

The future of data integration: Leveraging panels and big data

GfK Insight roundtable 2024

© 2024 GfK. All rights reserved.

Our mission



© 2024 GfK. All rights reserved.

Our mission

Over the last few years, a broad industry consensus has emerged that hybrid solutions offer the best way forward for audience measurement systems, with a role for both panels and ‘big data’ such as set top box, streaming, ad server and connected TV data.

There is far less consensus around an optimum or standard approach to combining panels and big data, with a variety of approaches being pursued. These variously involve the panel as lead technology, placing big data sets at the forefront, or the creation of synthetic respondents and panels.

In the latest of its annual industry roundtables, GfK set out to understand what the key drivers of the debate are when it comes to hybrid measurement solutions. To discuss the issues involved, we asked the views of ten thought leaders in audience measurement, who came together to share and debate views on the future of hybrid measurement.



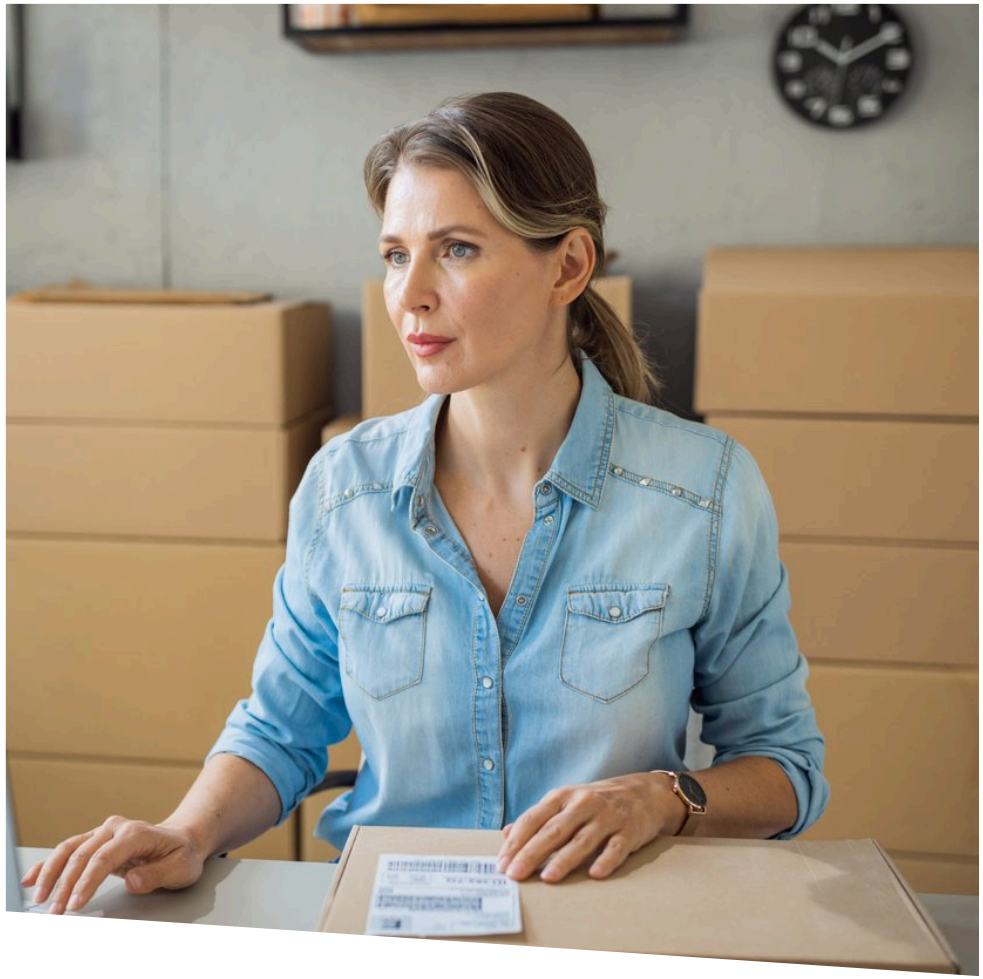
This white paper summarises the key findings. The roundtable took place on 19th September 2024 and was moderated once again by media industry expert Richard Marks.

The participants were:

- Alfonso Calatrava - Meta
- Josh Chasin - KnotSimpler
- Ian Dowds - UKOM
- Michael Debels - CIM
- Matt Hill – Thinkbox
- Dr Mirko Marr – Mediapulse
- Kerstin Niederauer-Kopf – AGF –
- Katherine Page - KP Media Research Consultancy Ltd
- Robert Schaeffner – RTL
- Jon Watts – CIMM



Management summary



© 2024 GfK. All rights reserved.

Management Summary

In this white paper, you'll find the thoughts of 10 audience measurement leaders who debated and evaluated three ways to combine panel and big data in a hybrid measurement system:

- **Panel-first** where an audience panel is enhanced with large data sets
- **Big data-led solutions** that combine a variety of data sets and use a calibration panel to fuse the data and provide demography
- **Synthetic datasets and panels** that blend virtual respondents with a panel, big data, and other datasets.

Our experts highlight several unique contributions that panels offer hybrid systems - total market coverage, persons-centric data, and independent third-party measurement. However, they recognise shortcomings such as the inability to reflect the longtail of media consumption, the high cost of quality panels, and representing harder-to-recruit sub-groups. They tackle the future relevance of reach and frequency metrics in a fully digital ecosystem in **Section 2**.

Big data offers the granularity that advertisers and content creators increasingly crave at a more achievable cost. However, it is no longer seen



as the solution to every measurement challenge. In **Section 3,** our experts address the limitations of big data sets – from representativeness to compatibility to definitions. In **Section 4** our roundtable members decide if they favour a panel first or big data first measurement system and in **Section 5**, the talk progresses to currency versus calibration panels. In the Politics of Measurement, the conversation turns to the bigger picture, considering the future influence of data set availability and privacy legislation on data integration. The panellists imagine what happens if data streams are turned off.

Section 6 covers the politics of measurement and addresses concerns around data set availability and evolving privacy legislation

In **Section 7** the experts investigate the opportunity for synthetic data sets - new, virtual, respondents that can be as large as the target population itself. There is concern about delivering the resulting datasets, both in terms of their size and compliance with privacy legislation. This is one attraction of the synthetic data approach where the source data are one step removed. As is discovered in **Section 8**, the future for hybrid datasets may lie in centralized processing systems, accessed via APIs.

In **Section 9** our participants examine the major obstacle of how to explain increasingly complex media measurement systems to data users. To maintain faith in currencies used to trade billions of dollars of ad spend, the industry needs to overcome an emerging 'crisis of understanding' where only a few experts understand how measurement systems work.

Informing users about the importance of quality data will be vital to the success of hybrid measurement systems. Education is one of the stated missions of our annual roundtables and the published papers that follow them. We see ourselves as one of the builders of the future of media measurement and we hope that this paper, and the previous ones, will be of help to our readers today and tomorrow. Until next year.



