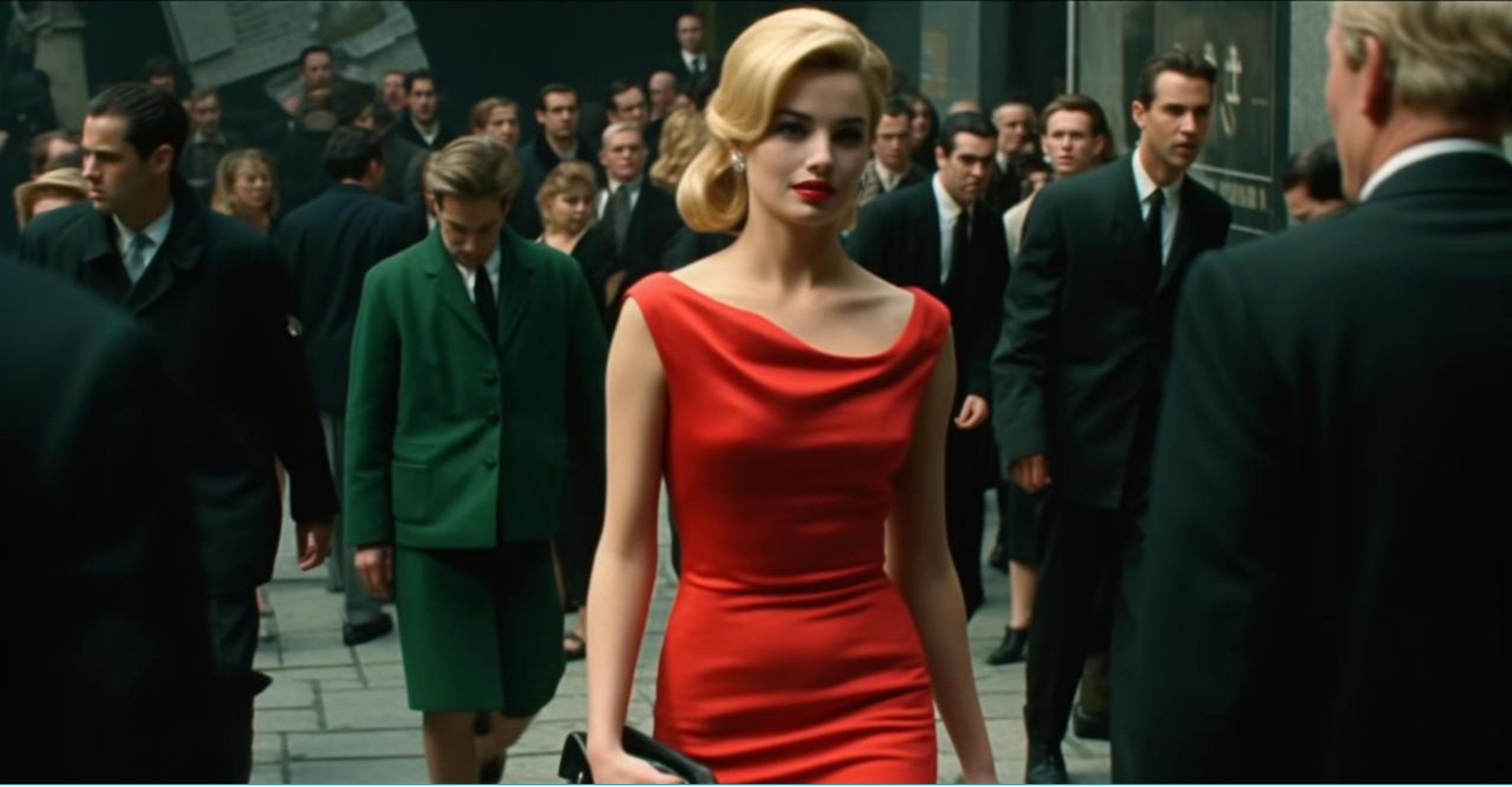
[MILTON DATA]

# *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 – 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

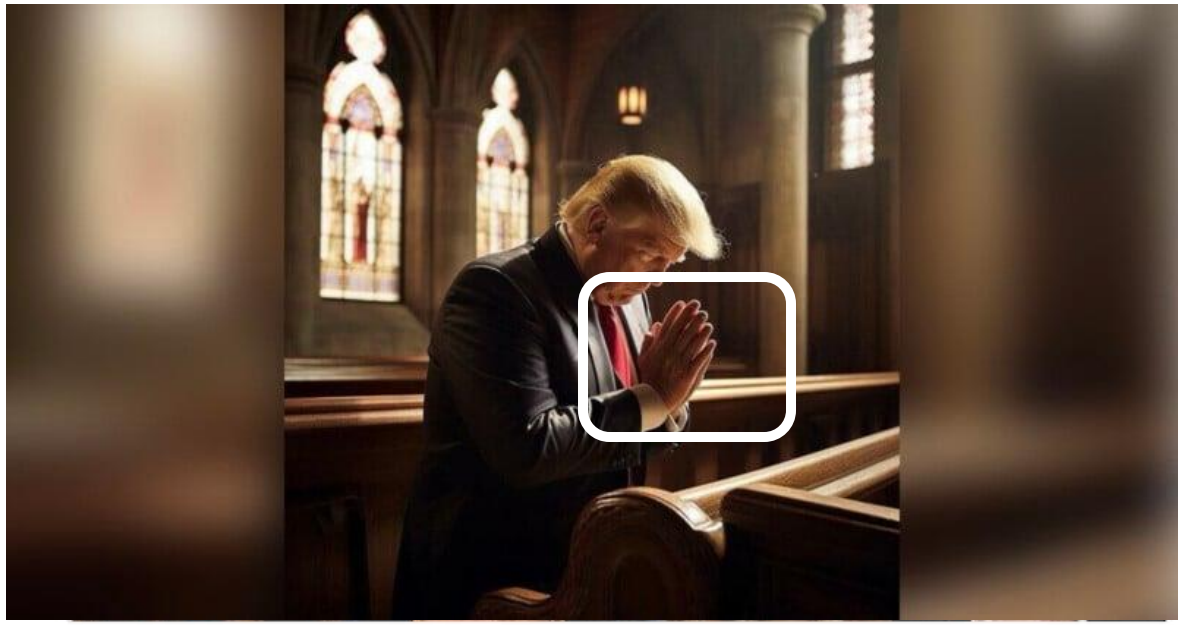
- When LLMs go wrong ….

