

## WHO WAS THERE? HOW MANY HAVE BEEN THERE? A NEW APPROACH TO MEASURING THE REACH AND USER STRUCTURE OF WEB SITES

**Rolf Pfleiderer, Intratest Burke InCom**

---

**This contribution will provide you with a suggestion for the "typification" of different measurement methods which are commonly used or may possibly be used to conduct reach analysis and user structure analyses of Web site offers. These measurement methods will be discussed (in brief). The basic approach developed by the author at Intratest in cooperation with the "Working Group Online Research", a group made up of members from various media associations, will be presented here in more detail.**

The fundamental questions for a measurement of the media audience have always been: How many users? and, What kind of users? The fact is that these "unassuming" questions are not actually so simple and that they cannot always be answered immediately, even by such "classical" and relevant advertising media as the major TV and radio stations, the major magazines or the daily newspapers. This insight does not have to be verified here any further. As media researchers, however, we can and must insist that these questions be answered for the no longer so completely new online media, as far as it comes into question as an advertising media. Of course, the answer to these questions must occur in a form which is similar and understandable for comparable media, and which is at least in a clearly documented form.

To this point, harmony is certain to be found among the media researchers and, as can easily be read in the "FAST Principles of Online Media Audience Measurement" ([www.fastinfo.org](http://www.fastinfo.org)), for instance, there is definitely a large degree of "general agreement". How these measurements are actually made – and, in individual cases, also what is measured here – is thoroughly controversial and, considering the rapid technical developments and the development of distribution of the online media, will continue to remain the subject of debates (and symposia!) of media researchers over the coming years. In this way, the online media research will not differ at all from the research of the "classical media".

Because of the volume which would otherwise have to be feared, the following contribution does not constantly nor even individually refer to the discussions carried out in the U.S., although they have, without doubt, been developed there to the largest degree. An intermediate balance was carried out in the FAST document which has already been mentioned above. It will present some "deviating" aspects, experiences and suggestions as to how the reach and user structure of the online media can be measured.

We have already reached the first problem at the "definitional" phase: The American colleagues have entitled their document the "FAST Principles of Online Media Audience Measurement". However, they have evidently related the terminus "online" both to the "online media" as well as to the "online measurements". Three types of online audience measurement have been distinguished, namely:

- Site-Centric, in which the Web site server log entries are the immediate subject of analysis
- Ad-Centric, in which the ad server log entries are the immediate subject of measurement
- User-Centric, in which the person using the online media is the immediate subject of measurement" (e.b.d.)

In addition to the fact that "site-centric" and "ad-centric" measurements may naturally only be conducted as online measurements, "user-centric measurement" is also explained as follows: "User-centric measurement readily measures the display of online media content by a browser" (e.b.d.). Together with the explicit reference that this kind of measurement is "generally based on panels", it is evident that this third type of audience measurement also demands the use of online measurements.

Here, I would like to ask the American colleagues to reconsider their definition of "user-centric": If we define "user-centric" as being a "display of online media content by a browser", we will be unable to avoid defining the browser as a user. The consequences could be quite considerable. Using the television research as an analog, this would mean that the "opportunity to see" is adequate when an instrument is tuned in, although no one is watching (or even if nobody is in the room). I don't even want to consider the dreadful analogs which might arise for the fields involving print media research. At any rate, I would find it to be sensible, also with regard to the online media, if "users" were always understood to be only those individuals actually making use of the media.

Furthermore, I also consider it to be at least unusual and highly related to the American market when the classical form of (ad-hoc) inquiry, be it face-to-face, performed by telephone or carried out in written form, is generally seen to be unfit for the performance of measurements in the online media, because of this description of the "three types of online media measurement".

In my opinion, one should instead distinguish between

- Measurements which are merely related to server log files, whereby one can (and must) additionally distinguish between Web site server log entries ("site-centric") and ad server log entries ("ad-centric"): **Log File Analyses**.
- Measurements which refer exclusively to individuals and their conduct (or statements about their own conduct): **Surveys**.

