

USING NON-REACTIVE OBSERVATION METHODS IN THE INTERNET: METHODOLOGICAL ASPECTS AND CHALLENGES

Wiedmann, K.-P./Buxel, H.
Chair for General Management and Marketing II
University of Hanover
Königsworther Platz 1
D-30167 Hannover, Germany
Tel.: 0049 / 511 / 762-4862
Fax: 0049 / 511 / 762-3142
Homepage: www.m2.uni-hannover.de
E-mail: wiedmann@m2.uni-hannover.de
holger_buxel@droege.de

The Internet presents a variety of new challenges for companies' market research. It requires conventional survey instruments to be adapted for online use, for example. More significantly, it provides new possibilities for collection of data on consumer behavior. Internet users automatically leave different types of tracks when they visit Web sites. Such tracks can be systematically recorded and collected and then used to prepare "customer profiles".

No observer, watching users' every step and precisely recording their behavior, is required for such data collection: customer-behavior data is generated largely automatically as part of information/data exchange in the Internet. In keeping with the intensive, consumer-oriented usage of the Internet as an information and transaction medium, such data can become very comprehensive. Its level of detail, with regard to consumer behavior, can far exceed that achieved by the traditional data-collection methods used in the non-virtual realm.

One example: observation in an online shop

When a customer visits an online shop, for example, his every step along the shop's virtual "shelves" can be observed automatically. Detailed, time-stamped records can be made of all of the customer's "pauses" and all of his actions at "forks in the road". And such detailed observation need not stop with the customer's preferred paths. Records can easily be kept of every product that the customer considers as he "stands" in front of a "shelf", as well as of the order in which he studies products and the time he spends on each product – and all of this can be recorded regardless of whether the customer ultimately purchases the products in question or not. When such information is linked with information about the customer's identity and demographic categories, a comprehensive picture of the customer can emerge, one so extensive that the customer becomes virtually "transparent". Such pictures can be used as a basis for customer-specific design of marketing instruments.

Observation in the Internet: what's really new about it?

In market research, there is nothing new about the idea of observing customers in order to gain a basis for designing

marketing instruments. For example, this idea is applied in studies of customer traffic in retail stores, of purchase behavior and of handling and use. But in spite of the unquestioned importance of observation in market research, this data-collection method is used relatively seldom – only in a few, selected problem areas – in the non-virtual world. The primary reason for this is that observation of customers, in many real-life situations, presents considerable problems [2]:

- Observation in the non-virtual world can be very time-consuming, involved, and expensive. The time and effort that must be invested in gaining the data grow as a function of the desired degree of observational detail.
- In the non-virtual world, an observer who wishes to obtain genuine, comprehensive and detailed insights into a consumer's behavior – such as insights into his professional life, recreational activities and purchasing behavior – must accompany the consumer everywhere he goes and record all of his behavior in detail. Obviously, any person would consider such observation very annoying and would hardly tolerate it in actual practice.

Outside of the Internet, therefore, consumer behavior has normally been observed only in order to answer small numbers of closely defined questions. Online observation, on the other hand, is not subject to the above-described limitations; the Internet is a platform that supports data collection with virtually no extra expense and effort. Consequently, the Internet permits detailed observation of consumer behavior, in many areas of daily life [2]:

- From a technical standpoint, the Internet makes it easy to determine what users visit what sites – and for how long. The main reason for this is that a great deal of relevant information about consumer behavior is generated and recorded automatically as a result of Internet data exchange between providers and consumers.
- The Internet's importance as a space for information exchange and for transactions, as a space covering more and more areas of life (professions, hobbies, recreation, etc.), has been growing very rapidly. As a result, such automatically generated data is not limited to small slices of consumers' lives, as is usually the case with data in comparable conventional (non-virtual) studies. Depending on how intensively con-

sumers use the Internet in various areas of their lives, such data can provide a relatively comprehensive picture about consumer behavior. In extreme cases, this picture can cover almost all areas of life.

