Forthcoming in the *European Journal of Marketing*: *Special Issue on Marketing in Cyberspace*

On the Analysis of Web Site Usage Data: How Much Can We Learn About the Consumer From Web Logfiles?

Shahana Sen, Balaji Padmanabhan, Alexander Tuzhilin, Norman White and Roger Stein

Abstract

In this paper, we investigate the information needs of marketers on the Web for consumer analysis purposes and examine how these can be met. Grounding our work in the existing bases of segmentation in the literature, we propose new Web-specific segmentation variables, by operationalizing and extending the existing variables for consumer analysis and segmentation. We discuss how marketers' information needs could be met from the different information sources available to a marketer, including Web logfiles, secondary sources, and information elicited from the consumer. We conclude that the readily available data in Web logfiles satisfies a significant part of the behavior segmentation information needs of marketers, while most of the demographic, geographic, psychographic and benefit segmentation information needs require information to be elicited from the visitors.

1. INTRODUCTION

The WWW is the latest entrant in the communication and marketing media alternatives available to marketers. In the words of Hoffman and Novak (1996), "For several years, a revolution has been developing that is dramatically altering traditional view of advertising and communication media. This revolution is the Internet..." Although originally the purview largely of technologists and academics, the Internet and its first and current networked global implementation, the WWW (Web), has become a medium utilized by a broad spectrum of individuals, from experienced business people to computer novices. Organizations have a variety of overlapping motivations for creating Web sites. Web sites allow firms to communicate directly with customers, to build brand awareness, position, and value, as well as create new sales channels for delivering on-line products and services. Large numbers of commercial organizations have put up Web sites to avail of the opportunities the Web promises. By one estimate, $312 million ad dollars were spent on advertising on the WWW in 1996, and this figure is estimated to grow to $5 billion by the year 2000 (Business Week 1996a). According to one survey, one fifth of randomly selected senior managers nationwide reported that their organizations currently had Web sites, and another one third indicated that they planned to launch a Web site within the next six months (Internet Marketing 1996).

Similar to other media, consumers on the Web have *different needs* and exhibit *heterogenous behavior* in their site usage (Hoffman et al.1997 and Venkatesh 1997). Marketers designing Web sites need to understand these differences through consumer analysis to be effective. While marketers benefit from a macro understanding of the differences between users on their overall

Web requirements and usage (as studied by Hoffman et al. 1997 and Venkatesh 1997), they also have a need for understanding their own consumers at a more micro level.

In this paper we investigate the information needs of marketers on the WWW for consumer analysis purposes, and examine how these can be met. We begin by briefly highlighting the attractiveness of the Web as a medium for communication. Next, we summarize the bases for consumer analysis and segmentation in the context of the traditional marketing communication media. Based on this framework, we examine consumer analysis needs in the context of the WWW medium and propose analysis variables relevant to this medium. Finally, we discuss how the information needs for consumer analysis may be met from the different information sources available to a marketer, including the Web logfiles which are generated as a result of tracking the interactions of visitors accessing information from a company's Web site.

WWW as a Communication Medium

An important distinguishing aspect of the Web as a medium is that consumers use it to *actively* seek information about products or brands in which they are interested, and can also provide *instantaneous feedback* to the marketer (Sterne 1995). The Web site establishes a one-to-many link between an organization and the receivers of the message (visitors to the site), whereby it is possible for both to communicate with each other in *real time*. Consumers therefore perceive this medium differently from traditional media, which are seen as vehicles for "push" strategies (Sissors & Bumba 1995). The one-to-one link with visitors allows companies to deliver targeted ads to them (see http://www.doubleclick.net/frames/adinfo/targoset.htm) and can potentially be used to deliver

customized products or other communication (see http://www.webdeveloper.com/ categories/cgi-perl/cgi_customized_web.html).

The WWW is a medium more in line with the fragmented nature of modern markets[1]. For the marketer, narrowcasting over the Web is a viable option given the relative inexpensiveness of delivering communication and content over the Web. Using the principle of Marginal $Cost_{additional\ segment}$ = Marginal $Revenue_{additional\ segment}$, and given heterogeneity in consumers' needs and preferences, we would expect to see marketers catering to much larger number of segments through their Web sites than through traditional media. In the extreme, the Web allows targeting each individual consumer, maximizing the producer's surplus.

This underscores the marketer's need for not only knowing who the Web site visitors or consumers are, but also analyzing and understanding how they differ in their behavior and in the benefits they may be seeking -- both from the product/brand and the Web medium/site itself. This type of analysis would ensure the consumer orientation that is critical to achieving the goals the company has for its Web site. In the next section we briefly discuss the traditional bases used for consumer analysis (and the corresponding segmentation variables under each), to set up the framework for analyzing a company's Web site visitors.