Introduction: the measurement ingredients



© 2024 GfK. All rights reserved.

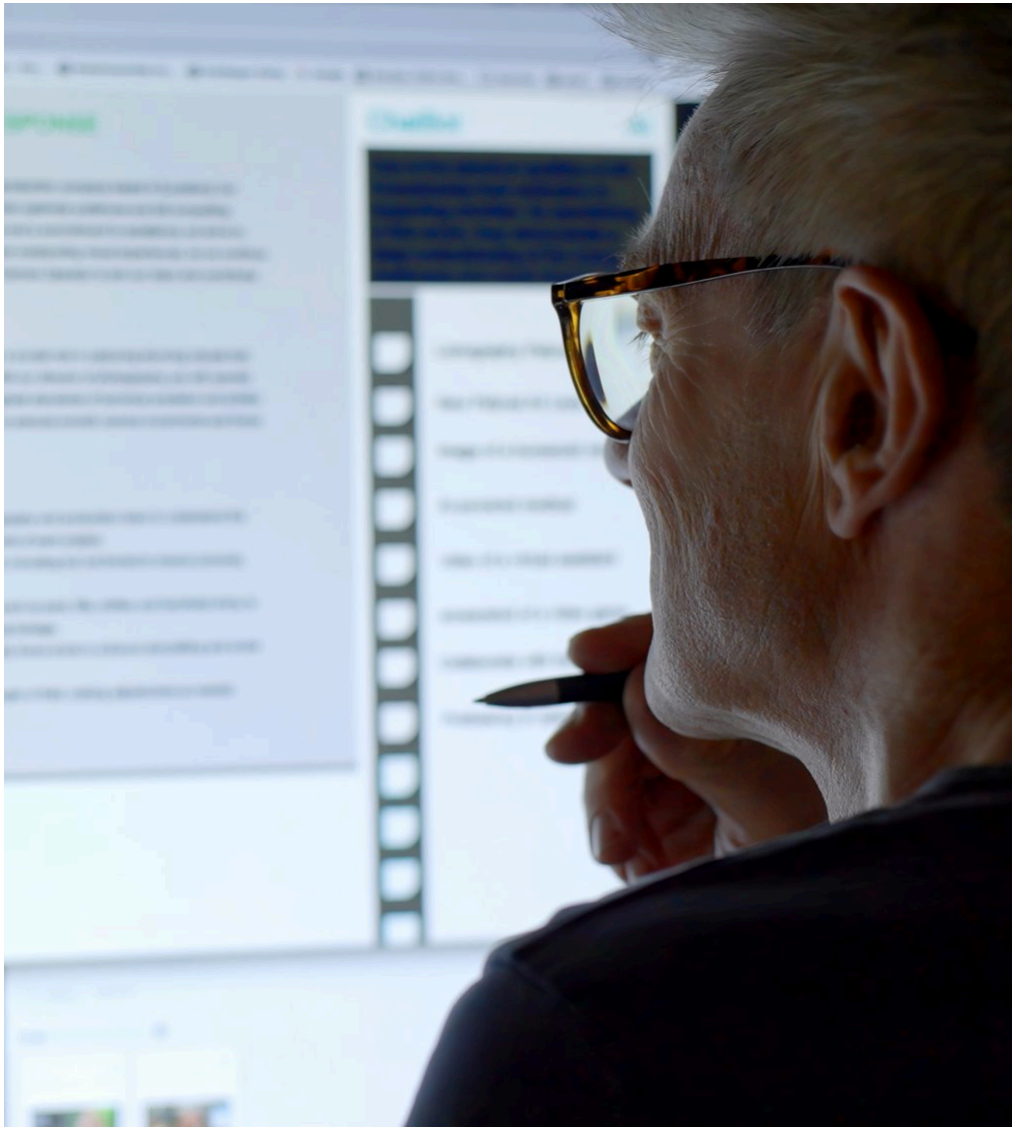
Introduction: the measurement ingredients

Our round table conversation focused specifically on the various methodological approaches to currency measurement.

There are three broad use cases for hybrid approaches:

- Improving currency audience measurement
- Media groups enhancing their own device level media usage data
- Using audience data to activate a brands own first party customer data to enhance media planning, improve targeting and evaluate outcomes

There are significant opportunities being pursued in the combination of audience data with first- party data sets tracking consumer behaviour, offering the ability to track outcomes (sales). The main focus of our debate was hybrid methodologies for audience measurement itself, with a primary – but not exclusive - focus on the measurement of television and video.



Four critical inputs for hybrid audience measurement can be identified, specifically:

Large digital data sets such as streaming data, Connected TV data, ACR data and ad server data. For ease of reading, this report will refer to the term 'Big Data' as a catch-all term for these forms of data.

Identity resolution techniques to attribute demographics to big data sets, such as log-in data collected for subscribers and modelling techniques.

Panels – Panels of households or individuals whose media usage is tracked using metering technology, recruited to be representative of a wider population.

Metadata – data which identifies and labels for analysis ad campaigns and content.

"I think the industry is largely aligned that the future of measurement is going to increasingly involve leveraging big TV data sets, which are hugely valuable for certain needs and use cases. However, it's not a panacea. The data sets have gaps and biases, they need to be projected to the wider universe, and so on."

Jon Watts CIMM

It's important to note that, whilst advertising was inevitably a core element of the conversation, audience measurement systems meet many use cases, including content evaluation, scheduling, competitive intelligence, viewer attraction and user retention to name but a few.

If we rewind ten years or more, when big data was first becoming available, strong arguments were being made that big data was the future and the availability of data at scale would make expensively recruited 'small' sample audience panels unnecessary: the future was big data. Debate raged between the proponents of big data and those who pointed out its limitations and argued that panels remained the most robust approach. Big data was the 'disruptor' technology with proponents of panels seen as defending a legacy business.

It's important to note that, whilst advertising was inevitably a core element of the conversation, audience measurement systems meet many use cases, including content evaluation, scheduling, competitive intelligence, viewer attraction and user retention to name but a few.

If we rewind ten years or more, when big data was first becoming available, strong arguments were being made that big data was the future and the availability of data at scale would make expensively recruited 'small' sample audience panels unnecessary: the future was big data. Debate raged between the proponents of big data and those who pointed out its limitations and argued that panels remained the most robust approach. Big data was the 'disruptor' technology with proponents of panels seen as defending a legacy business.

As is so often the case, the future has proven itself not to be a simple binary approach.

"From where I'm sitting, it's a little bit of a non-question: I think we need both things. We need the survey or panel framework, and we need that to be high quality. We need the big data because otherwise we simply won't be able to measure all these very fragmented and complex media vehicles and behaviours. So we absolutely need both and they sit alongside each other. I don't see it so much as, big data augmenting panels or panels calibrating big data. Both of those things obviously are true, but they're two truths in the same room, if you like."