Measurements which combine or integrate the monitoring of "electronic traces" (server log entries) with surveying methods: **Hybrid approaches**. With these hybrid approaches, one can also distinguish between:

- **Panel-based approaches** where the individuals being questioned (or the browser – see above) are optimally recruited as part of a representative random sampling of the population to participate in the surveys and in the ("automatic") continuous recording of their online activities;
- **Nth-intercept approaches** (general censuses occasionally have been performed on the registered users of some special sites, although these will not be discussed any further here) where the individuals who are to be surveyed (online) are recruited at random (respectively by a random selection of their visits, hits or whatever) and are questioned as to their sociodemographic data and their patterns of use.

In the following, these fundamentally different types of measurement approaches will be discussed in brief with regard to reach and users' structure data of Web sites, while the approach mentioned last, which has been developed by our institute, will be discussed somewhat more comprehensively.

### Log File Analyses

Log file analyses are especially "attractive" in this respect since analyzable log files are obviously produced by the individual sites or online services. In order to perform comparative media research, however, the following problems must be taken into consideration and/or solved:

- How "access" is attained and what information is written in the log file depends, in part, on the respective server and/or how it has been configured or programmed. Comparative analyses can only be accomplished on the basis of **homogeneous entries** in this regard. In practice, this means that the "media" involved in making comparative measurements must reach an agreement on the development of a "minimal standard". In Germany, the "IVW standard" is presently known best. (The IVW is the Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern e.V., the association for the determination of the circulation of advertising media which, among other things, is also responsible for the performance of controls in the press media. A more detailed explanation of this process here, however, goes beyond the scope of this discussion.)
- Web sites and/or Web pages which are frequently used often are not attained from the "site host", but are frequently stored temporarily (that is cached) with a "proxy server". No log file entry is found in the site host's server for such "cached" accesses. Since " **caching**" is not performed in the same manner for the different sites, particular arrangements must be made so that the log files of the different "media" can continue to be comparable. One possibility is to furnish every relevant "page" for the comparative analysis (or also every frame, etc. – which unfortunately cannot be discussed here more comprehensively) with a "dynamic" element which must be resent "anew" from the site host, even if a "cached" page may have already been delivered in individual cases. The IVW procedure in Germany functions with just such a technology.
- For the purposes of comparative ad-media research, the **measurement unit** must be defined comparably! In the online media, one can generally distinguish between "**visits**" and/or "**page impressions**" (although "ad impressions", "visit periods" and even more may be distinguished as well). "Visits" and "page impressions", which are also the most common units in the U.S., are not to be understood in themselves, but must each also be defined carefully and consistently. Particularly for the "visits" which can best be compared by the "readership" or "listeners" or "viewers" of the classical media, a problem develops since the log files of the machine accesses are recorded, and not the actual "visits" and/or contacts made by the individuals. Here as well, a definitional agreement is therefore once again necessary in order to know how to define a visit by an individual (which may be made up of one or more page impressions) and, if necessary, to define at which point in time a new visit begins by the same person. Also with regard to this, the online services connected with the IVW procedure in Germany have reached an understanding so that the verified "visits" and "page impressions" are comparable.

- Unfortunately, it is also the case that the technical developments constantly require revisions and restorations of the laboriously designed agreements concerned with the definition of "visits", "page impressions", etc.. One may therefore not expect to find a final solution to this problem in definition, since it is surely a "dynamic problem". The most difficult problem for media research in the classical sense is perhaps that the log file analyses always ultimately provide information on the technical access and not about the individuals themselves. Even if the number of "visits" per month was determined according to standards which have been designed jointly, for example, this number still says nothing at all about the **number of visitors** nor whether these visitors were men or women, old or young, etc., that is **who they actually are**. Whereas the information concerning the sociodemographic structure of the visitors cannot generally be derived from the log files, but only by way of additional investigations (surveys), there are technical possibilities for estimating the number of visitors without making such surveys. "Cookies" can be used, for example, which help to identify each individual performing a "request" according to whether they have ever been there before (if necessary, including the time of the last access and also possibly with information on the other pages accessed). Although the usage of cookies has meanwhile become more the rule than the exception worldwide, market research in Germany has dispensed with their use up to date because of ethical considerations and for reasons of data protection. I believe that we should maintain this conduct with the use of cookies, at least as long as the use of a cookie has not explicitly been conceded by the specific "individual" who is affected.

