Ownership is not the only issue with such information or other sensor feeds; integrity is equally important. It is vital that such information be not only correct but also accurate and authenticatable. The notion of someone "simulating" a major earthquake through the network, for example, is clearly unacceptable. In this connection, it is interesting to note that most of the existing cryptography based authentication technology could be quite problematic in these applications, unless considerable care is applied to address scaling issues; for example, imagine every workstation in Northern California simultaneously trying to obtain the public key of the earthquake information server to validate an earthquake warning, throwing the entire Internet into overload at precisely the time that smooth operation is most needed.[46]

8. Privacy issues in access in the networked information environment

Confidentiality and Anonymity of Access

Within the library community, confidentiality of collection use information is a well established principle. Library practice, as defined by the American Library Association, defines circulation records as highly private, to be revealed only under a court order—if then. Indeed, practice goes further-typically a library will only store the information that a given patron has borrowed a given book during the period while the book is on loan; once returned, only statistical information based on the patron's statistical categories is retained for subsequent management analysis. Most libraries, even under court order, can provide little or no history about the books borrowed by a given patron. Through statistical analysis, they may be able (if their circulation system is well designed and well implemented) to provide lists of the hundred most popular (in the sense of most frequently borrowed) books, or the ten books borrowed most often by high school students in the past year. It is also usually possible to find out how often a given book has circulated, or how many items a given patron has borrowed in the past year. In fact, such information is very important for the tuning of collection development strategies, for deacquistions decisions, and for overall management and budgetary planning.

Similar principles about privacy in the networked environment are far less clear; there is no general consensus about norms of behavior. Most users have a tendency to assume that their privacy is protected-either by legislation or generally accepted practice-to a greater extent than it probably really is, perhaps making this assumption by analogy to library practices and other situations, such as video rental records.

---

46 Another example of the use of the network to provide information to control machines is provided by the "smart power" technologies that are under discussion in projects such as the Blacksburg Electronic Village effort in Virginia [Bull, Hill, Guyre, & Sigmon, 1991]. The basic idea here is that under heavy load the power company must purchase additional power from other power companies on the national electrical grid at very high prices and it is very advantageous to them to be able to reduce loading during those times; additionally, their pricing, particularly to residential customers, does not let them recover these premium costs directly during periods of very heavy load. Instead, residential costs are to some extent averaged over long periods of time in setting rates. The proposal is that consumers would install smart appliances and controls (thermostats, refrigerators, air conditioners, etc.) that would be connected to the network. During periods of heavy power demand, the power companies would broadcast alerts through the network and these devices would reduce power consumption temporarily. Apparently, preliminary studies on the Blacksburg project have suggested that if the power company actually paid for smart thermostats (assuming that the network infrastructure was in place) they would recover their costs within two years.

Service providers, including libraries operating online catalogs and institutions supporting anonymous FTP archives, have little legal or policy guidance and view the situation with a considerable degree of unease.[47]

If one considers libraries, which at least have a historical policy context that might help them to develop policies for privacy in access to new online information services, one finds a variety of practices and motivations behind them. Many online catalogs provide anonymous access simply because it is easier than having to maintain user files for a large, and, in a university, rapidly changing user community, and not because of any policy commitment to the right of anonymous access to the online catalog (as distinct from a possible policy position on the right to privacy of searches; in other words, the library is saying that it will protect privacy, perhaps, but not providing the absolute guarantee of privacy that anonymous access gives to the user). Some institutions controlled access simply as a means of regulating resource utilization; these controls sometimes required users to identify themselves through user IDs and in other cases preserved anonymity by using controls such as originating network address to determine whether a user had access. As online catalogs have expanded to include abstracting and indexing databases and other electronic resources licensed to specific user communities, it has become necessary for many systems to implement some type of user identification mechanism in order to control access to these licensed resources in accordance with the license agreements. A few institutions, such as the University of California, have developed approaches to user identification that provide for patron anonymity, but many have simply gone to a user ID mechanism, often based upon the library card number. To some extent the questions about accommodating anonymous access tie back to the library's overall priorities; in a period of intense pressure on budgets and library resources, many libraries are articulating strategies that place priority on serving their primary clientele (for example, members of a given university community) and provide access to other users on a lower-priority basis.[48] Members of the primary clientele will typically be registered with the library and this registration process provides them with user IDs that can be used for authentication (thus shifting the issue from guaranteed confidentiality through anonymity to policy confidentiality provided by the library). As online catalogs grow into ever richer and more complex mixtures of public and restricted access licensed information resources, it is much simpler to abandon the attempt to provide anonymous access when feasible and move towards a uniform authenticated access model, which is less problematic for the

