



Data-Integratie Uitdagingen in Consumenten en Media Onderzoek

Ludo Daemen
Nielsen DataScience, Global R&D

NPSO -- 11 Februari, 2019

- **Marktonderzoeksbureau** (sinds 1923)
 - Activiteiten in >100 landen; 6.5 mld\$ inkomsten; >40K werknemers, >1000 'Data Scientists'
 - Gericht op **continue meting van gedrag** ... meningen/attitudes krijgen bij ons weinig aandacht
- **Twee domeinen :**
 - **Consumenten : wat wordt er gekocht** (focus op FMCG)
 - Vooral via gegevens over winkels (panels of census)
 - Grotere landen beschikken over huishoud panels
 - **Media : wat wordt er bekeken, gehoord op TV, radio, internet**
 - Vooral via gegevens over individuen (panels) ... met specifieke (hoog-technologische) meet-processen.
- **Gebruik van data door onze klanten :**
 - Monitoring/Tracking : objectieve data ter ondersteuning van processen/onderhandelingen
 - Analytische diensten (inzicht, ROI-evaluatie, ...)
- **Hoge mate van proces orientatie :**
 - Continue productie (a data factory) ter ondersteuning van hoog frequente aflevering



Vertrouwensfactor

- **Nielsen levert het cijfermateriaal dat bedrijven toelaat ...**
 - Om interne beslissingen (en belonings) processen te sturen.
 - Om op een feitelijke basis met elkaar te onderhandelen.
 - Om te controleren of andere de overeengekomen doelen bereiken.
- **Vertrouwen in het aangeleverde cijfermateriaal is cruciaal** (gelijkaardig aan NSO's)
 - Vooral aan de media-zijde geldt dat niemand de waarheid weet ...
 - vertrouwen in het resultaat is daardoor uiteindelijk gebaseerd op een vertrouwen in de methodologie
 - Rol van 'Media Research Council' (MRC)
 - Aanvaarding van methodologie/process (accreditatie, winnen maar ook verliezen)
 - Continue audit of de methodologie en het process inderdaad gevolgd worden.

Data-Integratie uitdagingen.

- **Deze sessie ... focus op data-integratie aspecten binnen het kader van onze continue diensten**
 - Data-integratie komt ook aan bod binnen de analytische diensten (veelal adhoc projecten met externe data)
 - ROI-analyse (wanneer ad gezien is, hoeveel hebben er uiteindelijk gekocht of interesse getoont (directe matching via 3de partijen)
 - Segmentatie-studies : Kijkgedrag naargelang een segmentatie ontwikkeld via een survey (statistical-matching, e.g. data fusion)
- **Nielsen's achtergrond zijn ... Panels**
 - Panels komen evenwel onder druk
 - Behoeftte aan steeds meer granulariteit (markt fragmentatie + gebruik van data wordt meer operationeel en tactisch)
 - Beperkingen mbt wat/hoeveel er betrouwbaar kan worden gemeten.
 - 'Big Data' worden dikwijls gepresenteerd als alternatief
 - Digitale meting mbt grote populaties
- **Centraal thema : Big-Data zijn zelden een pasklaar alternatief**
 - Meet andere dingen, heeft een beperkte dekking, privacy beperkingen, kwaliteitsproblemen, etc.
 - Meestal alleen betekenisvol in te zetten in combinatie met een panel (er zijn uitzonderingen)
 - Complexe processen

Digital Ad Ratings

- Hoeveel mensen zien met welke frequentie een digitale advertentie campagne (per dag, per site, per demo ...)
?
- Opvolging of gestelde doelen van de campagne bereikt worden (niet 'nice-to-know', opvolging van contractuele verplichtingen).

• Kan vrijwel niet gevat worden met een panel ... niet haalbaar met beperkte steekproeven
Hoe werkt dit ... **Klaar? Nee ... nog niet helemaal**

- Advertenties bevatten een digitale tag/code
- Nielsen server ontvangt een signaal telkens iemand in zijn browser (of in een app) de advertentie te zien krijgt
 census meting, maar geen demo-info, alleen impressies, geen deduplicatie over toestellen.
- Enter Facebook ... heeft demo-info mbt al hun gebruikers
 - Forward de tag naar FB (via de FB-cookie)
 - FB doet de demo-lookup en deduplicatie
 - FB levert ons dagelijks de totalen.

- Niet iedereen zit op FB (met al zijn toestellen)
- Niet iedereen geeft op FB zijn correcte leeftijd/geslacht
- Misattributie (meerdere FB-gebruikers op een toestel)

Correctie via modellen en ... panel gegevens

- Modellen @FB (implementatie in hun processes)
- Modellen @Nielsen (toegepast op de totalen)
- Maar ... al die modellen worden ingeschat op basis van onze panel gegevens.

Digital Ad Ratings

- **Vele componenten**

- + Census-measurement van 'ad impressions'

- + Partnership met een sociaal-netwerk operator

- Toegang tot relatie tussen toestellen en individuen
- Kenmerken van individuen

- + Eigen panels ter ondersteuning van de correcties (compositie + meetfout)

- + Flinke dosis tech/math om het allemaal aan elkaar te rijgen

Surveillance Economy @work

- Desalniettemin ... een consequente keuze voor een privacy-beschermende aanpak

Aanvaarding van de methode ... een complex produkt

- jje moet het allemaal uitgelegd krijgen
- e moet het spontane wantrouwen mbt complexiteit overbruggen.

DAR heeft een MRC-accreditatie (VS)

Globaal produkt

Afhankelijkheden ... Future-Proof ?

