

Magdalena Szpunar

## Internet Research vs. Research in the Internet, or How to Research the New Medium – Basic Methodological Issues

słowa kluczowe: badania online, badania offline, badania site centric, badania user centric, FGI, CAPI, SAPI, e-mailing, web surveys, site intercept

key words: on-line research, site centric research, user centric research, FGI, CAPI, SAPI, e-mailing, web surveys, site intercept

Streszczenie: Artykuł podejmuje kwestię badań internetowych. Autorka dokonuje zasadniczego rozróżnienia na badania prowadzone poprzez Internet (badania online), jak i badania Internetu jako badania specyficznego medium. Przedstawione zostają podstawowe problemy metodologiczne w związku z prowadzeniem badań przez Internet wraz ze wskazaniem na podstawowe typy badań prowadzonych w Sieci. Zaprezentowane zostają sposoby zachęty internautów do uczestnictwa w badaniu prowadzonym drogą online, jak i elementy ochrony praw osoby badanej poprzez Internet.

Abstract: This article discusses the subject of Internet research. The author differentiates between research on-line and research of the Internet as a specific medium. Presented here are the basic methodological issues regarding on-line research including a presentation of the various types of research on the Web. Also described are ways of encouraging internauts to take part in the research conducted on-line as well as the protection of personal data of the person surveyed this way.

The Word Wide Web (WWW), also known as the Internet, has become a true field for research exploration. This new medium allows us to register in a new way and consequently to analyse the registered content. A website can be considered a unit of analysis as well as electronic mail, discussion forum and mailing information and blogs. Since so much of people's activity has been transferred online, it should not be ignored. Online surveys have become a successful alternative to the labour intensive and expensive *face-to-face* interviews. Understanding the activity taking place online allows us to better understand the mechanisms governing this environment as well as to better manage the distribution of information in the virtual world<sup>1</sup>.

Polish sociologists are beginning to realize that in order to make progress in survey research it has become necessary to go beyond the present research schemes. K. Słomczyński states that the Internet and mobile phones, due to their popularity, enable methodologists to use these new means in gathering materials about respondents<sup>2</sup>. At the same time, he notes that in many cases the researchers are not prepared well enough so as to take full advantage of the possibilities that computer assisted interviews give<sup>3</sup>. Regarding the Internet, there are two methodological issues, one – adapting the present research techniques to the new, interactive environment, two – the creation of entirely new tools, specifically designed to fit the new medium. Although the amount of research carried out through the Internet constantly grows,

---

<sup>1</sup> P. Ejdyś, M. Grzanka, *Badania Internetu w Polsce (przykład badań typu site-centric)*, in: B. Bednarek-Michalska, EBIB Elektroniczny Biuletyn Informacyjny Bibliotekarzy, 9/2002, <http://ebib.osp.wroc.pl/2002/38/gemiup.php>

<sup>2</sup> K. Słomczyński, *Polska metodologia socjologicznych badań empirycznych na tle międzynarodowym*, in: P. Sztabiński, F. Sztabiński, Z. Sawiński (ed.) *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, IFiS PAN, 2004, p. 20

<sup>3</sup> Z. Sawiński, *Źródła rozwoju metodologii badań marketingowych*, in: P. Sztabiński, F. Sztabiński, Z. Sawiński (ed.) *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, IFiS PAN, 2004, p. 25

there have been written very few methodological approaches on how to work with this medium<sup>4</sup>.

### *Research on and off-line*

With regard to Internet research, an imperative distinction needs to be made. That is research of the Internet as a medium as well as its content versus research done through the Internet or research online. The first can be done two ways – using traditional offline methods or online, through the Internet. It all depends on what we wish to research. Only then the most appropriate technique can be selected to in order to obtain the answers we are looking for.

There is so far little literature with information on how effective web survey is although there have been reported astronomical sums of respondents filling in questionnaires within short hours or days.<sup>5</sup> Regarding Internet research, a classic differentiation can be made between on and offline research. In case of offline research there a quantity and quality analyses as well as tracking research which can provide us with data on Internet access, aims for use, user advancement, most popular sites, their functionality and a general socio-demographic user profile<sup>6</sup>.

