# MEASURING INTERNET AND PRESS AUDIENCE IN THE MEDIA CONVERGENCE ERA. IN SEARCH OF A NEW PARADIGM IN RESEARCHING CROSS-MEDIA

**Jaroslaw Pawlak and Malgorzata Póltorak, GEMIUS S.A.**

### Introduction

The evolution of media induced by the ongoing technological development necessitates a new approach and novel research tools. It forces researches to seek new paradigms and methodology to analyse and describe the changing reality. The Technopoly – to refer to the famous book by N. Postman (1993) – has ravaged the traditional media market. Influenced by contemporary technology, the consumption patterns are shifting. The internet has grown out of being just another means of communication to become a space where the traditional media are struggling to mark their presence. It has also become a catalyst of changes in the traditional media consumption patterns and led to remodelling their strive for consumer awareness. It is the development of new technology that made the traditional media face the challenge of redefining their role in the contemporary world, the way to get through to the customer or even compel a radical change in their offer and language. The transition is further fostered by the economic crisis, drop in sales of press and reduction of advertising spending, which hit the press in a particularly painful way. Owners of all sorts of media adopt a range of varied strategies to cope with the expansion of the internet: from simple accommodation methods to waging a war with the new medium. Some are trying to merge the new with the old in a possibly most natural way. The fact is, however, that the internet is dragging the readers away from the print media. Research shows that press websites audience is rising while the sales of paper editions is dropping. The key issue here seems to be the knowledge of reader behaviour and new consumption patterns of both the internet and paper media. It is hence worthwhile to come back to the sources and look again – this time through a technological filter – at the media consumer: a reader, or part of TV or radio audience.

### A reader redefined

The current changes in the world of traditional media, expansion of the technopoly and progressive convergence of the media must lead to reviewing of the basic media related nomenclature, including the "media consumer" definition. The technological development warrants a new definition of a reader. A contemporary press reader is not only a consumer of a paper edition of a press title. Bearing in mind that a vast majority of press titles have their counterparts on the internet, such new reader is also a person visiting, more or less regularly, the online versions of a paper. The technological innovation has thus redefined the touchpoints. A newspaper may now be read on paper, on the internet or a mobile phone. There soon might appear a new channel that would allow to enjoy the consumption of press in yet another manner. Hence, definition-wise, the channel through which a reader approaches a press title becomes insignificant. What is important is the fact that the reader has interacted with the title. This assumption must also be employed in researching media consumption.

### Effects of the new technology – media as a tool to collect data

The advances in technology and communication channels also precipitates changes in media research methodology. The 21st century media require effective and suitable research tools. It appears that so far the traditional and new media research projects were conducted independently of each other. The internet is not only another communication channel or one more kind of media, but also a "place" where the media and market research is conducted. It serves as a source of data. Needless to say, press titles were used for the very same purpose with the dawn of readership research. Even if it is not the history repeating itself, the issue has certainly come back in a new light.

We are now facing a problem of redefining our approach to media measurement and the currently recognised research methods. Let me quote S.McDonald and J. Collins (2007): "The opportunities afforded to media by rapid technological innovation continue to challenge audience measurement in fundamental ways". This technological advancement influences and will influence even stronger the media measurement paradigms. The traditional media have entered the new era, so should the research methods. This also pertains to the internet itself and the way it is studied.

### Internet Audience Measurement – user centric approach and site-centric factor

Merely a few years ago, internet audience research was based on two types of studies only: site-centric and user-centric. They were seen as two completely separate measurements serving different needs of publishers. We can now observe a trend towards merging these methods to get the best of online audience research. The so-called "hybrid model" i.e. combined site-centric and user-centric measurement is gaining ground. This comes as a natural answer to the changing needs of the market. One could even say that the site-centric component is crucial in panel-based research. This stems from the impact the technological developments have on the research. Internet market researchers had to respond to these changes briskly, modifying their approach to analysing this medium. The research methods in internet audience measurement have undergone convergence. We are still strongly involved in panel-based research but site-centric measurement is now their crucial element.

There are two components to the user-centric measurement: a cookie panel and a software panel.