2. EXISTING BASES FOR ANALYSIS AND SEGMENTATION OF THE TRADITIONAL MEDIA USERS

Demographic and psychographic differences between consumers lead them to exhibit heterogenous needs and preferences for products as well as media preferences and usage (Lehmann & Winer, 1997). Consumer analysis aims at identifying groups or segments[2] of consumers, such that this heterogeneity is minimized within and maximized between the segments, and a unique marketing strategy can be designed for each of the segments targeted (Kotler, 1997; Churchill & Peter, 1995).

Broadly, in understanding the users of each medium in general and within each for the specific vehicles, marketers' objectives are twofold:

(A) *Where (media/vehicle) to find the target consumers:* Which newspapers, magazines, radio stations, T.V. channels or types of direct mail etc. attract the consumers they are seeking;

(B) *How to get their attention and target the right message:* To achieve an understanding of the media users' behaviors and the benefits that they may be seeking, as consumers of the product/brand and of the media/vehicle; to understand the differences between consumers on these, to be able to design the right product and message for the desired segment(s).

The relevant bases (and the corresponding variables) used by marketers to analyze media users to achieve the above objectives are as follows:

(A) Where (media/vehicle) to find the target consumers: To identify whether to use a certain media and vehicle, information on its users is analyzed using *demographic, geographic and psychographic* segmentation bases (Aaker & Myers 1987). Examples of variables under these are (Kotler 1997):

(i) Demographic: Name, Age, Sex, Education, Occupation, Family size, etc.
(ii) Geographic: Region, City or Metro size, Climate, etc.
(iii) Psychographic: Lifestyle (i.e. activities, interests, opinions), Personality, etc.

(B) How to get their attention and target the right message: Consumers' *behavior* and the *benefits* they seek, are traditionally analyzed by understanding the differences between consumers with respect to the following variables:

(i) *Behaviors with respect to the product category/brand and the media/vehicle:*

1. Consumption/Usage Occasions
2. User status: nonuser/ex-user/potential user/first-time user /regular user
3. Usage rate: light/medium/heavy product user
4. Loyalty status: None/medium/strong /absolute
5. Buyer-readiness stage with respect to product/brand: Prefer/Intend to buy
6. Attitude toward product/brand or media/vehicle: enthusiastic/positive/ indifferent/negative/hostile

(ii) *Benefits sought from consuming the product/brand and media/vehicle and why they prefer these over others:* involves the analysis of what values the customers seek and whether these are being fulfilled. The *benefits sought from the media/vehicle by the users* conveys useful information to the marketer about how the media, and in turn the marketer's communication contained in it may be perceived by users. Customer value of a brand (and the media/vehicle) is composed of three basic elements (Lehmann & Winer 1997) as below:

1. Importance of the usage situation.
2. Effectiveness of the product category in the situation
3. Relative effectiveness of the brand in the situation (through its functional attributes, customer service, image, or brand equity)

Since differences on any of the above variables may have an impact on the choice of the media/vehicle and/or the message to reach the target audience, these variables may be used by marketers for analyzing the users, depending upon their appropriateness to the particular situation.

In the next section we discuss the consumer analysis needs of marketers with Web sites using the above framework.

3. ANALYSIS AND SEGMENTATION OF WEB SITE VISITORS

In order to design their marketing strategies on the Web marketers are interested in learning about the aspects which unite or differentiate visitors on both their product/brand and Web/site usage needs and behavior -- such as the nature of information sought at the site, their usage patterns of the Web/site, their navigation within the Web/site and their purchase behavior at the site, etc.. This information is useful in designing the site as well as deciding on the linkages with other sites. Marketers' information Web site consumer analysis needs are detailed in Tables 2.1-2.3. In Sections A of the tables, we list the traditional information needs about the visitors as consumers of the product/brand. In Sections B we analyze marketers' information needs on their Web site visitors as consumers of the Web/site, using the framework from Section 2. As part of the analysis, we:

(i) operationalize the segmentation variables for the Web, and

(ii) suggest new Web specific segmentation variables under the existing segmentation bases, under the heading "Additional Web-specific Variables".

The list of the analysis variables is not meant to be exhaustive, but representative of the broad types of information marketers would desire. Next, we briefly discuss the usefulness of this information to marketers. Tables 2.1-2.3 contain additional explanations and examples to demonstrate the usefulness of the specific variables in understanding Web site visitors.

Section B of Table 2.1 details the *demographic, geographic and psychographic* Web specific variables that help identify the Web site visitors. The information needs consist of being able to identify the visitor for learning about them and being able to send them communication if desired. For example, the extent of the visitor's Web savvyness and the browser used are important factors in decisions about designing the site. Similarly understanding their Web usage personality – e.g. the extent of their willingness to give information about themselves and to buy products using credit cards over the Web, is important for site design decisions. The more information on the visitor, the better marketers can understand their needs and tailor the site to those they want to target.

The Web specific *behavior* variables will indicate the differences between visitors based on their usage occasions, interest, familiarity, buyer-readiness stage, competitive brands considered, ease of use and usefulness of the site to their needs, etc.. These variables have been detailed in Section B of Table 2.2 with discussions on their usefulness to the marketer.