On line research, on the other hand, can be divided into two categories. These are *site/server centric* (from the server point of view) and *user centric* (from the user). *Site/server centric* research encompasses the entire user population (Internauts) as it registers all connections to WWW sites. It is *opt-in* type of research which means that it analyses only those servers who have opted to take part in it. The information obtained is: number of guests, number of users (including those visiting again), average amount of time spent, frequency of visits, links, first and last visited site as well as a spatial localisation of users<sup>7</sup>. This type of information allows for the description of Internet user habits and a trend analysis.

Table 1. *Site- and User-centric Research Differences*

<b>Site-centric Research</b>	<b>User-centric Research</b>
Website (or group of sites) is the subject of research	Internet user (or group of users) is the subject of research
Hard data about each event (calculated is every site entry and every click on an advertisement)	Research is done based on a select sample. Quality of data based on a proper method of population analysis
Research answers the questions how many? and how?	Research answers the question who?
Server and tracking systems are a source of data	Internet users (their socio-demographic profile) and hard data measured on user computers (behavioural profile) are a source of data

Source: <http://www.tezmedium.pl/12041.xml>

*User centric* type of research is similar to classic telemetric research. A panel<sup>8</sup> of Internet users is selected thanks to who constant monitoring of their activity on the Web is

<sup>4</sup> K. Sobieszek, *Problem błędu braku odpowiedzi w badaniach internetowych – rozważania teoretyczne*, in: Ł. Jonak, P. Mazurek, M. Olcoń and others. (ed.), *Re: internet – społeczne aspekty medium. Polskie konteksty i interpretacje*, Warszawa, WAIp, 2006.

<sup>5</sup> C. Smith, *Casting the Net: Surveying an Internet Population*, 1997, <http://jcmc.indiana.edu/vol3/issue1/smith.html>

<sup>6</sup> T. Żmijewska-Jędrzejczyk, *Badania internetowe*, in: P. Sztabiński, F. Sztabiński, Z. Sawiński (ed.) *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, IFiS PAN, 2004, p. 245

<sup>7</sup> Ibidem, p. 242.

<sup>8</sup> Panel is surveying the same group of people, thanks to which we can discover certain regularities within the sample.

possible. This is done by installing special software on their computers. One advantage of using this type of research is that it encompasses an actual population of Internet users and not the servers as is the case with *site/server centric research*<sup>9</sup>.

The most popular quantity research on the Internet is through *e-mailing* and *web surveys*. In case of *e-mailing*, a survey is attached to an e-mail inviting the recipient to take part in the research. In case of *web surveys* we are dealing with surveys edited and available on WWW sites<sup>10</sup>.

Regarding quality research, the most popular is the FGI (*focus group interview*) method. Some advantages of this method carried out online are<sup>11</sup>:

- geographic independence, access to people in remote areas,
- no individual dominance, every person has an equal opportunity to speak out,
- anonymity encourages frank and unreserved opinions, therefore it is possible to discuss difficult and controversial issues,
- cost and time savings, no necessity of additional hire and handling costs,
- author of research can take part in it.

There are also some disadvantages<sup>12</sup>:

- no possibility to observe people's non-verbal reactions,
- difficulties with keeping up group dynamism and stimulation of interaction between participants,
- maintaining concentration by author and participants,
- technical problems.

In Poland, the I-Metria S.A. company was the precursor in this type of research<sup>13</sup> with their own Internet panel, I-Metryka<sup>14</sup>.

Chart 1. I-Metryka panel outline

**Recruitment CATI → registration on a panel → [www.imetryka.com](http://www.imetryka.com) → main server → data base → results → analysis → reports**  
→ I-Metryka WEB  
→ I-Metryka Windows

Source: P. Ejdyś, M. Grzanka, *Badania Internetu w Polsce (przykład badań typu site-centric)*, in: B. Bednarek-Michalska, EBIB Elektroniczny Biuletyn Informacyjny Bibliotekarzy, 9/2002, <http://ebib.oss.wroc.pl/2002/38/gemius.php>

*Site-centric* type of research answers the question how many? while *user-centric* research answers the question who? It should be noted that both of these research types are characterized by anonymity and therefore no information can be obtained about the individual users such as their sex, income, or place of living<sup>15</sup>. In case of *site-centric* research, there are two types of Internet mobility measurement. One is a passive audit of log-ins. It analyses files

<sup>9</sup> T. Żmijewska-Jędrzejczyk, *Badania internetowe*, in: P. Sztabiński, F. Sztabiński, Z. Sawiński (ed.) *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, IFiS PAN, 2004, p. 244.