[47] T. a great extent library patrons are protected more by practice than by law; libraries do not collect information such as what books a patron has borrowed beyond the point that he or she returns them. There may be some statistical information used for collection development that links demographic characteristics of patrons to borrowing records, but the old sheet of paper or card in the back of the book in which the names of those people who have borrowed the book over the years has largely been eliminated by libraries, at least in part in response to growing concerns about patron privacy. In the electronic environment, we may see a clash of cultures; telephone companies, for example, typically gather and retain very detailed records of who each customer has talked to and when; these **are** easily accessible with a court order.

48 This is also being done for networked information resources; for example, some FTP sites limit access during the day by users outside of a specific set of network addresses that are viewed as defining the primary user community, or the limit the amount of anonymous traffic to ensure that resources are available to serve the primary clientele.

primary clientele than to other outside users who wish to make occasional, casual use of the library's information system through the network.

There are other reasons why online catalog designers are moving towards (at least optionally) authenticated access as online catalogs become more sophisticated [Lynch, 1992]. This is needed for current awareness services that electronically mail notification of the arrival of interesting new materials to users, for intelligent interfaces that track a user's history with the system and his or her preferences, or that tailor the dialog to the user's familiarity with the system based on how long and how often he or she uses it. Again, it is certainly possible to support both authenticated and anonymous access modes, and even to permit users to store preference files external to the library information system, importing them when they start an anonymous session and exporting them again at the close of the session, but all of these options add considerable complexity to the system design, the cost of which is certainly subject to question, particularly in the absence of any policy or community consensus that underscores the importance of offering an anonymous access mode.

Matters are complicated by several conflicting sets of demands on service providers. Indeed, this conflict goes beyond operational needs to a basic conflict of values between the library community's tradition of free access to information and the computer community's emphasis on tracking, auditibility and accountability. Computer and network security practices stress the importance of audit trails and other monitoring tools to protect systems from intruders, and system security and integrity are major issues for any service provider on the Internet. In fact, there seems to be consensus among many of the regional network service providers that anonymous access to the Internet is unacceptable (for example, providing access to terminal servers that can TELNET to any Internet host without first identifying the user at the terminal server so that attempts to break into system can be tracked back to an individual at that institution—but note here that there is no requirement that the information be propagated outwards from the source terminal server, only that it be maintained so that by cooperation among organizations a trail can be defined back to an account at the first institution). Certainly, there are many systems that permit anonymous incoming access, but in order to satisfy these restrictions they limit access going back out to the network to specific, limited sets of hosts that have agreed to permit anonymous incoming access. For applications where there is recharge for information access, careful tracking of users is needed to allow discrimination among user groups. This question of anonymous access to the network has sparked bitter arguments between the library community and the networking community, as many libraries view themselves as potential access points to the Internet, and at least some libraries have taken the position that they should not have to require users to identify themselves in order to access resources on the net that are willing to accept these incoming connections. It seems likely that as "public-access" resources on the network multiply, and particularly as federal, state[49] and local government information becomes more

[49] In California, Assembly Bill 1624 is currently under consideration, which, if adopted, would require that various legislative information be made available at little or no cost to the general public through the Internet. One serious proposal by some members of the legislative staff is that the identity of those members of the public requesting this information be tracked for various reasons.

commonplace that the conflict between security and the right to anonymous access will continue to be troublesome.