Katherine Page, KP Media Research Consultancy Ltd



When it comes to panels and big data, there is clear consensus that it is a not a case of either /or.

So, what do these two components – panels and big data – bring to the party when it comes to hybrid measurement?

Panels in a hybrid system



© 2024 GfK. All rights reserved.

Panels in a hybrid system

The benefits of panels

Why are measurement panels still seen as a vital component of a measurement system? Our experts highlighted a number of unique contributions that panels make to hybrid systems:

Total market coverage: The ability to cover the whole eco-system, as opposed to just subscribers to individual platforms or services. This is particularly important when it comes to services that may not have a return path such as free-to-air television. A panel can identify and track the overlap in usage between walled garden services, particularly vital for tying big data sets together in the correct way.

Persons-centric measurement. Advertisers want to reach people, not just devices. Panels play a key role in establishing who is using devices, in terms of their demography and also the number of people viewing a device, particularly important for television screens. Panels can track the same people across different screens. This is essential for establishing reach and frequency, which is becoming harder and harder due to privacy legislation and the gradual demise of cookie-based solutions. Panels offer a consented approach to tracking people across devices and platforms.
Independent measurement. A well-balanced panel with high quality recruitment provides a trusted and transparent core measurement for the market.

"I think the strength of a panel lies in its ability to represent the entire population. You create a small sample size that is representative of the total universe. You need a stable, robust basis on which to build your models, allowing you to cover different target groups, not just segments from specific parts of the population or particular technology owners. Therefore, you need a very good panel, a high-quality sample in order to avoid creating bias."

Kerstin Niederaurer-Kopf, AGF

The challenges of panels

Nonetheless, panels as a methodology do face some limitations and challenges, which necessitate the incorporation of big data sets. Our experts identified the following limitations:

Panels are by their very nature sample-centric, and their ability to accurately profile the increasingly long tail of niche content and track campaigns at a granular level is inhibited. As linear audiences fall relative to on-demand, panels can be expanded, but there is a law of diminishing returns when it comes to making panels larger and larger. Panels offer a 'satellite view' of the whole market, but tend to lose granularity when the user wants to zoom in on a specific part. That is where big data offers a cost-effective and more granular alternative.

"The balance is all a question about use cases. Are you interested in the big picture of viewing like overlap between media types and so? Then a panel would be excellent because it can cover exactly that in a very transparent and credible way. Are you interested in ad campaign management? Well then, the panel will probably struggle. And you have to find another solution."

Robert Schaeffner, RTL

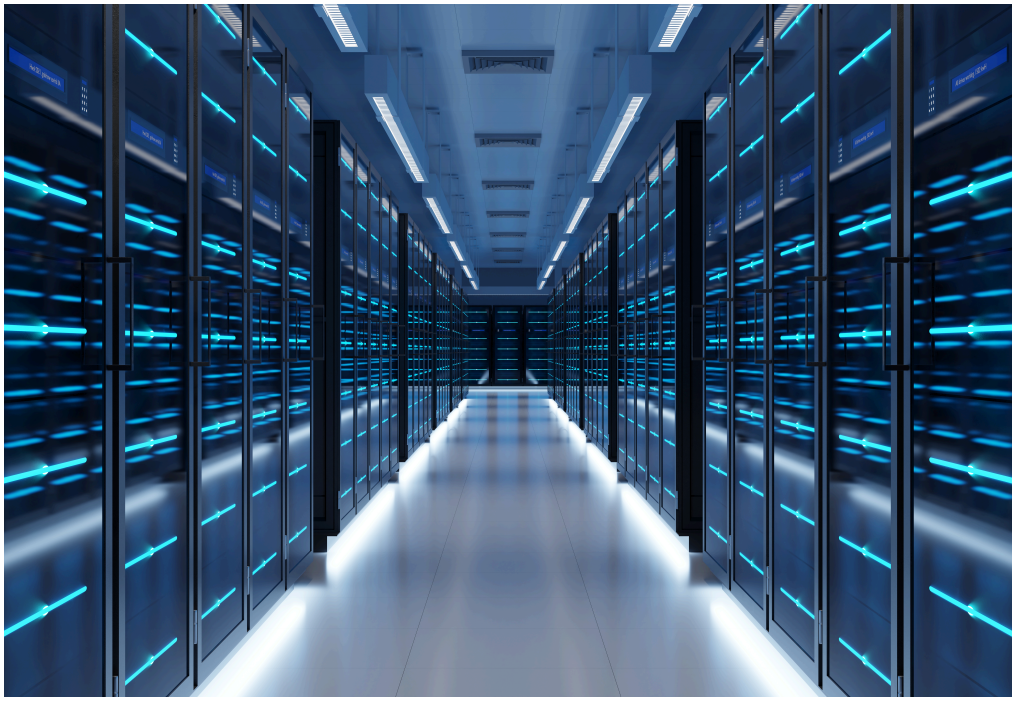


The cost of panels is also an important limitation. For panels to be accurate and fully representative, then high quality recruitment and panel management are needed, alongside robust measurement technology such as people meters and router meters. For these reasons, whether panels are used as the lead technology in a hybrid solution, or as a more basic element for calibration and personification, they are usually the single most expensive element within a hybrid system.

Doubts were also expressed by some of our experts about the longer-term ability of panels to remain fully representative of the whole market when it comes to demographic sub-groups. It can be challenging to have proper representation of sub-groups, not just age, gender and ethnicity, but social groups that may be less likely to be recruited such as more active or mobile parts of society. This can lead to a research bias in the panel itself.

