SINGLE SOURCE PASSIVE AND CLAIMED DATA: THE TGI EXPERIENCE

David Greene, TGI, Kantar Media

Introduction

TGI Clickstream was launched in June 2012 after two years in development. It is a complex product, which represented at the time, a significant departure for TGI, in that it brought a long and traditional self-completion survey into the world of digital. TGI Clickstream combined online browsing data, which was obtained via a meter placed onto the respondents' computer, with the recall data from the standard 25000 sample TGI media and product consumption survey. At its core is a 4000 strong, single source sample of respondents, who had completed the TGI survey and downloaded a meter. The result was a product that combined the depth of the standard TGI survey with detailed, web browsing behaviour across approximately one thousand websites and subdomains.

This paper looks at the background to the development of TGI Clickstream, how the product was put together, a comparison of clickstream results with recall data and how it has helped clients. The paper goes on to assess how TGI Clickstream will adapt in the future as a result of the changing market.

Background

The demand for TGI Clickstream became apparent some three years ago. Its predecessor, TGI Net, had served the market well for many years but was losing impact in a fast moving digital environment. TGI Net consisted of fusing the main TGI survey with a supplementary, shorter survey, which included questions about general online browsing behaviour as well as the recall of visiting some three hundred websites. The shortcomings of TGI Net included:

- Respondents were visiting, in many instances, a wide range of websites, thus tracking smaller sites was difficult. Moreover, websites are increasingly reached through non direct means e.g. via social media reducing, we suspect, the likelihood of respondents remembering all sites visited.

- TGI Net was limited to the analyses of no more than 300 websites as it would have been impractical to list any more on the questionnaire.

- The number of questions, which could be asked about each website, was limited. The meter had a lot more to offer, in terms of the metrics that were available including page views, time on site, subdomain visits etc.

Given that print media tracking was a mainstay of the TGI survey, it was important that it stayed relevant to clients' needs. If clients were embracing the digital revolution, it was logical for TGI to come up with a product which could help them take advantage of the new revenue opportunities that were available. Furthermore, we believed that a product such as TGI Clickstream could open up a range of potential new clients for TGI, notably pure-play Internet companies. It was our view that TGI Clickstream could offer advantages to both existing and new clients, not least the ability to link brands with online behaviour and also to plan and measure digital in the context of other media, including print.

TGI Clickstream Design

The design of TGI Clickstream is best illustrated by the diagram 1. In short, the TGI survey consisted of a 25000 representative sample, of which approximately 17,700 (71%) claimed to use the Internet. Some 4000 of these respondents undertook an additional twenty minute online survey about Internet usage (TGI part 4) and installed the meter that tracked their web browsing behaviour. This 4000 sample was then fused with the remainder of the online sample and the non- Internet population was added in to bring the total base back to the 25000.

Diagram 1
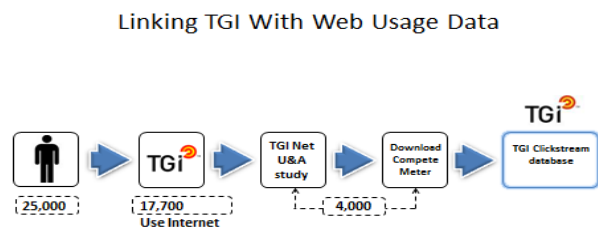
In designing the product, there were a number of issues to consider:

- Incentivising respondents to take the meter without negatively impacting response rates for the main TGI survey

- Establishing a process to handle the data coming from the meter, a data format of which TGI possessed no previous experience.

- Finding the optimum way of combining TGI claimed data with passive online meter data. Moreover, structuring the final product in such a way that clients would find it easy to use on the analysis software.

Meter Recruitment

The meter was operated by Compete, a Kantar Media company. It was a proprietary browser plug-in which acted as an intermediary between the web servers and the respondent browser, sending data to the Compete servers as the respondent browsed the Internet.