Many of the information services being offered on the Internet are viewed as somewhat experimental; indeed, we are all still learning how to build user friendly and effective information retrieval and navigation tools, and analysis of user sessions is a key tool in improving the quality of such systems, as well as more routine tuning and capacity planning efforts that are part of the operation of any large scale service. Finally, it is important to recognize that not all information providers on the Internet are institutional; for example, it is quite common to find academic departments, research groups or even individual faculty members setting up anonymous FTP directories to permit people to obtain copies of their papers and research reports. They view this as not much different than responding to postcards asking for offprints or orders for technical reports, and retain a natural curiosity about who is reading their work (which was evident in the days when they responded to requests for printed copies).

Ironically, part of the problem is the development of the distributed computing infrastructure. Ten years ago, when online catalogs were initially being deployed by most libraries, access was primarily from within the library, or perhaps from a few large timesharing hosts on a university campus; if the library was recording searches, it would typically only know that a given search came from terminal 43 within the library or from machine X (which might have 500 registered users). The identity of individual users accessing resources on the network was effectively hidden behind these large multi-user timeshared hosts, and, while a given network resource might require a user to identify him or herself in order to use that resource, the user was aware when such a request was issued by the remote system-one was asked to log in, or provide a password. Very little information about the identity of individuals accessing a remote service could be determined autonomously by the remote service; if the service offered anonymous access (that is, it did not ask for information from the user accessing it) then the user could have a reasonable degree of confidence that access really was anonymous (barring collusion between the user's local host and the remote host; statistical analysis of who was logged onto the user's local timeshared host in comparison to when a remote service was accessed from that timeshared host could, over time, probably allow a sufficiently interested analyst to trace accesses back to individuals, but such activities are rare, and most users view them as too much trouble to represent a serious threat to anonymous access). As we have migrated to a world of personal workstations, the origin address for a search (or a request to fetch a copy of an electronic document) is linked to a specific host address, and increasingly this host, which is now a workstation, is now in the service of a single master. In the new networked environment, the source of a query or a request suddenly provides a great deal of information about the identity of the individual issuing that query or request. This should not be narrowly viewed as a matter of personal privacy; in fact, in the network environment, it is often hard to identify an individual but easy to identify the individual's organizational affiliation by the network number in the incoming Internet address. While people outside of organization X may find it hard to determine that a given address is person Y's workstation, everybody can tell that the access has come from organization X. This may be a matter of competitive intelligence rather than personal privacy.

Certainly there are technological solutions to the problem of one's address revealing one's identity. The simplest is to carry forward the time honored method of mail drops (post office boxes, or the mail forwarding services that various newspapers have long offered in conjunction with personal advertising). Electronic mail based dating services offering such anonymity through the agency of a mutually trusted third party are already operating on the network; a similar service could easily be set up for TELNET. But, as the number of protocols multiply and distributed system architectures become more complex, the development of general purpose anonymity services will become quite problematic. Further, one must wonder whether the vast majority of users will recognize when their use might be appropriate; the example of dating services is a good one since it is simply a recreation of existing practice in the electronic environment in a fairly direct way, and consequently its use in the electronic environment is appealing to the same people who would likely have used it in a non-electronic world. Whether users will recognize the new risks introduced by the development of new electronic information services, or the redesign of old services for the electronic environment remains an open question.

We are only beginning to explore the challenges that distributed computing raises for individual privacy in the context of "anonymous" remote terminal access becoming increasingly easy to trace back to an individual as more users use their own personal workstations rather than large timeshared hosts. At least in the remote terminal emulation environment—be it TELNET or more modern X Window system based applications-the user employs widely available, well documented, industry standard utility software that is written according to publicly available specifications and which can be used with a very wide range of remote services. Often, software to implement protocols like TELNET and the X Window system is available from multiple sources for a given platform (both commercial software suppliers and public domain or "shareware" sources). While there are some true distributed client-server protocols that are well documented national or international standards, such as the Z39.50 information retrieval protocol, and these protocols are implemented in multiple client software applications that can again be used with a wide variety of remote servers, in the developing client-server oriented distributed environment we will see providers of information services implementing custom software clients. These clients will be distributed to users in executable form only; they will employ proprietary protocols, and will be needed to obtain access to specific information servers. In essence, the user of such a service is expected to execute a program of largely unknown function *which typically has full access to the files on his or her personal workstation, given the current state of the* art *in the operating system* software *that runs on most of these workstations,* and which opens and uses a communications channel to a remote service as part of its normal, expected behavior. This is already the case in some commercial services, such as Prodigy [Burke, 1991].