**Cookie Panel Approach**

Site-centric systems provide the most accurate measurement of page traffic for websites but they do not provide socio-demographic data for visitors. A frequently used method for determining audience profiles is to display on-site surveys (overlay format) to random samples of visitors. However, with this simplistic approach there is no integration with usage data. Such surveys cannot be used to derive measures of reach and frequency that can be used in online media planning.

The gemiusAudience system uses on-site surveys to generate a sample of cookies that can be used to track internet users. These on-site surveys are integrated with the site centric measurement system gemiusTraffic (see below). This gives us the ability to:

- **Turn a series of separate samples into a cookie panel –** aggregate the data from separate surveys executed on separate websites into an internet-wide 'cookie panel.'

- **Control and adjust the representativeness of the on-site survey sample –** by monitoring the display of on-site surveys, Gemius is able to control whether the sample recruited using on-site surveys is statistically representative (behaviourally) and then adjust as needed.

The gemiusTraffic platform is a web analytic tool that issues a unique cookie to each internet user who visits a website that is monitored. If the user completes a pop-up survey describing his or her demographic characteristics, then this demographic profile is 'remembered' when the user visits other websites that are monitored by the gemiusTraffic system. There is no need to fill out the survey more than once.

The tags implemented into web pages have three functions:

- To measure the site-centric performance of the website

- To display the on-site survey which will collect the socio-demographical profile of visitors

- To track the internet activity of panel members

It should be noted that, in line with data protection legislation, users are required to give their permission before they can be included in the cookie panel.

On-site surveys are distributed to randomly-selected samples of visitors to each website taking part in the cookie panel study. The questionnaires are displayed at random to every $n$th visitor and the samples are managed so as to control self-selection bias and to limit the frequency of exposure. In order for data on a website to be reported, at least fifty visitors must have filled out a cookie panel survey. However, there is no upper limit on the sample size (because the sampling is a continuous process).

The likelihood of a visitor filling out a survey may be different based on his or her psychology and there is a clear relationship between the likelihood of survey participation and patterns of online behaviour (Ejdys P., Cisek T., Modzelewski C., 2003). To compensate for this, Gemius has developed an innovative technique of 'behavioural weighting.' The data from the sub-sample of panellists that visited each website are weighted to the number of real users on the same website. In addition the data obtained from the cookie panel has to be adjusted to the demographic structure of the internet user population. This is because self-selection bias may skew the demographic representation of the cookie panel. To adjust for this, Gemius has developed a methodology referred to as 'structural weighting,' which weights the cookie panel data to the structure of the internet user population.

The end result of this process is a series of 'events' for each monitored website. These 'events' are similar in concept to the events recorded by PeopleMeters in traditional television audience measurement. More importantly, they are identical in structure and content to the events generated by a classic or large-scale user-centric software panel. This sample data is then projected to the total population of internet users in the market (provided that gemiusTraffic monitors over 30% of Page Views).

Of course cookies can be deleted and the total number of cookies in existence exceeds the universe of internet users. The lifespan of a cookie can vary according to how the browser is configured: some can be deleted every time the browser is closed, others can be deleted daily, weekly or monthly, or they might not be deleted at all. The estimation of Real Users (real people using the internet) takes into account the cookie deletion process.

The algorithm for calculating Real Users is based on counting the number of cookies that exist across all the measured websites every millisecond. The maximum number of cookies observed can then be weighted to the known internet universe thereby calculating a correction factor that can be applied to each website.

**Software Panel Approach**

The software panel is a group of individuals who have installed dedicated monitoring software on to their PCs. The panel is recruited using an offline recruitment survey (e.g. via face-to-face interviews) or on-line recruitment survey displayed on participating websites. After filling in the recruitment questionnaire, respondents are invited to install the monitoring software on their computers. This tracks and stores the website usage by all computer users in the household. (Each user declares his or her presence at the computer when logging in).

When installed on the PC of a panel member, Gemius NetSoftware can track internet behaviour and usage of applications. In principle, the use of any application can be tracked, this includes instant messenger services, media players and other programs.

Visitors who agree to download the software become members of the software panel. All the sites they visit (whether tagged or not) are recorded. They are encouraged to download the software on to all the computers they use. Moreover, in case of software panels, a special algorithm is used to estimate the number of Real Users for particular sites. There is also a range of data weighting methods employed: by demographic characteristics and behavioural weighing (e.g. to reduce over-representativeness of heavy users) for which the parameters are taken from site-centric research.