<sup>10</sup> Ibidem, p. 247.

<sup>11</sup> Ibidem, p. 251.

<sup>12</sup> Ibidem, p. 252.

<sup>13</sup> Regarding constant panel, recruitment is geared toward one specific goal, the research is continuous in character as it relates to one subject- cited from B. Gregor and M. Stawiszyński.

<sup>14</sup> B. Gregor, M. Stawiszyński, *Wykorzystanie internetu w badaniach panelowych rynku*, in: M. Sokołowski (ed.) *Oblicza Internetu. Internet a globalne społeczeństwo informacyjne*, Elbląg, 2005, p. 334-335.

<sup>15</sup> <http://www.tezmedium.pl/12041.xml>

generated through WWW servers. It can be done on an individual level but it does not register the Internaut's entire online activity<sup>16</sup>.

## Chart 2. Log-in analysis

WWW server → log-in files → log-in analysis → results

Source: P. Ejdys, M. Grzanka, dz.cyt.

Generally Internet research methods are quite similar to traditional media research methods such as quantity and quality analyses. However, in case of Internet research in comparison to regular media research the proportion of quantity to quality research is much greater. The first ratio is 20:1 while the second is only 4:1<sup>17</sup>. This is because Internet information is digital in form – it is directly accessible to the analyst as well as to the respondent. Traditional information, on the other hand, is in analogue technology (not accessible by computer). The costs involved in traditional research mainly deal with processing analogue information into digital form and reaching the source of information (respondents).

Polish Internet Research (PBI)<sup>18</sup> uses slightly different methodology which combines user-centric research with site-centric. A combination of these two research types enables a presentation of information about Internet traffic as well as the socio-demographic profile of its users. PBI research is done through filling out a website questionnaire, downloading and installing the netPanel programme which registers addresses of the sites visited. However, there are some shortcomings to this method. Lack of IT knowledge makes a number of users reluctant to install this type of programming as they are afraid for the privacy of their e-mails or passwords. Others treat these programmes as a type of surveillance and therefore may not like to be observed in their Internet endeavours. The PBI programme attempts to create a sense of uniqueness and mission. On their website we can read, "If you see this questionnaire, it means that we have invited **you particularly** to take part in the Megapanel PBI/Gemius research".

Internet users are encouraged to fill out the questionnaires several ways. One of them is the *site intercept* technique based on *banners*, or advertising spots with links to surveys. Another, more active technique is the *interstitial window* based on a pop-up window or an automatically opening sub-site with an invitation to take part in the research and an included questionnaire<sup>19</sup>. Surely, there is a series of factors which may influence people's decisions to fill out the Internet questionnaire such as<sup>20</sup>:

- the use of formal/informal style,
- a private/company address inviting to take part,
- information/lack of information regarding the research,
- author's photo/lack of it,

---

<sup>16</sup> P. Ejdys, M. Grzanka, *Badania Internetu w Polsce (przykład badań typu site-centric)*, in: B. Bednarek-Michalska, EBIB Elektroniczny Biuletyn Informacyjny Bibliotekarzy, 9/2002, <http://ebib.osp.wroc.pl/2002/38/gemiup.php>

<sup>17</sup> <http://www.tezmedium.pl/12041.xml>

<sup>18</sup> Polskie Badania Internetu Sp. z o.o. (PBI) is a company created by the owners of the largest domestic Internet websites - Gazeta.pl, INTERIA.PL, Onet.pl, Wirtualna Polska and Rzeczpospolita OnLine. According to its initiators, the goal was to create a standard Internet research procedure in Poland and to aid Internet development.

<sup>19</sup> T. Żmijewska-Jędrzejczyk, *Badania internetowe*, in: P. Sztabiński, F. Sztabiński, Z. Sawiński (ed.) *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, IFiS PAN, 2004, p. 247-248.

