Customer profiling in e-commerce: Methodological aspects and challenges

Received: 3rd December, 2001

Klaus-Peter Wiedmann

Professor Dr Wiedmann holds the Chair of General Management and Marketing II at the University of Hannover, Germany.

Holger Buxel

was a research associate for the Department of Marketing II at the University of Hannover, Germany from 1999 to 2001. In 2001, he became a consultant for Droegge & Comp. AG, Dusseldorf, Germany.

Gianfranco Walsh*

is Senior Lecturer at the University of Hannover's Department of Marketing II, Germany. He also consults on a variety of marketing challenges, especially relating to empirical marketing research and cross-cultural consumer behaviour.

Abstract Online profiling is the collection of information about Internet surfing behaviour across many different websites for the purpose of formulating a profile of users' habits and interests. While marketers have long profiled large consumer groups on the basis of demographics, online profiling allows companies to collect information from individuals across a wide range of traits. A company's ability to build and strengthen long-term customer relationships via individualised e-commerce offers will depend on its ability to use customer data to plan, develop and control interactions with its customers. Despite the importance of customer profiles in e-commerce, the methods for producing them have rarely been investigated in marketing research. This paper aims to close this gap by providing an overview of key areas and issues of customer profiling in e-commerce. More recently, this controversial practice has seized the public's attention because ethics and privacy are concerned.

INTRODUCTION

Since the early 1990s, and especially in connection with the one-to-one-marketing concept, discussion has been increasing on how to manage relationships between companies and individual customers. Long-term, revenue-maximising relationships with attractive customers are especially relevant as the focus of marketing interest shifts from companies' share of market to their share of customer. So called 'one-to-one' concepts focus on

learning from relationships with customers. Over time, companies can obtain comprehensive pictures of customers that can be used to cater better for customer needs. As companies are eager to increase the duration of a customer relationship, actions to increase the completeness and sharpness of such pictures, and to apply such pictures to individualised transactions, automatically lead to development of customised offers that have a positive effect on customer loyalty.

Dr Gianfranco Walsh*
Department of Marketing II,
University of Hannover,
Koenigsworther Platz 1,
30167 Hannover, Germany.
Tel: +49 511 762 4540
Fax: +49 511 762 3142
e-mail:
walsh@m2.uni-hannover.de

In e-commerce, the quality of one-to-one marketing grows with further developments in information technology, which help to improve quality of customer data and to build lasting relationships with its customers. The Internet allows marketers to obtain detailed customer-related information and enables sellers to individualise their marketing instruments on a large scale and in real time. This process is referred to as personalisation, which involves offering individualised products. The ability to do so depends very much on whether the company has data with which it can selectively plan, shape and monitor its individual interactions with its customers. To make use of this possibility, a company needs to collect, compile and continually update extensive information about its (potential) customers and their needs and preferences.

Customers leave a wide range of tracks as they move through the Internet. Such tracks, if systematically recorded and collected, can be used for formulating a 'customer profile' or representation of users' habits and interests that can support one-to-one marketing concepts in e-commerce and become useful with respect to a broad range of different purposes. Companies have been showing growing interest in creating customer profiles in e-commerce. It is estimated that more than 100 million Internet users have been profiled by DoubleClick (a large network advertiser) alone. As consumers and other interest groups have become increasingly aware that companies may be watching and recording their online behaviour, however, they have begun criticising profiling activities on ethical, moral and privacy-related grounds.

This paper constitutes a threefold contribution. First, it describes the basic methods used to create customer profiles.

Secondly, it discusses central problems in using profiling techniques in e-commerce, from the perspective of affected and interested groups. Thirdly, management implications are offered and discussed.