The meter was designed to collect information only from the computer account within which it was installed. Around 60% of the sample was identified as either the sole-user of the metered computer, or as having an individual log-in to that computer. For the remainder of the sample, Compete algorithmically associated user sessions to individuals in the household, using online behaviour patterns from each session. Therefore, each time a respondent opened the browser they did not have to identify themselves. Respondent identification was trialed for a while, but was found to have a negative impact on retention.

Recruitment for the meter took two forms. Firstly, Compete in conjunction with Lightspeed, (the Kantar online panel provider), had already established a panel of 10,000 meter enabled panelists in the UK. The TGI questionnaire was placed with these panelists delivering 75% of the final clickstream sample. At three hours, the TGI questionnaire would normally be too long to place with an online panel. However, because the survey had been divided into three parts, this made it more acceptable for potential respondents. In the event, the target set for Lightspeed of 3000 (out of 4000) was comfortably achieved.

The Lightspeed sample was supplemented by additional metered respondents using a random recruitment method. Respondents were initially recruited via the face to face Omnibus to complete the full TGI survey online and were subsequently, sent an email or letter inviting them to undertake an additional internet survey, as well as download the meter. A brochure and website were produced to describe the survey and ease any data privacy concerns. If respondents were interested in pursuing the project further, they were directed towards a log-in website.

At the log-in stage, respondents' browsers were checked to ensure they were compatible with the meter. Specifically, respondents had to use Internet Explorer, Google Chrome, or Firefox. Unfortunately, Apple's Safari was not compatible with the meter, but from TGI research, the three accepted browsers represented some 80% of all browsing behaviour.

In terms of incentives, for completing the two tasks of answering the additional online survey and downloading the meter, respondents were paid £40. The recruitment option of using the Omnibus provided a further 1000 respondents and the remaining 25% of the TGI Clickstream sample.

Meter data were collected for at least one month following activation, but a monthly prize draw was offered to encourage respondents to keep the meter on for a longer period. In general, the meter download process worked well. That said, due to the variety of computers owned by TGI respondents, or the various software packages which they had already installed, or simply that some respondents were not particularly knowledgeable about using their computers, the download process was not straightforward for everybody and problems would arise from time to time. On occasion, the meter downloaded but it would not install, or it might install but would not become active, or it might become inactive before the month was out and so on. Therefore, a help desk staffed by people with a reasonably high level of knowledge of the process was maintained.

As far as response rates are concerned, of all those respondents who completed TGI online, approximately 20% completed the U and A survey and downloaded the meter. To qualify for the incentive, it was essential to complete both tasks. Various

initiatives have been introduced to increase this figure including; variable incentives, targeted recruitment and improved marketing materials.

Data Privacy

Given the breadth of data which the meter was capable of collecting, protecting respondent privacy was of paramount importance. TGI ensured that respondents were clear from the outset the type of project in which they would be participating. From the introduction to the survey, right through to the detailed terms and conditions, the respondents were not in any doubt what they were agreeing to, and felt within their rights to terminate the process at any time. Respondents had to comply with the Kantar Media privacy policy and the Compete privacy policy before commencing the download.

It was stated that the measurement of web browsing would happen automatically without respondents having to do anything on a regular basis. Furthermore, the meter would not affect the speed or any other aspects of the browsing experience.

Data Preparation

The first task was to establish the top 1000 websites from Compete and remove any that were deemed not suitable. Sub domains were also identified, principally of the largest websites, but also of some smaller sites, especially when they could be of interest to potential clients. Categorising subdomains was not always straightforward and the rules varied from site to site. Yahoo News for example, was classified as uk.news.yahoo.com, whereas the news from the Guardian could fall under guardian.co.uk/uk or guardian.co.uk/world. So each subdomain had to be carefully scrutinised to ensure that the correct information was being gathered.

It is worth drawing a distinction between the raw data as collected by the meter, and the aggregated data supplied to TGI. The raw data, which consisted of thousands of page records for each respondent, was aggregated to respondent- level metrics including number of visits, time on site, page views etc.

Meter data was a new experience for TGI. The size and complexity of it meant that careful thought had to be given in terms of how it might be handled. The explosion of so called Big Data is putting pressure on many companies to assess how all this data can be stored and importantly, processed in a meaningful way.

