Ipsos MediaCT
The Media, Content and Technology Research Specialists

DATA FUSION
A White Paper by Ipsos MediaCT

2011

Written by:
Trevor Sharot
Consultant Statistician, Ipsos MediaCT

Trevor has been in media research for 35 years, formerly with AGB, TNS and Nielsen. Now consulting as Red Research, he is engaged full time with Ipsos MediaCT, where he has most recently developed the modelling for POSTAR. His paper on Fusion in IJMR in 2007 was highly commended for the MRS David Winton award for innovation in research methodology.

Content

1. **Introduction**

What do the following have in common:
rock and roll, the Sun, a slipped disc and chicken tikka masala?

The answer is of course 'fusion', an adaptable word that conveys quite different meanings to a musician, a physicist, an orthopaedic surgeon and the man on the Boris bike. But, playing rock, nuclear fusion, back surgery and fusion cookery share one similarity with data fusion: none of them is easy to do well!

Data fusion denotes merging two or more surveys by matching up respondents on common variables or 'hooks', such as demographics, lifestyle, and media and product preferences. It has proven most useful for combining media exposure with product consumption data, providing valuable insights for media planning.

The main attraction of fusion is to avoid the commercial and technical difficulties in capturing all the data of interest in a single-source survey.

Fusion has a long and chequered history, which provides the structure for this paper: how it started, how it stumbled (but recovered), leading on to the present, and its future potential. On this clothes horse will be hung a glimpse of fusion's technical undergarments, discussion of the importance of hooks, fusion's changing fashions and other material aspects...

2. **Fusion or Delusion?**

One thing is certain: on a global scale, fusion as a tool has not been fully exploited. We put forward several reasons for this – indeed, it is the presence of so many obstacles to creating a fusion-based service that seems to scupper most attempts. Yet it is a powerful, and now well understood, technique that we think deserves wider application.

3. **In the beginning**
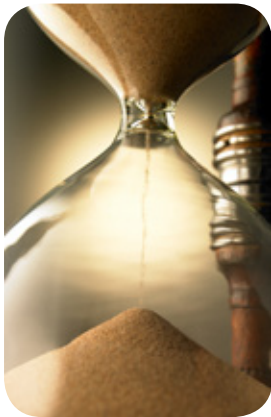
The term data fusion originated far outside market research, in the fields of defence and subsequently medical diagnosis. In both cases, data from multiple sensors are combined to create a single indicator. The military learned how to combine data from dispersed anti-missile radar sites for target detection and tracking, while medics adapted and developed these methods to combine their data clusters, such as from an array of ECG sensors. Googling 'data fusion' leads to a vast literature on these areas dating back to the 1960s. Now, data fusion permeates many diverse activities, including finance, traffic control, law enforcement and robotics.

Its adoption into market research is not well documented, but one of the earliest developers was Friedrich Wendt at AG.MA (Arbeitsgemeinschaft Media-Analyse: Association for Media Analysis) who was testing methods by 1976, and their partners at KONPRESS religious press and in The Netherlands. Wendt (1983) uses many terms that have since become essential terminology: single-source, donor and recipient surveys, common and specific variables, distance, marrying individuals. His context is also topical today: media planning, and his vision was presciently like the current IPA TouchPoints service:

'Our idea is to do all surveys in such a manner that they may serve as donors. We envisage the construction of a basic databank with more or less socio-demographic characteristics, and thus to have at our disposal the potential of a databank from which we may carry out combinations of surveys, combinations of partial vectors, to meet the particular needs of media planning and other purposes.'

$4.$ **Early methodology**

Wendt records using linear programming to perform fusion, perhaps also as a result of its roots in military research. A French school, led by Gilles Santini, explored other methods, emphasising group matching rather than at the individual level. Though valuable as explorations and for widening interest in fusion, use of these methodologies has lapsed. (See Santini, 1986).

An English school emerged next, with the Market Research Development Fund sponsoring an experiment in which one half of the Target Group Index (TGI) was fused onto the other. This validation process, known as folding, allows the predicted media selectivity indices to be directly compared to the actual indices. It taught the community a number of valuable lessons for creating a data fusion:

■ pay attention to details such as ensuring the two samples are using common question wording and code frames;

■ the importance of the choice of matching variables (hooks);

■ ways to measure (and not to measure) the validity of a fusion, particular using regression to the mean: the dilution of selectivity indices due to poor hooks or matching;

■ be aware of the many pitfalls that lie in wait!

Perhaps most important, the methodology employed for that study has survived largely intact and is still most commonly in use. No doubt this is due largely to its (relative) simplicity, easily expressed without recourse to mathematics:

Find the best suited couple and marry them, then the next best, and so on. If there are too many maidens or bachelors, allow limited polygamy or polyandry...

In fact, some polygamy or polyandry is usual, because the donor and recipient sample sizes are never the same. If the smaller sample forms the donors, it is inevitable that some donors will be married to more than one recipient. If the donor survey has the larger sample, polygamy is not necessary but may still be desirable to avoid unsuitable marriages.

The appeal of this method has not stopped many other variations being tried. One that must be mentioned is the transportation algorithm (borrowed, like linear programming, from operations research). In contrast to the above stepwise procedure, this provides a globally optimum solution (something like a North Korean mass wedding). Its first major commercial application was the Panorama service developed in the nineties by Nielsen New Zealand and subsequently applied also in Australia. Its major advantage forms the subject of the next section.