METHODS OF CUSTOMER PROFILING

When users move through the Internet, their every step through the virtual pathways can be followed and can be recorded in detail and combined with timing data. Recording what information and services a user seeks or uses, the order in which they are used and the amount of time spent on reviewing information present no particular technical problems. When such data are combined with information such as the user's identity and demographic profile, a comprehensive picture of the user can result, a picture that can closely approximate the fanciful image of the 'glass customer'. The possibilities for creating and using detailed profiles are centrally enhanced by the following factors:

- 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 behaviour 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 rapidly. Depending on how intensively consumers use the Internet in various areas of their lives, such data can provide a relatively

- comprehensive picture about consumer behaviour.
- because the data are 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.¹ This is accomplished with the help of suitable profiling tools that automatically collect and analyse data and make them available for marketing purposes. By conducting profiling and updating profiles continuously, companies can obtain profiles of very high quality in a very short time. Since data are already in digital form, they can be used directly and quickly in companies' analytical processes, which makes it possible to shape interaction individually and responsively, as a function of observed behaviours.

The process of using customer profiles begins with the production of a suitable database from which conclusions about customers' needs and interests can be drawn. In e-commerce, such data can be obtained by various procedures that differ in the types of data they can generate, in their methods of function and in the privacy-related problems linked with their use.

Much of the customer-related data available in e-commerce are in a form that is not conducive to efficient analysis and interpretation. This is due to technical factors and those relating directly to the data-collection procedures themselves. On the other hand, various customer behaviour factors can often render collected data unsuitable for direct production of customer profiles. As a rule, therefore, collected data have to be processed before they can be used.

Once they have been suitably processed, customer data have to be transferred to a data warehouse, which

supports profiling. In the data warehouse, data are stored in keeping with the relevant profiling objectives — and modified over time, since detailed customer profiles normally become available only through long-term observation of customer behaviour.

Data collection procedures

Internet-based data collection can be reactive and non-reactive. The term reactive implies that a customer is aware that his/her behaviour is being recorded and, possibly, used, which may cause the customer to react. Gathering of consumer-behaviour data in the Internet via non-reactive processes is, thus 'quasi-biotic' — it can proceed relatively free of disruptive factors, while a customer exposed to reactive collection procedures will be aware that his/her behaviour (ie data about him/her) is being recorded.

Non-reactive data collection procedures are especially suitable for collecting data about observable 'surf' behaviour in the Internet, while reactive data collection methods are best used in situations in which sociodemographic and psychographic characteristics need to be recorded, as well as characteristics that could not be derived from observing the usage behaviour of customers.

Figure 1 provides an overview of the most important data collection procedures used in customer profiling.

Non-reactive data collection

Non-reactive procedures focus on recording data resulting from customers' usage behaviour in websites. Common procedures and technologies include collection via log files, via Common Gateway Interface-based (CGI-based) files and via applications designed especially for Internet data collection.

Figure 1: The most important data collection procedures used in customer profiling

Data collection via log files

Data transport in the World Wide Web is based on the client server principle. Software programs that access services are referred to as 'clients', while programs that provide services are referred to as 'servers'. Log files automatically record information about exchanges of data between servers and clients and use codes to describe specific types of website access. Since they record every requested instance of data transmission and reception, they are relatively comprehensive data sources that can provide valuable information for profiling. Log files come in various formats, although the formats differ only insignificantly in terms of the information they include. Currently, the extended common log format (ECLF) is the most common.

The basic spaces of an ECLF file (see Table 1) provide the following

information: IP address (Internet address of the computer from which the user retrieves data from the server); User ID (data that identify a user when website access is subject to user authentication); Time (shows the time at which the client retrieves data from the server); Request (shows what data were retrieved from the server); Status (shows whether data exchange between the client and server was successful); Bytes (shows the number of bytes sent in response to the request); Referrer (shows the previous URL (Uniform Resource Locator)) and Agent (text string that can contain information about the client's operating system and browser software).

In profiling, such data can be used to reach conclusions about the Internet usage behaviour and, thus, characteristics of individual customers or customer groups (Table 2).