The example of this approach illustrates the growing popularity of the hybrid model involving merging site-centric and user-centric methodologies. This is also an answer to the needs of the markets for which the research is conducted.

In the internet audience panel-based research, press websites are simply a part of the online universe. Studying this type of sites is in no way different than studying any other websites. One could argue that such method of measurement does not provide the full picture of press consumption (or of online versions of press titles). Similarly, the traditional print media readership research provides incomplete knowledge on consumption of this medium.

**Web and Print cross media research – passive or declarative measurement?**

As mentioned previously, the internet is not only a medium or a "place" but also a channel for collecting data. It seems that a more appropriate way to conduct cross-media research of printed press and its online versions is via the internet. So the question at hand is how to measure a press brand strength vie the internet.

As far as measuring printed press readership online is concerned, it is obvious that we must base on declarations of users. Due to the above, it seems reasonable to employ the scheme of interview and the form of questions similar to this used in offline readership research, however with some modifications involving reduction in the length and number of questions.

**We believe that online declarative printed press readership measurement should be combined with passive monitoring of respondent activity in the internet**. There are at least two possible ways to perform such passive measurement: with software installed by users on their computers (dedicated software-panel), or with scripts placed on sites by publishers and identification of cookies.

An unquestionable advantage of the software way (software panel) over the cookie file method is the possibility to differentiate concrete users (*real users*) irrespectively of them deleting cookies, sharing the computer (and browser) with others and using more than one machine. Moreover, this method does not require publishers embedding scripts in their sites. An obstacle, on the other hand, may be the fact that the behaviour of individuals who use a PC other than their own or office one cannot be monitored; this stems from their unwillingness or lack of authority to install the necessary software on their computers.

The benefit of basing on cookie data is the fact that the participants need not install any software, potentially resulting in more response to the invitation to take part in the research. The most significant and key disadvantage of this variant is the possible confusion of a real internet user with a cookie. Note that computers and browsers are frequently co-used or an individual may use more than one PC on everyday basis. This means that a history of visits to websites recreated basing on cookies should not always be deemed an accurate reflection of an individual's online activity.

**Web & Print Readership case study**

Below we present the preliminary results of a pilot study conducted by Gemius on the Danish market. The study was conducted in cooperation with IPSOS and FDIM. The main aim of the project was to test the method of researching printed press readership via the internet. Another aim of the study was to analyse the audience duplication and reach of the printed press (hard copies) and of the websites of internet magazines.

The study lasted 6 months – from November 1, 2008 to April 30, 2009. It was carried out with the help of the CAWI method (Computer Aided Web Interviewing) RTS (Real Time Sampling). The data was collected continuously, day by day, evenly. The total size of the analysed samples in particular months ranged between 2671 and 3756. The response rate of the study (although declining slightly over the subsequent months) was satisfactory, that is, it did not diverge significantly from the typical values obtained in the studies carried out in this way.

Invitations (in the pop-up format) to participate in the study were randomly displayed to the users of nearly 300 sites associated in FDIM. The total reach of these sites among the Danish internet users is 95%, while internet penetration in Denmark, among people aged 15 years or more, is 85%. Because of the high level of target group coverage (81%), it was considered justified and acceptable to generalise the results of the study on the entire population of Danes aged 15 years or more, and not only on the Danish population of internet users at this age. To make sure that the results obtained from the sample are representative of the population, the structure of the sample was analyzed using the analytical weight. The first step was the validation and the adjustment of internet usage frequency in order to eliminate the burden being the consequence of sampling according to RTS scheme. The next step was the consideration of the information about the respondents gender distribution, age and region of residence. Weight values were calculated separately for each month of the study.

Because of the way the study was conducted, only the most popular of the Sunday editions of newspapers, dailies, regional newspapers and magazines (in total 28 titles) were included. The questionnaire asked about the recent contact with the title (recency) and frequency of reading – based on these indicators the estimation of, inter alia, the value of the average issue readership (AIR) was conducted.

The data on the audience of the websites of the considered magazines (in total 19 sites) among the questionnaire respondents was obtained based on the information stored in cookies thanks to gemiusTraffic research scripts inserted by the publishers on their websites. The history of site visits from 30 days prior to filling in the questionnaire was taken into account as well. Based on the fact and the time of visits on individual pages, the estimated value of the site reach was calculated, which is basically similar to the AIR (Average Issue Readership) indicator.

