  - Survey data (establishment survey)
  - Panel data (consumer behaviours)
  - Census data (site tagging data)

- Alternate approach
  - Project Origin/WFA Halo (North Star)
  - OzTAM VOZ

- Virtual Id Model
  - Train on census/panel data
  - Create Identity Hub (different to id graph)



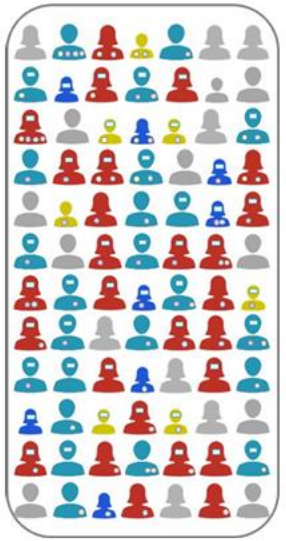
Making Synthetic Data – Media/Marketing Version

- Synthetic data in media more about integration
  - Survey data (establishment survey)
  - Panel data (consumer behaviours)
  - Census data (site tagging data)

- Alternate approach
  - Project Origin/WFA Halo (North Star)
  - OzTAM VOZ

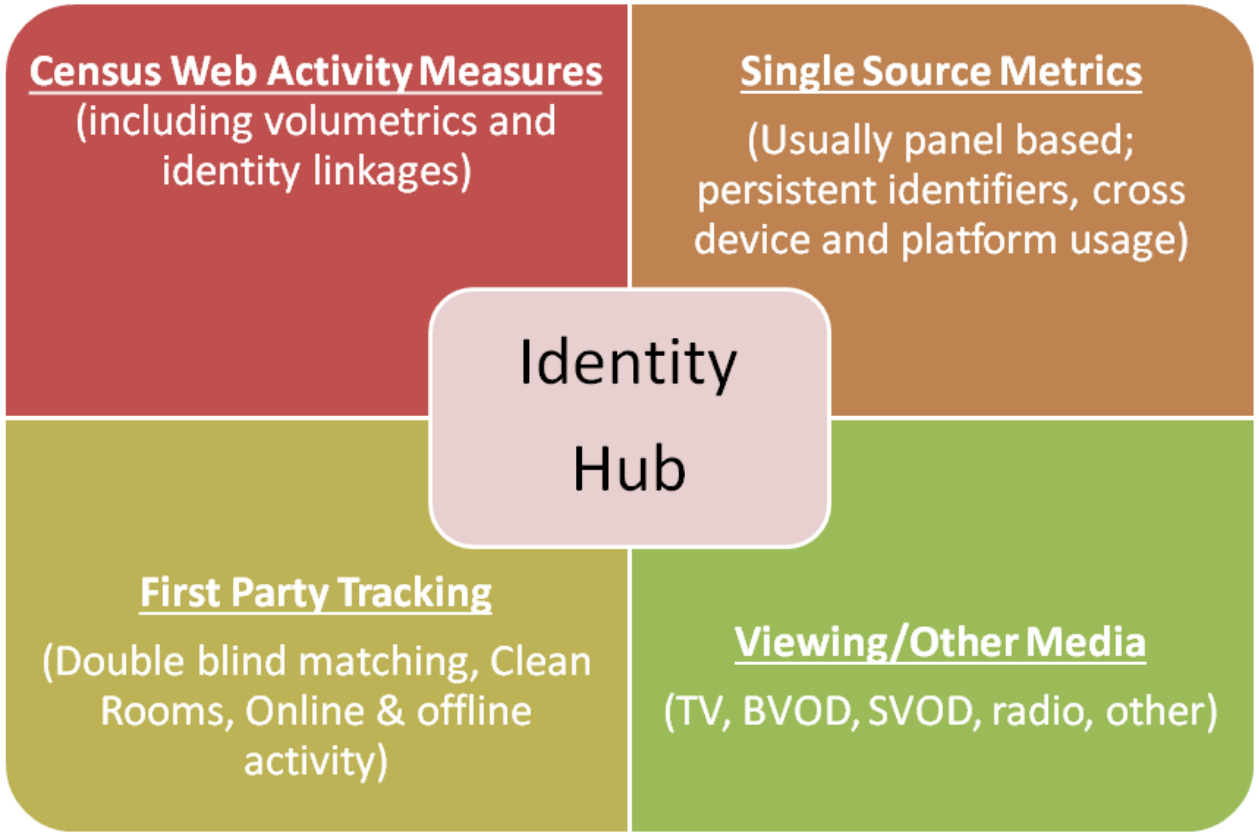
- Virtual Id Model
  - Train on census/panel data
  - Create Identity Hub (different to id graph)