Information gathered from the Web on the *value* the visitor is seeking from the product/brand and the Web/site will help the marketer design their marketing strategy for the brand and the strategy for their site and its contents, to meet the needs of their target segments and provide maximum possible value. These variables have been detailed in Section B of Table 2.3.

In a technical report (XYZ 1996) we have illustrated the uses of the above information in a retail setting for marketing and Web site strategy determination. The next section discusses to what

extent these marketing information needs can be satisfied from the various sources of information available to marketers with Web sites, including the data contained in their Web logfiles.

4. INFORMATION SOURCES

In this section we identify the various sources of information relevant to meeting the information needs discussed in Section 3. We illustrate these sources in Figure 4.1. We consider three main information sources: (i) data that is automatically recorded at Web sites in logfiles, (ii) secondary data from publicly available sources, and (iii) data elicited from Web site visitors. We conclude this section by providing a classification of these sources based on the ease of gathering the information. We begin with a discussion on Web logfiles.

4.1 Data Automatically Recorded at Web Sites

Each time a user interacts with a Web site (e.g. clicks on a link to or at the site), the server on the Web site automatically adds entries to a series of computer files called *logfiles*. The four common logfiles generated by Web servers are *access_log*, *error_log*, *referrer_log* and *agent_log*.

The access_log file records the name of the file transmitted, a timestamp and the IP address of the client that requested the page. An example of an entry (hit) that is recorded in an access_log file at a Web site http://www.yoursite.com, is:

```
eldorado.stern.nyu.edu unknown - [09/Aug/1996:20:13:13 -
0400] "GET promotions.html HTTP/1.0" 200 1876
```

This entry indicates that a user whose Internet address is “eldorado.stern.nyu.edu” accessed the file “promotions.html” on 08/09/96 at 20:13:13 hours EST. The name of the file accessed is a concise version of the full URL (e.g. http://www.yoursite.com/promotions.html) that appears on the screen of the user’s browser. The other fields in the example (“unknown” etc.) are less relevant here. Explanations of these entries may be found in (XYZ 1996).

The error_log file records errors that occur during an interaction between a user and the Web server. Agent_log files contain the name of the Web browser (such as Netscape, Spry, Mosaic etc.) that was used in each access to the site.

The referrer_log file lists the Web page that a user accessed prior to accessing any Web page on the site. The information tracked in the referrer_log file can be used to determine from *where* (from which site or Web page) a user came to this site. We present a detailed description of the information automatically tracked in logfiles in [XYZ 1996].

While Web logfiles contain certain information about visitors’ accesses to a Web site, they do not provide some important information that marketers would need. These limitations relate primarily to the inability of identifying each access to a specific user, and the difficulty in tracking when a user's web "session" begins and ends[3]. We discuss these two problems and partial solutions below.

Inability to trace each access to a specific user:

While entries in the access_log indicate the "hostname" (computer address) from where a request originated, this is seldom enough to identify a particular user. A hostname may be as general as "xyz.unilever.com." Typically there are a large number of users who may access a Web site from the same host. As a result, the hostname cannot be used to uniquely identify individual users. Nonetheless, being able to track a unique user at a Web site is central to satisfying many of the marketing information needs outlined in Section 2.

This limitation may be surmounted to some extent, as discussed later in this section. There are two different levels at which a unique user could be tracked: (a) anonymously, where a user is nothing more than a unique user ID (which remains the same whenever this user accesses the site) or (b) in addition to a user ID, some knowledge about the user (such as his/her name, email address etc.) is also maintained. The technical solution to this problem is described in (XYZ 1996).

Inability to track client sessions: Session identification issues arise for slightly different reasons than user identification issues. A session consists of a set of Web pages that a user accesses during a continuous period of time that the user spends at the Web site without leaving it[4] . For example, say that a certain company's Web site allows users to purchase products at their site while browsing through an online catalog. In order to keep track of the products purchased by each user during their respective "shopping sessions", it is essential that the site keep track of all individual user sessions (including identification of the beginning and the end of a user's session at

the site, the list of products that a user has added to his or her personal “shopping basket” in that session, etc.). In this case it is not necessarily important to identify the users, but only to be able to track specific sessions using a unique session ID.

Though plain logfile data does not permit the tracking of sessions or users, there are certain technical methods, such as the use of *tokens* and *cookies* [http://www.netscape.com; XYZ 1996], that can be used to attempt both session-tracking and user-tracking. Tracking user-sessions, however, is “easier” than tracking individual users (since it is a sub-problem of tracking users). In either case, these methods can be used to supplement every entry in the logfile with a unique user or session ID. We use the term *extended logfiles* to refer to automatically tracked information comprising of logfile data and data on user/session IDs obtained by using additional technical methods. The detailed description of cookies and tokens and how they can be used in marketing applications can be found in [XYZ 1996].