Table 1: Example of an ECLF log file

IP Address	User id	Time	Request (Method/URL/Protocol)	Status	Size	Referrer	Agent
123.456.78.8	--	[09/May/2001:03:04:41 - 0500]	"Get Buxel.html HTTP/1.0"	200	3290	--	Mozilla/3.04 (Win95,i)
123.456.78.8	--	[09/May/2001:03:04:51 - 0500]	"Get Wiedmann.html HTTP/1.0"	200	5450	Buxel.html	Mozilla/3.04 (Win95,i)
123.456.78.8	--	[09/May/2001:03:05:32 - 0500]	"POST/cgi-bin/p1HTTP/1.0"	200	5096	Wiedmann.html	Mozilla/3.04 (Win95,i)
123.456.78.8	--	[09/May/2001:03:05:41 - 0500]	"Get Buxel.html HTTP/1.0"	200	3290	--	Mozilla (IE4.2,WinNT)
123.456.78.8	--	[09/May/2001:03:05:59 - 0500]	"Get Wiedmann.html HTTP/1.0"	200	5450	Buxel.html	Mozilla (IE4.2,WinNT)
123.456.78.8	--	[09/May/2001:03:06:30 - 0500]	"Get Frenzel.html HTTP/1.0"	200	1000	Wiedmann.html	Mozilla (IE4.2,WinNT)
123.456.78.8	--	[09/May/2001:03:07:11 - 0500]	"Get Buckler.html HTTP/1.0"	200	2020	F.html	Mozilla/3.04 (Win95,i)
123.456.78.8	--	[09/May/2001:03:07:45 - 0500]	"Get Halstrup.html HTTP/1.0"	200	3030	Frenzel.html	Mozilla (IE4.2,WinNT)
123.456.78.2	--	[09/May/2001:03:12:23 - 0500]	"Get Meissner.html HTTP/1.0"	200	4040	Wiedmann.html	Mozilla/3.04 (Win95,i)
123.456.78.3	--	[09/May/2001:05:05:11 - 0500]	"Get Buxel.html HTTP/1.0"	200	3290	--	Mozilla/3.04 (Win95,i)
123.456.78.5	--	[09/May/2001:05:06:03 - 0500]	"Get Walsh.html HTTP/1.0"	200	4040	Buxel.html	Mozilla/3.04 (Win95,i)
233.999.79.4	--	[09/May/2001:05:06:05 - 0500]	"Get robots.txt"	200	1020	--	Mozilla/3.04 (Win95,i)
	--	[09/May/2001:05:06:07 - 0500]	"Get Buxel.html HTTP/1.0"	200	3290	--	Ultraseek

Table 2: Information contained in server logs

Information	Explanation
User identification	Analysis of IP addresses and user IDs can provide information about who the users of a particular service are.
Interests	URLs provide information about users' interest in particular services. Analysis of request sequences can also show important relationships between individual services.
Usage periods	Analysis of time spaces can reveal the times of day when users access services and the amounts of time they spend with services when they access them. The duration of users' visits to websites can serve as an indicator for the interest being shown in the relevant offers. In addition, time-space analysis can also show distribution, over time, of access to the relevant website (when and how often).
Related interests	The referrer provides information about the user's visit before they access the relevant site. Analysis of such data can support conclusions about significant relationships to other websites, and about interests of relevant consumers.

Data collection via CGI-based files

Instead of retrieving static HTML documents, a client computer can run a program known as a CGI script on the web server. This is protocol that can be used to communicate between web forms and a program. Like log files, CGI-based files keep records of events that are caused by user actions. CGI-based files can provide information about users' access patterns, consisting of the services that users request and the time users spend with each service, hence supplementing log files.

Data collection via special applications

Other procedures can also be used for non-reactive data collection, such as cookies (software agents that providers can use for data collection, special browsers with source code modified for enhanced data collection), packet-sniffing technologies and web bugs, invisible images that produce usage information.

Reactive data collection

Reactive data collection is concerned with consumer characteristics that cannot be revealed with the behaviour tracking technologies outlined above. While

almost any sort of data might be available to reactive data collection, the following are of particular interest with regard to the creation of customer profiles.

