EDP ANALYZER

NOVEMBER, 1975
VOL. 13, NO. 11

THE DEBATE ON INFORMATION PRIVACY: PART 1

Existing and pending "information privacy" legislation—both in the U.S. and in other countries—has been receiving much attention in the trade press. Perhaps you have been wondering how this legislation might affect your data processing operations. Or you might have heard widely conflicting opinions about the costs this legislation might impose on your company. The question is clouded because the privacy legislation applies to many aspects of an organization's operation in addition to data processing. In this report, we will concentrate on the possible impact of the proposed legislation in the U.S. on the data processing function, including the changes that will be needed in collecting and maintaining personal information on individuals. This is only one part of the total impact, of course, but it is a part in which you are likely to be very interested.

On December 31, 1974, the U.S. Congress passed landmark privacy legislation and President Ford signed the Privacy Act of 1974 into law on January 1, 1975. This Act imposes constraints and requirements on how U.S. federal agencies handle information about people. The provisions of the Act became effective on September 27, 1975.

In early April 1975, the U.S. National Bureau of Standards, in cooperation with the Mitre Corporation, sponsored a three day conference on "the privacy mandate." A good part of the discussion at that conference dealt with the problems of implementing the new law in federal agencies. But the conference took a broader viewpoint than just that. It also considered the proposed legislation, at both federal and state levels, that would extend the privacy mandate into the private sector. Reference 4 is a summary report of that conference.

A history of the information privacy program of the IBM Corporation was discussed at the above mentioned conference. We will give the highlights.

IBM's experience

An early indication if IBM's concern with improving the handling of information about their employees occurred in 1965 when Thomas J. Watson, Jr., then Chairman of the Board, wrote a letter to management defining an employee's right of access to his or her personnel folder. In 1968, a similar letter dealt with the separation of personal and business life. And in 1971 and again in 1973, selected groups of people made a thorough review of where personal data about employees was being stored and how it was handled. Then in 1974, the first formal management training was conducted on privacy measures for information about employees.

Through April 1975, the major changes that were made had to do with the collection and release of personal information about employees.

Each data item on the employment application forms, the employment records, the career records, and so on was scrutinized and questioned. Some data items, such as the history of nervous disorders and prior arrests, were deleted from the employment application forms; however, a question about convictions in the past five years was retained. The company stopped using personality tests and outside investigatory reports on applicants for employment. The use of the employee's social security number as an identification number was stopped; this number was removed from employee identification badges and is no longer provided, for instance, to insurance companies in connection with insurance claims.

IBM's previous policy had been that an employee's immediate manager was the one responsible for hiring, promotion, firing, aid in solving personal problems, and so on. But with the concern for individual privacy, some personal information about employees—such as a history of past illnesses, insurance beneficiaries, and history of prior convictions—no longer is given to immediate managers as they have no "need to know" such information.

The "need to know" criterion is now much more strictly enforced with regard to the release of personal information. An immediate manager, plus his manager, have authorized access to an employee's employment folder—but only for the information needed for the question at hand. When an outside company calls to verify that someone is indeed an IBM employee, the amount of information disclosed is very limited unless the employee agrees in writing to the release of additional information.

These are the salient points of what IBM has accomplished to date in the handling of employees' personal data. So far, costs have been incurred on a running basis. In many instances, new forms have replaced outdated forms only when supplies of the old forms were exhausted. Also, interviewers simply stopped asking some questions that they had asked previously. Interestingly, relatively few employees asked to see their employment folders when informed of their rights to do so. One of the problems encountered has been in the feelings and attitudes of managers, who in some instances thought that the company was endorsing a lowered level of concern for employees.

IBM's privacy program has only just begun. The company is in the process of establishing practices more in harmony with the pending privacy legislation in the following areas.

1. Identification of all files containing personal information about employees.
2. Determining the kinds of personal information collected and retained.
3. Informing employees how the personal information is used.
4. Limiting the dissemination of personal information.
5. Providing an employee access to his own personal records.
6. Providing ways for an employee to correct his own personal records.
7. Controlling the release of personal information outside the company.
8. Identifying the sources of personal information.