5. **Preserving the currency**

On September 19, 1931, the UK left the Gold Standard for good. The pound would no longer be worth its weight in gold. (It probably never was, though possibly 240 silver pennies, 'sterlings', originally weighed a pound.) But gold standards remain important in the media marketplace, allowing buyers and sellers to trade on agreed audience estimates. Several consortia – including BARB, NRS and RAJAR – have become established to regulate and administer this state of affairs.

One disadvantage of the usual 'shotgun marriage' approach is that it does not preserve both currencies, only the recipient data. Donated data become separated from the donors' survey weights and some of their demographics, and no longer add to the original audience sizes.

For this reason, the Target Group Ratings are created by fusing the TGI onto BARB rather than the other way round (though as we shall see later, Bob Hulks' decision in this regard has proven to be sound on a technical level too). However, the transportation algorithm automatically preserves both currencies, at least at total population level and for significant breakdowns too; for this reason, it is sometimes called constrained fusion.

6. **Current services**

To list all past and current fusions is not possible since several research companies conduct proprietary exercises for their clients. However, fusions of syndicated measurement vehicles include – in the UK - Target Group Ratings and TouchPoints already mentioned, plus the NRS/Nielsen fusion currently under development. Also Nielsen fuse TV viewing data with online usage (Netview), readership (MRI), multimedia (Media Index), purchase data (Homescan) and with other services, both in the USA and elsewhere. In Canada, ComScore has been fused with PMB (Print Measurement Bureau) data.

Fusions conducted by Ipsos MORI include:

■ Europe 2003 Print & TV Surveys

■ Hungary: National Media Analysis (print) and Internet Audience Research

■ The Economist Subscriber Study.

This list, though not comprehensive, seems substantial. However, it falls far short of fusion's true potential. Market penetration is low and efforts to establish fusions in other major markets such as South Africa have not been successful. Where fusion is conducted, it is very much led by suppliers rather than driven by demand. There are few case studies published or presented at international conferences.

Discussion of the possible reasons for this will form the bridge in this paper from the past to the future.

7. **Fear of the known, fear of the unknown**

The list of reasons we think have stood in the way of widespread application of fusion includes:

Technical

- There is no dominant methodology, like say k-means for cluster analysis. Indeed, Roland Soong – formerly with KMR - gives examples that suggest that there can be no single best method, and he is probably right (Soong, 2004).

- There is no out-of-the-box algorithm for any methodology. Much care is always required in preparing the datasets, choosing the hooks, deciding their weights, acceptable levels of polygamy, performing folding or other validation etc.

- Fusion has technical limitations: it only preserves that part of the relationship between the unique donor variables and the unique recipient variables that are accounted for by the common matching variables (hooks). For some combinations of variables, this may not be a large part.

The previous point means that every effort should be made to find the best possible hooks and to perform validation to be sure that, even then, the best is not the enemy of the good.

Commercial

- By definition, a fusion requires purveyors in different fields to come together and agree on the need for the project, the objectives, the methodology and funding.

- Some of the methodologies are difficult to describe for the layman, so appear to be black boxes.

- Fusions are not automatically successful. This leads to uncertainty among buyers and the potential need for independent assessment.

- Rival methodologies that claim to offer the same benefits with less complication and cost have emerged.

8. **The future for fusion**

We think these concerns have been overstated. The methodology has come a long way since the early days. The metrics of validity are now well-known, particularly regression to the mean. The means to quantify them are in place, such as folding and limited single-source comparative data. A long-running debate as to whether it is better to fuse 'small samples onto large' or 'large onto small' was recently resolved by a paper in IJMR by Trevor Sharot, Consultant Statistician to Ipsos MORI, that allows the margins of error for any potential design to be calculated in advance. Among other results, this paper established that it is always better to fuse the larger survey onto the smaller one, or to use the transportation algorithm rather than sequential matching (Sharot, 2007).

In short, while constructing a fusion remains hard work, it is no more fraught than any other form of data-modelling. In these times of media convergence, with a wide and growing range of devices and channels being offered to the public and a growing need to see the whole picture, fusion's time has surely come.

References

SANTINI, Gilles (1986) *Méthodes de fusion : nouvelles réflexions, nouvelles expériences, nouveaux enseignements*, Les médias, expériences et recherches, Séminaire de l'I.R.E.P.

SHAROT, Trevor (2007) *The design and precision of data-fusion studies.* International Journal of Market Research, 49,4, pp.449-470.

SOONG, Roland and DE MONTIGNY, Michelle (2004), *No Free Lunch In Data Fusion / Integration*, WAM 2004 and http://www.zonalatina.com/WAM2004.pdf

WENDT, Friedrich (1983) *The AG.MA model.* Proceedings of the Montreal Readership Symposium.

FURTHER INFORMATION

For more information please contact:

Trevor Sharot
e: trevor.sharot@ipsos.com

Ipsos MORI
Kings House
Kymberley Road
Harrow HA1 1PT

t: +44 (0)20 8861 8000
f: +44 (0)20 8861 5515

www.ipsos-mori.com

ABOUT IPSOS MEDIACT

Ipsos MediaCT plays a prominent role within media and communications research, holding key industry audience measurement contracts and conducting bespoke research to assist their clients in informing their strategic decisions. We help clients make connections in the digital age, as leaders in providing research solutions for companies in the fast-moving and rapidly converging worlds of media, content, telecom and technology. Using a wide variety of research techniques, we help individual media owners, content owner studios, technology companies, agencies and advertisers address issues such as editorial and programming, advertising, audience profiling and music tastes, market positioning, piracy, high definition and theatrical markets, new product and programme development and license applications.

Winner - 2010 MRG Awards: Best Research Team - Research Supplier