- Each Virtual Id is one person in population
  - Assignment via demographic models (calibration)
  - Assignment via data fusion



Using Synthetic Data – Data Integration

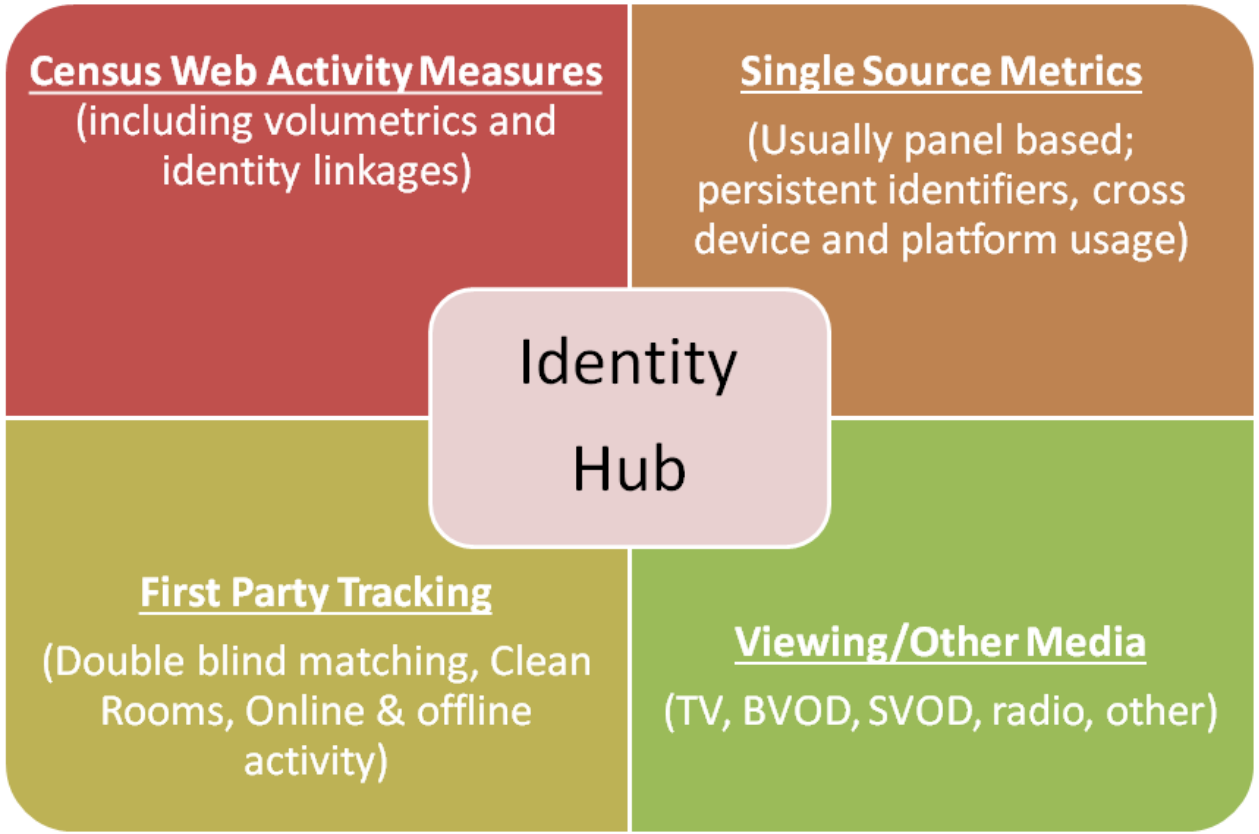
- Identity Hub alternatives
  - Completely anonymous (audience applications)
  - 1:1 mapping (activation applications)



Using Synthetic Data – Data Integration

- Identity Hub alternatives
  - Completely anonymous (audience applications)
  - 1:1 mapping (activation applications)

- Critical issues
  - Preservation of source metrics (currency)
  - Filling in data gaps (missing cookies)
  - Obtaining cross-platform usage estimates
  - Conforming data sources
  - Quality of inputs