In studies to date, IBM has found that there were some 128 files containing personal information about employees at corporate headquarters alone. Much more employee information is stored at IBM's twelve divisions, one subsidiary, and at other locations within the U.S. In addition, IBM World Trade Corporation and its subsidiaries have similar employee data files throughout the world. It is likely that the company will adopt somewhat different practices for the different countries, depending on individual laws and customs.

As mentioned, the significant progress that IBM has made to date in establishing these practices has proved to be in harmony with privacy legislation. Also, it has involved no major costs, we were told. Much of it has involved the change in design of forms, no longer asking certain questions, and no longer releasing information as in the past. It looks as though the major impacts, from a cost standpoint, still lie ahead. Further, as IBM points out, so far they have dealt mainly with employee information. They still must consider stockholder information, customer information, vendor information, and so on, for the many types of files that carry personal information.

For more information on IBM's approach to the privacy question, see Reference 9.

Discussion of the IBM program

At the conclusion of the presentation of the IBM program at the NBS/Mitre conference, conference attendees raised a number of questions. In general, they dealt with areas which IBM was in

the process of considering and as yet had no opinions or answers.

One challenging problem is that of defining an authorized use of information. A company must be able to perform planning studies, statistical analyses, and so on, without having to get the consent of employees each time. At the same time, the company should protect the privacy of each individual. Just what uses will be "authorized"?

Another problem has to do with the *mixture* of personal and business data—and in particular, company-confidential business data. One example of the latter is planning information on future activities which might include the names of employees who will be involved. Another example is promotability lists, and another is recommendations for reassignments. Information such as this undoubtedly should not be disclosed to the employees; it could lead to false hopes, etc., on the part of the employees and to a higher risk of disclosure of confidential information for the company.

Still another problem area concerns the employee's "right of access" to records containing personal information about him. Does this mean that the employee can discuss his records with a member of the personnel department? Or can the employee read a copy of the records in the personnel office? Or can the employee carry away a copy of the records? In the last two cases, how will misunderstandings by the employee be prevented? Note that personal information is often scattered through many files at numerous locations within large organizations.

Also, if the Privacy Act of 1974 were to apply to the private sector, it would allow the employee to bring a third party and to discuss the records in the presence of this third party. The problems that this provision might cause are still unknown. For instance, it is possible that other persons—such as landlords, union officials, prospective employers, etc.—might pressure the data subjects to allow them to be present in such cases.

Yet another problem area has to do with the possibly burdensome record keeping that privacy legislation in the private sector might require. Specifically, it may be necessary to have an accounting of all "non-routine" accesses to personal data records and retaining such accountings for a period of at least five years. This provision is in the Privacy Act of 1974. However, the concept of "routine" versus "non-routine" access is still ambiguous in the legislation, so it is hard to say what accesses will qualify as "non-routine." We will have more to say on this subject later in these two reports.

What is the privacy problem?

For this discussion of "the debate on information privacy," we have had access to a number of sources of information. These include:

Dr. Willis Ware, of the Rand Corporation, Santa Monica, California, who reviewed the initial draft of this report. Dr. Ware was the chairman of the Special Advisory Committee on Automated Personal Data Systems, reporting to the Secretary of the U.S. Department of Health, Education, and Welfare. The report of this committee is the well-known HEW report, "Records, Computers, and the Rights of Citizens" (Reference 1). This past summer, Dr. Ware was appointed by President Ford as a member of the Privacy Protection Study Commission, under the Privacy Act of 1974; Dr. Ware is Vice Chairman of the commission.

Mr. William A. Fenwick, of Davis, Stafford, Kellman & Fenwick, Palo Alto, California, and New York City, who also reviewed the draft of this report. Mr. Fenwick is a lawyer who has studied existing and proposed privacy legislation in depth. He has written and spoken extensively on the subject (Reference 2).