- Because the data is generated automatically, data collection and processing can be largely automatic as well. As a result, consumer profiles can be produced at virtually no extra cost and effort. Suitable profiling tools are available that can automatically collect and analyze data, in keeping with pre-defined criteria, and make it available for marketing purposes.
- Since automatically generated data is already in digital form, it can be used directly and quickly in analytical processes. Depending on the speed with which relevant analysis can be carried out, it can even be possible to generate results in real time and apply them directly to transactions with customers. This, in turn, makes it possible to respond to all of consumers' movements – and thus to shape communications individually and responsively, as a function of observed behaviors.

How observation in the Internet works

Most of the procedures and options for observational collection of data in the Internet can be classified as "non-reactive". The term "non-reactive" means that the consumer does not influence or distort the data collection by reacting to the observation, simply because he is not aware that his behavior is being recorded and that the relevant data is being analyzed and used. Collection of consumer-behavior data in the Internet, via non-reactive processes, thus is "quasi-biotic"; it can proceed relatively free of disruptive factors (cf. in greater detail [3, p.157ff.], as well as [1]).

Data collection via log files

One way of observing consumer behavior in the Internet is to use log files. Log files automatically record information about exchanges of data between consumers and Web sites; they do this by recording every requested instance of data transmission and reception. They thus are relatively comprehensive data sources that can provide valuable information about consumer behavior during visits to Web sites. The task of log-file analysis is to link user information and attributes with requests recorded in such files, in order to derive specific behavior patterns. The methods for doing this are not always trivial, however. Nonetheless, such data, within the framework of observation, can support conclusions about usage and, thus, behavior, of individual consumers or of consumer groups as they access Web sites. It can be used to answer questions such as:

- When did the person in question visit a given Web site?
- What areas in the Web site did the person visit?
- How much time did he spend in these areas?
- What information did he request, and in what order?

123.456.78.8	--	[09/May/2001:03:04:41 - 0500]	"Get Buxel.html HTTP/1.0"	
123.456.78.8	--	[09/May/2001:03:04:51 - 0500]	"Get Wiedmann.html HTTP/1.0"	
123.456.78.8	--	[09/May/2001:03:05:32 - 0500]	"POST /cgi-bin/p1 HTTP/1.0"	
123.456.78.8	--	[09/May/2001:03:05:41 - 0500]	"Get Buxel.html HTTP /1.0"	
123.456.78.8	--	[09/May/2001:03:05:59 - 0500]	"Get Wiedmann.html HTTP /1.0"	
123.456.78.8	--	[09/May/2001:03:06:30 - 0500]	"Get Frenzel.html HTTP /1.0"	
123.456.78.8	--	[09/May/2001:03:07:11 - 0500]	"Get Buckler.html HTTP /1.0"	
123.456.78.8	--	[09/May/2001:03:07:45 - 0500]	"Get Halstrup.html HTTP /1.0"	
123.456.78.8	--	[09/May/2001:03:12:23 - 0500]	"Get Meissner.html HTTP /1.0"	
123.456.78.2	--	[09/May/2001:05:05:11 - 0500]	"Get Buxel.html HTTP /1.0"	
123.456.78.3	--	[09/May/2001:05:06:03 - 0500]	"Get Walsh.html HTTP /1.0"	
123.456.78.5	--	[09/May/2001:05:06:05 - 0500]	"Get robots.txt"	
233.999.79.4	--	[09/May/2001:05:06:07 - 0500]	"Get Buxel.html HTTP /1.0"	
IP Address	Status	Size	Referrer	Agent
123.456.78.8	200	3290	-	Mozilla/3.04 (Win95,I)
123.456.78.8	200	5450	Buxel.html	Mozilla/3.04 (Win95,I)
123.456.78.8	200	5096	Wiedmann.html	Mozilla/3.04 (Win95,I)
123.456.78.8	200	3290	-	Mozilla (IE4.2,WinNT)
123.456.78.8	200	5450	Buxel.html	Mozilla (IE4.2,WinNT)
123.456.78.8	200	1000	Wiedmann.html	Mozilla (IE4.2,WinNT)
123.456.78.8	200	2020	F.html	Mozilla/3.04 (Win95,I)
123.456.78.8	200	3030	Frenzel.html	Mozilla (IE4.2,WinNT)
123.456.78.8	200	4040	Wiedmann.html	Mozilla/3.04 (Win95,I)
123.456.78.2	200	3290	-	Mozilla/3.04 (Win95,I)
123.456.78.3	200	4040	Buxel.html	Mozilla/3.04 (Win95,I)
123.456.78.5	200	1020	-	Mozilla/3.04 (Win95,I)
233.999.79.4	200	3290	-	Ultrasseek