However, solving the technological challenges is only one aspect of what will determine the ease of availability of information in the future. Privacy issues on the Internet is an important and complex question, and technology that keeps track of individual users’ Web accesses and personal information would have to address several privacy concerns, including ethical, social and legal issues (Business Week, 1996b). The Privacy Working Group of the United States government Information Infrastructure Task Force (IITF) outlines two fundamental privacy principles (see Clinton Administration's "The Framework for Global Electronic Commerce", 1997):

> "*Data-gatherers should inform consumers what information they are collecting, and how they intend to use such data; and*

Data-gatherers should provide consumers with a meaningful way to limit use and re-use of personal information."

However, privacy on the Web is a complex and multifaceted issue and cannot be fully addressed as a part of this paper.

4.2 Publicly Available Secondary Sources of Data

This includes any information that is available through secondary data sources (e.g. Industry Directories, Company Annual Reports, the 10-K and 10-Q filings of companies made publicly available by the Securities and Exchange Commission (SEC) in print or at the Electronic Data Gathering and Retrieval (EDGAR) site (http://www.sec.gov)). This information is "freely available" and could be used to supplement the extended logfile information. For example, the "hostname" tracked in the logfile could be used to infer an organization's name, and secondary information sources could now be used to get additional information on the organization. In this manner, a visitor's affiliation may be approximately inferred.

4.3 Data Elicited from Web Site Visitors

This class of information contains sources of information that are not readily available and require the visitor's, or the consumer's involvement. This data may be gathered using both traditional as well as Web-specific methods by the company itself or by marketing research firms. Sources of this data include:

(i) Visitor Registration Data: The information which users provide if the organization that owns the Web site asks them to register for the use of their site. A company may limit itself

to asking for the user's name, physical and email address, or it may ask the visitors for more detailed information on any of the segmentation variables depending on their needs and visitors willingness to furnish the information. Marketers have been successful in getting substantial information from registrants by providing suitable incentives (typically entry into some sweepstakes).

(ii) Marketing Research Data: Traditional marketing research data (from interviews, focus groups, cross-sectional or panel surveys) may be used in conjunction with the logfile data to build a more complete picture of the visitor. The ease of communication with individuals over the Web, allows the use of this new medium for collecting data through Web-based surveys or even focus groups (Advertising Research Foundation Media Research Summit II 1996). This method is typically faster, less expensive and can have a wider geographical reach as compared to the traditional data collection methods (see Hoffman and Novak 1996). Beyond simply asking for information once during registration, marketers have now begun conducting periodic on-line surveys of visitors who come to their site for the purpose of tracking segment level trends (Gupta 1996). Companies may also form on-line consumer panels and elicit information from them on a periodic basis, by providing them suitable incentives. By building a long term relationship with the group of consumers in the panel, the company can track individual level trends in needs, preferences and behavior.

The potential of this medium has also attracted some marketing research firms to start conducting online surveys about topics Web marketers are interested in (some of these firms are PC-Meter LP, Cyber Dialogue, Perception Research Services, Inc. and Modem Media). These companies are finding that with suitable incentives, consumers are willing to give demographic, psychographic and other information if it is used anonymously (Kevin Mabley 1996). However, users may not want to share information if they are unclear how it will be used or how it may benefit them (GVU's 6th WWW User Survey, November 1996).

(iii) Data from Web Site Traffic Auditors and Information Brokers: Though currently auditing of site traffic is fraught with complex and unresolved issues[5] (Hoffman & Novak

1996; Advertising Research Foundation Media Research Summit II 1996), site-traffic auditors (e.g. I/PRO, NetCount, etc.) and information brokers (e.g. DoubleClick) collect various types of information on visitors. Such firms provide anonymous demographic and psychographic information about individuals who have registered with them and provided them with this information. From these information brokers, businesses may get a more complete profile of visitors to their site in terms of their usage, demographic and psychographic descriptors.

4.4 A Classification of Data Driven by Ease of Gathering

Whether or not the data has to be specifically *elicited* from a Web site determines how easily the marketer may be able to gather information. Automatically tracked logfile data and data from publicly available secondary sources can be obtained at little or no cost. Such data does not require any interactions with customers and are, therefore, typically easy to gather.

We assume that all other relevant data which needs the visitor's involvement and cooperation (e.g. demographics, motivations, attitudes, intentions, feedback etc.) require additional effort and are more difficult to gather than the data automatically tracked or publicly available. For example, a consumer may browse a company's Web site but may be reluctant to fill out its questionnaire on-line. Or, a visitor might be willing to provide product feedback at a company's Web site but may be less inclined to provide personal demographic information. However, in this article we do not distinguish between potentially different levels of difficulty in gathering consumer-specific data, but treat the entire category as being more difficult to obtain. Thus, we distinguish between two broad categories of data -- *augmented Web logfile data* and the data *elicited from visitors and consumers*[6] . These two broad categories of data are illustrated in Figure 4.1.

5. SATISFYING INFORMATION NEEDS

Table 5.1 lists the nature of information required to satisfy the specific categories of information needs of marketers outlined in Section 3. The table summarizes how these information needs are satisfied using current state-of-the-art technology (including cookies and tokens) described in Section 4.

