Journal of medical ethics, 1977, 3, 174-178

Privacy, confidentiality and automated health information systems

Hannu Vuori * *Department of Community Health, University of Kuopio, Finland*

Professor Vuori's paper, first presented at the fourth Medico-legal Conference in Prague in the spring of this year, deals with the problem of the maintenance of confidentiality in computerized health records. Although more and more information is required, the hardware of the computer systems is so sophisticated that it would be very expensive indeed to 'break in' and steal from a modern data bank. Those concerned with programming computers are becoming more aware of their responsibilities concerning confidentiality and privacy, to the extent that a legal code of ethics for programmers is being formulated. They are also aware that the most sensitive of all relationships – the doctor-patient relationship – could be in danger if they failed to maintain high standards of integrity. An area of danger is where administrative boundaries between systems must be crossed – say between those of health and employment. Protection of privacy must be ensured by releasing full information about the type of data being stored, and by maintaining democratic control over the establishment of information systems.

In 1965 the Social Science Research Council of the United States recommended that a national data centre be created. Warnings against intrusions into privacy had certainly been voiced earlier, but this recommendation can be considered to be a major impetus for the current lively discussion on privacy and confidentiality. The nature of the recommendation readily gave the impression that it is computerized information processing that poses the greatest threat to privacy.

Three fallacies concerning privacy

The discussion on privacy seems to be plagued by three fallacies: 1) that the possibility of violating an individual's privacy is a new problem, somehow related to the current mode of functioning of industrialized societies; 2) the privacy problem is almost exclusively related to, and perhaps even created by, the increasing use of computers; and 3) the problems and their solutions are largely technological.

Because the industrialized countries are orientated towards a specific goal and try to base their activities on the prudent use of information, the privacy issue is characteristic of them. But it is definitely not a new issue. It has existed as long as information pertaining to individuals has been needed, for example, by the authorities for taxation or for military purposes, nor is the problem exclusively related to the use of computers. The ability to reproduce *en masse* the likeness of an individual (photography) or to distribute to large audiences (the press) information concerning him, such as facts or fabrications about his behaviour, sayings, opinions and possessions, constitutes as important a landmark in the development of the privacy issue as the production of the digital computer. It is interesting to note that one of the really pioneering articles elucidating the ramifications of the privacy issue was published as far back as 1890 by Warren and Brandeis,[1] long before the advent of computers. It can safely be maintained therefore that the privacy issue is not the result of the development of computer techniques. The increasing concentration of information relating to individuals would have developed in any event.

The issue of privacy is clearly one of values and not of technicalities. In a society with certain aims, a certain amount of information is necessary, and the right of that society to gather information about its members is an acknowledged principle. In addition, public access to such information is considered to be an invaluable method for preserving democratic control over the bureaucracy. To centralize information has become a goal of people engaged in almost every kind of activity that depends heavily on the analysis of data. Much of the data does not relate to the circumstances of the lives of individuals, although certain demands for centralized information do concern information about individuals. Research in the behavioural sciences, information for policy making and planning at various levels and the operations of law-enforcement agencies could be greatly enhanced by comprehensive information about every citizen. On the other hand, the right to privacy is also a cherished value. As a consequence, the issue of privacy is one of a conflict of values: the value of a greater understanding of society and of more intelligent methods of formulating policy and the value of allowing an individual to keep information about himself and his life private and unknown to others may be contradictory.[2]

* Currently at the National Center for Health Service Research, Hyattsville, Md, USA.

Privacy, confidentiality and data security

The discussion usually revolves around the concept of privacy which is primarily concerned with the basic issue whether society, an organization or an individual has the right to gather information about a person. The vagueness of the concept is illustrated by the fact that at least four disciplines have their own definition of privacy: psychology, the engineering sciences, political science and law. The various definitions emphasize, however, certain common features. The right to privacy is the right of the individual to decide for himself how much he will share with others his thoughts, his feelings and the facts about his personal life.[3] Psychological studies with hospital patients have indicated that there are three areas of privacy: privacy of life style, privacy of event and privacy of personality.[4] Typical of these areas is that for each of them a boundary can be defined – but a flexible boundary. The individual does, however, try to maintain control and autonomy over the areas that he has defined as private. In legal discussions concerning the protection of privacy, the following specific areas have been mentioned:[5] privacy means the protection of the individual from 1) unreasonable observation by auditory, visual or psychological means; 2) unreasonable usurpation of his name or likeness; 3) unauthorized interference or interception of private communications by any means whatsoever; and 4) the unauthorized access to personal or confidential information. To these can be added 5) protection against public disclosure of embarrassing private information about an individual; 6) protection against publicity which places that person in a false light.[2]

According to Westin,[6] privacy has the following functions for an individual: it safeguards personal autonomy; it provides emotional release; it constitutes a basis for self evaluation; and it enables protected communication. Because of the vital importance of these functions to the wellbeing of an individual, the right to privacy has been increasingly recognized by the law and given attention by the public and by the decision makers.