Fig. 1: Example of a log file (ECLF)

Data collection via environment variables

Data collection via environment variables is another form of non-reactive data collection. Like log files, such variables keep records of events that are caused by user actions. A Web server records environment variables whenever a user executes a program on the server by running a CGI script. The information provided by environment variables is similar to that provided by log files.

Data collection via cookies

In Internet consumer-data collection, "cookies" can be used as complements to log files. Cookies are files that Web sites place on consumers' computers, in order to store information that they (the Web sites) can check again later. Where cookies contain ID codes that identify consumers, the cookies can be used to observe how specific consumers access Web sites.

Use of cookies in profiling can present some problems:

- In general, cookies function as applications that can be used "secretly", i.e. even without the consumer's explicit consent. As a result, their use often presupposes their acceptance by consumers, and this has sparked concerns about this technology with regard to privacy issues.
- Most users do not understand how cookies work. What is more, users do not have the opportunity to assess cookies' content – users are never told what information about them is going into their cookie files.

In spite of these problems and others, cookies, in combination with log files, are an important instrument for observation. They are thus now used extensively.

Data collection via software agents

Another client-side data-collection method is use of applications or programs that support data collection on the consumer's computer, in non-reactive form, as well as transmission of the relevant data to a server or other desti-

IP Address	Userid	Time	Request (Method/URL/Protocol)
------------	--------	------	-------------------------------

nation addresses. Examples include e-mail attachments and Web-site programs that self-extract and/or self-execute upon arriving at the client computer. The most common representatives of this class are software agents such as Java scripts, Java applets and Active X components. In data collection, all such software agents access and collect data that is generated in the client computer. Consequently, their use, like use of cookies, presupposes the consumer's implicit consent – and, once again, such consent may not really be present.

Data collection via modified browsers

Another data-collection technique involves modifying the browsers that consumers use to surf the Internet. In the modification, browsers are given an additional log-function component that causes them to automatically record the behavior of test consumers and send the resulting data to an external source for evaluation. To date, such modification has been carried out for research purposes only; no such browsers have been made available for commercial use.

Data processing: the challenge of converting the data into a usable form

Once it has been collected, data must be processed in such a way that it can be used to prepare observation profiles. Due to the manner in which it is recorded and presented, non-reactively collected data has to undergo a number of transformations before it can be used to generate profiles. The relevant data processing is a complex process whose design, of course, influences the structure, quality and (ultimately) usefulness of profiles. Once data has been properly processed, it must be stored within a suitable data structure that supports the relevant study objectives. Buxel [2] presents a description of the relevant procedures.