Mr. Dahl A. Gerberick, chairman of the Ombudsman Committee of the Los Angeles Chapter of the Association for Computing Machinery, who also reviewed the draft. Mr. Gerberick and his committee have studied the privacy question in depth and have written a report of their findings and recommendations (Reference 3).

We are most appreciative of the comments and helpful criticisms of these three individuals.

Also, we attended *The Privacy Mandate,* a conference jointly sponsored by The Mitre Corporation and the U.S. National Bureau of Standards (NBS) and held in April 1975. This conference was attended by many of the leading figures in the U.S. on privacy legislation. Reference 4 is a summary of the results of the conference.

We have drawn on a number of published works, several of which should be singled out since we reference them frequently in these two reports. One is "Privacy Legislation: Analysis of

Alternatives," (Reference 5), a report prepared for clients by McCaffery, Seligman, & von Simson, Inc. Another is *The Cost of Privacy,* a book by Dr. Robert C. Goldstein (Reference 6). Still another is "A Briefing on the Impact of Privacy Legislation," the report of a seminar held in May 1975 and sponsored by the Data Processing Management Association (Reference 7). And another is "Exploring Privacy and Security Costs," the report of a workshop held in February 1975 and sponsored by NBS (Reference 8a).

In addition, we have drawn on numerous other sources of information which we will cite in the discussion.

What is privacy?

Goldstein (Reference 6) gives a good discussion of the historical development of the concept of privacy. The right of privacy apparently was first proposed by Warren and Brandeis in 1890; they defined it as "the right to be left alone."

Professor Alan Westin of Columbia University has identified four types of privacy: solitude (to be left alone), intimacy (to be a part of a small group), anonymity (to be lost in a crowd), and the right to refuse to answer. The first three of these really involve physical privacy.

Former U.S. Attorney General Elliott Richardson (in Reference 9, June 1974) has said that the objective of privacy legislation should not be to assume that the individual is let alone but rather to assume that he can participate in determining how he will be affected by the creation and use of records. The concept here is information privacy.

Turn et al (Reference 10) state a good definition of privacy: "An individual's rights regarding the collection, processing, storage, dissemination, and use of information about his personal attributes and activities. These include his right to know about the existence of a record on him; the right to inspect, challenge, and amend information in the record; and the right to prevent unauthorized use of information for purposes other than that for which it was obtained." This definition might be expanded to include three other aspects: "To prevent the use of the information for purposes to which the individual may object; to prevent unintended harm to the individual from abuse of the information; and to prevent unintended harm to legitimate public and private activities by constraints placed on the use of the information."

Dr. Ware has developed a set of definitions that he feels is representative of the way he hears the terms being used. Two of these are: *Privacy,* the right of an individual to be left alone; to withdraw from the influence of his environment; or to be secluded, not annoyed, or not intruded upon; by extension, the right to be protected against physical or psychological invasion or against misuse or abuse of something legally owned by an individual or normally considered by society to be implicitly his property, e.g. one's home, one's solitude in a public place; the right to maintain something for private use or not available to others. *Information privacy,* (1) the claim on individuals, groups, or institutions to determine for themselves when, how, and to what extent information data about them is communicated to or used by others; (2) the protection of an individual against harm or damage as a result of the operation of an information system; (3) the protection of an individual against unwelcome, unfair, improper, or excessive collection or dissemination of information or data about himself. His definitions cover other related terms such as invasion of privacy, confidentiality, computer security, network security, data security, integrity, and access control.

Goldstein states very clearly what he considers the "most sophisticated" concept of privacy, and which we believe lies at the heart of the debate on information privacy:

Personal information is the property of the individual and he has the right to say what it shall be used for and by whom; this right is limited only by the need to balance it against the right of society to provide for the common security and welfare.

The thrust of this concept is: personal information is the property of the individual; it is *not* the property of the file owners. Goldstein says that people give personal information about themselves for two reasons. One reason is in return for some benefit, such as the issuance of insurance or credit. The other reason is that the law requires it. Gerberick says that personal information is given (and collected) for the mutual benefit of both parties. They each have a say in how much data and what kind of data is to be collected. The data is given for a specified purpose and should not be used for an unspecified purpose.