In summary, one can say that log file analyses may indeed provide a great deal of interesting information. These analyses, even on the basis of far-reaching definitional agreements as have been created for the IVW procedure in Germany, may provide a starting point for the determination of the results of the actual reach and the users' structure, although this cannot be the final result.

### Surveys

The classical means of comparative media research, the survey, can of course also be used for the determination of the reach and user structure of online media. Irregardless of whether or not one considers face-to-face, telephone or written inquiries, the advantage of such procedures is that the subjects in these investigations are actually those individuals whose response to media consumption behavior is to be explored. Surveys with regard to the online media, however, demonstrate all problems which are also observed in surveys made related to the offline media:

- **Sampling problems** and/or problems with the correct representation of the respective universe (households and/or individuals) – in practice, most problems are related to the sample response rate;
- problems found with the correct **identification** of individuals in a survey of the media involved – here, it must be intensified even further since many of the relevant online media also involve offline versions which must clearly be distinguished from the online versions;
- **recall problems**, particularly with regard to determine the time period in which a particular medium was last used, and also regarding the frequency of using a medium or the number of contacts within a defined time period;
- **economic problems**; because population-representative surveys with high sample-size are expensive. An extremely expensive aspect of the surveys involving the online media is the fact that a representative random sampling of the population making use of online services reveals that only 10 percent or perhaps 45 percent (depends on the country) are actually active in the online media. Furthermore, because of the extremely diverse, but also fragmented landscape of online media, for the performance of reach and structural analyses sufficient sample sizes are only available for a few of the larger sites and/or only through the use of unusually large random samples (in the range of more than 100,000 persons).

A (negative) example for the fact that particular care must be taken in the measurement of online-media coverage is to be seen in the results of the ACTA '98 (Allensbacher computer and telecommunication analysis). This investigation has been performed using (disproportionate) quota sampling of the German residential population between 14 and 64 years of age. A total of 9,558 individuals were questioned face-to-face from February to August '98. With the performance of an extremely extensive questionnaire, the broadest audience of a couple of online media, the use during the last month, the use per week as well as other things were inquired (this and the following information have been taken from a summary of the department Marketing and Research Services of Gruner & Jahr publishing house, which has been presented, among elsewhere, also at the AG Online Research; I would like to take this opportunity to thank them for allowing me to use these documents!). The following table shows the comparison of the ACTA results concerning the number of users and the weekly results determined for the relevant time period by the (technical) IVW counting (compare above):

	(I) Users per week ACTA '98	(II) Visits per week (mean) IVW 2-8/98	Visits per user II : I
Online – site of ...			
Wirtschaftswoche	220,000	49,591	0.23
Chip	300,000	101,646	0.34
Handelsblatt	230,000	79,071	0.34
PC-Welt	230,000	112,136	0.49
AOL Homepage	590,000	346,738	0.59
Spiegel	350,000	252,151	0.72
SAT1	400,000	409,350	1.02
Focus	430,000	442,821	1.03
Dino	250,000	281,567	1.13
TV Spielfilm	280,000	371,339	1.33
Stern	150,000	252,384	1.68
Bild	200,000	379,797	1.90
Fireball	480,000	982,516	2.05

Even if one does not consider the results of the technical IVW measurements to be 100 % accurate for the reasons implied above, this comparison demonstrates the total uselessness of the application of ACTA figures in "media planning": The figures are much too high and the relationships of the individual media to one another are also extremely unlikely. One can only speculate here concerning the causes for these errors. It would not be wrong, however, to search for these causes among the above mentioned general problems of surveys (random sampling/representativeness, media identification, problems in recall). It has also been pointed out, however, that surveys which do not attempt to consider hundreds of the most varied media simultaneously, and to additionally include dozens of attitude and behavioral questions, but which instead concentrate on reduced facts and fewer media sites, lead to plausible reach values for the online media (compare below).

