PRIVACY AND CONFIDENTIALITY

The process of gathering data for later statistical measurement should place emphasis on the aggregated data, not on tracing the items accessed by a given person. When collecting information about the use of a library's electronic content, focus on striking a balance between the detail of information collected and analyzed and encroaching on user privacy.

Libraries concerned with protecting user privacy should think carefully about how they record user activity information. Another concern should be how they provide data to people outside the library, which would likely only be under limited and carefully considered circumstances.

The general ethic common within the library community holds that the gathering of statistics must not betray the privacy of the user. Although understanding the online behavior of any given class of user is useful, do not maintain data in such a way that divulges the materials accessed by a particular person.

Most access logs contain an IP address, the time and date of access, and the item requested. In most cases a network IP address cannot be linked to a particular user. The following types of access scenarios do not directly link an IP address with a given user:

- In-library public access workstations that provide unauthenticated anonymous access
- Workstations on the organization's network that rely on dynamically allocated IP addresses
- Most remote access. Few Internet service providers (ISPs) provide static IP addresses to home users.

Some conditions exist where an IP address can infer use by a particular person. Anytime a computer maintains a fixed IP address and the use of that computer is limited to a particular person, then a strong likelihood exists that any use of materials linked to that IP address is associated with that person. Even in these cases, the link between the IP address and the person can be discerned only by those with access to detailed configuration information of the organization's network. In cases where concern exists, libraries can increase user privacy by modifying any system log files held for long-term analysis by obscuring the last octet of the IP address so the log file indicates only the network or subnet of the activity but not the individual user.

Increased personalization of library services in turn creates opportunities for recording personally identifiable use data in system logs. To offer customized services, a common practice within the next-generation Web online pubic access catalogs (OPACs), involves saving and storing a lot of information related to the user's preferences, subject interests, and items consulted or checked out. These practices have become common in online book selling and other consumer-oriented Web sites. Online vendors want to tempt customers with highly targeted offers, such as, "If you like this book, then you might also want to consider this other one." This type of highly targeted advertising can happen only when the site collects specific information about each customers' buying habits.

Libraries likely want to use some of the same techniques with their patrons. Personalized Web OPACs can deliver some of the same features as the online bookstores. To support such services, systems need to rely on prefer-

ences set by the user and possibly to track what content the user has previously selected. Libraries need to understand well how their system stores and makes use of personally identifiable use data within these new personalized systems and weigh their comfort level accordingly. Commercial companies have a reputation for sharing personal information about their customers with partners and advertisers or making use of this information in other revenue-generating activities. Libraries tend to follow strict policies regarding privacy and confidentiality of their patrons' online behavior.