On a more limited scale, this was a Big Data moment for TGI in terms of how such a large amount of data would be processed. The route taken was for the meter data to be first uploaded to the Compete servers, then Compete would aggregate the data into twenty eight day packages for each respondent. Even with this level of aggregation, the file from Compete was still exceptionally large at some 30 million cells. Furthermore, the data was entirely numeric and not the usual type of data to which TGI was accustomed.

The data files were processed in the TGI proprietary software, PTK. The software was required to undertake a number of tasks including create the original online U and A survey, and combine the survey data and meter data together. PTK also allowed the creation of additional metrics such as page views per visit and time spent per visit, as well as managing averages and ranges. For example, time on site was banded into percentile groups.

Websites were grouped into the Compete taxonomy consisting of 120 classifications across 24 broad categories. To avoid having to create multiple groupings for different clients, TGI developed a tagging function which facilitated the grouping of websites using key words e.g. sport, travel, or even by company.

In terms of the TGI Clickstream database, it was created by fusing the Internet Usage and Attitude survey and the Compete meter data (4000 respondents) onto the Internet users who did not complete the U and A survey or install the meter (13700 respondents). The fusion operated by matching respondents at an individual level, and then passing all the data from a donor respondent (U and A Survey / meter) to a matching recipient respondent (TGI internet user but no U and A survey or meter). Donor and recipient respondents were paired to match exactly on gender, region and fieldwork period. Additional hooks centred on Internet usage and media consumption.

METER COVERAGE

CLICKSTREAM Metrics
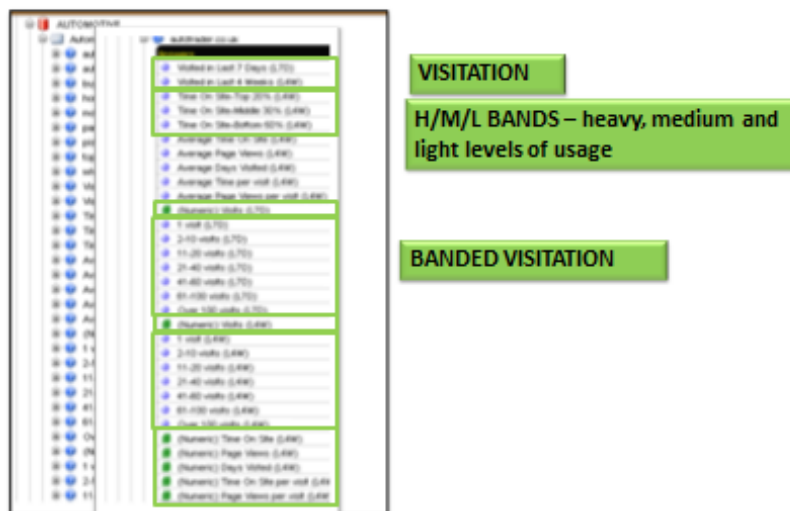
Data Comparisons

Given that Clickstream was a new type of survey for TGI, detailed quality checks were put in place. This included demographics checks to establish the profile of the metered sample. In addition, the results from the meter were assessed against the results from its predecessor, TGI Net. The meter measured online browsing behaviour only from the main computer used at home. That is, it excluded browsing behaviour from any other computer (or indeed any other device which was capable of surfing the web), browsing at work, or browsing on the go through a smartphone or tablet. However, in order to establish a total browsing figure for at least 300 websites, the level of recall of visiting these sites (for the last 7 days and last 4 weeks) from any device other than the main computer was asked on the questionnaire. The meter data was combined with the questionnaire recall data to produce a total website reach figure. When comparing meter data or meter data plus recall data against TGI Net (all recall) in terms of last 4 weeks, there were some differences, but by and large these were acceptable.