The opportunities for collection of information are endless; for example, such client software might upload a list of what software the user has installed on his or her hard disk,[50] or the list of USENET newsgroups to which the user is subscribed.[51] Unlike

[50] Lists of software installed on machines is useful not only for marketplace research or marketing demographics (for example, to identify people who might be interested in add-on software to an existing product or in competing products) but for other purposes like identifying illegal copies of software: a

general purpose utility software (for example a TELNET-based terminal emulator), the covert information collection activities of specialized client software may be very difficult to identify and monitor,[52] and while very sophisticated users or institutions may be able to address this problem legally with the supplier of the service (and the client software), most users will likely remain unaware that the problem even exists. We may see organizations giving away client software and access to certain remote services through that client software just to be able to get users to run the client and unwittingly export information that the service provider can use directly or resell to others. We may find a direct contradiction between realization of the distributed computing potentials of the Internet and individual user privacy.

There is another interesting relationship among pricing, privacy and the capabilities of systems supported by information providers in the distributed computing environment. Currently, information providers frequently charge based on the amount of information that is exported from their systems; they offer filtering tools of varying degrees of sophistication. On a purely technical basis, some users of some system choose to do fairly unselective extractions from the information providers and then do ranking and filtering on their local machines; this has the effect of preserving some privacy (since the fine details of what the user is interested in are not conveyed to the information provider) but also tends to run up a large bill since the information provider assumes that everything that is exported is of value to the user and will probably actually be examined by the user, rather than filtered by a computer program running on the user's machine. As information provider capabilities improve, the decision as to how much information to give the information provider in order to permit the provider to perform filtering will likely be based in part on how specifically the user is willing to reveal his or her interests to the information provider; privacy (gained by the method of asking vague questions) will have a price. Balancing this, however, we should note that the trends in technology are towards user clients that act as integrators for *multiple* information providers, not just one providers, and such an integration function obviously cannot be pushed outwards to the providers, since no individual provider has the full range of information necessary to do the ranking and filtering of information from multiple sources.

---

company making multiple products could use one to scan for the presence of copies of others, and then check its registered user files.

51 The suggestion that a local client could exploit information about a user's subscriptions to USENET newsgroups is due to Simon Spero, although he proposed it in the context of client software using this as hints in developing a user profile which could be used to help tune information retrieval applications, and not as a mechanism for invasion of privacy.

52 Many personal workstation operating systems can now be equipped with virus protection software which can detect and warn the user of unexpected *modifications* to **critical** files on the user's machine, but I have never seen one which monitors access. The user does have some countermeasures, such as keeping critical files on a separate disk and never mounting that disk while running software that he or she does *not* trust, or encrypting critical information when it is not being used, but the cumbersome nature of these measures makes them impractical outside of very high security environments with very security-conscious users.

Who Owns Access Histories?: Privacy and Market Research

The analysis of consumer behavior has become a major focus of attention in the business world. Supermarkets have on the one hand implemented laser scanners that track the products being purchased by shoppers (and linked them to systems that automatically issue a set of custom tailored discount coupons at the checkout register) and on the other hand now encourage payment with credit cards, allowing the development of databases that track consumer purchases in tremendous detail [Mayer, 1990]. Companies like American Express that have access to extensive histories of customer spending practices and preferences are now marketing finely tuned customer lists to their business partners-for example, I might receive mailings from American Express that offer me airline upgrades on airlines that I don't fly regularly, based on statistical analysis of my purchasing profile which indicates that I spend over $25,000 per year on airline tickets and that none of these charges go to certain airlines. Similarly, in many industries there is now an intense focus on what goods are selling, and in what marketplaces, and this information is employed in very complex pricing decisions (consider again airline seats as an example.) The practice of "data mining" from customer histories has begun to be viewed **as** simply effective exploitation of previously untapped corporate assets [Piatetsky-Shapiro & Frawley, 1991]. In addition, we are now seeing considerable use of multi-source data fusion: the matching and aggregation of credit, consumer, employment, medical and other data about individuals. I expect that we will recapitulate the development of these secondary markets in customer behavior histories for information seeking in the 1990s; we will also see information-seeking consumer histories integrated with a wide range of other sources of data on individual behavior.