# Making Synthetic Data – General Version

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

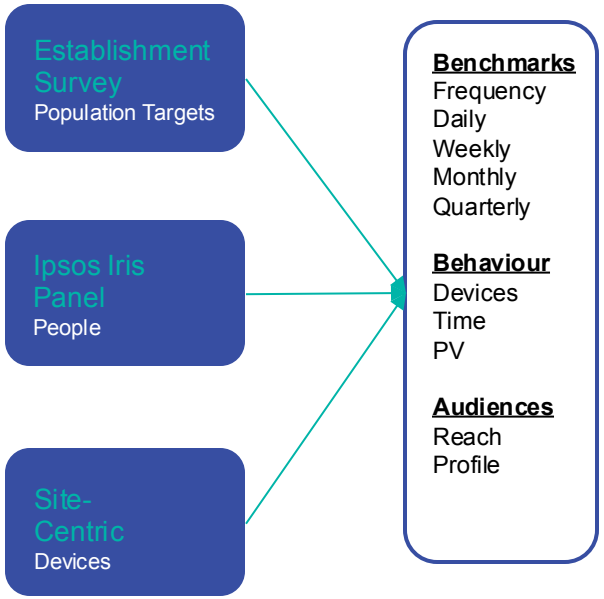
- Problems for audience measurement applications
  - LLMs revealing training data
  - No control on representativeness

- When LLMs go wrong ….
  - It's not always so obvious



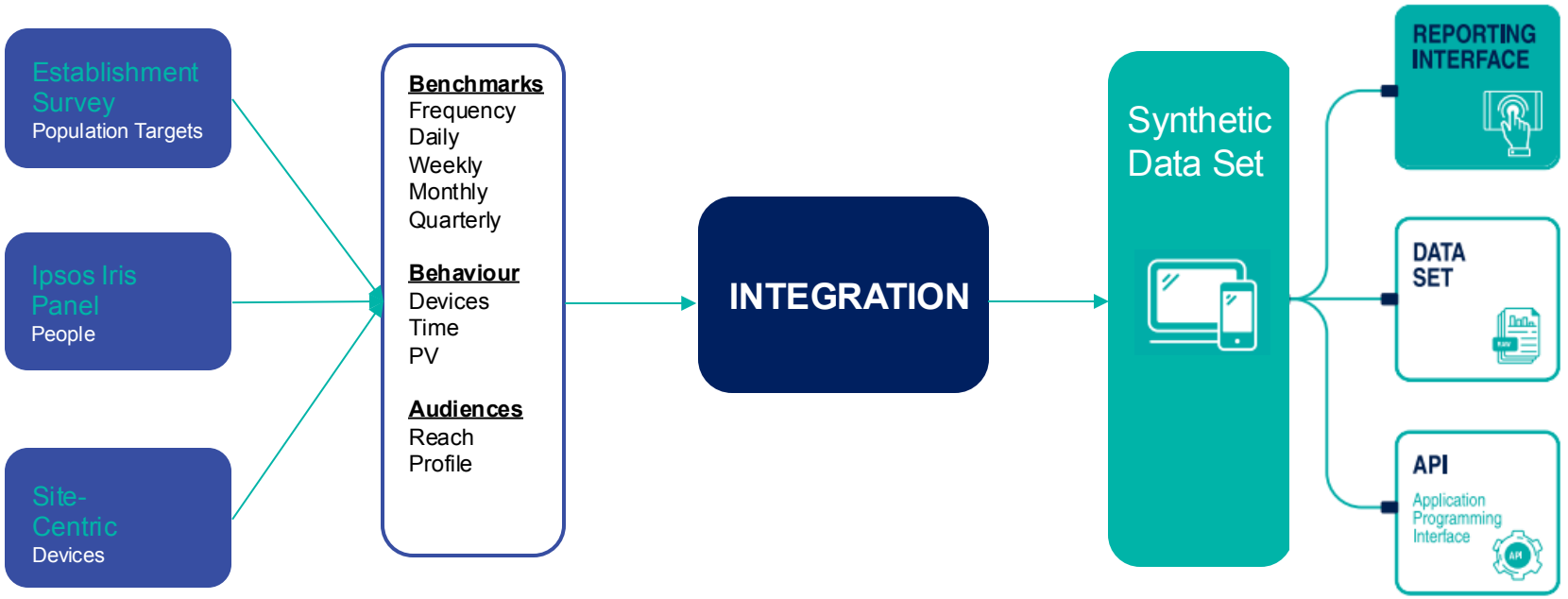
# Making Synthetic Data – Media/Marketing Version