Legal aspects of observation in the Internet

German lawmakers have recognized that consumers can be observed non-reactively while they are in the Internet, and they have amended Germany's laws to take account of this fact. In Germany, market researchers who use non-reactive observation methods in the Internet must conform to the data-protection provisions of the TKG (Telecommunications Act), TDDSG (Act on Data Protection in Teleservices), MDStV (State Agreement on Media Services), BDSG (Federal Data Protection Act) and TDG (Act on Use of Teleservices). The following general statements can be made in this area ([5], provides greater detail):

In general, non-reactive data collection and usage is permissible, without the user's (i.e. consumer's) consent, where none of the collected data is personal data or personally correlatable data. Where usage profiles are prepared solely on the basis of pseudonyms that cannot be correlated to any specific, identified person, the relevant observation methods do not violate any data-protection provisions.

Where the observation collects personal data or personally correlatable data, profiling is subject to the consumer's/user's explicit consent. All efforts to obtain such

consent must comply with the provisions of Art. 2 (2) TDDSG (or Art. 11 (2) MDStV) in conjunction with Art. 3 TDDSG (or Art. 12 MDStV).

In Germany, collection of personal data is not subject to the consumer's consent when it is carried out by an Internet-services provider who is located abroad and whose activities do not fall within the scope of application of German data-protection provisions. On the other hand, in such cases the legitimacy of the data collection is governed by the legal provisions of the country in which the provider is located.

Privacy Issues

Use of non-reactive procedures to observe behavior in the Internet is seen as problematic by consumer advocates, who maintain that such observation could violate consumers' rights to privacy, and related rights. The concerns expressed about computer-based observation in online environments in general, and about development and use of observation in electronic commerce in particular, can be divided into three different levels [6, p.106]:

- In general, consumers and the public may consider data collected for the purpose of customer profiling to be private (or at least confidential) information. Where this is the case, non-reactive collection of such data is considered inherently unethical or immoral and is seen as an invasion of privacy.
- Even where the data itself is not seen as private information, the non-reactive collection of which would be inherently unethical or immoral, consumers and the public may still have reservations about the methods used to collect the data. In such cases, criticism of observation focuses on the non-reactive aspects of the data collection.
- And even when the data collection and the observation methods are seen as acceptable, consumers may still be concerned about the many ways in which the data is used.

Specific concerns about comprehensive observation in the Internet focus on the ethical and moral aspects related to violation of individual privacy. "Privacy is an essential component of individual autonomy and dignity. Our sense of liberty is partly defined by the ability to control our own lives – whether this be the kind of work we undertake, whom we choose to associate with, where we live, the kind of religious and political beliefs we hold, or the information we wish to divulge about ourselves." [4, p.398f.]

Criticism of data-collection methods and concepts concentrates on three main aspects [2]:

Secrecy

Many Internet users are not aware that they leave tracks that can be recorded non-reactively – that Web-site operators can collect such data from consumers without consumers' being aware of this. Providers of Internet eCommerce services can prepare profiles of consumers without

the affected consumers' knowing that such profiling is taking place. Because of the "invisibility" of such data collection, many consumers do not know what is happening with their data. Relevant criticism thus considers use of cookies, for example, to pose a danger to consumers' right to privacy, since surfers normally are unaware of cookies' arrival unless their browsers are set to provide relevant warnings. Recent studies give an indication of the extent to which criticism of cookies can be considered justified:

- A 2000 study of the *Pew Internet & American Life Project* found that only 43 % of all Internet users are aware that cookies can be used to observe activities in the Internet.
- These findings were confirmed by a 2000 study of *CyberDialogue*, which found that 42 % of all Internet users do not know what cookies are and how they work.

Involuntary participation

Another focus of criticism, closely related to the focus on secrecy, is that many non-reactive collection procedures, such as "Web bugs" or "packet-sniffing" technologies, can theoretically be used to obtain and record data about consumers without the consumers' consent. Consumers' resulting loss of control over their own data is also seen as a potential invasion of privacy.