<sup>20</sup> K. Sobieszek, *Problem błędu braku odpowiedzi w badaniach internetowych – rozważania teoretyczne*, in: Ł. Jonak, P. Mazurek, M. Olcoń and others. (ed.), *Re: internet – społeczne aspekty medium. Polskie konteksty i interpretacje*, Warszawa, WAiP, 2006, p. 384-385.

- information on how much time is needed to fill out the questionnaire/lack of personalized/general letter to the respondent
- note about respondent's uniqueness/lack of it

### *Internet Research Advantages*

Among Internet research advantages there are<sup>21</sup>:

- lower costs in comparison to traditional research methods
- shorter amount of time necessary to carry out the research and obtain results
- interactivity and instant access to results
- easy access to multimedia
- high elasticity ie. possibility to change questions in progress

Computer supported data collection, such as CAPI (*computer assisted personal interview*) and SAPI (*self-assisted personal interview*) allows for individualizing the questionnaire, adapting the questions to situations and the respondents' expectations. Computer assisted interviews eliminate asking unnecessary questions and enable a focus on key to research issues. Computer assisted surveys brings researchers closer to interviews conducted via the *face-to-face* method<sup>22</sup>.

The CAPI technique offers a wider variety of unused previously opportunities ie. the introduction of audiovisual elements where the researcher can present fragments of films such as a politician's speech or an election spot on a computer screen<sup>23</sup>. This method is particularly useful if the questionnaire is complex and includes complicated rotations, such as the earlier mentioned presentation of audiovisual materials<sup>24</sup>. Especially important in Internet research is the elimination (or minimising) of the survey effect which significantly distorts the results.

There are, however, some reservations as far as online Internet research. One is their representative character. Some people claim that people taking part in this kind of research cannot do not represent any kind of specific population sample. E. Babbie states that until recently the same was thought of telephone surveys which these days are considered reliable. H. Taylor and G. Terhanian attempted to evaluate the results of 1998 US governor and Senate elections. In order to do that they carried out four Internet surveys<sup>25</sup>. Their results were quite surprising – 21 out of 22 (95%) winners were correctly predicted by the respondents. The survey results differed from the actual by 6.8% (in comparison the telephone surveys differed from the actual by 6.2%). According to C. Smith, regarding research sample size, similarly to traditional survey research, is the questionnaire's length and time needed to fill it out. Also respondents were quite interested in where the researchers obtained their e-mail addresses<sup>26</sup>.

Electronic mail has quickly become an essential tool in order to obtain detailed information about selected population samples in spite of earlier reservations as far as its usefulness in research. L. Parker states that using e-mail in data collection is much more effective than through the traditional mail. In his research of AT&T employees, 68% of responses came by e-mail, while only 38% arrived by traditional mail<sup>27</sup>.

<sup>21</sup> B. Gregor, M. Stawiszyński, *Wykorzystanie internetu w badaniach panelowych rynku*, in: M. Sokołowski (ed.) *Oblicza Internetu. Internet a globalne społeczeństwo informacyjne*, Elbląg, 2005, p. 333-334.

<sup>22</sup> Z. Sawiński, *Źródła rozwoju metodologii badań marketingowych*, in: P. Sztabiński, F. Sztabiński, Z. Sawiński (ed.) *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, IFiS PAN, 2004, p. 25.

<sup>23</sup> U. Krassowska, *Badania opinii publicznej w Polsce: trendy, metody i doświadczenia lat dziewięćdziesiątych na przykładzie OBOP*, in: P. Sztabiński, F. Sztabiński, Z. Sawiński (ed.) *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, IFiS PAN, 2004, p. 31.

<sup>24</sup> *Ibidem*.

<sup>25</sup> E. Babbie, *Badania społeczne w praktyce*, Warszawa, PWN, 2004, p. 299.

<sup>26</sup> C. Smith, *Casting the Net: Surveying an Internet Population*, 1997, <http://jcmc.indiana.edu/vol3/issue1/smith.html>

<sup>27</sup> L. Parker, *Collecting data the e-mail way*, Training and Development, 1992, p. 52-54.