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



# Making Synthetic Data – Media/Marketing Version

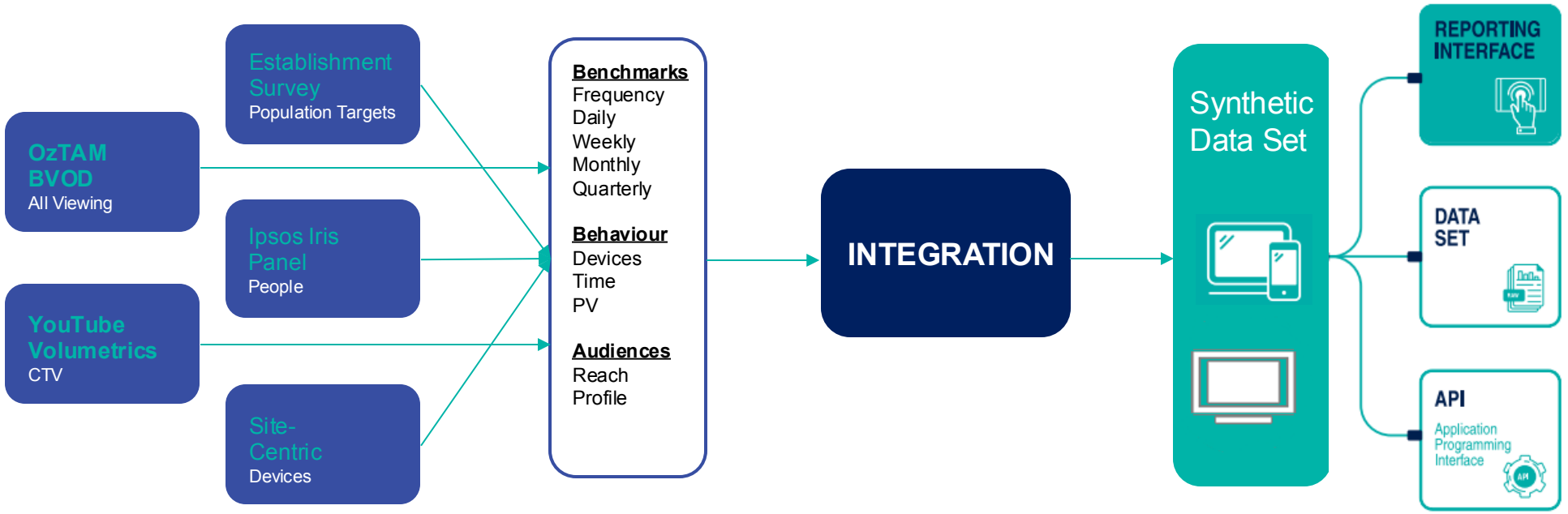
- Synthetic data in media more about controlled integration (Ipsos Iris Australia)
  - Survey data (establishment survey)
  - Panel data (consumer behaviours)
  - Census data (site tagging data)



# Making Synthetic Data – Media/Marketing Version

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

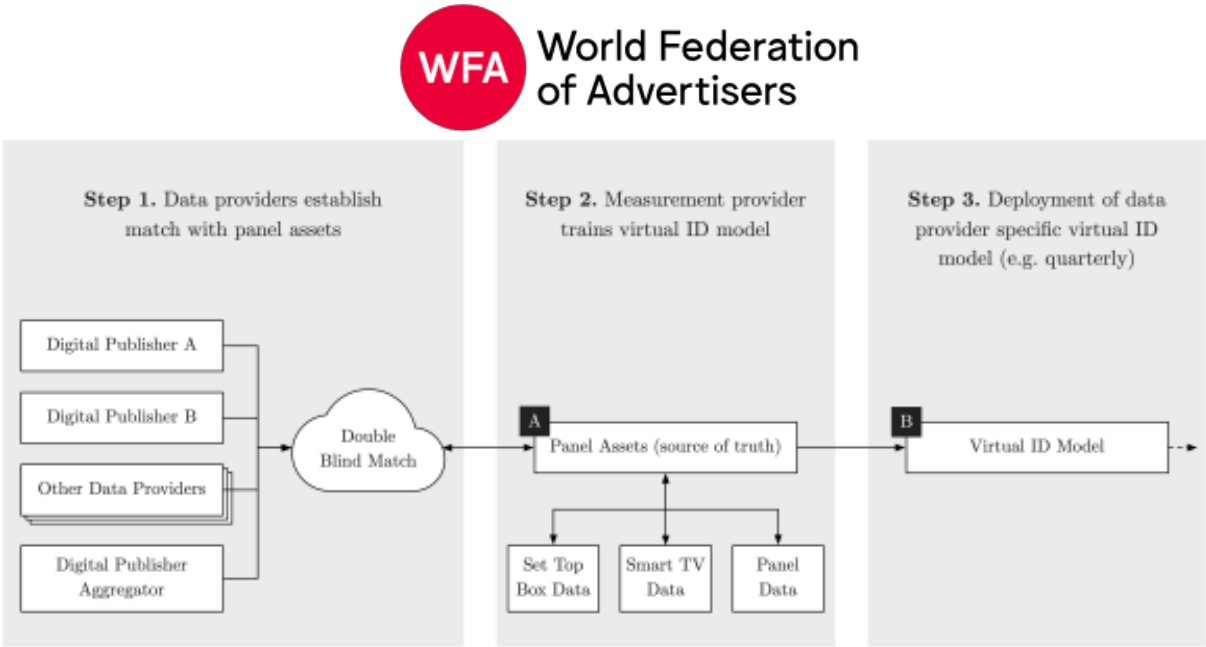
Can be extended to include additional data sets