Lack of protection

Yet another point of criticism, closely related to those focusing on secrecy and consumers' involuntary participation, is that consumers lack ways to protect themselves against most non-reactive data-collection procedures. There are three reasons for the perceived lack of protection:

- Before consumers can protect themselves against undesired data collection, they must be aware that they can be observed. Where data-collection procedures are used secretly, or where their method of function is unclear, consumers, logically enough, have little opportunity to protect themselves.
- What is more, no protection is even available for many data-collection procedures, due to system-inherent reasons – i.e. even consumers who are aware of the collection cannot stop it (except by avoiding Web sites where they presume profiling to be taking place). This is especially the case with data collection via log files and environment variables. On the other hand, such data-collection procedures cannot be used, by themselves, to collect personal data.
- Consumers are able, theoretically, to block some data-collection procedures – for example, by setting their browsers to reject cookies or certain applications. On the other hand, consumers must have a certain amount of know-how in order to make use of the available protection, and not all Internet users are so well-informed. For example, consumers can protect themselves against cookies only if they know what cookies are and/or have learned how to destroy or modify the content of cookie files. A 2000 study of the *Personal-*

ization Consortium revealed that 57 % of all Internet users do not know how cookies can be rejected.

Challenges in use of non-reactive observation procedures

Non-reactive behavior observation in the Internet provides many opportunities and new business areas for market research. To date, market researchers have given little attention to the relevant methods. The extent to which the new opportunities can be tapped will depend on how quickly market researchers can develop (improve) suitable approaches for non-reactive collection, processing, storing and analyzing of data (i.e. data that is suitable for non-reactive collection) in the Internet and how fast market researchers acquire the necessary know-how to apply the new methods in their work. In addition, non-reactive data collection in online environments touches upon privacy issues that will require development of, and compliance with, privacy guidelines – as well as efforts to engender trust in test persons. The market-research community is now challenged to develop and implement relevant guidelines and measures programs of the sort currently being discussed in the U.S..

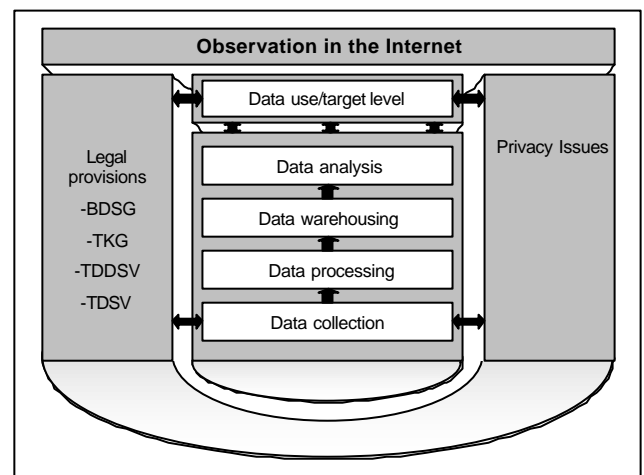


Fig 2: Challenges in use of non-reactive observation procedures in the Internet

REFERENCES

- [1] Batinic, B./Bosnjak, M. "Fragebogenuntersuchung im Internet", in: Batinic, B. (ed.) *Internet für Psychologen*, Göttingen 1997, p. 221-243.
- [2] Buxel, H. *Customer Profiling im Electronic Commerce: Methodische Grundlagen, Anwendungsprobleme und Managementimplikationen*, Aachen 2001.
- [3] Janetzko, D. *Statistische Anwendungen im Internet*, Munich 1999.
- [4] Marx, G.T./Sherizan, S. "Monitoring on the job", in: Forester, T. (ed.) *Computers in the human context: Information technology, productivity, and people*, Cambridge 1989, p. 397-406.
- [5] Moos, F. "Datenschutz im Internet", in: Kröger, D./Gimmy, M.A. (ed.) *Handbuch zum Internetrecht: Electronic Commerce – Informations-, Kommunikations- und Mediendienste*, Berlin u.a. 2000, p. 411-441.

Wiedmann, K.-P./Buxel, H.

- [6] Peters, T.A. *Computerized monitoring and online privacy*, Jefferson u.a. 1999.