	Visited in Last 4 weeks (%)		
	Meter Data	Meter data + Part 4 (recall)	TGI Net (recall)
BBC	44.8	51.9	52.5
Guardian	15.5	18.7	12.5
Trip Advisor	9.0	11.6	9.8
AOL	7.8	10.9	9.5
Youtube	50.2	59.1	51.2

An important area to monitor were the occasions companies would change the names of their websites and subdomains, as it would have a significant effect on recorded trends. Such changes tended to be well publicised in the media in advance, or alternatively, they were spotted in data checks by Compete or TGI. Compete would also send TGI a monthly report that identified significant trends. For example, the TV station Channel 5 rebranded from Five.tv to the current Channel5.com. This led to a sharp fall in recorded views for the former and results of the two had to be combined to provide an accurate reading.

Client Reaction

TGI Clickstream has been well received by the market and was seen as an improvement on the previous TGI Net survey. Clients have signed up from a range of sectors including traditional media companies, as well as pure play Internet companies.

TGI Clickstream was viewed as a useful tool to understand the in-depth profiles of a vast range of websites and subdomains. The central purpose of the survey was not to measure the penetration of websites in the market, although we certainly hoped to be as near the mark as possible. In essence, TGI Clickstream allowed clients to link the detailed metrics of the 1,000+ sites and subdomains to the product, media and attitudinal data in TGI. What brands are specific site users consuming? What motivates them? What are their attitudes and their interests? The product also allowed for cross-media comparisons, with companies keen to look at how their online brand complemented their more traditional offers and how the products worked in conjunction with each other.

One daily newspaper publisher referred to TGI Clickstream as its 'go to survey' instead of the standard TGI survey, in that it covered increasingly important digital metrics combined with an offline world. The database allowed the client to identify targets to substantiate sales arguments based on online properties and traditional paper formats, in isolation and in combination. The data proved useful for both the insight and sales teams within the client company.

Base: All Internet Users with a Volkswagen in household

Source: GB TGI Clickstream 2013 R3

Especially pleasing was the fact that pure-play Internet companies have seen potential value in the product. One website, for example, identified that Clickstream could assist it in understanding the profile of, not only who was visiting its website as a whole, but who was using each section of the website and thus build a sales story around the profile of each section for potential

advertisers. This was something that the company had never been able to do before, and it gave the sales teams compelling sales arguments in each category, as well as the ability to sell against a much wider range of 1,000 sites and subdomains.

Another significant application was the ability to look at more detailed online metrics such as time spent on the website in an average month. Whilst most companies will have access to site analytics sources measuring session dwell time, TGI Clickstream allowed a client website to look at the average time on sites across many sites and subdomains and analyse that against the whole of TGI. This meant that a website could establish whether a particularly valuable target audience (such as high earning groups for example) spent longer on the site than the average visitor or visitors to a competitive site.

Future Developments

TGI Clickstream has continued to develop since its launch, and improvements over the last 12 months have included adding more websites and subdomains, custom categorisations, adding new website metrics, identifying which sites carry banner advertising and introducing a ranking tool. Establishing which sites have been purchased from is also being considered.

In addition, TGI intends to increase the meter sample size over the next two years, and thereby deliver more robust sample sizes across all websites. Not only will this allow detailed analysis of domains and subdomains, but also bring more websites above the sample threshold for analysis. The larger sample size overall will also facilitate the increase in the number of websites featured in the database.

Tracking online behaviour on the PC is all well and good, but Internet browsing via a PC represents a decreasing proportion of total browsing behaviour, as mobile devices such as tablets and smartphones have come to the fore. The slump in PC sales recorded by the major PC manufacturers has been dramatic, and this has only served to add to the urgency for TGI Clickstream to move to the next stage of development. Therefore, the meter monitoring programme will be expanded into mobile devices as soon as is practicable.

TGI Clickstream in Other Countries

The launch of TGI Clickstream in GB also sparked interest in other TGI countries, in particular France and Turkey.

France

In France, the first Clickstream release is scheduled for mid-November 2013. Here, a slightly different approach was taken in terms of meter recruitment.