# Making Synthetic Data – Media/Marketing Version

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

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

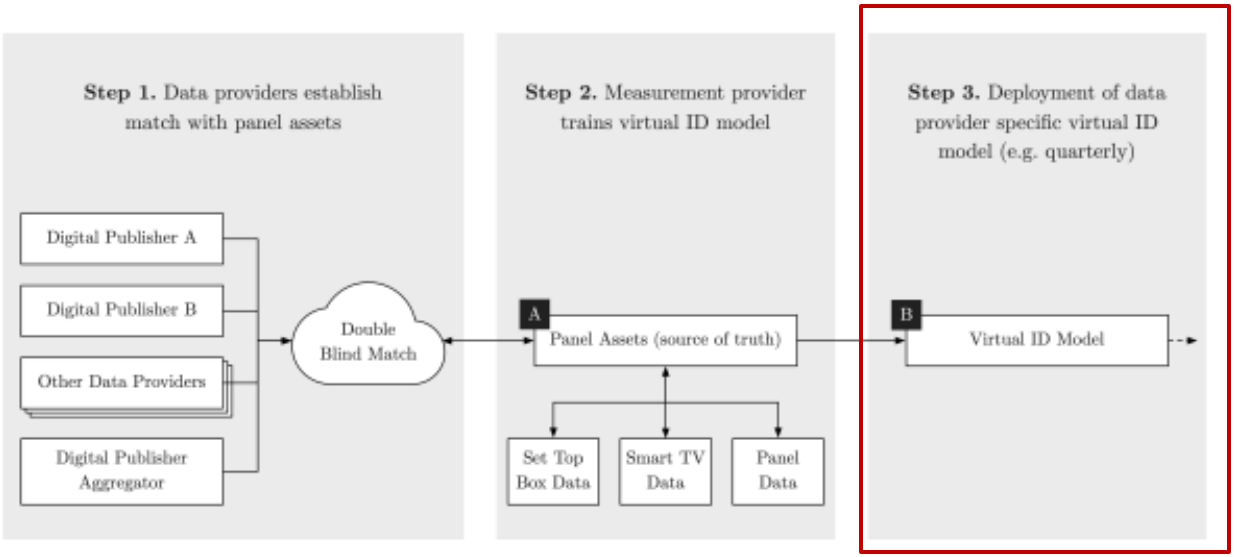


# Making Synthetic Data – Media/Marketing Version

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

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

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



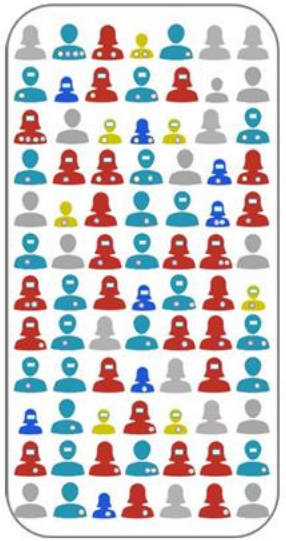
# Making Synthetic Data – Media/Marketing Version