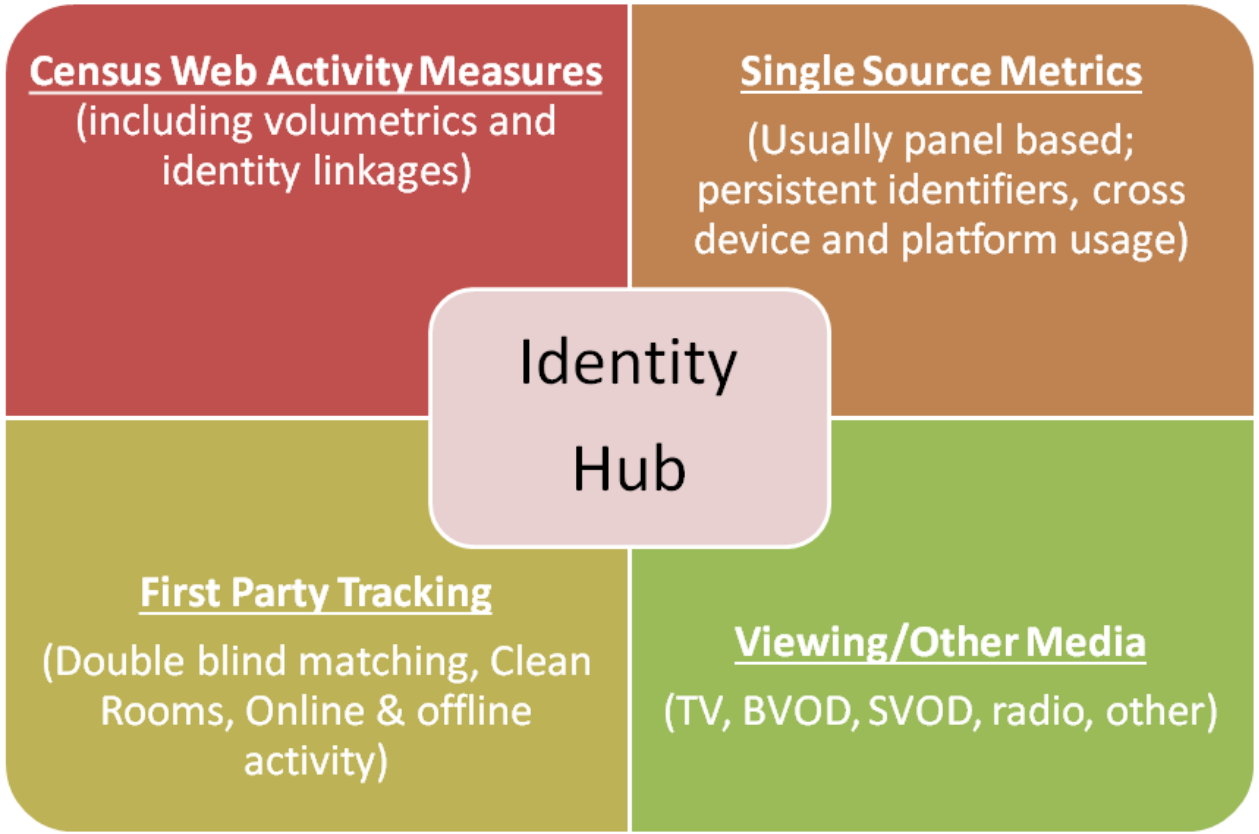


Using Synthetic Data – Data Integration

- Identity Hub alternatives
  - Completely anonymous (audience applications)
  - 1:1 mapping (activation applications)

- Critical issues
  - Preservation of source metrics (currency)
  - Filling in data gaps (missing cookies)
  - Obtaining cross-platform usage estimates
  - Conforming data sources
  - Quality of inputs

- Importance of First Party Data
  - Owning your viewer/user/customer
  - Improvements in data matching processes
  - Offline measures



Conclusions

- Australia well served by synthetic data approach
  - Ipsos Iris
  - OzTAM VOZ
- Complementary approaches
  - Privacy enabled
  - Vs
  - Links to activation
- Core takeaways
  - Quality of inputs essential
  - Importance of matching algorithm(s)
- How to improve
  - Richer data sets
  - Link to real-world

[MILTON DATA]