

FACING THE CHALLENGE - MEASURING THE COMBINED AUDIENCE OF MULTI PLATFORM TITLES

Jim Ford and Tamás Perjés, Ipsos MediaCT

The Objective

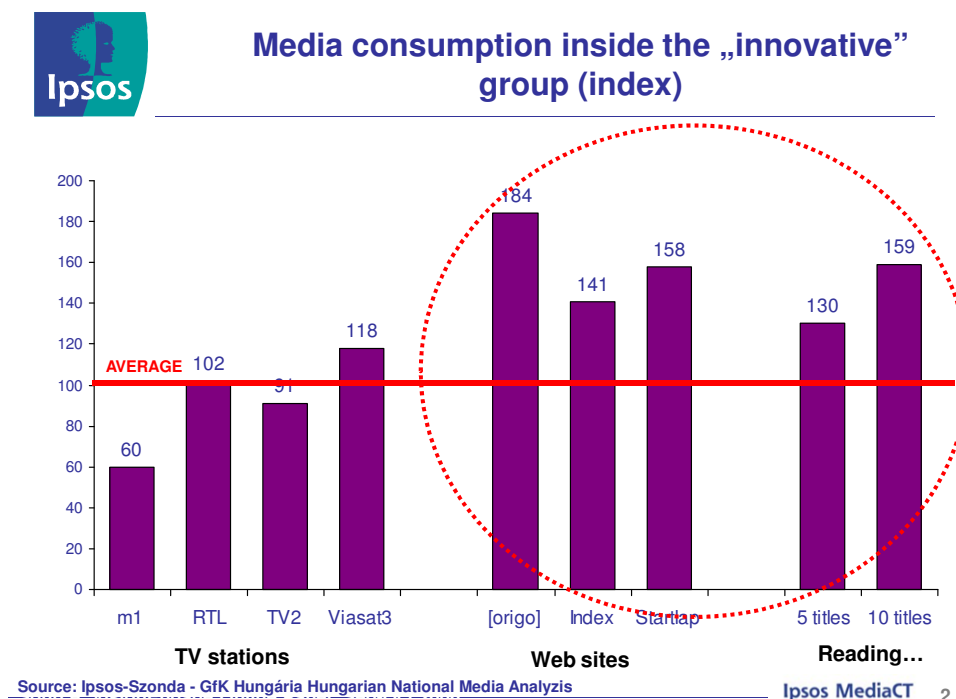
The crucial issue in media research for the more traditional media (TV, Radio and Press), is the impact of online. In particular, for the Print market, it is measuring the effect of online on its readers. Are Print titles being “cannibalised” by their own digital offerings, or do these platforms offer a complimentary service to the printed title? Does online provide a new readership for print titles?

This example from Hungary attempts to shed light on the issue. Although Hungary is a relatively small market compared to Western Europe and North America, it has a diverse media marketplace, mirroring all the key issues facing media owners, planners and buyers globally.

For example, in line with global trends, the share of online in total advertising expenditure in Hungary has grown rapidly in recent years. In 2003 it was 1%, whilst 2009 estimates show that the online share is now close to 10%.

At the same time the corresponding share for print has decreased from 41% in 2003 to 2009 estimates of 35%. It is too simplistic to explain the changes as simply as a response to the belief that print readers are simply swapping directly to online content.

In an attempt to explain the dynamics behind the shifts in media consumption across all media in Hungary, GfK created a cluster based solution to examine media consumption in a fragmenting marketplace. They created a 6 cluster attitudinal based solution in the Ipsos/GfK NMA data. One of the solutions was a group of “innovators”, defined by liking to try new things, searching for new information and wanting to purchase new products before everyone else. The findings of this analysis show that there is not this simple shift from readership to online, rather they identified an “innovative” and sizeable segment of media consumers in Hungary who are voracious consumers of all media. Importantly they identified that heavy online users are associated with intensive offline readership. As the chart demonstrates these innovators are consumers of all media and are readers of a large variety of print titles.