Once data about an individual have been collected, the confidentiality of information becomes the central concern. This refers to the improper use of information voluntarily and confidentially given by the individual to a user of information for a certain purpose.

The last problem related to this area is that of the security of data. Besides improper use, many other things can happen to data. Data can be lost or accidentally or purposely destroyed. The manipulation of data in ways not intended when the data were collected can be classified either under this category or that of confidentiality, depending on the nature of the data and of the manipulation.

The distinction between these three types of problem is helpful as the actual breaches related to the problem complex can be classified in the same way. Furthermore, the protective measures available largely depend on the kind of problem we are dealing with. As, however, privacy, ie, the right to gather information about an individual is the overriding issue, the whole problem area is often treated under the single heading 'privacy'.

Computers and privacy

The introduction of computers cannot be blamed for the concentration of information relating to individuals. In fact, none of the operations performed by a computer is different in kind from those which could at least in principle be carried out by traditional methods although the use of computers has undoubtedly enhanced some of the existing problems of privacy. The following features characteristic of computers are relevant in this respect: 1) they facilitate the maintenance of extensive record systems; 2) they make data easily and quickly available from many distant points; 3) they make it possible for data to be transferred quickly from one information system to another; 4) they make it possible for data to be combined in ways which might not otherwise be practicable; and 5) because the data are stored, processed and often transmitted in a form which is not directly intelligible, few people may know what is in the records.[7]

In spite of the potential danger these capabilities and features constitute, it should be emphasized that the privacy issue is not primarily related to the method for handling information but to the kind of information to be handled. Similarly, it should not be forgotten that the use of computers offers many safeguards that are not available in the case of traditional information-processing methods. However the problem is rendered insidious by the fact that people are relatively poorly informed about the existence of records containing information about them.

The Committee on the Protection of Privacy *vis à vis* Electronic Data Banks (of the Organization for Economic Cooperation and Development) very pertinently remarked at its third meeting in 1972 that besides EDP-based information systems, other ways of collecting information might also create problems of privacy. Therefore, care should be taken not to increase and facilitate the improper use of other types of information system while trying legally to control the use of computerized information. The protective policy to be adopted should reflect a balance between the need for freedom of information and the need for privacy, result from the consensus of all affected parties, cover both manual and automated systems and provide for a uniform approach to its implementation.[8]

Threats to privacy

The most commonly cited dangers to privacy in connexion with the use of computerized health information systems stem from three main sources: 1) inaccurate, incomplete or irrelevant information; 2) the possibility of access to information by people who should not or need not have it; and 3) the use of information in a context or for a purpose other than that for which it was collected.[9]

The major problem of privacy created by the use of computers is the fact that information is becoming a mass product. Large amounts of information can easily be assembled, processed, regrouped and distributed. Alongside this technical development there are two trends concerning the content of the information systems: they tend to become broader (contain data on more variables than before) and deeper (contain more detailed, intimate and far-reaching information about an individual than before). As a consequence, we have two major interrelated concerns: the access to data and their accuracy. We are worried about easy accessibility because we suspect the accuracy, and we are concerned about accuracy because we are afraid that the attempts to limit accessibility may fail.[9]

A third concern is related to the computer's ability to remember. According to the Judeo-Christian way of thinking, man should be given a second chance. The computer cannot forget and is incapable of forgiving. As a consequence, there may be a danger that he has to drag with him the past although it may have lost its relevance. This fear is closely related to the Orwellian 'Big Brother' syndrome: a man's past can be used against him if sensitive information has been stored. It has been claimed that although the present safeguards against such use of information may be adequate, there is no guarantee that there will always be a government that respects the privacy of its citizens.

The nature and magnitude of the threat is partly determined by the type of system under scrutiny: isolated systems vs an integrated computer network; batch processing vs on-line processing; dedicated single-purpose systems vs multipurpose systems; a system falling under one authority vs a system serving several authorities. As a general principle it can be said that the more complex and larger the system, the more vulnerable it is. Some of the proposed health information systems, especially those serving national planning, tend to be large.