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

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

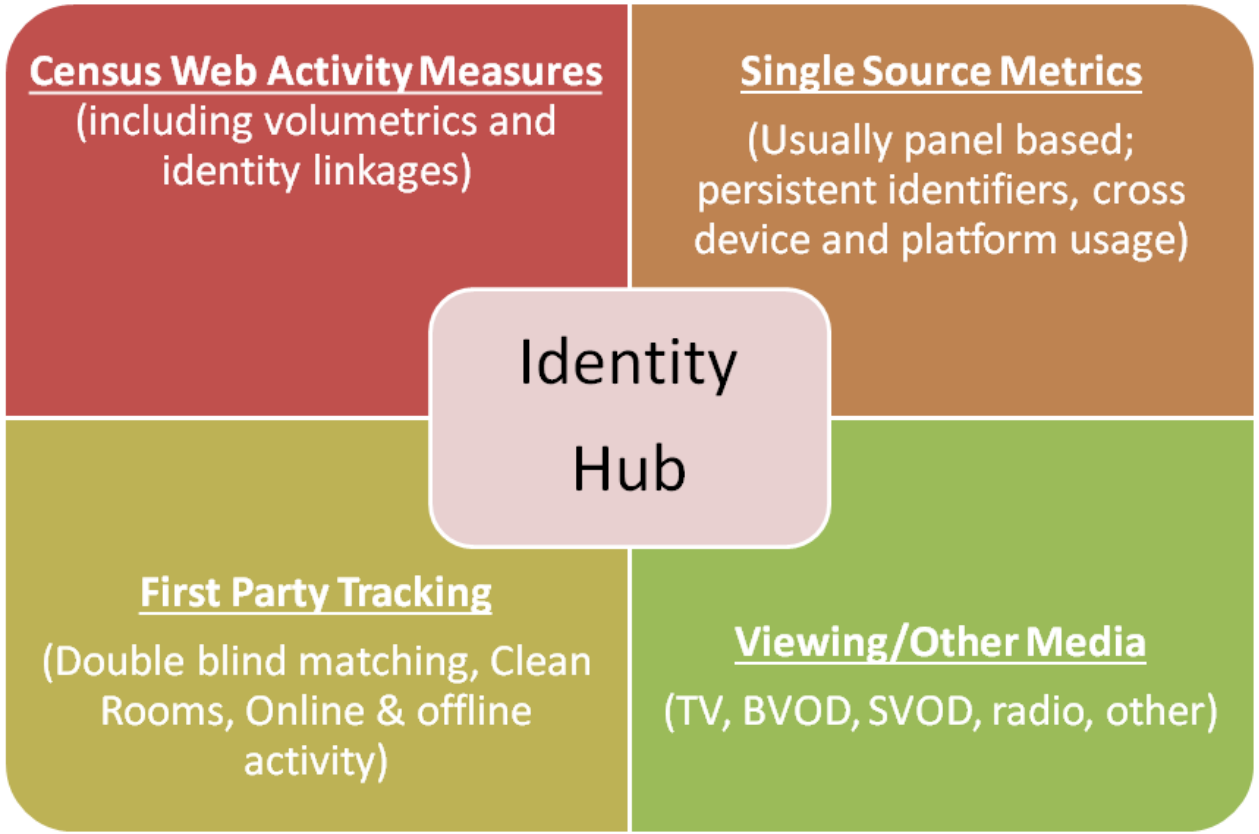
- Each Virtual Id is one person in population
  - Assignment via demographic models
  - Assignment via data fusion

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



# Using Synthetic Data – Data Integration

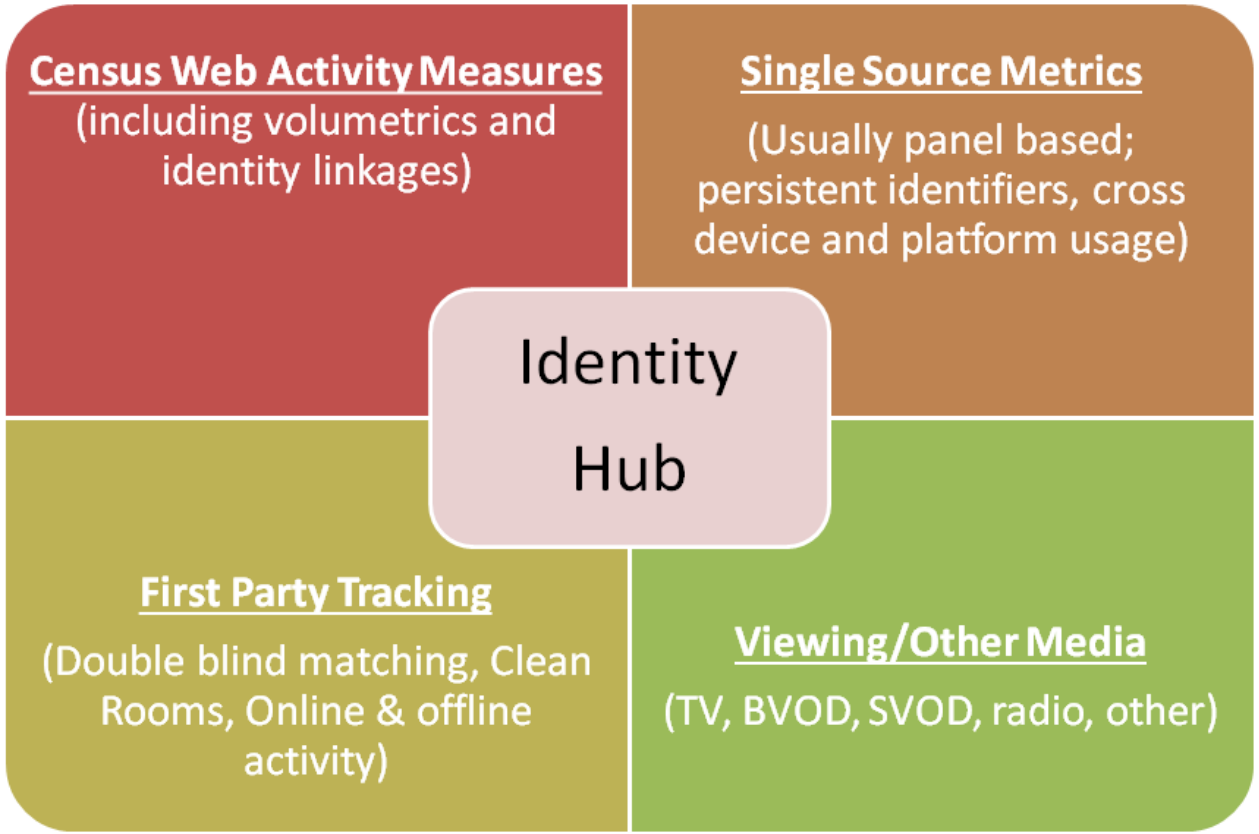
- Identity Hub approaches
  - Completely anonymous (audience applications)
  - 1:1 mapping (activation applications)