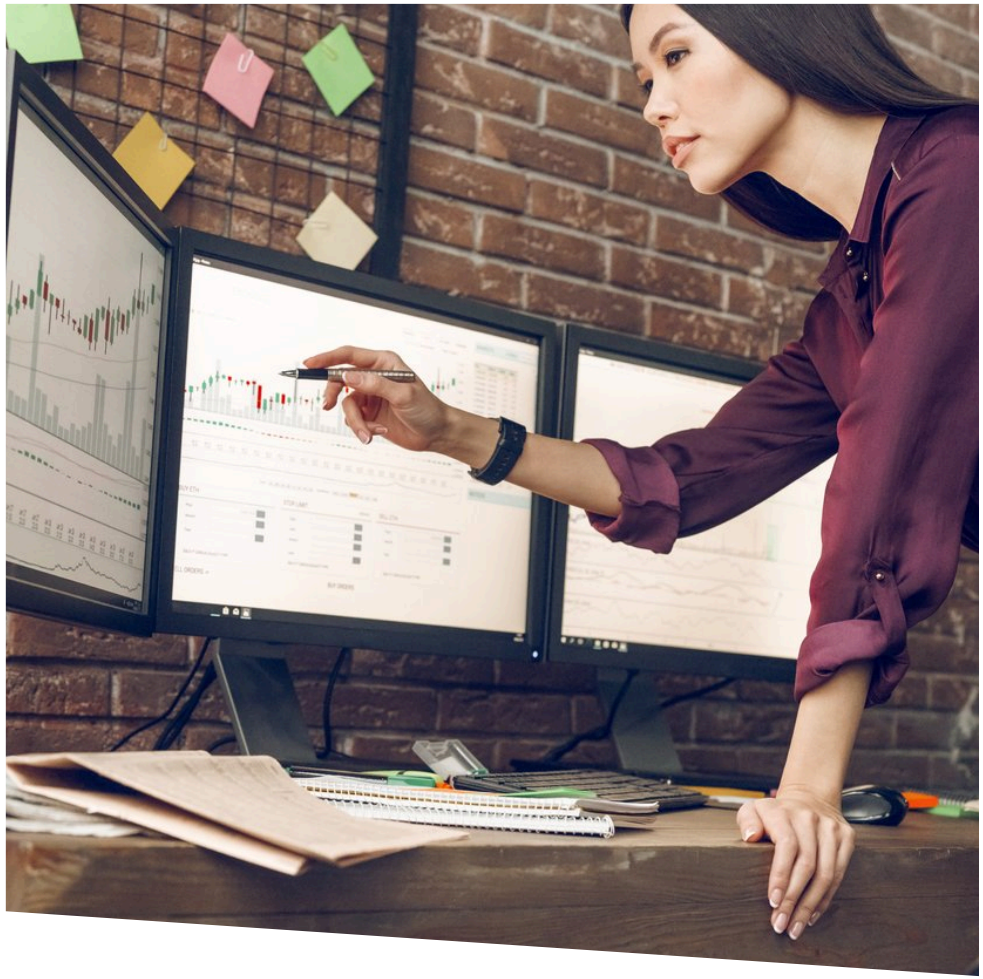
“There's a recruitment challenge and a price challenge. It’s impossible to take away a research bias in a panel. There's only a certain type of person that is willing to take part and be a part of a panel, so that's always a problem. There's the cost in terms of recruiting it and maintaining it. You know the churn that you get on panels. People get bored of doing it. Then there's the fact that it's trying to scale it to a size. Very quickly, the math stops working from a cost effectiveness point of view.”

Matt Hill, Thinkbox



Katherine Page also noted that one of the key rationales for the continued existence of single-source panels is the traditional planning focus on reach and frequency going forward. It can be argued that many online trading mechanisms have moved beyond reach and frequency. One of the determinants of the industry’s willingness to continue to fund panels may be the continued relevance of reach and frequency and therefore the need to track individuals. This should not be taken for granted in a world of impressions and outcomes.

Big data in a hybrid system



© 2024 GfK. All rights reserved.

Big data in a hybrid system

The benefits of big data

If there is one word that best describes the core strength of big data – streaming, CTV, ACR, ad server, set top box data et al - it is 'granularity'. Not only is the tail of content becoming longer and longer, presenting a challenge to sample-centric measurement, but ad campaigns are increasingly using addressable techniques, meaning that the links between content and advertising are becoming severed. This means that adverts have to be tracked at an individual spot level, as opposed to just being linked to the content they accompany.

There has been growing concern about the growth in the number of zero-rated spots reported by solus panel solutions, spots with audiences too small for a panel sample to measure. When added together these comprise an increasing and worrying proportion of all advertising. This is where big data – particularly census streaming and ad server data – comes into its own. Systems that incorporate census data sets are seeing dramatic improvements and reductions in the proportions of zero cells.

There are also benefits for content measurement. As on-demand usage becomes more prevalent, being able to identify content in large scale VOD archives is important for media owners to understand content demand, with audiences accumulating over extended periods of time.

"We have a diverse media landscape full of various offerings in the video segment. If you want to rebuild everything out of a panel, this would not be feasible due to the fragmentation of the market. Not everything can be measured by only using a panel. I think you have to be cost effective as well, and that's why, in order to keep your costs under control, you need to enrich your insights through big data sets."

Kerstin Niederaurer-Kopf, AGF

The challenges of big data

A primary and obvious limitation of big data sets is that they typically reflect device or platform usage as opposed to persons usage. In some cases, a user may be logged in to an account which has demographics linked to it and this can be a useful start point, but this does need to be verified. For example, one of the participants in our round table was using the Teams account of their daughter. Meanwhile, shared-use devices, such as TV screens may have different users at different times of the day and multiple users. In many cases big data sets may have no demographics available at all. This is why panels are essential for turning devices into people.

"At some point our stakeholders were thinking that big data could solve everything. An easy way, maybe that they could do themselves and that they didn't need us anymore. But the fact is that everybody realized, indeed, that there are gaps in the big data and that they still need some other companies, for example like JICs, to organize, to structure, the exchange of the data, to audit, to evaluate the quality of the data, to identify the gaps, and to find also solution to combine them, to have a sort of synergy and add to the value of both types of data set."