The ability to accurately, cheaply and easily count the amount of use that an electronic information resource receives (file accesses, database queries, viewings of a document, etc.) coupled with the ability to frequently alter prices in a computer-based marketplace (particularly in acquire on demand systems that operate on small units of information such as journal articles or database records, but even, to a lesser extent, by renegotiating license agreements annually) may give rise to a number of radical changes. These potentials are threatening for all involved. Consider just a few examples:

- For the first time, libraries should be able to **easily** collect reliable data on how often specific journals are read, or even the pattern of access to specific articles within these journals. This information can be used to decide not to subscribe to journals, which worries the publishers.

- Publishers can employ this usage information to set prices on journals or even specific journal articles based on popularity. This leads to price instability, which worries the libraries.

- While citation data as a measure of the impact of a publication has been controversial (though it is already considered in tenure and promotion decisions at some institutions) usage data is less ambiguous; if nobody reads a publication, it is unlikely to have had much impact. This is of great concern to authors.

- Usage data makes it much easier for authors, publishers and libraries to rapidly reflect the short-term interests of the user community by keeping track of what is popular and trying to produce or obtain more of it.[53] To some extent, this is at odds with the development of the scholarly record and the integrity of scholarship. Archival publications are not necessarily read a great deal, but some would argue that it is of vital importance that they continue to exist.

- There is a tendency in systems that stress popularity to ultimately reduce diversity; if everybody else is reading something, then one concludes that one needs to read it also. The temptation to select as one's reading the ten most popular articles of the week is very dangerous to the development of a diverse body of ideas. It is also worth noting that producers of abstracting and indexing databases are increasingly considering the subscription patterns of libraries in deciding what journals to cover; this seems to make the databases more marketable. If these producers were to emphasize heavily read journals, these abstracting and indexing databases will tend to become less comprehensive guides to the literature (and, indeed, pathways to material in less well known journals).

- There is a danger that the system of statistics collection can be manipulated by those that understand it. This can range from authors repeatedly accessing their own works to get their statistics up through more sophisticated approaches (for example, including many popular keywords in an abstract even if they have little to do with the actual subject of the work so that many people will retrieve and view the work).

These examples have emphasized applications of data about the use of information resources such as viewing or downloading journal articles. However, the availability of searches is also of great value: it tells information product designers about the kinds of information that people are looking for, and also the means that they are using to locate it. This is invaluable market research data for designing new information products, and for marketing and improving existing ones.

The ability to collect not only information on what is being sought out or used but also who is doing the seeking or using is potentially very valuable information that could readily be resold, since it can be used both for market analysis (who is buying what) and also for directed marketing (people who fit a certain interest profile, as defined by their information access decisions, would likely also be interested in new product X or

---

53 It is interesting to note how each technological development that undermines privacy seems to be complemented by a technological countermeasure that supports privacy. Consider the case of pay telephones. At one time, these were a wonderful way to obtain anonymity; one simply deposited cash. and the source of the call was untraceable. Now, of course, most **pay** phone users are using credit cards because they are so much more convenient, not realizing that if they use these cards all their calls can be tracked in great detail. (In fact, many public phones will not even take cash anymore, due in part to the expense of collecting the cash and the fact that the cash is an invitation to vandalism.) In France, vendors now offer a phone card which has a specific "cash" value; one pays cash for it, and it is debited as one makes phone calls using it. This is a form of "electronic cash" which facilitates anonymity. (It also has some other important advantages; for example, while one **can** loose the card, one cannot incur the virtually unlimited bills against one's account that can be caused by a stolen phone credit card number.) Another example of this technological balance is the development of Caller ID facilities by the phone companies; these were quickly complemented by facilities that allowed a caller to block the display of the Caller ID to preserve anonymity.

special offer Y). While such usage (without the informed consent of the recipient of the advertising) may well offend strong advocates of privacy, in many cases the consumers are actually quite grateful to hear of new products that closely match their interests. And libraries and similar institutions, strapped for revenue, may have to recognize that usage data can be a valuable potential revenue source, no matter how unattractive they find collecting, repackaging and reselling this information.