The rows of the table correspond to the categories of Web-site specific information needs (each row corresponds to one table in Tables 2.1 through 2.3). The columns of the table split the information needs into two sub-categories: (i) those that can be satisfied using augmented logfile information (shaded area in Figure 4.2), and (ii) those that need (additional) elicited information (white area in Figure 4.2). The asterisks besides some of the categories signify that this category is only partially or conditionally satisfied and therefore involves some assumptions. Due to space limitations we do not describe *how* each information need was mapped onto a specific column in Table 5.1. The specifics of the mapping are described in detail in [XYZ 1996].

We do not list traditional market segmentation variables in Table 5.1 since almost all traditional variables can only be satisfied using elicited information [XYZ 1996]. This is not surprising since Web logfiles provide us primarily with information on a visitor's *behavioral* patterns at the Web

site, whereas traditional needs include much broader types of information that have to be elicited from a consumer (e.g. family income, why they purchase a product, etc.)

However Table 5.1 indicates that many of the Web-site specific information needs of marketers can be satisfied using automatically tracked data at Web site augmented with technical methods of gathering additional information. The second row of Table 5.1 indicates that most of the needs pertaining to a customer's *behavior* at a Web site can to a significant extent be satisfied using automatically tracked information. (Due to space limitations, we have discussed this in more detail in our technical report.) From a marketer's perspective, this is encouraging for the following reasons that:

- They are able to *easily* track customer behavior patterns at a site (or within a "virtual outlet") and use it for designing marketing strategies, in contrast to traditional outlets where it is more difficult to track a customer behavior patterns within a store.
- They can use this rich information on visitors' behaviors, to deliver custom-tailored site content to target customers.

However, elicited information is still required to satisfy several of the demographic, psychographic and benefit segmentation information needs. The good news, though, is that the interactive nature of the Web makes it technologically feasible to obtain this information electronically at a Web site, instead of having to conduct surveys or polls in person. As we discussed earlier, not only can elicited information be gathered faster, cheaper and more efficiently over the Web, marketers have been finding that, if Web site visitors can be shown the advantage of responding, their cooperation can be expected.

6. CONCLUSIONS

In this paper, we investigated the information needs of marketers with Web sites for consumer analysis purposes. Grounding our work in the existing bases of segmentation in the literature, we proposed new Web-specific segmentation variables. We discussed how these information needs could be met from the different information sources available to a marketer, including Web logfiles. We conclude as in Section 5 above, that the readily available data in Web logfiles satisfies a significant part of the behavior segmentation information needs of marketers, while most of the demographic, geographic, psychographic and benefit segmentation information needs require information to be elicited from the visitors.

Table 2.1
Web Demographic Geographic and Psychographic Segmentation Variables

A. Traditional Consumer Analysis and Segmentation Variables (Kotler 1997)

1. Demographic information: Name, Age, Sex, Education, Occupation, Family size, etc.
2. Geographic Information: Region, City or Metro size, Climate, etc.
3. Psychographic information: Lifestyle (i.e. activities, interests, opinions), Personality, etc.

B. Web-SiteConsumer Analysis and Segmentation Variables

Additional Web-specific Variables

1. Demographic information: **"Variables relevant for identifying the visitor on the Internet"**

(i) Visitor's email address, visitor's homepage address

- e.g. the last part of the email address (.edu/.com/.org/.mil/.net) helps in identifying the type of organization the user is affiliated with, viz. university/company/public institution/military/subscriber of an internet service provider, etc. Information from the visitor's homepage could be used to know the visitor better; An email address permits correspondence to the visitor, if desired.

(ii) Web savvyness

- e.g. Web savvy visitors may desire state-of-the-art technology to present web site content (e.g. frames) - however, novices to the WWW may find such features discouraging or simply unable to use them; Web savvy visitors may be more (or less) inclined to purchase products electronically compared to novices because of what they know about the safety features incorporated during the transaction process.

(iii) Browser software used by the visitor

- e.g. important to know since the browser software could limit the nature of interaction a user may be able to have with the site.

2. Geographic Information: **"Physical locational variables relevant to accessing the site"**

(iv) Where from do they access the Internet/Web (which *network*)?

- WWW users need access to the Web from a network service provider. Knowing where visitors come from can be used to plan advertising campaigns for the site at various network/content providers' sites.

(v) Where from do they access the Internet/Web (home/office/internet cafe/etc.)?

- This will provide information on the type of needs the user has.

3. Psychographic information: **"Attitude variables influencing the use of the site"**

(vi) ***Attitude towards privacy:*** i.e. willingness to give information on themselves over the WWW
(vii) ***Attitude towards safety:*** i.e. willingness to buy products using a credit card over the WWW;
(viii) ***Attitude towards delays and other technical problems with accessing information on the WWW***
(ix) ***Attitude towards innovations on the Web*** (such as attitude toward sites powered by new technology such as java applets, frames, etc.)