At the same time, it is not so that the Internet does not allow the possibility to choose respondents according to criteria of interest. This can be done through the pop-up window option. It is a graphic element which appears as a separate window after a user has entered a WWW site or a server. Through this window the respondent may be asked his data such as sex, age, address, education, etc. If he meets the criteria we are looking for, he may proceed to the main part where the questionnaire is located. If he does not, the window closes itself and the user does not fill out the questionnaire.

### *Internet Research Shortcomings*

Although the Internet allows the researchers exploration, there is the issue of the quality of the acquired data and a generalization of the results obtained. It appears that the population of internet users significantly differs demographically, socially and psychologically from those who do not. In the US, Internet users are usually while, young and have children. The sample, therefore, may be biased. Additionally, the internet users may either agree or not when it comes to filling out the questionnaire<sup>28</sup>. It turns out that the percentage of filled out surveys on-line are lower than those via mail or telephone. Another shortcoming regards longitudinal research. Internet users often change their e-mail addresses which makes them much harder to contact subsequently<sup>29</sup>. In Internet surveys there is lack of control over who the respondent really is, which affects the quality of the data acquired<sup>30</sup>. Due to low penetration of Internet users, there is a problem with representativeness and the possibility to extrapolate the results onto the overall population. We are dealing here with a so called *coverage error* which means that not all individuals within the population have a chance to be selected for the research. For example, not all people in Poland have access to the Internet, in fact only a minority has<sup>31</sup>, hence too many individuals have no chance to be in the sample<sup>32</sup>.

### *Research Subjects' Protection of Rights*

Researchers point out two potential sources of risk regarding Internet research. These are:

- damage as a result of users' direct actions ie. emotionality or manipulation
- damage as a result of breach of confidentiality

Protection of Internet users taking part in the research rights should be based on assurance of freedom of choice in taking part, anonymity, guarantee of withdrawal from the research at any time, possibility of contacting the researcher in case of any doubts, parent/guardian consent (in case of minors), and not sending survey to people who have previously refused to take part in the research<sup>33</sup>.

### *Problematic Issues*

---

<sup>28</sup> R.E. Kraut, J. Olson, M. Manaji, A. Bruckman, J. Cohen, i M. Couper, *Psychological Research Online: Opportunities and Challengep*. American Psychologist, 2004, p. 107-108.

<sup>29</sup> Ibidem, p. 108.

<sup>30</sup> T. Żmijewska-Jędrzejczyk, *Badania internetowe*, in: P. Sztabiński, F. Sztabiński, Z. Sawiński (ed.) *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, IFiS PAN, 2004, p. 245.

<sup>31</sup> In March 2006, 26% of respondents declared Access to the Internet, data according to CBOS, "Internet i komputery: wyposażenie gospodarstw domowych, sposoby i cele korzystania".

<sup>32</sup> K. Sobieszek, *Problem błędu braku odpowiedzi w badaniach internetowych – rozważania teoretyczne*, in: Ł. Jonak, P. Mazurek, M. Olcoń and others. (ed.), *Re: internet – społeczne aspekty medium. Polskie konteksty i interpretacje*, Warszawa, WAiP, 2006, p. 369.

<sup>33</sup> T. Żmijewska-Jędrzejczyk, *Badania internetowe*, in: P. Sztabiński, F. Sztabiński, Z. Sawiński (ed.) *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, IFiS PAN, 2004, p. 254-255.

—

When discussing Internet research, we should take into account the definition which most research companies use to describe Internet users. According to a generally accepted definition, an internaut is a person who uses the Internet at least once a month, a definition first used by the *Internet Advertising Bureau* (USA). This definition should, however, does require redefining as it should not encompass people who have only used the Internet once in their lifetime and therefore cannot be defined as active users.

Additionally, the term *Internet usage* can also be problematic. It may seem trivial initially but it does not mean the same thing to different people. For some it may be, for example, active usage of the Web, ie. looking through WWW sites or using e-mail, for others it may be just passive access to communicators signalling that the person is connected and available to users of the given communicator. Frequently, people working in offices at computer terminals have Internet access turned on all day but will only actually use it for a few minutes. Therefore, it is impossible to estimate real usage.

In reality there are many more questions than answers, however, posing them may eventually lead to coming up with solutions. Science cannot be indifferent to winds of change of which the Internet and computers definitely are.