Competitive intelligence is a burgeoning field promoted by any number of consultants. One aspect of competitive intelligence is knowing in what areas competing corporations (or, in academic world, research rivals) are seeking information. For example, it is valuable to know, if one is a coporation in the pharmaceutical industry, that a competing corporation is seeking articles about the effects of a given drug. Of course, once one recognizes that one may be a target of competitive intelligence, it is possible to deliberately offer disinformation that will lead the competition to an incorrect assessment of one's interest, and even deliberately send a competitor down false trails. Clearly, the type of information that can be collected about information seeking and use in the networked environment is invaluable for competitive intelligence. And it is worth noting that even fairly highly aggregated information can be of value in a competitive intelligence activity: for example, from the aggregated article access information for a given university (without any indication of who within that university accessed the material) it is quite reasonable to draw conclusions about the research directions of specific research groups that are very likely to be correct.

Some of these examples seem farfetched. But consider a number of trends. As electronic information providers license information rather than simply selling it, they can *require* usage reporting as a condition of license.[54] This is done in other areas. Libraries and universities are both aggressively seeking new ways to generate revenue; the resale of statistics about electronic collection use and/or searches, particularly if they can satisfy themselves that some level of privacy is being maintained by not including the identity of the user (if they know it) could be a very attractive revenue source. It is unclear whether these institutions have any legal obligation to ensure confidentiality of this information.[55] If one signs a contract with a commercial information service such as Dialog, issues of confidentiality can be negotiated in advance as part of the contract; but when one is accessing (anonymously or otherwise) a public-access information service, it is unclear what to expect, and in fact at present there is no way to even learn what the policy of the information service provider is. As the secondary markets develop it is even conceivable that when accessing a for-fee resource one might pay more for privacy, and that when accessing a public-access resource the user's client and the server for the public-access system might well negotiate various levels of confidentiality (no logging, statistical compilation only, actual

54 Such conditions could be imposed either directly on a library licensing the information for local mounting, or, less visibly, thorough contractual constraints on a third party such as Dialog or OCLC that makes the information available to the library community.

55 Resale of information about who is searching what type of information is not the only issue. A financial information service might be a good investment for a brokerage house or a merchant bank for example; they might internally exploit knowledge that could be gained from records of what customers were searching information about what companies or products.

text logged for searches but without ID, no resale or reuse outside of the internal operations of the information provider, etc.)

A final aspect that should be mentioned is that in the print world the library served as a shield for its user community in the sense that it purchased materials such as journal subscriptions. The act of purchasing and the cost of purchase might well be an act of public record discoverable under a Freedom of Information Act. But only the library knew who was using the material, and that information (in the form of circulation records) was protected. Further, because information was acquired in highly aggregated forms such as an annual subscription to a journal, the act of purchase revealed very little about the interests of the library patrons—indeed, there is no a priori reason to assume that purchasing decisions are always directly driven by the short term needs of specific patrons in the case of a research library. Now, consider the electronic world, where a library frequently acquires rather specific information (such as a single article from a journal) in response to the specific request of a user. This purchase, as an external business transaction, may be a matter of public record. Further, if the end user rather than the library as intermediary acquires the article, it may be possible to rather directly link information use to individuals or departments. The electronic information environment may well call for considerable reassessment of the definition of public records, particularly in the context of state universities, as these records are defined by federal and state Freedom of Information Acts.

The uses described for information about searches and access patterns are simply extensions of well established practices such as market demographic analysis and competitive intelligence. New uses for this information, unique to the networked information environment, are also being researched. For example, Professor Mike Schwartz at the University of Colorado has been exploring the concept of resource discovery-automated methods of discovering or locating potentially interesting network resources [Schwartz, 1989]. One of the techniques that he has studied is the examination of access pattern of other members of a user% affinity groups; for example, a botany professor might be interested in resources that other members of the botany faculty are utilizing regularly but that he or she is unfamiliar with. Such research may ultimately lead to new tools for locating information resources which will call into question the appropriate balance between privacy, competition, and cooperation in various communities.