Identification data include such data as user name, address, telephone and fax number, e-mail address and, possibly, the URL of the user's own homepage. With such data, a provider is in a position to send the user specially tailored offers or even to provide special support for real-life (non-virtual) offers.

Descriptive Data can be used to determine customers' basic goods and services preferences and, thus, to generate relevant, applicable conclusions about customers (Table 3). Apart from data that link user behaviour data with specific persons, there are many types of (user-) descriptive data that are very difficult to obtain through non-reactive data collection. Relevant data of particular interest in connection with profiling includes information about the customer's current purchasing behaviour and purchasing history, sociographic and psychographic data. Information about purchasing behaviour, for example, indicates which products a customer prefers at which prices, and which methods of payment. Typical sociographic data are, for instance, date

Table 3: Collection characteristics of reactive procedures

Procedures	Defining characteristic
Closed form spaces	Space for manual entries; no latitude on the topic level
Open form spaces	Space for manual entries; latitude on the topic level
Selection menus	Presentation of defined, fixed answer categories

of birth, sex, education, occupation, and income. Psychographic data include information about customers' interests, hobbies, recreational activities, opinions and lifestyles.

Communications data provide such information as when, and in what form, a customer contacts a provider. They thus can be used in the generation of company-related data. Communications data can be collected reactively whenever a customer enters directly into dialogue with a provider, ie not only when a customer simply surfs through a website. Relevant data in this context include information about the subject of communications (product, negotiation, order, etc.) or the occasion for communications (eg reaction to certain matters, spontaneous query, issue of an order), etc.

On a concept level, the means by which identification data, descriptive data and communications data can be obtained in e-commerce can be broken down into closed and open form spaces and selection menus (see Table 3).

Data processing as a preliminary phase in profile formation

After the data are collected, they must be processed to permit creation of profiles. With reactively collected data, the primary data processing problems have to do with useful conversion of data from open form spaces, filtering out of non-relevant data and wrong entries and treatment of cases in which data missing.

With non-reactively obtained data, data processing challenges normally arise through the relevant data exchange between client and server. The reason for this is that the means by which non-reactively collected data is recorded and represented require that the data undergo a number of transformations before it can be used in profiling. Data processing then becomes a complex process that influences the formation, quality and, ultimately, usefulness of resulting customer profiles.

Non-reactively obtained data consist of records of exchange (requesting and sending) of page files between clients and servers. A user's actions in a website (eg mouse clicks) normally lead to a change in the web content displayed and thus to the generation of a new 'picture' in a user's browser windows. In the ECLF (eg data collected with other procedures), such actions appear as records of all page files exchanged between client and server in processing the client's request. In practice, a user's single click on a website link generates a new browser image consisting of several new graphic elements and HTML files, each of which leaves a separate entry in the log.

The provider's problem in interpreting such records is that their data do not readily support conclusions regarding the usage behaviour, especially in the light of specific questions of interest to the provider. In an analysis of usage behaviour, interest must focus not on the various individual transmitted files, but on aggregates, since a user normally does not request 'x' page files with texts and

'z' graphics for their browser; instead the user requests a website with specific information. The key data for interpretation of online behaviour include: user, server sessions, page views and episodes:

- a user is a person who uses a browser to retrieve page files from one or more servers
- a page view (also referred to as page impression) is a set of page files needed to generate a web browser display at a given time. A page view consists of all those elements that the user is presented in their browser window in response to a request via a mouse click
- over longer periods of time, it becomes very likely that some users will visit a given website more than once. For this reason, all data requests that are recorded in a log and that can be assigned to a specific user have to be divided into individual sessions, if conclusions are to be drawn about the way the user's behaviour is distributed over time. A 'server session' (also referred to as visits) refers to a process of continuous usage of a website, by a client, expressed in terms of related page views. Usage processes thus include every instance of technically successful and external access, by an Internet browser, to the website
- in addition to page views, 'episodes' are relevant. Episodes are semantically meaningful sub-areas of server sessions. Examples of episodes include website pages devoted to particular categories of goods, eg sporting goods or food, as well as pages that present company-oriented data.