The need to be online is an accepted strategy for all the traditional print companies in Hungary, the reasons are clear:

- It is another competitive distribution method to deliver their content
- It maximises their total reach of their readership / audience

The Hungarian Marketplace

Print titles are measured by the National Media Analysis (NMA). This project has been conducted by Ipsos and GfK since 1995. It uses a face to face CAPI methodology, capturing the recent reading of 32,000 respondents per year. The study covers c.170 titles and from 2000, the NMA covered website usage as well, with 30 sites covered within the survey. This allows a “net offline / online” measure of usage for key print titles and their websites.

Running parallel to this study, the Internet Audience Research study measured the online attitudes to c.100 websites. The data collection method was also face to face CAPI and captured the web site usage of 3,000 respondents per month.

To improve the data available in the market, these two databases were fused together. This meant that the online measure was broader and more robust, however the data was being captured with an offline methodology.

In 2005 an online audience measurement project that utilised both site centric and user centric web analytics was launched in Hungary. In their 2007 paper at the 2007 WRRS Scott McDonald and James Collins discuss these approaches in detail. The Hungarian project was a joint venture between Ipsos and their technology partner Gemius. The measurement service now covers 300 websites, amongst them are the online versions of the key Hungarian media companies. The approach uses a site centric panel of 55,000 panellists and a user centric panel of 5,000. Both are continuous reporting.

The Challenge

The request for the delivery of a measure of both online and offline versions of print media was made by the Hungarian media market a number of years ago. The market did stipulate that any “measure” must maintain the gold standards of the TWO existing offline and online measures. So a fusion was proposed as this was the only way of maintaining the individual quality and effectiveness of the survey designs for both offline and online Print measures. There had been international precedents. Most notably in the US where Nielsen’s Netratings Netview was fused with the National Television Index (also Nielsen).

The pilot for the Hungarian fusion began 2 years ago. Ipsos and Gemius appointed independent academic mathematicians. The aim of the pilot was to tackle a number of crucial areas:

- How to define readership? In particular, developing an online readership measure from behavioural rather than survey data.
- How to combine the offline and online readers?

So in the process we defined the two datasets:

- The offline readership used a six month sample, with readership defined by the Recent Reading method.
- The online database was a completely different structure, containing valuable dynamic information such as “time spent visiting, time of visits, downloading etc”

The test for the pilot was to define from this behavioural database “online readers” that could be used as a sample to merge with the offline readership data. So the initial steps were to:

1. Identify panellists who were daily visitors to websites.
2. Create a probability that would provide an “average monthly visit score” for each online panellist. This figure was between 0 and 1, so that a figure of 0.14 meant that the probability for a panellist to visit a site was 14%.
3. Step three was to convert the probability to visit scores to dichotomous variables (visit or not visit). This was done by applying a special random statistical model which took into account the range of the daily probability to visit figures.
4. Finally we had to select and define the linking variables / fusion hooks (demographics, media consumption etc), from common variables in the offline readership survey and panellist information from the online panellists.

As in many cases, the necessity to run a pilot proved its worth. This was because the results were not satisfactory, principally because:

- the probability model was not prescriptive enough
- The fusion lost the “dynamic” element of the online data. When moving from a daily measure to a monthly assumption figure, the process loses valuable information such as page impression, time spent on a site, number of visits etc.

So at the end of year 1 after examining the findings of the pilot, we decided to start afresh, this time removing online and offline readership as fusion “hook” variables. This meant that the fusion between the 2 databases was done via the following common variables:

- Socio-demographic data
- Media consumption (e.g. weekly consumption of TV, Online, Print and Radio)
- Internet specific data (usage, access, frequency etc)
- Ownership of specific goods / usership of particular services (bank cards, games consoles, cars etc)

In total there were 45 variables used for the fusion proper, which this time allowed us to maintain:

- the offline readership currency
- the richness of the online audience data (page impressions, time spent etc)

The results of this fusion were then subjected to rigorous statistical examinations to demonstrate that it complied with acceptable statistical tolerances.

The Fusion

This time the fusion was developed by professors from the mathematical and statistical department of the Eötvös Lóránd University in Hungary. They defined the donor database as the online dataset of 60,000 panellists. The recipient database was 6 months of the offline Readership study of 16,000 respondents.

Stage one was to remove respondents on the offline readership measure who were not online, this left 50% of the recipient database.