By the way, very large "access panels" which have been employed to date, especially for ad-hoc inquiries on consumer habits, are suitable for accessing the reach and online users' structure data either via the normal postal services or via E-mail, or through the performance of Web surveys. Decisive for the suitability of the comparative media research have always been the aspects of the sampling quality, the media identification and the question models, which should not overstress the recall capabilities. Examples for sufficiently large access panels are to be found in the U.S., especially in the NFO panel, with more than 200,000 online households, and in the access panel of Market Facts. Online panels, like that from Harris Black, are more likely not to come into question for comparing media research because of the sampling procedure which is based on self-selection. These panels may only be taken into consideration after evidence has been provided that all of the online users are represented free of bias; not only with regard to their sociodemographic, but also with regard to their use of classical media as well as to their use in the online area. Such evidence, however, will only be produced through the use of large, parallel studies.

### Hybrid approaches: Panels

At least in the U.S., panel institutes like Mediametrix and also Nielsen's Netratings claim to offer **the instrument** for reach measurements of the Web sites. These panels offer a combination which makes use of the advantages of the "traditionally" recruited random samples and provide the possibility of solving the identification and recall problems of surveys via technical measurements with the aid of special programs installed on the respective PCs. In practice, these panels are unfortunately a long way from being as perfect as they might theoretically be:

- As with log file analyses, the **definition of "visits"**, "page impressions", etc. principally is a problem which could be solved as described above for the log file analyses through the use of conventions.
- A much more difficult problem is the **willingness** of target individuals **to participate** in the panel when they are asked for participation (e.g. in a population representative telephone survey). After all, these individuals should install a program onto their PC themselves (or let one be installed) and must additionally undergo routine questionnaire actions which may prove to be more or less extensive. Our own experiences, as made with our institute, as well as the little information we know or have heard from other institutes, are not very encouraging in this respect.
- A perhaps even unsolvable or at least presently still very relevant problem of these panel approaches is that a large proportion of **usage** of the relevant online media comes from **companies, universities, school PCs** and/or PC networks – in Germany, for instance, more than half. The willingness of private and public organizations or institutions to allow the installation of "foreign" software which is designed to monitor the use behavior of their employees into their network and/or onto the PCs of staff members, however, is minimal. Members of the works council – at least in Germany – are regularly shocked by such ideas.

- Ultimately, there is still an **economic problem** derived from the mandatory **size of the panel** which is required to adequately portray all of the relevant online media. Already now, complaints in the U.S. related to this subject, and to the fact that the Mediametrix panel includes "only" 40,000 individuals, are being sounded very loudly and clearly. Based upon the conditions of competition of the panel providers, and their obligation of getting out of the red and into the black, the pressure to increase the size of the panel will not necessarily lead to an improvement in the quality of the random samples nor to a more complete portrayal of this online use.

All in all, such panels represent an elegant, but neither economical nor unproblematic solution for comparative media research. Particularly the systematic neglect of company-based usage appears to be unacceptable.

### Hybrid approaches: N Viz

The "Working Group Online Research" was formed as a forum for the discussion on research for and about internet, the WWW and other online media by the DMMV (Deutsche Multi Media Verband, the German multimedia association), the VDZ (Verband Deutscher Zeitschriftenverleger, an association of German magazine publishers), the BDZV (Bundesverband Deutscher Zeitungsverleger, a German federal association of newspaper publishers), the VPRT (Verband der Privaten Rundfunk- und TV-Anstalten, an association of the private radio and TV networks), as well as by representatives of the advertising agencies. Since 1997, as one of the most important questions of the industry, this working group has been discussing how one can perform comparative user research in the new online media and how one should develop the generally accepted base of media planning.

Our institute has developed a suggestion for the determination of users' structure data and has already performed empirical tests at the end of 1997 and presented to the AGOF. Last year we amended our suggestions on **users structure analyses** by including a suggestion concerning the measurement of reach data of Web sites.

Our suggestion is essentially based on the fact that **each n-th access**/retrieval of a specific Web page and/or a specific Web offer becomes rerouted with the aid of a relatively simple software program to a page on which the "request" is made to participate in a short survey: The "**Survey Request Page**". The software module must therefore be installed in the respective test site (which usually is done remote without any problems, but is naturally only performed with the knowledge of and under control of the respective Web site). The N can be chosen arbitrarily, although it should, on the one hand, be large enough to "disturb" the "total population" of the site as little as possible and should, on the other hand, be small enough to enable a sufficient number of interviews for the time period to be analyzed. Furthermore, the software module (compare the remarks on the log file analysis to be found above) must "count" in such a manner that the various programmed sites/pages are counted comparably and are also taken into account when found on the "cached" sites of proxy computers.