Processing of non-reactively obtained data can be described as a four-step process (see Figure 2).² The first step

consists of filtering out irrelevant elements to identification of usage behaviour and to later analysis ('data cleaning'). In a second step, the data records must be assigned to individual clients or users and then divided into server sessions. In a third step, the identified server sessions are subjected to still more analysis, leading to identification of individual page views. Finally, individual episodes are identified.

Processing of information that describes user behaviour must be supported by processing of such structural and content data that describe the company's website.

Storage in a data warehouse

Processing of non-reactively collected data produces information about a user's 'hits' on the company's website (possibly also about hits on other websites) and the amount of time spent with each item (good/service), which can be used for creation of usage profiles. These data, in turn, can be of two basic types: personal (ie data that can be correlated with individual, identified customers) and non-personal (that describe the usage behaviour of anonymous customers).

Processing of reactively collected data yields information about user characteristics that cannot be directly obtained from usage behaviour. Five different types of marketing-relevant customer profiles can be produced by combining reactively and non-reactively collected data (Table 4).

Once data, either reactively or non-reactively obtained, have been processed for creation of customer profiles, the data are stored in a suitable structure, by either of two basic procedures: a) establishment of a standalone database system that comprises only data from Internet-based transactions; or b) integration of the data

Figure 2: Processing of non-reactively obtained data

within a comprehensive data warehouse environment that is linked with other department-specific databases, via an integrated approach aimed at creating a general management support system.

A standalone database system can be useful when information relative to customer profiles does not have to be combined, for further purposes, with data from other areas of the company. For example, this could be the case if non-reactively collected Type 4 profile data are to be used solely in personalisation tools. The primary advantages of a standalone database system are that it has very few IT interfaces, to other areas. The resulting simplicity can positively affect the system's installation and management costs.

Data available to the company from different areas of the company can have much greater strategic usefulness if combined within a decision-oriented data

model that makes it possible to observe, analyse and optimise the company's market-oriented activities across all areas, in a general manner.

Apart from the general problems of setting up data warehouse concepts, the major challenge in integrating profile data from Internet-based activities involves representing masses of non-reactively collected information in a suitable form. Log files of very popular websites can grow by several hundred megabytes daily.³ Complete, detailed presentations of all relevant non-reactively obtained usage data in a data warehouse can reach orders of magnitude that considerably constrain storage and/or subsequent usage within the framework of analysis. In particular, the extremely calculation-intensive analysis used in 'data mining' can create problems in terms of computer capacity and calculation time (it can take too long to produce a result). Thus, it may be

Table 4: Typology of customer profiles

		Reactively collected data	
		Relevant	Not relevant
Non-reactively collected data	Personal	Type 1	Type 4
	Non-personal	Type 2	Type 5
	Not relevant	Type 3	—

Table 5: Presentation of profile data in a data warehouse, using a weighting system

ID	Reactively obtained data		Non-reactively obtained data			
	Name	...	Product group A	Product group B	...	Episode n
1	Aaker	...	W_{1A}	W_{1B}	...	W_{1n}
2	Abend	...	W_{2A}	W_{2B}	...	W_{2n}
...	...	...	...	...	W_{ij}	...
m	Zacher	...	W_{mA}	W_{mB}	...	W_{mn}

useful to pare data before they are transferred to a data warehouse or used for analysis. For instance, it can be useful to combine data to form aggregate packages. This approach can make the information in all relevant data available for profiling and yet largely eliminate the need for fine-grained storage of all non-reactively obtained data.