Michael Debels, CIM

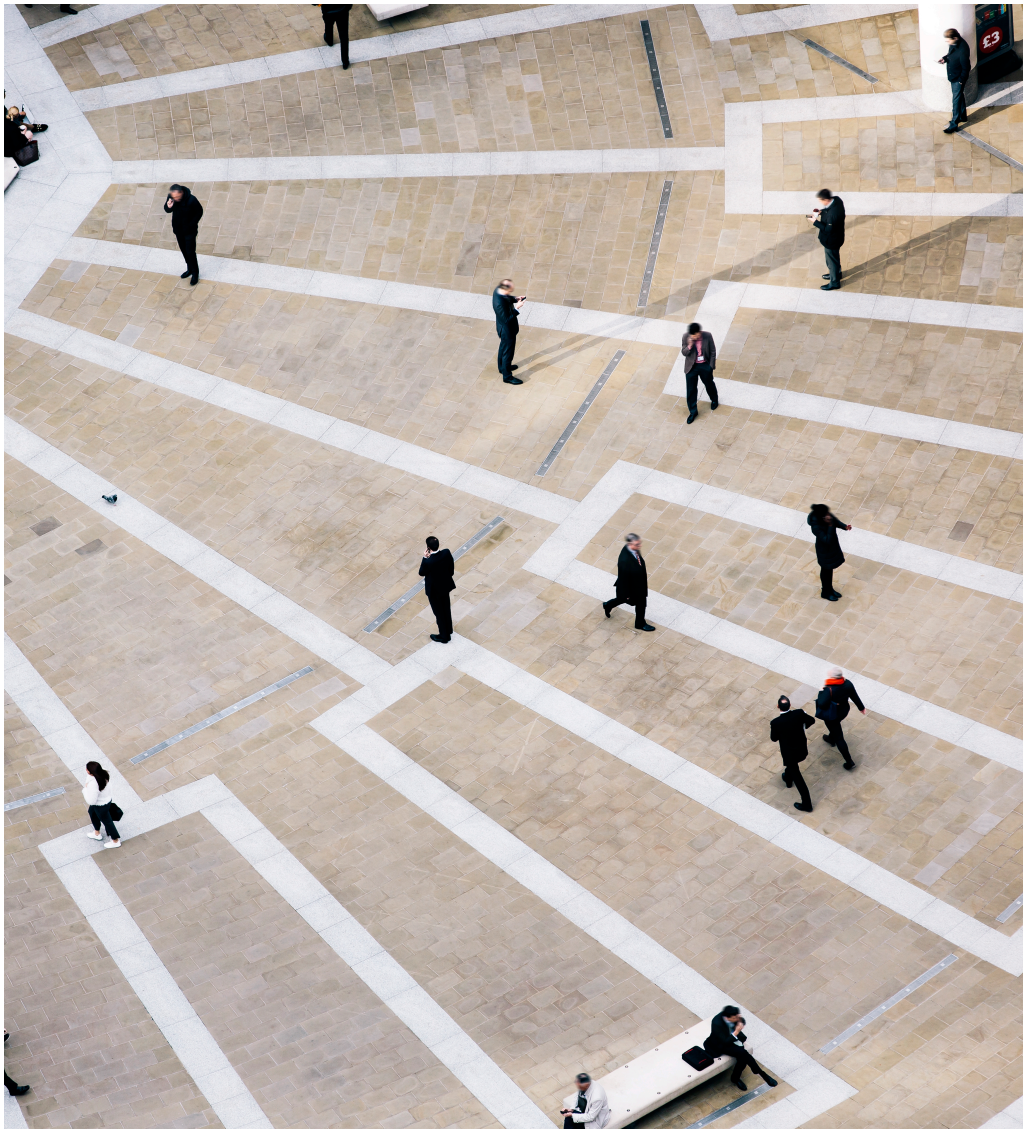


A further limitation of big data sets is that they typically reflect at a very granular level a subset of the total media eco-system, a detailed view of subscribers to a particular platform, or of devices that a particular ad campaign was served to. They are a deep, detailed dive into a specific part of the market, but need to be connected to other walled garden data sets to provide the bigger picture and understand the context. This is particularly challenging when attempting to track individuals across devices and services. It's a challenge to tie together ad campaigns to evaluate the contribution of different platforms, but it's also difficult for media companies to understand the competitive context in which they operate, what their users do when they are not using the platform.

"A regular "path to purchase" nowadays involves multiple touchpoints and different connected devices, which makes it much harder to get an individual-level understanding of actual consumer behaviour and media consumption patterns from a cross-media perspective. This is a relevant information gap that panel-based solutions can effectively help fill in."

Alfonso Calatrava, Meta

Jon Watts of CIMM highlights that even when multiple big data sets are available, their compatibility cannot be taken for granted. Different data sets may have different compositions and different definitions. The way these issues are addressed can produce very different results, to the extent that there is a fine line between data science and just 'guessing'. This highlights the importance of the industry collaborating to agree common definitions and shared resources like agreed universe figures to project to. Josh Chasin also highlights that there can be issues with correctly attributing ethnicity to large data sets. Initiatives are now underway to address potential issues in the error rate in demography in large data sets.



Panels first or big data first?



© 2024 GfK. All rights reserved.

Panels first or big data first?

There is now general agreement in the industry about the benefits of hybrid methods for audience measurement, leveraging the benefits we discussed of both large data sets and of panels. The debate has arguably moved on from whether to create hybrid data sets to the 'how to do it' and most specifically the new role of the panel. Is the panel the lead element, with big data sets used to enhance the panel in terms of granularity and precision? Or is the panel there to play a role in the background, providing demography for personification and calibrating the way in which the big data sets are tied together, training the models that combine big data? Put simply, is the future about audience panels, or calibration panels? Is it about panels first or big data first?

Panels first?

'Panel first' solutions retain the measurement panel at the heart of the system and ingest big data sets where available– in whole or in part – into a respondent-level database to increase granularity and precision across specific platforms and services. The panel may be expanded virtually to allow matches to be made within big data or other calibration techniques used, but typically the output is a respondent level database.

"Yes, no doubt the future is hybrid. I'm a 'panel first' fan and our experiences have made me even more of a fan of the 'panel first' approach. The big data we use, they are a beast and you have to domesticate that beast. It's really, really important to have a panel as a source of truth to use these data and to bring them into a structure, to bring them into a proper relation to the universe, to the figures of the of the population you are measuring."

Mirko Marr, Mediapulse

One of the main arguments for panel-first solutions is that often larger data sets may only be available for specific parts of the market, particularly those without a return path, or those within walled gardens that may not be made available. However, this may gradually change as more and more media usage is via connected devices, such as Connected TVs and an increasing proportion of the eco-system has census level data available for use. This will need to be monitored as it can't be taken for granted that a panel first approach will always be fit for purpose.

"Our position currently is people first and therefore it's panel based, but that's not just set in stone philosophically that's the best thing to do. The growing importance of big data and the need to have a consistent or transparent evaluation of the inputs into that big data could change our position on that. So we would never say it's one over the other, but the need for some element of consistency and transparency and auditing of the value of the inputs, and the consistency of the inputs is critical."

Ian Dowds, UKOM

A critical determinant will also be the quality and size of the panel itself. As the roundtable discussed earlier, high quality representative panels are increasingly challenging to recruit and maintain. As the media landscape becomes even more fragmented, audience panels may become more and more stretched. So, what are the benefits of using panels for calibration as opposed to the lead technology?



Big data first?

An alternative approach to ingesting big data into a currency audience panel is to move away from a sample-centric approach and effectively have the big data as the lead technology. In this approach multiple large datasets are joined or fused together directly. A panel is used as a building block, providing critical demographic information, but also informing the way in which the data sets are joined together by providing a 'source of truth' for the overlap in usage between different services and platforms. The panel can be used to train the models that are used to fuse the different datasets together. Various approaches exist to achieve this, including the VID (Virtual ID) approach being used by the Origin (UK) and Aquila (US) cross-media projects and outlined in the WFA's open-source Project Halo guidelines.

This approach is proving particularly popular amongst disruptor services such as the 'alt currency' suppliers that are competing in US measurement. One of the main attractions is that in reducing the role of the panel, or even sharing a panel across multiple vendors, the cost and barriers to entry are reduced.

"Over the last, say 7-8 years, I've become a convert from panel first to big data first. The economics of being able to make a rating service out of big data are different than a panel-based system. If having big data enables you to compete in the TV measurement business or let's call it cross-platform video measurement business then the possibility emerges that there can be a shared panel that each of these companies accesses."

Josh Chasin, KnotSimpler

Currency panels vs calibration panels



© 2024 GfK. All rights reserved.

Currency panels vs calibration panels

In the conversation about the relative merits of panel-led and big data-led solution an obvious methodological question becomes important, one that is implicit in Josh's comment above: in terms of design (and therefore cost) what is the difference between a currency panel and a calibration panel? What are the implications, for sample quality and panel size in particular?