In a large system connecting several remote terminals, there is no guarantee that the remote access locations will have any semblance of physical security although it may be vigorously maintained at the central computer facility. The communication lines themselves are vulnerable to tapping. The more extensive the networks, the greater the probability of error and the vulnerability of intrusion. It is very difficult to verify that any large software system is completely free of errors and anomalies. Errors, compounded with frequent change in the system, can cause great security problems when multiplied rapidly over a large network. The familiarity with other personnel, typical of small and isolated units and conducive to preventing misuse, is lacking in the big networks.[10] At the same time the big networks do offer some safeguards against misuse. Many of these are commercial. The companies are not selling hardware but simultaneous and multiple access to central systems. These companies would not stay in business for very long if they could not protect the privacy and integrity of their customers' files and programs. The multitude of customers not only necessitates but also enables extensive security measures by allowing heavy investments in security.[5]

An additional way of counteracting some of the dangers to privacy inherent in big systems is the fact that it is relatively easy to develop various authorization schemes for people using the system via terminals.

Medical and health information systems

When discussing the issue of privacy, health has often been mentioned as an especially sensitive area where the misuse of information may have grave consequences. The problem here, however, is much more one of confidentiality than of privacy. The patient has usually consented to give the needed information. In fact, the disclosure of even the most private kind of information is considered to be one of the essential elements of a successful doctor-patient relationship. As a consequence, privacy does not constitute a special problem in the treatment of patients. The situation may, however, be entirely different when information is being collected not for direct patient management but for epidemiological studies or for planning purposes. In patient management, the real problem is that of confidentiality. To maintain the confidential doctor-patient relationship the patient has to be able to be certain that the information he has confided to the physician or other health personnel will indeed be kept confidential and not used for any other purposes that the one for which it was originally disclosed. Only in cases where the patient has in some way been under coercion when submitting the information can there be said to be an intrusion into his privacy. Such instances can of course, occur, eg, when applying for an insurance policy or a certificate for absence from work.

The confidentiality of medical information contained in health information systems may be in jeopardy exactly for the same reason as any other kind of information stored, processed and distributed by computerized methods. In fact it has often been claimed that medical information can be in much better custody when in the computer files than when kept in manually operated records. The physical security of records is more often than not far from optimal, and medical information may actually be safer than many other types of information. It is treated by staff who, during training, have absorbed a code of ethics which strongly emphasized the importance of confidentiality. Also medical information is in most cases handled within a closed system which renders it impossible or at least difficult to combine it in an improper way with other information. The fraudulent use of medical information very seldom carries the same kind of financial incentives as that of, say credit information or accounts. The greater sensitivity of certain types of medical information, eg, of psychiatric conditions, venereal disease, reproductive behaviour (including abortions) does, however, put medical information in a separate category that warrants special attention.

An additional reason for paying special attention to medical information is the tendency to develop large-scale computerized health information systems. These can be defined as systems for the collection, storing, processing, analysis and distribution of information for planning health care, management and treatment as well as for research and teaching.[11, 12] The wide range of uses and the broad distribution of the data indicated by the definition should render these systems subject to special vigilance.

Methods for protecting privacy

When pondering over safeguards against potential threats to privacy, one should first try to anticipate the frequency with which various such threats occur. According to the experience of the computer industry by far the most probable mishap to occur to data is the occurrence of errors and omissions. The second commonest mishap is related to the perhaps most feared aspect of the breakdown of privacy – the misuse of data. In the great majority of cases this is done by the personnel employed in computer installations. Furthermore, these employees tend to use dishonestly those data which they have been authorized to use in order to perform their duties. Third on the list is fire, fourth disgruntled employees and fifth come damages from water. Last, and accounting for a very small proportion, is the improper use of data by outsiders, which features in such a prominent role in the apocalyptic views of those who are afraid of a computer-controlled 'big brother' society.[8]

The available methods of securing information that has been collected can be classified on two bases. According to the nature of protection, one can distinguish between legal measures, organizational measures, administrative measures and technical measures.

Legal protection consists of legislation defining the purposes for which information may be collected and the agencies that may collect information relating to individuals.

The administrative measures pertain to the procedure that has to be followed before an agency that is legally entitled to establish an information system can actually implement such a system. In several countries there are already supervisory bodies that have been granted the right to authorize information gathering. These bodies may issue their own regulations and directives concerning the kind of information and the application forms that have to be submitted before a data-gathering operation can be launched. These bodies may also exercise continuous supervision by requesting the data-gathering agencies to allow access to the computer premises to check the security measures, the contents of files and the format of outputs.

The organizational safeguards pertain to those measures that an agency that has been authorized to gather data adheres to in its own internal functioning. These may cover policies relating to employment, the allocation of duties, access to various premises, files and outputs, etc. There may be a code of ethics developed and endorsed by the agency.