Sociaal-netwerk partnership (penetratie, regulering)

Technologische basis (browser settings)

Local Markets TV measurement

- Vele lokale markten in de VS zijn te klein om er volwaardige panels in op te bouwen
 - Traditioneel werden deze markten gemeten met 'dagboeken'. Duur, vrij onbetrouwbaar (response rates; compliance), granulariteit
- Zijn SetTop-box data van kabel/sateliët verdelers van tv-signalen een alternatief

- Belofte : Machine registered viewing data on a near-census basis.

Dekking :

- Niet alle huishoudens hebben kabel/sateliët
- Niet alle tv-toestellen in een huishouden hebben kabel/sateliët
- Niet alle verdelers delen hun data met ons
- Niet alle SetTop-boxen zenden kijkgegevens terug

Data : digital breadcrumbs

- Geeft de tuning van de set-top-box maar niet van de TV
- Kwaliteits problemen moeten er nog uitgefilterd worden
- Geen informatie over de huishoudens (privacy)

- Geen informatie over wie kijkt

Dekking :

- panels blijven nodig om sommige delen van de populatie af te dekken, en om het projectie schema op te zetten.
- panels bieden uitkomst voor ongemeten aspecten (niet aangesloten toestellen, VCR, ...) ... dit kan verschillen per leverancier.

Modellen op basis van tuning-info (in te schatten via de panels)

- Onderscheiden van reëel kijkgedrag en machine-eigenaardigheden
- On/Off model
- Gezins-samenstelling
- Locatie van het toestel binnen het huis

Local Markets TV measurement

- **Opnieuw ... een patchwork van vele componenten**
- + Delen v/d populatie worden nog steeds ingeschat via panels
- + Projectie-schema is afgeleid van de panel gegevens
- + Panel-gebaseerde modellen ter overbrugging
 - van wat settop-box meet en wat je eigenlijk wil meten
 - van wat is weggelaten omwille van privacy-overwegingen
- + Flinke dosis tech/ML om het allemaal aan elkaar te rijgen

Niet voldoende om het te doen ... je moet ook nog het vertrouwen wekken in de methodologische opzet

- Recent op de markt ... maar niet evident
- Panel-gegevens zijn opnieuw onmisbaar als referentiepunt

'Total Consumer' Measurement

- Differentiatie in Retail ... naast super/hypermarkten ... Opkomst van gespecialiseerde kanalen ... en ... Ecommerce
- Traditionele huishoud-panels hebben het steeds moeilijker om verscheidenheid aan consumptie-gedrag te vatten.
 - Data verzameling vraagt te veel inspanning compliance recruitering onderhoud

Opkomst van nieuwe methodes van data verzameling :

- Receipt-capture
- Email-scraping
- Account-scraping

Big Data vs Small Data:

- Big-Data aanpak heeft zijn problemen
 - representativiteit, volatiliteit, beperkte achtergrondinformatie
 - duurzaamheid mbt privacy
- Panel aanpak is wellicht vereist om tot een fundament te komen

Data integratie complexiteit

- Parsing van kassabonnen, bestellings-info, account-info
- (de)codage van de produkt informatie
 - Essentieel om aan te sluiten met de produkt-referentials
 - Enorme schaal (the everything store) ... ML/AI niet meer een kwestie van efficiëntie ... maar een kwestie van haalbaarheid
- Koppeling, deduplicatie

Een nieuw model voor data-verzameling ?

- Meer en meer gedrag wordt digitaal geregistreerd ... moeilijk om als 'big-data' te exploiteren (privacy)
- Komen we op het punt waar je in panels nauwelijks nog informatie verzameld, maar eerder de toelating/credentials om de data elders op te pikken ?

Conclusies

- **'Digital Data' zijn essentieel** ... om op een betrouwbare manier een ruim palet aan gedrag te meten
- **'Big Data' zijn essentieel** om te voldoen aan de roep naar granulariteit
- **'Big Data' zijn echter ... zelden leefbaar zonder panel-informatie**
 - Zelden representatief ... Panels blijven nodig voor aanvulling van de dekking
 - Meestal een beperkte inhoud ... Panels blijven nodig voor aanvulling van aspecten die niet gevat werden
 - Meet niet altijd wat je eigenlijk wil meten ... Model-matige vertaalslag ... Maar panels zijn enige bron om de modellen te voeden
 - Panels blijven essentieel voor validatie en controle
- **Privacy-overwegingen**
 - Niet enkel een kwestie van naleving van wetgeving
 - Ook ... het behouden van het vertrouwen in Nielsen als een bedrijf dat verantwoord omgaat met persoonlijke informatie
 - Essentieel voor data procurement (panels maar ook data-providers) en duurzaamheid van de dienst.

- **Complexe interne data integratie processen :**
 - Op vlak van methodologie, technologie, operations, maar ook op het vlak van de inkoop van data
 - Stelt uitdagingen op het vlak van talent ...
 - Nood aan een zeer gespecialiseerde staf ...
 - Maar anderzijds ook profielen die het geheel op een voldoende niveau van diepte kunnen overzien.
 - Relaties met andere functies veranderden ... Methodologie treedt binnen in technologie, product ontwerp, data inkoop,
- **Vertrouwen in de resultaten :**
 - Vertrouwen is essentieel voor een vlot gebruik van de data door onze klanten
 - Methodologie moet kunnen worden uitgelegd, beargumenteerd en 'bewezen' ... Ook dat vraagt nieuwe vaardigheden.

The background of the entire image consists of a series of overlapping, wavy, blue lines that create a sense of depth and movement, resembling a topographical map or a liquid surface. The lines are in various shades of blue, from light to dark, and are oriented horizontally, creating a rhythmic, undulating pattern.

nielsen

• • • • • • • •

THE SCIENCE BEHIND WHAT'S NEXT