Our experts were not necessarily of one mind on this. There are strong arguments that there should not be a difference, that the quality and size of the panel has to be of paramount importance, no matter how it is being used. It is vital for data users to keep pressing the paramount importance of quality and that any system can only be as good as the quality of the data on which it is built.

However, some, like Josh Chasin acknowledge that there are arguments that a calibration panel can be relatively smaller in size if its primary role is to track the reach and overlap at a service or platform level, as opposed to specific pieces of content or ad spots: that the panel is an architectural plan for joining the data together, with the big data providing the brickwork. The MRC in the US does have a framework for accrediting panels recruited using techniques other than random selection.

This is an important argument where more concrete proof is needed – if the primary purpose is to train fusion models, then how good is good enough? There are already cost pressures on panels when used as a lead technology, those pressures will be even greater when a panel is simply a component of a measurement system, part of the recipe, as opposed to the main meal itself. If it becomes ever harder for a panel alone to measure at a granular level, does big data relieve that pressure and refocus panels on measuring at a more macro level?



Data availability – the politics of measurement.



© 2024 GfK. All rights reserved.

Data availability – the politics of measurement.

"As a neutral and independent company responsible for audience measurement, we have to remain independent from data providers. That's why the panel is very important for us. We won't say 'OK, we will absolutely focus on a particular big data set', because what will we do if that provider is no longer able to provide the data? We have to deliver market data on a daily basis. If this data is somehow unavailable, we should at a minimum, be able to report from an independent representative panel."

Kerstin Niederauer-Kopf, AGF

Whilst our conversation understandably focused on methodological issues, there was important debate about the wider 'real politik 'context in which measurement exists. Two external factors were felt to be particularly important in differentiating what is practical and sustainable as opposed to methodologically possible: data set availability and evolving privacy legislation.



Both considerations have a particular impact on the feasibility of big data-led services.

To get access to large data sets, agreements have to be made, with contracts in place, Data Protection Agreements and APIs accessed.

Taking the US market as an example, various measurement companies have deals in place, exclusive or not, to process data from cable and satellite companies, CTV vendors and online platforms. They build a product, combine it with other datasets and go to market, but what happens if the data is switched off – if a partner demands more money, has a change of ownership and decides to take their data away.

How does a big data-first approach deal with the politics of the data suddenly disappearing? How is that hole filled? How is that particular service reported? With a panel first approach, ratings would still be published, albeit on a smaller sample as an audience grade panel is still in place to produce TV ratings.

Equally if there were no political issues, but technical issues arise, for example a change to a different ad server or a temporary blackout from that platform, data can become temporarily unavailable.

"The killer argument for 'panel' first, at least from a JIC perspective, is the stable availability of the data and control over the data. If you have to deliver viewing data day by day, everyday, you have to rely on the fact that you can get the data and most of our data we are talking about are in the hands of other people. In the panel we have control over the data delivered by companies like GfK and all the other data we build. So we are ready for those days when we don't get set top box data. This security you can only get from a panel."

Mirko Marr, Mediapulse

It was also pointed out that it may not just be the absence of a particular data set that causes disruption, but also the addition of a new data set could affect the continuity of ratings provided.

This concern about dependency on external data sets does seem to be informing the strategy of measurement JICs when it comes to using big data sets: that big data can enhance elements of the panel measurement, but cannot replace them entirely. Systems that lead on big data will need to have systems in place to cope with permanent or temporary losses of data. This can also lead to a more controversial area which is using models to take a large dataset from one platform and use modelling techniques to adjust and project it across missing data sets to impute ratings: for example adjusting big data from platform A to work as a proxy for platform B. The question is at what point do the limits of data science lie? At what point does it turn into just guessing?

Either way, it's clear that any successful system, whether panel first or big data led has to be able to cope with unavailability of specific data sets, whether temporary or permanent.

"There are definitely problems, or issues you confront, when making assets out of big data. And the obvious answer is you need to have - and I've believed this for a while - you need to have enough redundancy in your data assets that you can recover from losing anyone."

Josh Chasin, KnotSimpler



The opportunity for Synthetic data



© 2024 GfK. All rights reserved.

The opportunity for Synthetic data

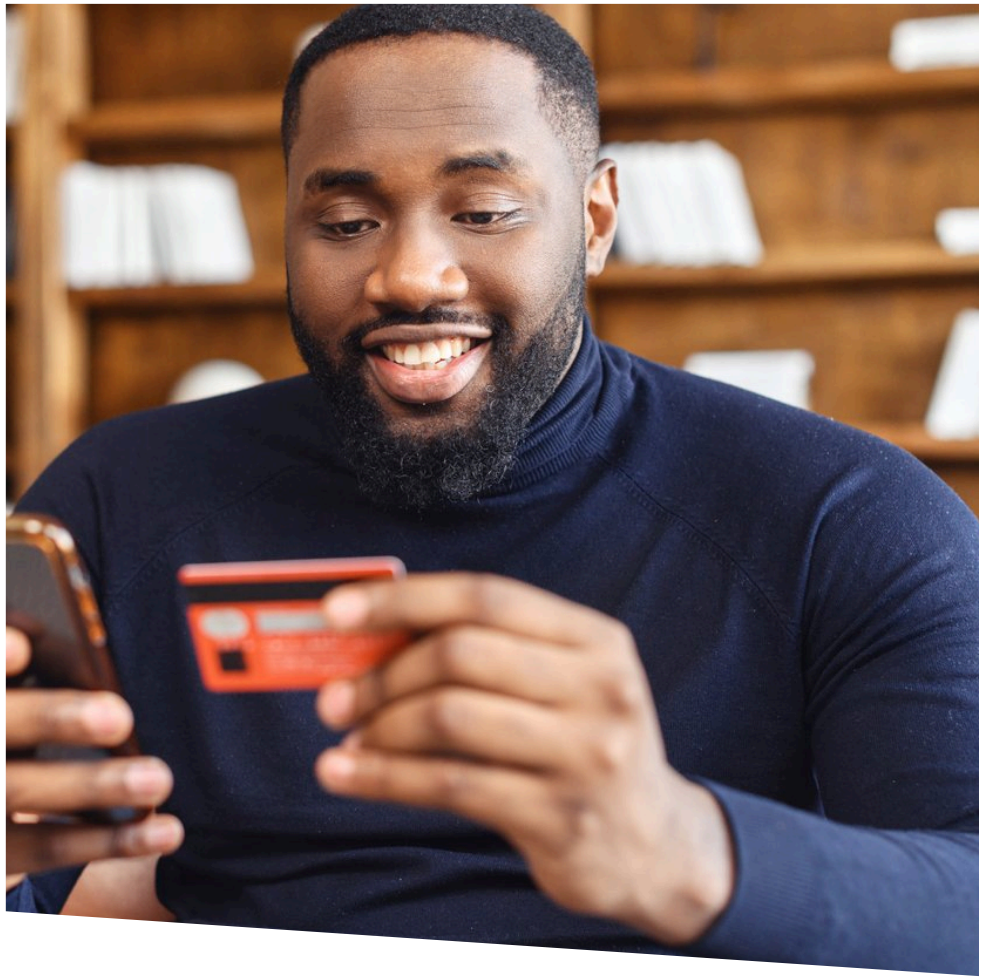
Whilst the conversation did focus on the relative merits and challenges of panel-led and big data first solutions, the growth in the popularity of synthetic data and panels was also highlighted by our experts. Whilst it would be a grave over-simplification to describe synthetic panels as a 'third way' approach, they do provide an alternative to either ingesting big data into a panel, or fusing big data sets together at a census level. In this methodological approach, a set of new, virtual, respondents are created. This panel can even be as large as the target population itself. Relevant information from various data sets, including panels, big data and personification techniques can then be ascribed onto these virtual respondents to create a synthetic respondent level data set.