The fusion model was based on the so called ‘constrained fusion’ approach. The units of the donor data base were “transported” to the recipient data base, noting that the distance between the paired units must be the smallest possible value.

The definition of the distance is based on a special main component analysis. During the fusion 45 hooks variables were selected, such as: gender, age, type of settlement, education, profession, lifestage, social status, size of the household, number of children, ownership of some selected goods, general media consumption – usage frequency of TV, radio, print, online access and usage. These were the key variables and were used during the fusion process.

The next step was to carry out a Principal Component Analysis (PCA) with these 45 variables. A Principal Component Analysis is similar to Factor Analysis, and was carried out to reduce the total number of possibly correlated variables into a smaller number of uncorrelated variables, or Principal Components.

This produced 5 Principal Components and for each unit we calculated the distance in this 5 dimension Euclidian space. The main challenge was the calculation and storage of the distance matrix.

In the end we had to reduce the number of components and with the PCA approach we ranked and scoped the most important variables. Three Principal Components were defined:

<u>Principal Components</u>
genre (male/female)
education (basic or skilled worker / graduation / degree)
type of settlement (Budapest / county capital / other towns / villages)

The next step was examining the online audience data. We created an average working day and an average week-end day data base. The key issue was that we maintained the 3 most relevant online consumption data dynamics:

- page impression
- time spent
- number of visit occasions

The fusion based on the 45 hook variables and the distance matrix (see above) meant that at the end of the process we got a merged data base with:

- the offline readership figure (yes/no)
- the online audience data with the 3 Principal Components

The end result was that we were able to produce a “fused dataset” that maintained the gold standard of both the original currencies in a robust format that can be interrogated by subscribers to the database.

The Validation

The model developed by the mathematicians was designed to be a “flexible model” which meant that each time the fusion process is run, we get slightly different results.

The greatest variance in the results came when examining Visitors (defined as someone who has at least one page impression on an average working day and/or an average week-end day). The difference being less than 2%, as the table below demonstrates:

Table 1. Number of visitors (N) in the fusion model

website	days	Fusion - run1	Fusion - run2	Difference
index	working-day	981 822	989 090	+1%
index	week-end	518 303	520 470	=
index	total	1 088 808	1 097 261	+1%
port	working-day	945 537	943 494	=
port	week-end	677 901	666 951	-2%
port	total	1 131 654	1 126 877	=
hvg	working-day	418 356	423 551	+1%
hvg	week-end	169 957	167 258	-2%
hvg	total	482 208	485 367	+1%
iwiw	working-day	2 579 166	2 587 626	=
iwiw	week-end	1 932 986	1 942 500	=
iwiw	total	2 667 602	2 676 052	=

When comparing the original online audience measurement figures with the merged data base, the results are still acceptable.

Table 2. Number of visitors (N) in the original and combined data base

website	days	combined data base	Original Online Audience data	relative error	absolute error
index	working day	1 006 365	988 437	+2%	17 928
index	week-end	529 934	592 664	-11%	- 62 730
index	total	1 112 814	1 076 138	+3%	36 676
port	working day	959 594	980 512	-2%	- 20 918
port	week-end	678 710	694 992	-2%	- 16 282
port	total	1 146 022	1 116 362	+3%	29 660
iwiw	working day	2 647 057	2 507 622	+6%	139 435
iwiw	week-end	1 988 526	2 027 371	-2%	- 38 845
iwiw	total	2 736 441	2 570 722	+6%	165 719

