

QUANTITATIVE DATA GATHERING AND ANALYSIS ON THE NET | 141

Demographics provided by a Web log such as type of browser used, domain name location. The maintenance issues include:

- Number of errors of any kind encountered by users
- Link verification (should support a wide variety of links—http, ftp, mailto, image, applet, etc.)
- Site maps generated in Rich Description Format (RDF)
- Orphaned pages—those that are no longer connected to other pages on the Web site
- A graphical view of individual and summary statistics of navigation through your site
- Pages with slow download times
- Site reliability—when and for how long the site was not responding to requests for information
- A log of search items found in help or directory searches

QUANTITATIVE DATA GATHERING AND ANALYSIS ON THE NET | 141

These machine-gathered data are frequently combined with information provided by the user—typically when they first register at the site or through a standard educational registration process. In educational applications the researcher may have access to other demographic information including grades, prerequisite accomplishments, and scores on pre- and post-tests. Access to this personal information is of course controlled through ethical constraints and the e-researcher must obtain informed consent from participants (see Chapter 5). If e-research is being conducted on sites where personal information is not being gathered, it is still important to inform users of what information is being gathered and for what purposes. This information should be posted prominently in a privacy policy accessible from the first page a participant is likely to encounter.

QUANTITATIVE DATA GATHERING AND ANALYSIS ON THE NET | 141

The process of analyzing Web logs can be tedious as the volume and amount of irrelevant data translates into a great deal of preprocessing before analysis can commence. Zaiane (2001) lists the major steps in the analysis of educational Web logs: • Remove irrelevant entries. Identify access sessions (to determine individual users). Map access log entries to learning activities. Complete traversal paths (what pages did the user request and in what order). Group access sessions by learner to identify learning sessions. • Integrate data with other data about learners and groups of learners. (p. 61) Fortunately, many applications require participants to log in, so that the activities of different users can be uniquely identified. This login identification is kept on the user's machine and information is passed to the Web server through the appendage of a cookie