- The design of the site should be in harmony with all of the above psychographic characteristics of Web site visitors. For example, if most of the users who access the site are inherently unwilling to provide personal information over the WWW, the site may lose visitors by asking them to do so. Or, if they are concerned about the safety of their credit card information during a Web transaction, the marketer will need to provide them alternative ways of completing their order (e.g. One wine merchant asks for telephone numbers as part of the order placed on the Web instead of credit card information, so that he can call the customer and get this information over the phone.)

Table 2.2
Web Behavior Segmentation Variables

A. Traditional Consumer Analysis and Segmentation Variables (Kotler 1997)

1. Consumption Occasions for Product/Brand
2. User status: nonuser/ex-user/potential user/first-time user /regular user
3. Usage rate: light/medium/heavy product user
4. Loyalty status: None/medium/strong /absolute
5. Buyer-readiness stage: Prefer/Intend to buy product
6. Attitude toward product: enthusiastic/positive/indifferent/negative/hostile

---

B. Web-Site-Consumer Analysis and Segmentation Variables

1. Consumption occasions:

(i) How did/do they reach the Web site?

- e.g. Visitors coming from links/banners at other sites may indicate successful advertising/ promotions. Visitors entering the site by directly typing in the URL may indicate "serious interest" or familiarity with the site.

(ii) What time (of the day and week) did/do they visit the site?

- Can be used to plan "time-based" advertising banners at other places on the WWW -- in order to attract more visitors at "low-load" periods, the site may consider negotiating with popular, relevant sites on placing more advertisements at *only* those times; Also whether the visit is during working hours/lunch/after-office hours/weekends may provide information on the nature of their needs -- e.g. visitors during office hours vs. lunch-break may be more likely looking for their work or business related information rather than for their personal needs.

(iii) Are their visits ir/regular?

- The company may want to have their regular visitors see only the revised or updated information on their product or services to save them time as compared to their first time/infrequent visitors (who may need be interested in reading the site overview, information on the company, etc.)

2. User status:

(iv) First time or repeat visitor?

- As in (v) above. For repeat visitors the site may choose to custom-tailor the content presentation based on the visitor's previous accesses, such as by providing "updates" upfront on products/information previously accessed.

3. Usage rate:

(v) How often do they visit the site?

(vi) How much time did/do they spend at site?

- The site could be designed in keeping with the needs demonstrated by variables (vii) & (viii) as per the explanations in (v) & (vi) above.

4. Loyalty status:

(vii) Do they visit or conduct transactions at competitors' sites?

- e.g. Appropriate strategies could be adopted for visitors known to conduct transactions at competitors' sites, such as by presenting comparative figures or by providing suitable incentives.

5. Buyer-readiness stage:

(viii) What files did/do they access?

(ix) Did they purchase a product?

- Is the information accessed related to product specifications, dealer or retailer locations, ordering product, customer service, sending mail to company, consumer interest information, etc. The information accessed could be used to infer "how close" a visitor is to making a purchase and decide on whether to communicate (email/phone/regular mail) with him to lead him to the desired outcome.

6. Attitude toward site:

(x) Did they add the site to their bookmark?
(xi) Did they download/print any files?
(xii) Did they give any feedback/send mail to company about the site? If so, what was the feedback?

- All of the above are needed for understanding the effectiveness of the site in meeting the needs of the visitor;

Additional Web-specific Variables:

7. Interactive feedback:

(xiii) Did they give any feedback or send email about the product to the company from the site?
(xiv) Did they give any feedback or send email about the site?

- The ease of using the Web to send feedback makes visitors more willing to provide feedback in real time (i.e. during the same visit/session when they have read the marketer's communication). This unique feature of instantaneous feedback on whether the visitor was satisfied with the product/service or any or all of the characteristics that are involved in designing the site and its contents, can help marketers in responding to the needs of their customers better and faster for increasing value for them.

8. Search pattern at the site:

(xv) What is their search pattern at the site? Is it similar to those on previous occasions?
(xvi) Did they conduct the navigation and search efficiently?

- Visitors' search patterns may indicate the ease with which information can be accessed at the site and thus effectiveness of the site's design. If discovered that an often requested document requires substantial navigation through the site, the site's layout could be redesigned. If the differing search patterns among visitors demonstrate significantly varying levels of Web savvyness, then the site may need to be redesigned to cater to the different groups. Based on the visitor's prior search patterns at the site, the firm may be able to infer which information the visitor deems relevant and try to customize the site to serve the customer better.

9. Search pattern or navigation across sites:

(xvii) What was their search or navigation pattern across sites?

Such information would be useful for making decisions on where to place ads/banners/links to the site as well as know the effectiveness of those which are in place. Pattern of search across sites would also reveal information on the decision making process of the visitor, e.g. which competitive brands are they interested in, or do they rely on the recommendation of some other site for their needs, etc.