By way of background, until the end of 2012, France TGI was a paper survey partly sourced through the TNS Sofres offline panel 'Metascope' (30% of the total sample) and partly through ad hoc recruited fresh sample (70% of the total). In 2013, the online methodology was introduced resulting in 1500 online sample (out of a 7500 total) for the first wave. This was achieved using two recruitment methods:-

- Ad-hoc telephone recruitment with subsequent web questionnaire and an invitation to download the Compete meter – total 380 respondents.

- The existing Lightspeed online panel which, similar to GB, represented the best opportunity to achieve the online methodology introduction, as well as the passive Internet measurement – total 1250 respondents.

In terms of the clickstream sample, there was a third component:-

- Re-contacting all paper respondents who were identified in advance as Internet users

An invitation to download the meter was sent out to 4200 qualifying respondents, who had already completed a paper TGI. They were directed towards a simple, clean, website to download the meter. The response using this methodology was 26%, which represented all the respondents who had successfully downloaded the meter and remained active for at least 28 days. This was a particularly good result, compared to the experience of recruiting paper respondents in GB.

The first release of clickstream in France will contain 2500 metered sample out of 12000 TGI Internet users, and the aim is to increase this to 4000 in the following release in 2014. The release cycle will be twice a year.

Turkey

Turkey currently has a successful TGI Net product, but in the light of strong interest in passive Internet measurement, the team has plans to launch a TGI clickstream product in 2014.

The approach is a simplified one compared to GB or France. Given that most interest in a meter solution has been shown by existing TGI Net clients, the preference was to add the passive Internet measurement to the TGI Net sample. A pilot is scheduled

to take place in November 2013 and the live launch in January 2014. The initial clickstream sample for the first year will be 2000 (as with the TGI Net) with a view to increasing it in future. The release cycle will be four times per year with a rolling sample of 500 clickstream respondents per wave, which will be fused with the latest release of the main TGI product.

There has also been interest in TGI Clickstream from other countries including China, Poland and Brazil. We are confident that interest will continue to grow amongst TGI Partners globally and this will help TGI broaden its experience in terms of methodologies and processes, as it respond to the different market requirements.

Conclusion

For some time, there was a strong feeling within the business that TGI needed to take advantage of meter technology and build a new database focussed on website profiling. The existing TGI Net website data was limited, and clients were demanding more detailed information, well beyond anything a standard survey could provide. However, there were clearly a number of potential hurdles to overcome. Would a sufficient number of people complete TGI (not an insubstantial task on its own), plus an additional U and A Internet survey and download a meter within the space of 12 months? Could data privacy concerns be overcome? How would this new data be processed and integrated within the existing TGI database? Lastly, would the meter data be robust enough?

It turned out to be a long journey, albeit one that has gone reasonably well. There have, of course, been problems along the way, but these have been overcome. The result is a product, which we believe, addresses client needs on website profiling. That said, the world has moved on very quickly since the first TGI Clickstream launch and there is more to be done. Demands are getting more urgent for bigger sample sizes to facilitate more detailed analyses, especially of subdomains, and to understand mobile browsing. However, TGI Clickstream now has a firm foundation from which to move to the next stages of development.

Appendix

TGI

TGI is a syndicated product and media consumption survey which operates in 65 countries around the world, through a network of wholly owned studies, licensees and marketing agreements. TGI is part of Kantar Media and sits alongside the TV audience measurement, news monitoring and advertising monitoring divisions. Kantar Media belongs to the Kantar Group which, in turn, is part of WPP. In GB TGI has been operating since 1969 and is currently a 25000 sample survey, representative of the GB population. The sample is partly recruited on the back of the group's face to face Omnibus and partly recruited via Kantar Group's own Lightspeed panel.

Compete

Kantar Media Compete, a sister company of TGI, delivers digital intelligence that helps the world's top brands improve their marketing based on the online behaviour of millions of consumers.

Compete's products and services are fueled by the largest integrated online consumer behaviour panel in the industry. Digital insights are interpreted by analytical experts in the automotive, financial services, media, mobile, online, retail, telecom and travel markets to deliver data and recommendations to create effective online marketing strategies and online experiences.