Although these demands can fundamentally be fulfilled by various different programs, we have proposed that the programs to be used in the performance of the comparative reach and user structure analyses should count similarly, as is to be seen in the procedure used by the IVW in assessing the visits and page impressions: Here, one can make use of a convention which has already been agreed upon. An "N-Viz module" which has been programmed accordingly has meanwhile proved to be successful.

The "Survey Request Page" cannot be "avoided" by the user, although the participation in the survey must naturally be voluntary: Thus, every user rerouted as an "n-th" has the following possibilities on the Survey Request Page:

- He/she can immediately participate in the survey. He/she is then forwarded to the server where the questionnaire is to be found and, after replying to the questions on use, sociodemography, etc., is automatically led back to the page where he/she wanted to go as his/her visit was interrupted by our software module.
- He/she may decide, however, that he/she would like to take part in the survey, although he/she has no time at the moment. The questionnaire will then either be made available to the browser immediately or he/she will receive a Web address where the questionnaire can be filled out; otherwise, he/she may indicate his/her E-mail address and the questionnaire will be "mailed" electronically.
- He/she may also indicate that he/she has already taken part in this survey (within a certain time period) and will consequently not be questioned again. (Model calculations with the claimed number of visits within a relevant period of time have shown, by the way, that this response is not used to any extensive degree as a "hidden refusal".)
- Nevertheless, he/she may also refuse to participate in the survey. In contrast to traditional survey methods, some additional information can be attained from these "refusers" in the log file analyses. For instance, compared with the other participants of the survey, we thereby know that the refusers demonstrate a higher proportion of "foreign" domain names.

To a certain extent, the Survey Request Page thereby represents the "contact phase" of traditional market research. Therefore its design and the improving information are correspondingly important to get high response rates. Examples of improving information are: Indications for how long the interview will last (naturally, of course, only if the duration announced is not too long) and the fact that one has been selected according to a random sampling procedure ("every 500th individual").

Our suggestion concerning the performance of a user structure analysis also indicates that the design of the **Survey Request Page should be standardized** for making comparative investigations and that the **questionnaire** which follows should also be **standardized**: In this way, the provocation of obtaining different break-off quotas as a result of questionnaires of highly differing lengths should be prevented so that the comparability of the results from different Web sites is not endangered. Practically, this does not mean that no other questions beyond the standard program of questions may be asked, but rather that questions may be posed as to whether or not one is willing to answer further n questions on topic y. The standard questionnaire includes inquiries concerning:

- Sex
- Age
- Education
- Size of the household/children in the household
- Professional status/profession pursued
- Household – net income
- Permanent residence/nationality
- Current online media use (privately/professionally)
- (Usual) use of the online media (this site) per session
- Frequency of using this site
- Last use of this site before the current access
- Frequency of online use, altogether
- Duration of online use per day

Information concerning the use and the last use of an online site can – and must! – be employed to calculate the different, individual selection probabilities as part of the n-th intercept procedure for the transference of visit-oriented data to visitor-related data. Naturally, this presupposes that one agrees upon (for example via convention) whether to conduct the user structure analysis of users per month, users per week or users per day: The transformation procedure, at any rate, appears to look different.

The tests with the procedure described here show that highly plausible differences are evident for the user structure of Web sites – and also for the different pages within a Web site. Aside from such "internal" validity evaluations, which will not be expanded upon any further here, however, interesting examples of "external" validation are already existent as well. The MGM (the marketing company for the TV programs of ProSieben, Kabel 1, DSF, etc.) has compared the results of an N-Viz user structure analysis with the results of a telephone survey which was carried out simultaneously (with a representative population of a large sample made up of over 10,000 interviews per month) and the results of a so-called "online survey" ("W3B survey") using click-me recruitment:

**ProSieben – Online Users (March 1999)**

	Representative telephone questionnaire in %	N-Viz in %	W3B in %	In comparison Population above 14- years-old in %
Men	67	69	80	48
Women	33	31	20	52
Age				
up to 19-years-old	29	32	14	8
20 to 29-years-old	34	34	43	14
30 to 49-years-old	32	30	37	38
50-years-old and over	5	4	5	40
Education				
low level	15	12	20	40
medium	29	29	28	33
high level	56	58	52	27

There is a great deal of agreement between the results of the telephone survey based on representative sampling and the structural data of users which has been determined using the N-Viz method. The results of the "click-me" survey demonstrate a clear deviation from these findings. The convincing results of this test and the much lower costs in comparison to the representative offline surveys have meanwhile inspired further large media Web sites to continuously survey their user's structure by N Viz.