Privacy, Intellectual Property and Electronic Mail Enabled Communication

Electronic mail based discussion groups—sometimes called electronic journals in academic circles if their editorial policies parallel those of traditional printed journals —have become extraordinarily popular on the Internet. These fall into two major categories— LISTSERVs and mail reflectors. Mail reflectors are simply special user IDs; when one sends electronic mail to such a user ID it is redistributed to the subscribers of the mail reflector automatically. Maintenance of the subscriber list is typically done manually or semi-automatically, with the convention being that if the mail reflector's address is of the form user@ hostname then there is an additional mailbox in the form user-request@ hostname to which requests to join or leave the mail reflector are directed. LISTSERVs are based on a program that was originally developed for the IBM

Conversational Monitor System (CMS) environment.[56] There are thousands of such mailing lists on the Internet; in addition, many such lists are reciprocally gatewayed to Usenet newsgroups, which are essentially very large collections of publicly readable electronic mail messages that are propagated through the Internet (and beyond), organized by topic (newsgroup name). Like electronic mail lists, some Usenet newsgroups are moderated; others are completely open. Unlike directly electronic mail enabled services (LISTSERVs, LISTSERV imitators, and mail reflectors), Usenet newsgroups do not appear to the reader as electronic mail, but rather as continually updated databases that are viewed through a program such as a newsgroup reader; electronic mail only comes into play when a reader wants to enter a Usenet Newsgroup discussion by posting a message. The privacy and intellectual property aspects of these mailing lists and newsgroups are very interesting, and probably largely ignored by most participants in them.

Some lists deal with topics that are controversial—for example, the Usenet news group ALT.SEX.BEASTIALITY or the recently established LISTSERV on gay and lesbian issues in librarianship, to mention only two examples.[57] Currently, USENET newsgroups offer a moderate degree of privacy; a newsreader running on a client machine uses a protocol called NNTP (Network News Transfer Protocol) to pull list postings down from a local NNTP server to the user's client. The list of newsgroups that the user is interested in is stored on the client, and one can find out what a given user is interested in only by looking at his or her preference files on that client machine or by monitoring data transfer from the NNTP server, both of which are relatively difficult.[58] LISTSERVs, on the other hand, require interested parties to actively subscribe in order to receive electronic mail that is posted and include options which allow anyone to look at who has subscribed to a given list (except for those users who have explicitly chosen to conceal their identities; these individuals are invisible except to the system administrators or list administrators). Of course, the vast majority of LISTSERV subscribers are blissfully unaware of the fact that their identities can be easily discovered, or of their option to conceal their identity. It is only a matter of time, in my

---

56 More recently, software has been developed for the UNIX environment which emulates most of the functions provided by the LISTSERV program in the IBM CMS environment.

57 There has been enormous controversy about the appropriateness of various universities making such newsgroups available to their communities; these have been well documented in the Computers and Freedom mailing list postings by Carl Kadie. To my mind, these controversies help to illustrate the gulf between the library tradition of not only intellectual freedom but of free access to information and the values of the computing establishment. It seems likely that if these were print works that were owned by the libraries of the Universities in question there would have been little debate about the right of these libraries to own them as part of their collection and to make them available to the university user community; this would have been a clear case of intellectual freedom on the part of libraries. But when such resources are made available thorough institutional computer systems (where there is little philosophical basis established for determining appropriate content) major controversies quickly erupt.

58 Interestingly, traffic analysis between client and NNTP server is probably easier than breaking into the file system on the client, and this can be viewed as another illustration of the way in which the deployment of the distributed computing environment has exposed individual's activities to much more scrutiny. If the client is on a large time shared host then it is not clear why specific newsgroups are being transferred to that timeshared host; if the client is on a personal workstation, however, it is relatively easy to *assign* responsibility for the transfer of material from a specific newsgroup. There is also a program called "arbitron" written by Brian Reid which publishes regular statistics about the usage levels of various newsgroups; this is again, I believe, based on traffic analysis techniques.