By combining the data from the questionnaire with hard data about website audience, the value of the indicator of audience duplication of an average issue of a hard copy and online version was calculated. In the analysis of the websites of the printed press, the word "readership" was used interchangeably with a visit on a website.

**Table 1. Dailies. Comparison of readership of the hard copies related to different methods of data collection - offline and online (November 2008); AIR indicator in thousands**

|  | Offline (TNS Gallup Index Denmark) (N=3 542) | Gemius CAWI study (N= 3 429) |
| --- | --- | --- |
| B.T | 383 | 298 |
| Berlingske Tidende | 328 | 257 |
| Dagbladet Borsen | 242 | 201 |
| Ekstra Bladet | 449 | 399 |
| Erhvervsbladet | 77 | 50 |
| Jyllands-Posten | 531 | 400 |
| Politiken | 432 | 387 |

**Table 2. Sunday Newspapers. Comparison of readership of the hard copies related to different methods of data collection - offline and online (November 2008); AIR indicator in thousands**

|  | Offline (TNS Gallup Index Denmark) (N=3 542) | Gemius CAWI study (N= 3 429) |
| --- | --- | --- |
| B.T. | 445 | 474 |
| Berlingske Tidende | 418 | 434 |
| Ekstra Bladet | 478 | 647 |
| Fyens Stiftstidende | 206 | 201 |
| Jydske Vestkysten | 276 | 240 |
| Jyllands-Posten | 728 | 573 |
| Nordjyske Stiftstidende | 240 | 127 |
| Politiken | 563 | 567 |

The comparison of readership data obtained by means of different methods shows the differences in the level of individual titles readership. Overall, the web study showed lower rates of readership of dailie.. Bigger differences between the obtained values of readership may be observed in most popular titles, and the smaller in less popular ones. The differences were also observed in case of the Sunday newspapers but the results seem to be more coherent. Of course, every method has an impact on the obtained results and here we observe the consequences of this. Thus, the way the data is collected is in this case crucial. We are fully aware that the indicators of readership fluctuate month to month and one should compare the results in a longer time span, at least for a few months.

Many publishers of the printed press are present in the internet -- printed newspapers have their mirror images in form of websites. As it had been mentioned earlier, another aim of the study was to show the audience duplication and the cumulative total reach of printed press titles and its counterparts in the internet. The analysis does not tell us only about only about consumer habits and consumer loyalty to a particular brand, regardless of where and how it is published (online versus print).

Such data also shows the pattern of consumption of a brand in the internet and the traditional world. The results of the described cross-media study may have a dual use: the sale of advertising campaigns (OTS in print and online) and management of titles and the content published there. We wanted to show the strength of media brands regardless of the communication channel or the published content. The Total Reach indicator in the below tables tells us about the strength of a media brand and, as a consequence, about its advertising potential.

**Table 2. Selected indicators for press titles and websites of these titles - dailies; (Gemius CAWI Study - November 2008; N= 3 429)**

|  | AIR hard copy | Audience Duplication | Web-Site Reach (based on AIR) | Total Reach – Print & Web |
|---|---|---|---|---|
| **B.T.** | 7,3% | 2,9% | 7,9% | 12,3% |
| **Berlingske Tidende** | 6,3% | 1,6% | 4,7% | 9,3% |
| **Dagbladet Borsen** | 4,9% | 1,6% | 3,2% | 6,5% |
| **Ekstra Bladet** | 9,7% | 5,7% | 13,4% | 17,4% |
| **Erhvervsbladet** | 1,2% | 0,2% | 1,0% | 2,0% |
| **Jyllands-Posten** | 9,7% | 2,7% | 6,3% | 13,3% |
| **Politiken** | 9,4% | 3,2% | 5,4% | 11,6% |