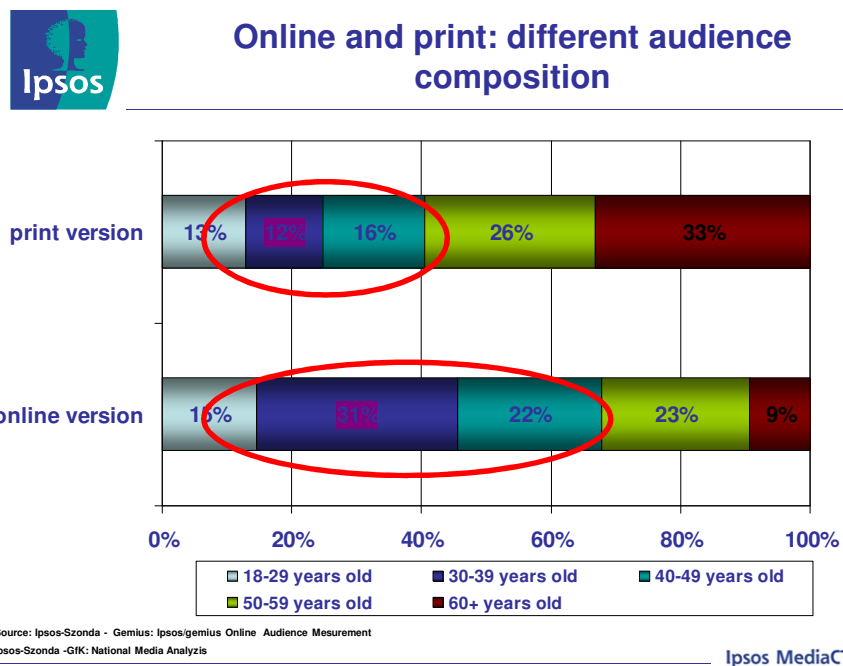
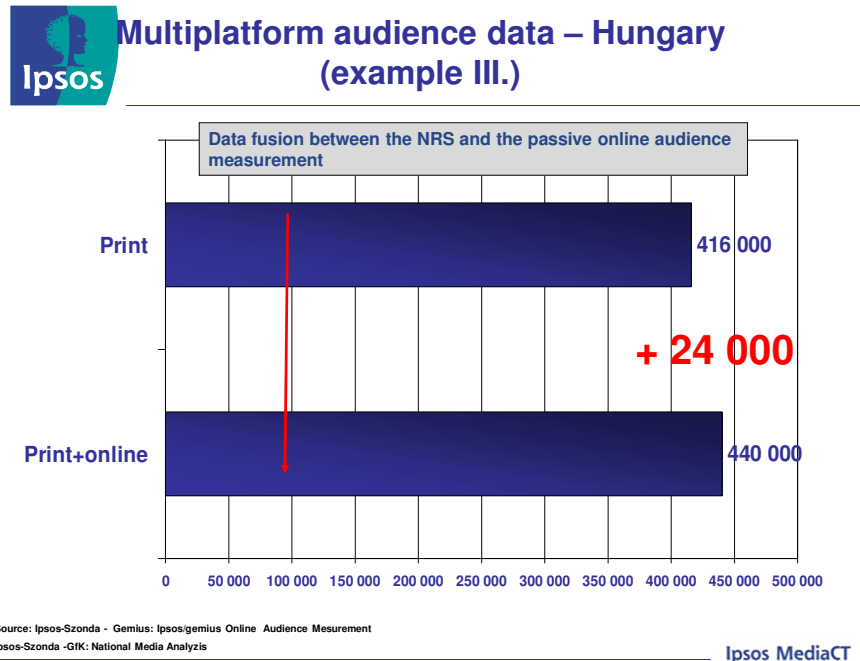
The Findings

The most important element of the fusion was that both gold standard currencies were preserved and the database was flexible enough to be used in day to day business planning. The market leader newspaper in Hungary, Népszabadság, was facing a number of marketing problems:

- It's offline readership showed signs of a steady decline over time
- The profile of these readers appeared to be becoming older in terms of age.

When they looked at the combined online and offline market penetration they were able to show:

- a 7% increase in "total" readership penetration
- The net online audience was younger, more urban and had a higher social status.



Conclusion

In conclusion, the approach we have adapted allows us to deliver the market expectations, namely the maintenance of the richness of the offline and online databases. We have been able to deliver a transparent, logical and functional fusion solution. The pilot allowed us to test one hypothesis and the findings of that work directed us to a more relevant solution and data that is now available to all the major players in the hungarian media marketplace (publishers, portals and media planning / buying agencies).

The next steps are to make the database more flexible, in particular to allow users to define their own website usership definition (e.g. visit in the last 7 days). The measure of success is that the fused database is now an essential part of the day to day planning in Hungary.