Table 2.3
Web Benefit Segmentation Variables

A. Traditional Consumer Analysis and Segmentation Variables (Lehmann & Winer, 1997)

1. Importance of the usage situation
2. Effectiveness of the product category in the situation
3. Relative effectiveness of the brand in the situation (through its functional attributes, customer service, image, or brand equity)

B. Web-Site-Consumer Analysis and Segmentation Variables

1. Importance of the usage situation

(i) Is the visitor a casual browser or a serious searcher
(ii) What are the benefits sought from the Web site for the specific occasion – is it product information gathering and evaluation, or purchasing products, or customer service etc.

2. Effectiveness of the product category (the WWW) in the situation

(iii) What, if any, are the economies associated with using the Web for the specific occasion compared to other alternatives?
(iv) What, if any, are the conveniences associated with using the Web for the specific occasion?

3. Relative effectiveness of the brand (the site) in the situation (through its functional attributes, customer service, image, or brand equity)

(v) Did they find what they were looking for at the site?
(vi) Did they encounter any problems accessing information at the site?

- Information gathered from the Web on all of the above will help the marketer design the site and its contents to meet the needs of their target segments and providing maximum possible value.

Figure 4.1 Levels of Web-Related Data

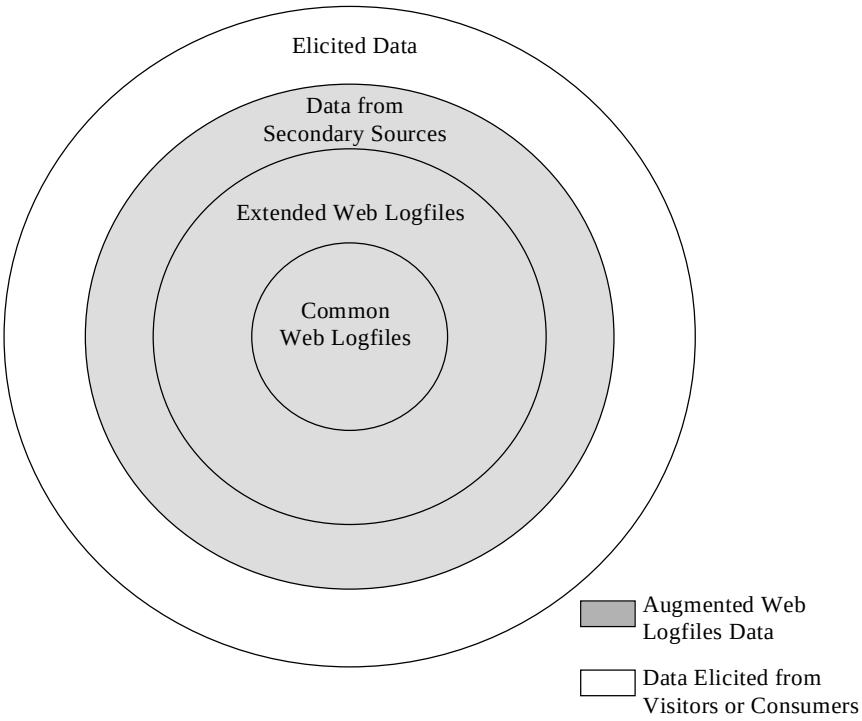

Table 5.1 Satisfying Information Needs With Augmented Logfile Information and Elicited Information using Current Techniques.

	Web-Site Specific Marketing Segmentation Needs	
Information needs category	**With augmented logfile information only**	**With additional elicited information from visitors**
Demographic and Psychographic Variables	(iii) browser software (iv) from which network?	(i) email/homepage (ii) Web expertise level (v) whether access was from home/office etc. (vi) willingness to give information. (vii) willingness to buy products (viii) attitude towards delays (ix) attitude towards innovations
Behavior Segmentation Variables	(i) where Web page access came from (direct /through links/ other). (ii) time of visit (iii)* regularity of visit. (iv)* first/repeat visit (v)* frequency of visits (vi)* time spent (viii) files accessed (ix) product purchased? (xii, xiii, xiv) feedback. (xv, xvi)* search pattern at the site	(vii) visit competitors' sites (x) bookmarked? (xi) download files? (xvii) search pattern across sites
Benefit variables	(vi)* problems accessing information	(i) casual/serious, (ii) benefits sought. (iii, iv) economies, conveniences of using the medium (v) find relevant information?

Note: The numbers in the table correspond to the specific information needs under each of the Web segmentation variable categories. Text beside the numbers are brief descriptions of the needs listed in Table 2.1 through Table 2.3 An asterisk beside a number indicates that the need is only partially or conditionally satisfied.

References

Aaker, David A., and John G. Myers (1987), *Advertising Management.* New Jersey:Prentice Hall, 3rd edition.

Advertising Research Foundation Interactive Media Research Summit II (1996), *Bringing Clarity to New Media Research*. New York: Advertising Research Foundation.

Business Week, (1996a), "Making Money on the Net," by Rebello, K. and Cortese, A. September 23, pp 104-118. New York:McGraw Hill.