# Using Synthetic Data – Data Integration

- Identity Hub approaches
  - 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

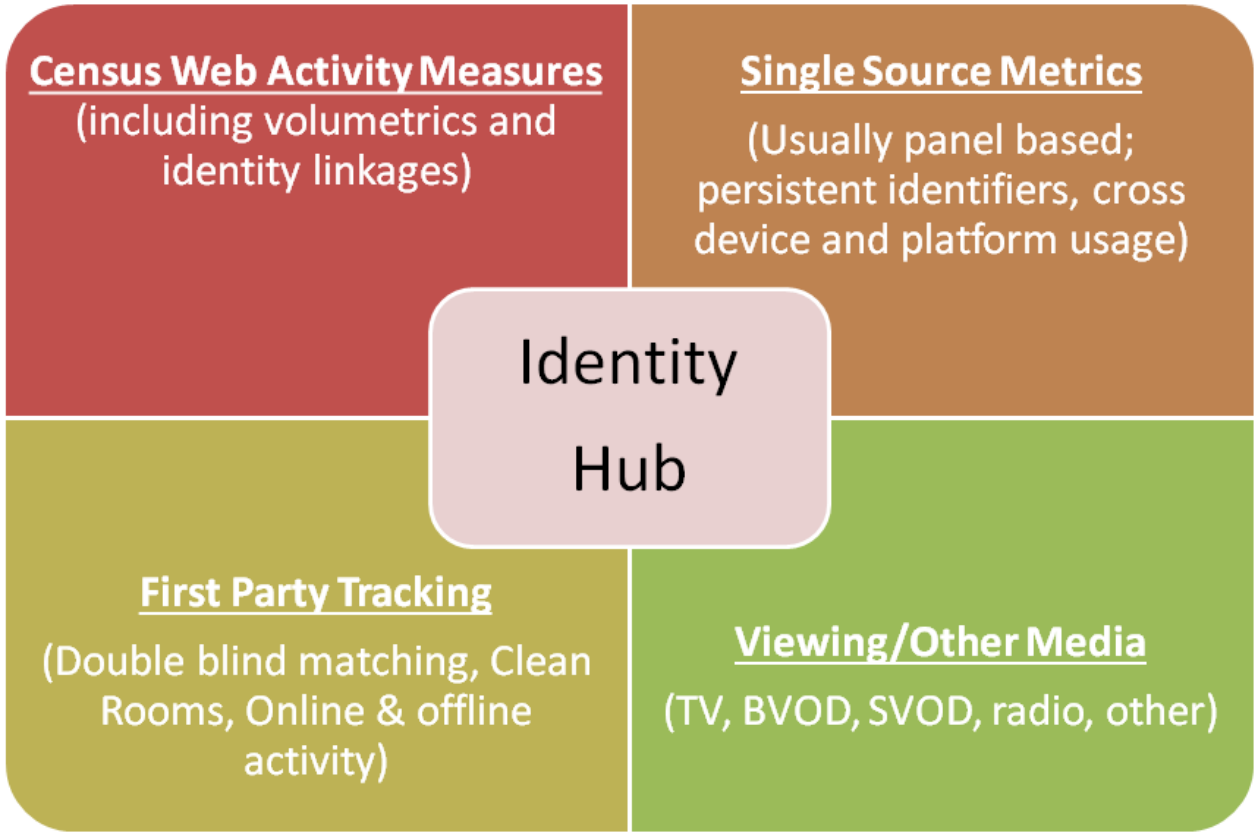


# Using Synthetic Data – Data Integration

- Identity Hub approaches
  - 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