One can even go a step further: Since the number of "visits" to the test site by each of those interviewed is known as a result of the information provided by each of the interviewees, and since the response rate and assessment interval are known, the **number of visitors to a site** within the respective time period can be calculated directly from the results of an N-Viz survey. Thereby, as with other "projections" made on the basis of demoscopic sampling investigations, however, one must assume that the number of those "refusing" to undergo the inquiry, at least with regard to the number of visits to the test site during the time period in question and the respective user definition, does not differ from that of the users taking part in the inquiry. There are indications that this assumption is not entirely false, and that it generally represents more of an underestimation of the "actual" coverage. We are able to look into this since the results of the "official", monthly IVW count are available for determining the number of "visits". This number can simply be divided by the average number of visits per visitor during the defined time period – and one also attains a "reach" which is again based upon the assumption of the use habits stated above. Such calculations are presently being carried out with the first comparable data files, although it is still not possible to report on this today.

What are the strengths, and what are the weaknesses and/or limitations of the N-Viz approach for comparative advertising media research?

The weaknesses and limitations include:

- As long as the method of counting the "visits" is related, and with this the intercept-interval N of the log-file entries as based upon the IVW method, the remarks concerning **possible technical problems** related to specific peculiarities in the programming of individual Web sites and of the servers from which they are made available, are also valid for the N-Viz. In practice, these are generally problems with very little relevance.
- The procedure of the survey is based on the voluntary willingness to participate in the respective random sample. Here, one can find **individuals who refuse to participate** (and naturally also interviews which have been broken off or which cannot be evaluated for other reasons), a feature which may lead to a systematic bias of the results. In our tests, the "response rate" defined as the proportion of analyzable responses from the total number of reactions to the "Survey Request Page" (that is, the analyzable interviews which have been given or which have been cited to have already taken part in the inquiry), amounts to between 40% and slightly over 50%. This is high, if one considers that no "follow-up actions" are possible, as are usually to be found in traditional offline market research surveys. This response, however, naturally allows a great deal of room for possible bias. Further work will continue to be necessary here: Through an improvement in the design and in the "persuasive powers" of the Survey Request Page, through the public promotion of this method and its use in the development of the online media, through technical solutions for the problem of multiple responses especially of the particularly frequent users, etc.
- Finally, with regard to the reach, **"external overlapping"** cannot be determined using N-Viz as with the classical offline surveys, but also with such online panels as Mediametrix, net ratings, etc. The use of sites different from the respective test site can only be inquired in the course of an interview, a process which is possible, if required, for only a very limited number of sites/competitors.

Indisputable advantages, in contrast, are also available for N Viz:

- It is **simple and fast**, and – compared with the costs required for large panels or representative offline surveys – it is **extremely economical**.
- It may be used **for sites with very large visitor numbers as well as for "small" sites**: There are no essential differences in quality between the results in cases where N=50 and cases where N=5000 have been used (in as far as that a comparable sample size is available for the structural analysis). It is thereby a procedure which is suitable for both the "general-interest media" in the online area, as well as for "special-interest media". Furthermore, it is a process which can even be easily employed without any problems in countries with less developed internet access!
- N-Viz can be used on **entire Web sites**, on the **combination** of jointly marketed Web sites, as well as on **individual parts of a Web site**. So it may also be used to examine the internal structure of a Web site (with its subsites).
- In Germany, N-Viz is based on the only visible and generally accepted "currency" for the reach of online media to date: It is **compatible with the IVW measurements** for the visits and page impressions.

In my point of view, there are presently no better alternatives available for user structure analyses nor for the performance of reach surveys than the N-Viz procedure.