Finally, the technical measures are related to the physical security of the premises and various technical procedures and devices related to the computer hardware, files and programs.

Another basis for the classification of data protection consists of the object of protection. These may consist of the surroundings of the information system, the computer hardware, the computer programs and the organization.

Protecting the environment of the information system can include the safe siting of the central processing unit and display devices as well as that of possible remote terminals. The same measures can be used partly to protect the hardware. Of specific importance is to separate the files so that only those are accessible to an individual which he needs for the performance of his duties. For this, two methods are useful: the identification of the user and the classification of the information. To identify the user, one can use something that he knows (password), something that he has (a card, a key), or something that is his (his voice or his fingerprints). When remote terminals are being used, it is fairly easy to develop quite reliable identification mechanisms ('extended hand shaking'). This identification may also include, besides the user, the terminal.

Information can be classified in several ways depending on the type of the system and the capacity of the users. In this context, the form of permitted output can also be related both to the kind of information and to the capacity of the user. In the case of medical information systems, the information can, for instance, be classified in the following ways: information required for administration (patient's name, age and sex); general medical information (length of stay in hospital, diagnosis, results of diagnostic tests); sensitive medical information (abortion, psychiatric diagnosis).

The precautions to regulate access to the output include, for instance, the following: reading the output only on a cathode ray display tube or in printed form only in a specific place; getting a printed output; obtaining information-storing media, eg, magnetic tape, punched cards; and direct data transmission between computers.

Conclusions

The computer industry is becoming more sensitive to the privacy issue, and as a consequence, the new hardware on the market contains built-in security devices that provide protection against most reasonably foreseeable risks. To break the current security systems, is so expensive that it may well exceed the potential gain to a penetrator from his intrusion. This is especially true in medical information systems where the money value of the information to be thus acquired is relatively small.

Protective legislation is also developing, in some countries to such an extent that there is already grave concern that completely legitimate and justifiable data-gathering activities are being hampered. For example, in Sweden epidemiologists are complaining that their research has been rendered tedious, time consuming and ineffective because of the regulations governing the establishment of information systems. There is also emerging among computer professionals a clear sense of their own responsibilities and even a code of ethics.

With regard to the specific area of health, the system seems to be fairly well under control. Those concerned in programming health data are aware of the importance of maintaining confidentiality which is of the utmost importance for the doctor-patient relationship. Even the public seem to be less concerned with the dangers posed by the health information systems than with those of many other information systems. Consequently, patients readily consent to provide even the most intimate information, so long as they are reassured that that information will be kept within the data bank relating only to the health services. But any crossing administrative boundaries between one system and another is being looked at with greater suspicion.

In the last analysis, perhaps the best protection is being provided by those concerns always explicitly stating the purpose for which the information is being collected and by limiting the amount of information to the really necessary items. The public should be told about the existence of records and systems containing information relating to individuals, and be provided with an opportunity to check the accuracy, relevance and timeliness of that information. Finally, the establishment of information systems should be controlled by democratic means, without, however, creating undue obstacles to legitimate and socially justifiable data gathering.

References

1 Warren, Samuel, and Brandeis, Louis, The right to privacy, *Harvard Law Review*, 1890, **4**, 193.
2 Meldman, J A, Centralized information systems and the legal right to privacy, *Marquette Law Review* 1969, **52**, 335–354.
3 Panel on Privacy and Behavioral Research: Privacy and behavioral research, *Science*, 1967, **155**, 535–538.
4 Schuster, Eleanor, A, Privacy, the patient and hospitalization, *Social and Scientific Medicine*, 1976, **10**, 245–248.
5 Lickson, C P, The right of privacy in the computer age, *Computer Group News*, 1968, **2**, 13–17.
6 Westen, Allan, F, Privacy and Freedom, Atheneum, New York, 1970.
7 Computers and privacy. Cmnd 6353. HM Stationery Office, London, 1975.
8 Thomas, R L, and Courtney, Robert H, A systematic approach to data security in approaches to privacy and security in computer systems, NBS Publications 404, Washington, 1974.
9 Karst, K L, 'The files: Legal controls over the accuracy and accessibility of stored personal data, *Law and Contemporary Problems*, 1966, **31**, 342–376.
10 Browne, Peter S, Security in Computer Networks in Approaches to Privacy and Security in Computer Systems. NBS Publications 404, Washington, 1974.
11 World Health Organization, Health Information Systems. WHO Regional Office for Europe, EURO 4914, Copenhagen, 1973.
12 Vuori, Hannu, The use of automated health information systems in the management and planning of health services, *Public Health*, 1977, **91**, 33–43.