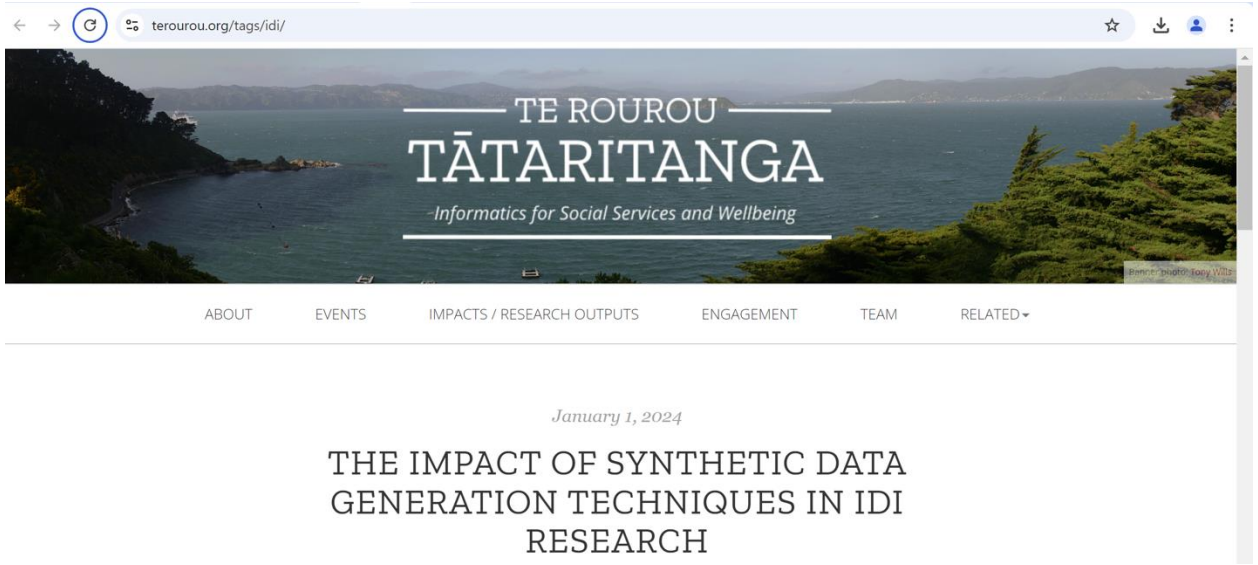


# New Zealand Perspective – Social and Policy

- Generating a synthetic version of New Zealand
  - Victoria University
  - University of Auckland
  - Stats New Zealand

- Based on Integrated Data Infrastructure
  - IDI links multiple data sources
  - Restrictive confidentiality requirements

- Synthetic data aim
  - Meets stringent privacy requirements
  - Minimise re-identification risk (differential privacy)
  - No ability to "reverse engineer" source data
  - Cognizant of Māori data sensitivities



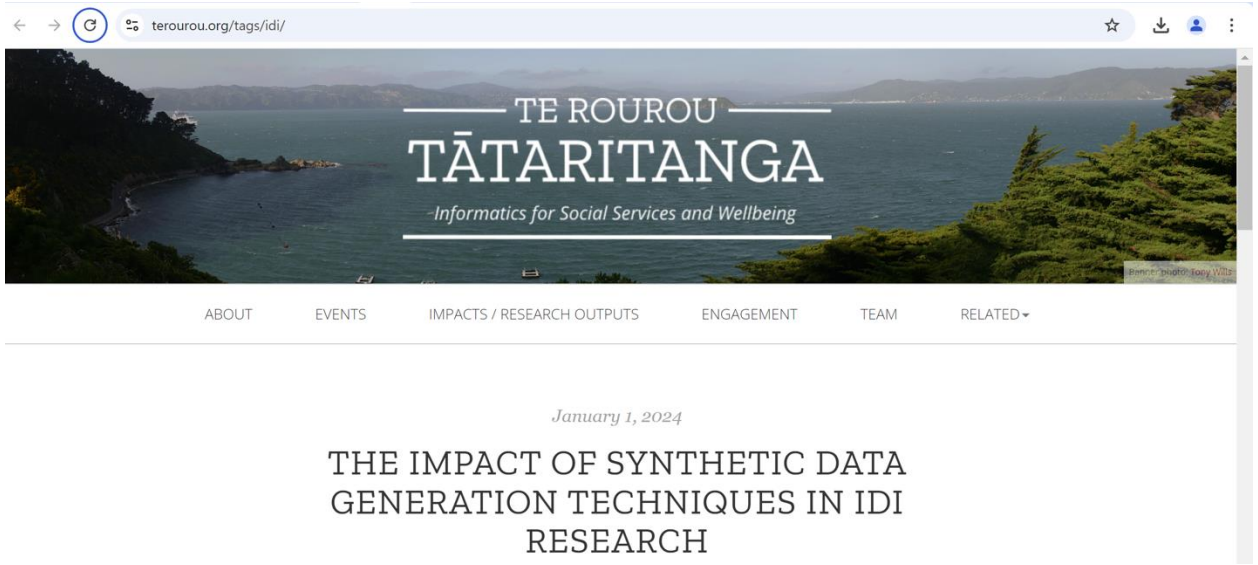
# New Zealand Perspective – Social and Policy

- Generating a synthetic version of New Zealand
  - Victoria University
  - University of Auckland
  - Stats New Zealand

- Based on Integrated Data Infrastructure
  - IDI links multiple data sources
  - Restrictive confidentiality requirements

- Synthetic data aim
  - Meets stringent privacy requirements
  - Minimise re-identification risk (differential privacy)
  - No ability to "reverse engineer" source data
  - Cognizant of Māori data sensitivities

- ***But what about marketing and media?***



# New Zealand Perspective – Marketing and Media

- Cookies were never a complete solution

# New Zealand Perspective – Marketing and Media

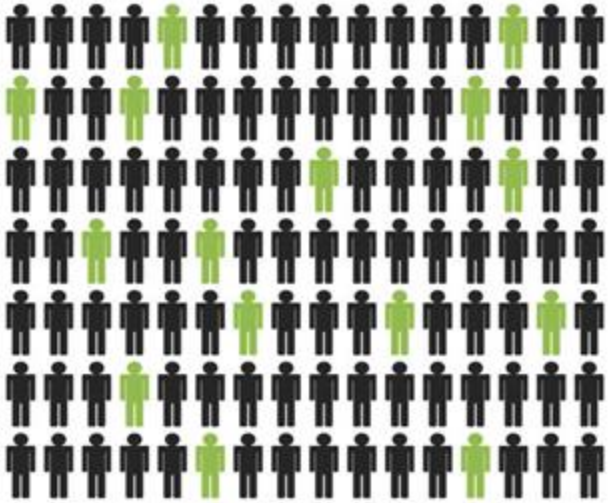
- Cookies were never a complete solution
- Synthetic data supports broader measurement objectives