As described above, appropriate combination of page views into episodes can make it possible for customer requested page views to be automatically classified and assigned to thematic areas. If episodes are formed in such a manner that they can be used to generate relevant conclusions about customer characteristics (eg interest shown in certain products or topics), then non-reactively obtained data can be represented in a data warehouse in the form of weightings. In a data warehouse a set of non-reactively obtained usage data for a single customer can be represented as a vector with each of the vector's dimensions corresponding to an area of interest (eg an episode assigned to

a certain product group). In evaluation of non-reactively obtained data, a weight w_{ij} can then be defined, for customer i and dimension j , that describes the strength of the interest shown in the episode in question (Table 5).

Weight w_{ij} , corresponding to customer i and episode j , can then be evaluated using information about all relevant page views for a given episode, as a function of three central figures⁴:

- the view time d_{ij} that a customer spends on page views for a relevant episode
- the relative frequency h_{ij} with which the page views for a given episode are retrieved
- the chronological distribution or proximity of accesses (hits). While the relative frequency h_{ij} serves as an indicator for the general interest shown in a given product, the importance of an area of interest can vary over time. Management of a product based on relative frequencies can produce poor results if hits on the

most frequently visited episode lie some time in the past. For example, a customer who requests current data about a new topic might continue to receive information or offers relative to the episode the customer visited most frequently, even if that episode does not reflect the customer's current interests. For this reason, current accesses (hits) have to be given greater weight in evaluation of weight w_{ij} than accesses that lie further in the past. Consequently, a factor v_{ij} , representing the chronological distribution of accesses, has to be included.

Logically enough, compression of non-reactively obtained data, to form weights, destroys some data. It should be noted that this could pose a problem in some applications. In general, it is possible to combine non-reactively obtained data with reactively obtained data in a weight.

Data analysis and use

Data stored in a data warehouse environment can be processed and used in many ways. In the simplest case, rule-based procedures can be applied to submit offers to customers with certain observed, fitting characteristics (rule-based matching). For example, banners advertising private health insurance might be directed only at customers of a certain minimum age (over 18), with a certain minimum income (over \$4,000/month) and with certain fitting hobbies (knitting instead of motorcycling).

PROBLEM AREAS IN USING CUSTOMER-PROFILING TECHNIQUES IN eCOMMERCE

In practice, and apart from any methodological complexities, the extent

to which the above-described data collection procedures and, ultimately, the profiles derived through them, can be used depends on a number of factors. Both data collection, for creation of customer profiles, and the use of such profiles in e-commerce are sensitive matters that an increasing number of customers and institutions are viewing sceptically. Concerns about development and use of customer profiling in electronic commerce tend to be related to the following areas:

- both customers and the public can consider data collected for customer profiling purposes as private (or at least confidential) information. In this view, collection of such data is considered inherently unethical or immoral and is seen as an 'invasion of privacy'. Privacy issues seem to be of particular interest because privacy is a fundamental human right
- even when the data themselves are not seen as private information, consumers and the public may still have reservations about the methods used to collect the data. In such cases, criticism focuses on the inherent methods and concepts of non-reactive data collection. Privacy issues seem to be of particular interest because privacy is a fundamental human right — also in the information environment. It is a basis of human freedom (of action) and human dignity. With regard to human freedom, it is argued that, in many areas of life, a person's anonymity is a basis for creation and protection of personal freedom and personal development. Ultimately, every human being has spheres, in his/her private life, that contain actions, statements and thoughts about which no other (or only few) people should have any knowledge. Most people

would not want their immediate, relevant environments (eg transaction partners, family, workplace) to know about their 'dark areas' (eg certain behaviours and preferences); armed with such knowledge, others would intervene in their personal freedom, for instance, by exerting social and moral pressure or by criticising (often, unjustifiably) their individual behaviour.