____(1996b). "Privacy and the Cookie Monster," by Wildstorm, S.H. December 16, pp 22. New York:McGraw Hill.

Blattberg, Robert C. and Deighton, John (1991), "Interactive Marketing: Exploiting the Age of Addressability," Sloan Management Review, (Fall).

CASIE Guiding Principles of Interactive Media Audience Measurement (1996), in *Getting Started on Interactive Media Measurement*. New York: Advertising Research Foundation.

Churchill, Gilbert A. Jr. and Paul J. Peter (1995), *Marketing: Creating Value for Customers*. Illinois: Austen Press.

Clinton Administration (1997). The Framework for Global Electronic Commerce. http://www.whitehouse.gov/WH/New/Commerce. 1997

GVU's 6th WWW User Survey, October 1996. Georgia Tech Research Corporation. http://www.cc.gatech.edu/gvu/user_surveys/survey-10-1996

Gupta, Sunil (1996), "Online Research: Matching Decisions and Data," paper presented at the Workshop on Internet Survey Methodology and Web Demographics, MIT, January 30, 1996.

Hoffman, Donna L., William D. Kalsbeek, and Thomas P. Novak (1995), "Internet Use in the United States: 1995 Baseline Estimates and Preliminary Market Segments". Project 2000 Working Paper, Owen Graduate School of Management, Vanderbilt University.

____, and Novak, Thomas P. (1996), "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations," Journal of Marketing, 60 (Fall), pp. 50-68.

____, Kalsbeek, William D., and Novak, Thomas P. (1997), "Internet and Web Use in the U.S.," Communications of the ACM, Special Issue on the Internet @ Home, 39 (December), 36-46.

Internet Marketing, (1996), "Rust-belt marketers plot Web strategies: How the Nation's

Manufacturers reach Cyberspace" by Joe Mullich. New York: Special Supplement to Advertising Age and Business Marketing, May 1996.

Kotler, Philip (1997), *Marketing Management: Analysis Planning Implementation and Control*. New Jersey: Prentice-Hall, 9th edition.

Lehman, Donald R. & Russell S. Winer (1997), *Analysis for Marketing Planning.* Illinois: Irwin, 4th edition.

Mabley, Kevin (1996), "How do Corporations Get Consumers to Reveal Valuable Information Without Turning Away Customers?" Talk presented at the Advertising Research Foundation Electronic Media and Research Technologies XV Conference, December 1996, New York.

Sissors, Jack Z. and Lincoln Bumba (1995) *Advertising Media Planning*. Illinois: NTC Business Books, 5th edition.

Sterne, J. (1995), WWW Marketing: Integrating the Internet into Your Marketing Strategy. New York: John Wiley & Sons.

Venkatesh, Alladi (1997), "Computers and Other Interactive Technologies for the Home," Communications of the ACM, Special Issue on the Internet @ Home, 39 (December), 47-54.

XYZ (1996), Technical Report by the authors: "Analysis of Web Site Usage Data: How Much Can We Learn About the Consumer from Web Logfiles?"

Footnotes

[1] To quote Russel I. Haley (1987), the developer of the benefit segmentation concept, "The *new (electronic) media*...make it increasingly possible to reach smaller and smaller segments of the market with special interests or needs ... or *narrowcasting*."

[2] We wish to clarify here that we use the term *segment* to mean any body of consumers for whom the marketer can create a unique marketing mix, and they may even be single consumers if the marketer wishes to target them individually, as in the case of one-to-one marketing.

[3] Hypertext transfer protocol (HTTP) defines the protocol for communication between a browser and a Web server. This protocol is "stateless" in that the server disconnects itself from the client after a single request is processed. If the same client accesses two different pages on a Web site, the statelessness of the protocol results in the server being unable to infer that the successive requests came from the same client

[4] In practice it is usually a non-trivial task to identify when a user's "session" ends. A heuristic used by some commercial packages assumes that a user's session ends if the time difference between successive accesses by the same user is greater than thirty minutes.

[5] The problem here is agreeing on a standard set of measurement units and their definitions that would assist in a fair comparison of traffic across different web sites. For example, "visits", "sessions" and "hits" could be interpreted differently by different packages. In order to address these concerns and arrive at a common understanding of the measurement vocabulary (or "measurement units"), and to develop objective guidelines for the measurement of these units, CASIE (a joint project of the Association of National Advertisers, Inc., and the American Association of Advertising Agencies), with the support of Advertising Research Foundation (ARF), has created the Guiding Principles of Interactive Media Audience Measurement (ARF, 1996). More recently Hoffman and Novak (1996) have proposed a set of interactivity and outcome metrics based upon the idea that the best measures of ad value in the Web medium are based upon the degree to which the visitor interacts with the ad.

[6] Our categorization of information is related to a classification, referred to by practitioners, as "customer-centric" and "site-centric". Our classification of information into "augmented logfile" and "elicited" is primarily motivated by the *ease* of gathering information, rather than *where* the information is gathered.