| | IAB Measurement Framework | | |
|---|---|---|---|
| | **Media Effectiveness (Improve Delivery)** | **Brand Effectiveness (Impact Brand)** | **Sales Effectiveness (Increase Sales** |
| **Techniques** | View ability/attention<br>Campaign metrics<br>Cross media R&F | Ad recall<br>Brand lift<br>Brand equity | Market Mix Modelling<br>Multi-touch attirbution<br>Sales lift |
| **Measurement** | Valid impressions<br>Viewed by humans<br>Attention | Consumer perceptions<br>Associations<br>Influences on brand choice | Consumer behaviour<br>Long term effects<br>Short term activation |
| **Metrics** | Impressions<br>Deduped R&F<br>Demo profile<br>Attention time | Brand awareness<br>Brand favourability<br>Purchase intent<br>Brand equity | Customer lifetime value<br>ROI Measures<br>Change in sales<br>Incrementality |

# New Zealand Perspective – Marketing and Media

- Cookies were never a complete solution
- Synthetic data supports broader measurement framework

- **Start curating your data sets**
  - Customers (current & potential)
  - Find out what you can about them
    - First party data
    - Second party data
    - Third party data

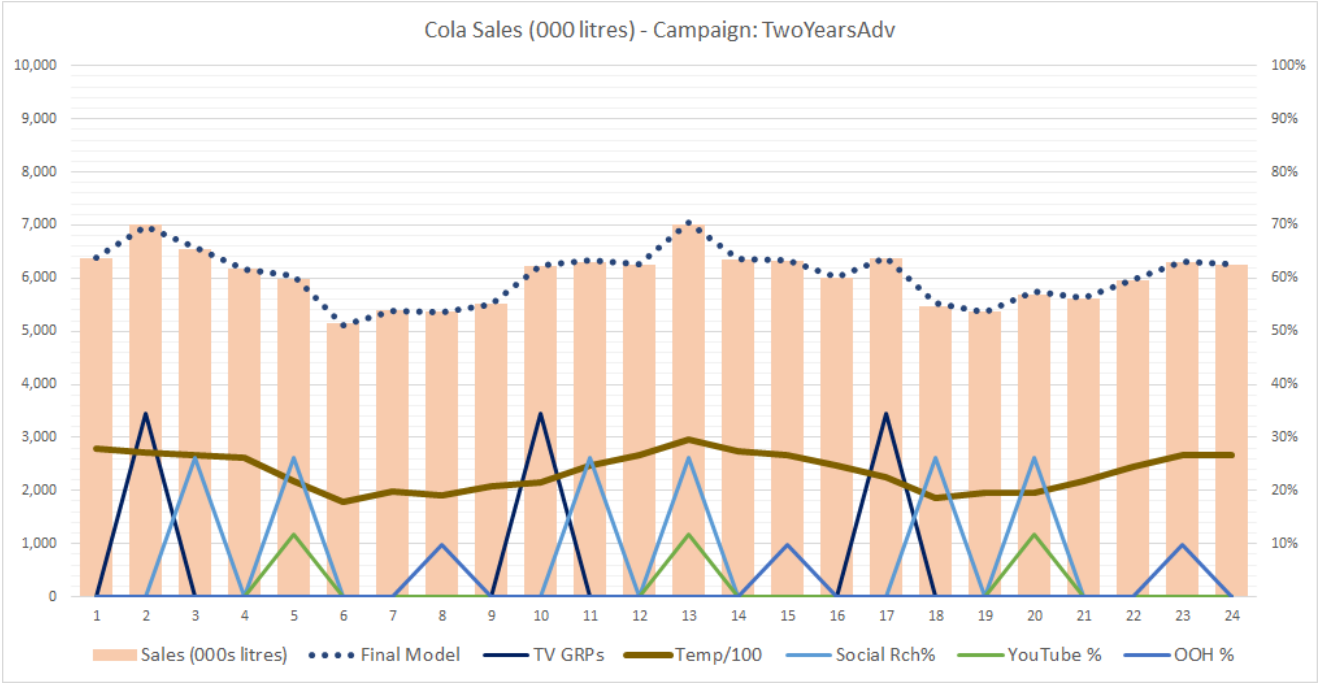


# New Zealand Perspective – Marketing and Media

- Cookies were never a complete solution
- Synthetic data supports broader measurement framework

- Start curating your data sets
  - Customers (current & potential)
  - Find out what you can about them
    - First party data
    - Second party data
    - Third party data

- **Enrich your data sets**
  - Campaign measures
  - Other media
- Other signals
  - Offline sales/leads
  - Environment (economy/weather)
  - Competitor activity



# Conclusions

- Synthetic data applications will grow
  - Complementary approaches
  - Privacy enabled
  - Vs
  - Links to activation

- Core takeaways
  - Start curating data now
  - Don’t limit yourself to online datasets only
  - Quality of inputs essential
  - Importance of matching algorithm(s)

- How to improve
  - Richer data sets
  - Link to real-world

- Keep learning

# [MILTON DATA]