- even when data collection and relevant customer profiling methods are seen as acceptable, consumers may still be concerned about the many ways in which the data are used. Advertisers usually use customer profiles in order to predict preferences and future needs of customers, thereby placing themselves in a position to influence existing needs and purchases efficiently. While criticism focused on the manipulative nature of advertising is not new, such criticism has grown into a new dimension as the possibilities for approaching individual customers via the Internet have developed. More than ever before, advertisers are able to reinforce customers' wants and to spark more or less 'uncontrolled' purchases; purchases that customers might not make under other circumstances (or might make only at different times and under different terms)
- privacy activists often stress that many Internet users are not aware that they leave tracks that can be used for creation of profiles. Others argue that customer profiling information, in principle, belongs to the affected customers and thus cannot become the property of other parties without the customers' consent.⁵ All non-reactive data collection procedures provide behaviour observation instruments with which

website operators can collect data from customers, and use such data in customer profiling, without customers' awareness. Another related complaint 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. Interestingly, almost 60 per cent of all Internet users do not know how cookies can be rejected.⁶

- while profiling techniques make it possible to enhance marketing efficiency through precise selection of customers for marketing approaches, they also give rise to concern about dangers of discrimination and selective advertising and transaction design.⁷ Profiles, which are used to assess customers and their future behaviour, can be used by companies as a way to avoid certain unattractive types of customers, to block such customers' access to information about products and services⁸ or to offer them poorer purchasing terms.⁹ For example, in online tests conducted in 2000, Amazon set DVD prices in a customer-specific, selective way as a function of the user's browser type, service provider and frequency of previous visits to the Amazon website. As a result, at a given time the price for a standardised DVD such as the film 'Planet of the Apes' differed by \$10 and more between customers.¹⁰

The criticism aimed at customer profiling methods and against the use of non-reactive data collection procedures does not apply in equal measure to all data collection procedures. Table 6 provides an overview of features of non-reactive data-collection procedures in the light of criticism targeting the various methods.

Table 6: Traits of non-reactive data collection methods

Procedures	Secrecy	Involuntary participation	Lack of protection
Log files	Yes	Yes	Yes
CGI-based files	Yes	Yes	Yes
Cookies	Yes*	No*	No*
Software agents	Yes*	No*	No*
Packet sniffers	Yes	Yes	Yes
Web bugs	Yes*	Yes	Yes
Modified browsers	No*	No	No

*Depends on the user's know-how

Development of consumer empowerment tools

One way of responding to customers' concerns, as well as to criticism of profiling, is to promote development and dissemination of privacy enhancing technologies¹¹ and procedures that enable customers to protect themselves as necessary against customer profiling and even to help shape the profiling process. Logically enough, this approach responds to criticism that customers lack protection and participate involuntarily, and it positively affects consumers' perceptions of the general risks resulting from customer profiling. Such tools relate to:

- controlling exchange of data (primarily reactively collected data) (control-enhancing tools, CET)
- making collected data (primarily non-reactively collected data) unusable (protection tools, PT).

Incentives

Yet, another way of responding to customers' concerns is to offer monetary (eg cash payments, discounts for online purchases) and/or non-monetary incentives (eg personalisation of web areas, provision of exclusive information) that enhance customers' willingness to participate voluntarily. Incentives can have a positive effect on acceptance of customer profiling efforts, can encourage

customers to overlook losses of privacy and risks of manipulation, discrimination, etc., and thus can turn customer profiling into a 'win-win' relationship from which providers and customers alike can profit.

While in theory all providers can use incentives in e-commerce, in practice such incentives will be a viable option only for companies whose core business is in management of customer profiles. At websites with relatively little user traffic, incentives will be unlikely to entice customers to participate in complicated log-in procedures. Furthermore, incentives create management overhead for providers, overhead that may not amortise itself in collection and use of profile data. For 'normal' e-commerce providers, therefore, incentives present efficiency problems, and this is likely to limit their use.