"Synthetic data is a way of bringing together a variety of data sources. Before synthetic data was starting to emerge as a solution, there were very, very clunky solutions where it was very difficult to align different data sources. Synthetic data is basically providing a much better way to do that. Absolutely I think synthetic data will be the way forward and we're seeing that that will become increasingly used. One of the brilliant things about synthetic data is that you can get around the privacy issues and have a respondent level database."

Katherine Page, KP Media Research Consultancy Ltd.

In particular, as our experts went on to discuss, the privacy challenges that Katherine raises may well be the single most important factor in how panel and big data come together in a future-proof way.

Data delivery, APIs and PII



© 2024 GfK. All rights reserved.

Data delivery, APIs and PII

“In an extremely competitive industry context, the need for advertisers to be strongly data-driven in their decision making is really now more critical than ever. With privacy regulations making access to personal level information increasingly challenging, panel-based solutions combined with privacy-safe integration approaches like “data clean rooms” represent a meaningful opportunity for businesses to drive well-informed decision making in a more sustainable and resilient way.”

Alfonso Calatrava, Meta

It is no surprise that privacy legislation was an important topic in our round table conversation. The definition of PII (personally Identifiable Information) is widening, not just to the obvious data like IP addresses and people's emails, but also potentially viewing patterns that may be enough to be able to identify someone in a way that's actually in contravention of legislation in some cases.

The strength of panels is that they are fully consented, but systems that use actual respondent- level first- party granular viewing behaviour may present PII issues if users can access individual viewing records and also panel security issues if those who donate the data can then use pattern recognition to work out which of their users are actually on the core measurement panel.



So one question going forward is the security large data set measurement. The second issue of how practical and efficient it is to distribute databases that may number millions rather than thousands of respondent records.

One approach is not to make the data available at respondent level. We've lived in a world of respondent level databases largely where databases are made available to bureaux, competitive companies who then have the software ingest the data and users analyse 'anything by anything' because they have a respondent level database. This may no longer be sustainable in a world of big data. There is awareness of this issue. We may be moving towards data being processed centrally and accessed via API to protect the integrity of the database itself.

One advantage of the 'synthetic data' approach is that it makes it significantly harder – although not impossible, some argue – to identify and relate the data to the original 'real world' donors themselves.



"I think that a synthetic population is also a part of the answer because they are also a technical way to protect and analyse data. It solves the problem of the size of the dataset because you can decide yourself on the size of your synthetic population and it certainly is a part of the solution for the anonymization of the data even viewing patterns. In a synthetic population the general modelling is making it nearly impossible to make a link between an original respondent in a census data set and the final modelling tool population."

Michael Debels, CIM

Are hybrid solutions leading to a crisis of understanding?



© 2024 GfK. All rights reserved.

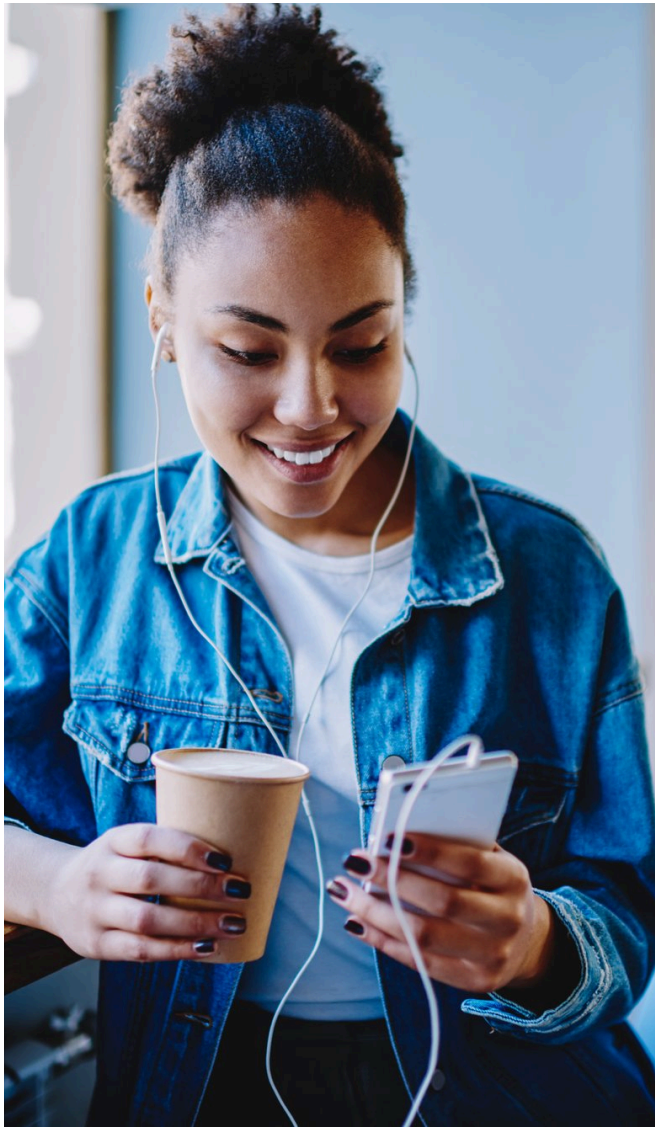
Are hybrid solutions leading to a crisis of understanding?

It's clear that hybrid solutions are the way forward for audience measurement but, as our experts discussed, these increasingly complex data integration solutions present significant challenges in terms of transparency.

There has been industry consensus on the need for transparency in measurement if the trust in currencies is going to be maintained. Users will need reassurance around the source and quality of the data inputs and the optimal nature of the processes needed to combine them and turn devices into people.

However, as these systems become ever more complex, how realistic is it to expect users to understand highly complex methodologies, data science and statistical principles? To comprehend techniques that may only be understood by specialists in each of these fields, let alone a small number who understand all the moving parts of a hybrid measurement system? Put simply, what point is transparency if few people understand what they are being shown? Might it backfire as people find the explanations incomprehensible?

As our moderator Richard Marks pointed out, we are quite happy to use advanced technology like iPhones that would appear like magic to past generations, without needing to understand how they work, as long as they do work. Are we now just going to have to rely on a handful of experts to police the data that we are using?