**Table 3. Readership for titles with one-week publication period (Sunday newspapers) - (Gemius CAWI Study - November 2008; N= 3 429)**

|  | AIR hard copy | Audience Duplication | Web-Site Reach (based on AIR) | Total Reach |
|---|---|---|---|---|
| **B.T.** | 11,6% | 6,1% | 17,5% | 23,0% |
| **Berlingske Tidende** | 10,6% | 4,6% | 11,1% | 17,1% |
| **Ekstra Bladet** | 15,8% | 10,2% | 24,5% | 30,1% |
| **Fyens Stiftstidende** | 4,9% | 3,1% | 5,0% | 6,8% |
| **Jydske Vestkysten** | 5,9% | 2,8% | 3,0% | 6,0% |
| **Jyllands-Posten** | 14,0% | 6,9% | 12,2% | 19,2% |
| **Nordjyske Stiftstidende** | 3,1% | 1,2% | 2,5% | 4,3% |
| **Politiken** | 13,8% | 7,1% | 12,9% | 19,6% |

**Table 4. Selected indicators for newspapers and Web sites of these titles - magazines; (Gemius CAWI Study - November 2008; N= 3 429)**

|  | AIR hard copy | Audience Duplication | Web-Site Reach (based on AIR) | Total Reach – Print & Web |
|---|---|---|---|---|
| **Billed Bladet** | 9,3% | 1,7% | 5,2% | 12,8% |
| **Computerworld** | 1,9% | 0,9% | 3,6% | 4,6% |
| **Den Bla Avis** | 4,9% | 3,6% | 23,8% | 25,1% |
| **Familie Journalen** | 5,1% | 0,5% | 0,7% | 5,3% |
| **Femina** | 3,1% | 0,3% | 0,4% | 3,2% |
| **Ingenioren** | 4,0% | 1,7% | 5,5% | 7,8% |
| **Se og Hor** | 10,9% | 2,6% | 7,4% | 15,7% |

The results of the study show some facts important from the perspective of the advertising potential of certain titles. First, a relatively low rate of audience duplication was observed. It can point to the variety of the audience of a hard copy and its equivalent in the net. The low rate of the audience duplication may indicate different needs of readers as regards the printed and internet titles. Secondly, in the case of several titles the reach of their online versions is bigger as compared with the printed versions. This applies both to some dailies and monthlies. On the other hand, the audience duplication indicator for Sunday newspapers and its web-sites were significantly higher than for dailies and monthlies. Finally, internet adds to the reach of hard copies, which increases the power of a media brand. This fact is of considerable importance in building cross-media advertising campaigns in the two media and in particular titles.

**Conclusions, or the web and print cross media research dilemma – problems to solve**

The changing world of media and consumption patterns confront the key market players, including researchers, with a need to revise their views on analysing media. The new brave 21$^{st}$ century world of media convergence forces us to re-think the way we perceive researching the media. We must not forget that at the end of a day, it is the consumer who drives these changes and hence it is the consumer who needs to be re-discovered. The "consumer is king" approach must be reflected in the research solutions offered by the research industry and the media themselves. In this article, we presented one of the possible ways to perform cross-media research in relation to two types of media: press titles and their internet versions.

This article also discusses the proposed "hybrid model" of the two media measurement, pointing to the fact that this very method best suits the research goals. The model should base on passive measurement of users' online activity. Monitoring the consumption of printed press requires a declarative measurement. The model in question may take two forms:

- cookie-panel and declarative measurement of press readership
- dedicated software-panel and declarative measurement of press readership

One of the above forms has been discussed in this article. In its assumptions, it resembles the "2 Rivers" model proposed by some researchers (Elliott K., Scionti R., Page M., 2003). Naturally, another possible solution would be merging data from the two research standards: press readership and internet audience measurement. However, regardless of the choice of the method, there are still some important issues to deal with, e.g.

- equivalence of indicators for press titles and websites – can AIR used in case of press be utilised for measuring audience of websites? Should a visit be treated as an equivalent of AIR?

- online vs. paper issue – should the temporary character of issues, particularly as regards updating the content of press websites, influence the definitions of indicators and comparisons of results?

These are just a few dilemma to be faced by the publishers and researchers. There may be, however, even more perplexing problems lurking just around the corner.

**Bibliography**

Ejdys, P; Cisek, T; Modzelewski, C: Real Profile: A new approach to online media planning (ESOMAR, WAM 2003
Elliott K., Scionti R., Page M.,The Confluence of Data Mining and market Research for Smarter CRM, SPSS & The Kantar Group white paper, 2003
McDonald S., Collins J. Internet Site Measurement Developments and print, WRRS 2007
Postman N., Technopoly, Vintage, USA  1993