Involvement of certification agencies

Rating services can also help build trust relationships between market players and participants. Via such services, companies can have their compliance with privacy policies certified by external organisations, thereby making such compliance available for marketing communications purposes. An especially important role in this context could be played by state organised certification via data protection audits, such as those now being discussed in connection with a

draft programme standard for amendment of the German Data Protection Act (*Bundesdatenschutzgesetz*), as well as by 'seal programs' carried out by private sector certification organisations. Prominent representatives of such private organisations include TRUSTe (www.truste.com) and WebTrust (www.cpawebtrust.com).

Development of industry codices

Trust of customers and the public in companies' privacy policies is centrally affected by the behaviour of 'other' providers. For this reason, companies must seek to influence the behaviour of other market participants, with the aim of reducing the risks that other participants will act unethically. One way of influencing competitors to act responsibly is to apply moral suasion by establishing suitable, binding industry codices of behaviour in the market.

IMPLICATIONS AND CONCLUSION

The main purpose of this paper is to shed some light on the emerging area of online marketing research and customer profiling. It provides an overview of online customer profiling methods and discusses caveats of such methods. This study has implications for international marketers targeting international Internet users.

The Internet provides companies with new opportunities for exploring customers' needs and characteristics in order to obtain a basis for selective planning, design and control of all interactions with customers. Such a basis can support efficient interaction with the market as well as effective stimulation and satisfaction of customers' needs. As to customer profiling methods, however, no 'best' approach to generating and using profiles exists.

Problems in using customer profiling can arise primarily through ethical and moral concerns and through criticism of non-reactive data collection procedures. Such problems can limit acceptance of profiling in e-commerce and can burden companies' relationships to their customers and to the public. Consequently, companies need to develop and implement profiling methods that, while producing useful, detailed profiles, also respond to customers' needs and concerns in connection with data collection and use. A range of different measures is available for enhancing affected groups' acceptance of profiling methods, measures that differ in terms of aims, function and relevant application problems. Regardless of whether they use such measures, companies should remember their social responsibility and carefully consider the general and specific ethical and moral boundaries for their own actions.

References

- 1 Peters, T. A. (1999) 'Computerized monitoring and online privacy', McFarland, Jefferson.
- 2 Buxel, H. (2001) 'Customer profiling in electronic commerce: Methodische Grundlagen, Anwendungsprobleme und Managementimplikationen', Shaker Verlag, Aachen.
- 3 Zaiane, O. R., Xin, M. and Han, J. (1999) 'Discovering web access patterns and trends by applying OLAP and data mining technology on web logs', online: www.cs.sfu.ca.com (last revision: 23rd March, 2000).
- 4 Buxel *op. cit.*
- 5 Clarke, R. (1999), 'Internet privacy concerns confirm the case of intervention', online: www.anu.edu.au/people/Roger.Clarke/DV/AC-M99.html (last revision: 24th July, 2000), also in *Communications of the ACM*, Vol. 2, pp. 60–67.
- 6 Personalization Consortium (2000), 'Personalization & Privacy Survey', online: www.personalization.org/survey.pdf (last revision 7th April, 2000).
- 7 Stepanek, M. (2000) 'Weblining: Companies are using your personal data to limit your choices — and force you to pay more for products', online: www.businessweek.com/2000/00_14/b3675027.htm (last revision 4th January, 2001).

- 8 Novek, E., Sinha, N. and Gandy, O. (1990) 'The value of your name' *Culture & Society*, Vol. 12, pp. 525–543.
- 9 Weichert, T. (2000) 'Zur Ökonomisierung des Rechts auf informationelle Selbstbestimmung [Economising the law on individual self-determination]', Bäuml, H. (ed.) 'E-Privacy: Datenschutz im Internet', Vieweg, Braunschweig pp. 158–184.
- 10 Rosencrance, L. (2000) 'Amazon charging different prices on some DVDs', online: www.computerworld.com/cwi/story/frame/0,121-3,NAV47_STO49569,00.html (last revision 1st October, 2000).
- 11 Wang, H., Lee, M. K. O. and Wang, C. (1998) 'Consumer privacy concerns about Internet marketing', *Communications of the ACM*, Vol. 3, pp. 63–70.