"What we're doing in the measurement industry is getting more complicated. You know the price of a data scientist these days is pretty high, isn't it? It's because there's not that many of them and there's not that many of them who are specialists in audience measurement. So, you do end up with just a handful of people who actually really know what the hell is actually happening in terms of that overall model. So, I think the value of credible businesses and their brand is going to increase. The importance of having a known brand and a known person who is respected for being able to do creditable work is going to increase. If I'm going to advise my kids what career to go into, I'm going to strongly advise that they stick to their statistics courses at school!"

Matt Hill, Thinkbox

As Katherine Page argues, it may not be a deliberate lack of transparency that is the issue going forward, or a deliberate intention to hide what's going on, but the need to inform and educate data users as to why quality matters. If people are making business decisions involving substantial amounts of money, they need to have some confidence in the data in which they're making those decisions. They need to understand the importance and value of quality data, otherwise the danger exists, as Jon Watts fears, that some executives may prioritise the attractiveness of the data tool over the quality within that tool when it comes to allocating budgets.



"I do think that we need to think carefully about how securing and strengthening the institutional arrangements underpinning transparency, governance and auditability in the future. In the US, there are processes in place – good, thorough processes –but there are industry participants who argue that these purposes aren't fit for purpose in the age of big data and that they discourage new entrants. How will we audit the AI-driven synthetic data models of the future? We know from historic experiences that independent third-party verification is hugely important, as there are inevitably problems and errors. How will we deliver transparency in the future? I think we're going to need a step change in the institutional arrangements."

Jon Watts, CIMM

In this new hybrid world, the roles of governance, auditing and data verification will become even more important.

"Governance is the key word that stands out when it comes to what UKOM does, and the governance and the understanding of both the quality of the inputs and the methodology by which they have been fused is key for the UK industry"

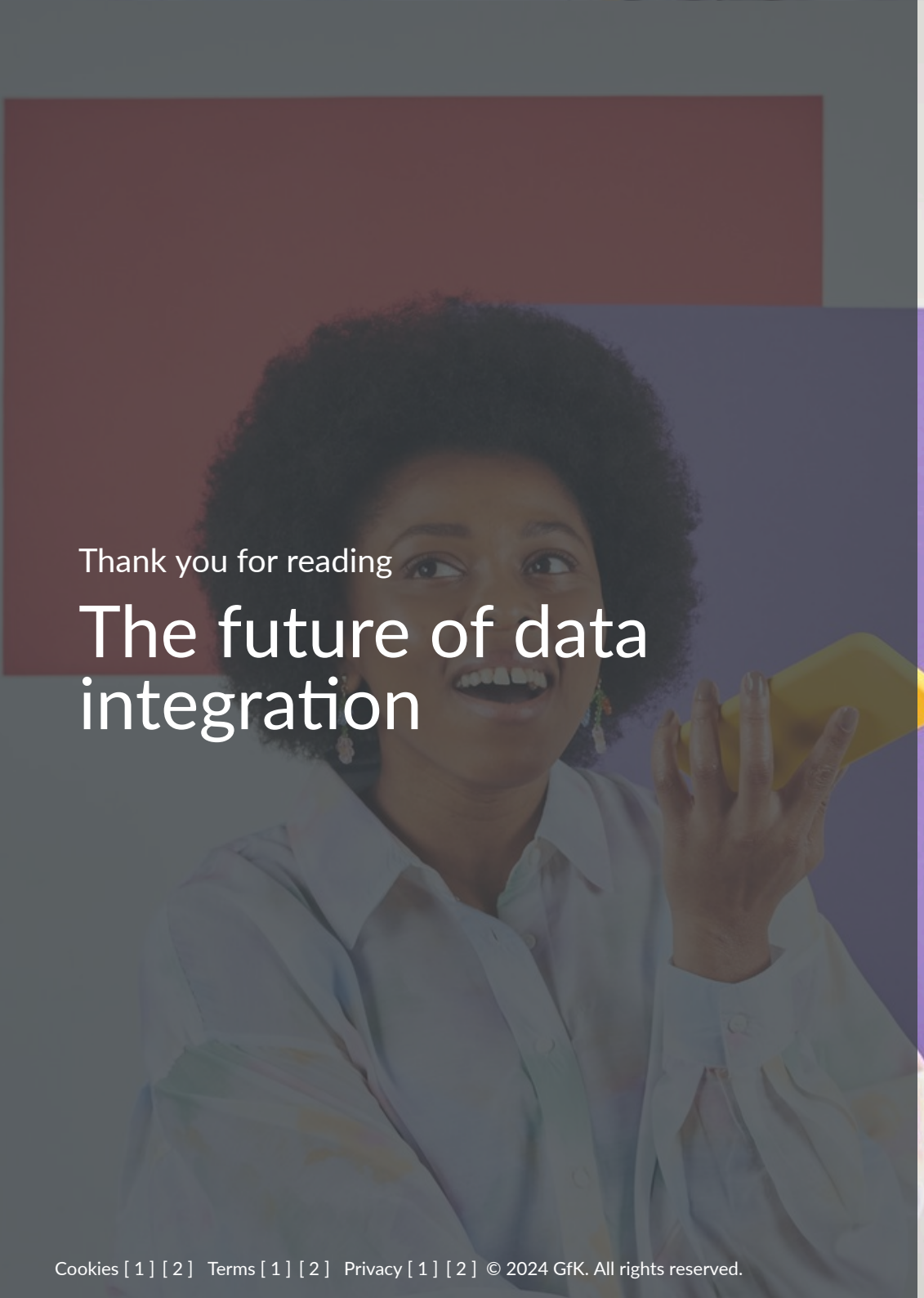
Ian Dowds, UKOM

If we look to the future of hybrid measurement, it's also becoming clear that, regardless of our industry's own desire to provide transparent and consistent measurement solutions, increasingly this will become a legal imperative as well. Legislative interest in audience measurement can be dated back as far as US Congress hearings in the 1960s which led to the formation of the Media Ratings Council (MRC). Right now, the EU is legislating on a number of aspects directly impacting audience measurement, including transparency, open access to data sets and consistency around metrics and standards:

"I'd like to draw your attention on a project from the European Union.
EMFA, the European Media Freedom Act, which actually states its principles for media measurement, and two of those principles are transparency and verifiability. So it is a big question: it is probably not ever as transparent as we would like to have it, but we need some professionals that perform audits to increase transparency."

Robert Schaeffner, RTL

It remains to be seen how this legislation will be applied and enforced in practice, but it does underline the vital economic impact of audience measurement and the current high profile it has across the media eco-system. Whatever the future holds, the efforts of industry experts like those at our round table are as important as ever.



Thank you for reading

The future of data integration

Cookies [1] [2] Terms [1] [2] Privacy [1] [2] © 2024 GfK. All rights reserved.

POWERED BY Turtl