So we believe that the debate will center on the question: who owns the personal information? In

the past, the concept has been that the "file owner" owns the information, once given by the individual. Now that concept is being challenged.

What is the fear?

Goldstein identifies three types of record keeping systems. One type is the *administrative system,* of which business data processing systems are examples. For these systems, says Goldstein, the data subject himself is often the source of much of the information. A second type is the *intelligence system.* Here the data subject is usually *not* the source of the information; instead, the information may come from a variety of third parties. Finally, there is the *statistical system.* The criterion here is that such a system does not release any information that can be related to an individual; there is always sufficient aggregation of data that the data applying to an individual cannot be segregated.

It is worth noting that much of the privacy legislation refers to "record keeping *systems*" but without defining the word "systems." In reading the legislation, we interpreted the word to mean "application system," such as a payroll system, a personnel records system, and so on. Fenwick, in a comment to us, said that the only meaning understandable to him in the context of data processing is to interpret the word to mean "file." And Panagacos, in Reference 7, says that his interpretation of the word as used in the Privacy Act of 1974 is "any program that has input and output." The definition of the word "system" becomes very important when, for example, privacy legislation imposes constraints upon the transfer of data among different systems.

The fear of the privacy advocates is that the data in any of these "systems" can be misused. For one thing, dossier information may be collected and used to threaten, influence, or harass individuals. Dossier information can be compiled in many ways, such as the following (some of which are already illegal and hence not the subject of privacy legislation):

DOSSIER INFORMATION

1. Accumulating transactions, as a history of an individual's actions.
2. Merging information on an individual from numerous sources and files.
3. Interviewing, obtaining investigative reports, and making security checks, using third party sources of information.
4. Bugging and wiretapping of private conversations.
5. Intercepting private mail of an individual.
6. Spying and using informants.
7. Taking photographic evidence of an individual's activities.
8. Gaining access to private records by impersonation.

The fear, of course, is that punitive measures will occur, that dossier information originally gathered for a perhaps benevolent purpose can lead to punitive purposes.

The advocates of privacy recognize that there are countervailing forces to be considered, for making certain personal information publicly known. One such force is the concept of public records, open for anyone to see. Another is freedom of the press. (Fenwick commented to us that all present indications are that the press will be exempted from the various privacy laws. However, he says there has been little indication that information in the public domain, such as information contained in newspapers and periodicals, when used by others, will in fact be exempt.) Still another is the Freedom of Information Act, which forces the disclosure of certain types of information that government agencies previously held confidential. Parsons, in Reference 7, says that the 1974 Privacy Act was worded so that there is no inherent conflict between it and the Freedom of Privacy Act. But explicit wording was needed to accomplish this; the two Acts tend to act in opposite directions.

So here is another aspect of the debate—the desire to restrict the use and disclosure of some personal information versus the desire to release perhaps the same personal information for the common security and welfare.

What threat is posed by the computer?

There is little evidence that the computer has been the cause of any *new* loss of privacy. Westin (Reference 8b) reports that studies have been conducted in several countries and all developed remarkably similar findings, considering the differences in laws and customs. Some of the significant common findings were:

RESULTS OF PRIVACY STUDIES

1. Computer technology increases the efficiency of record keeping.
2. There is a significant fear (of loss of privacy) on the part of the public.

3. None of the studies can document specific episodes where automated systems created *new* loss of privacy. Abuses that were uncovered had been carried over from manual systems.
4. The computer intensifies the problems that have existed with manual systems.
5. All reports recommend protective measures to protect individual rights.

Goldstein, in referring to studies conducted both in the U.S. and Canada, says "while both studies concluded that there was little evidence of widespread harm coming to individuals from the misuse of information systems, they seemed to feel that this was largely a fortuitous accident resulting from the slower than planned implementation of many large systems." Our belief is that this conclusion goes too far; application systems that use personal information have been implemented by the thousands, and many have been in operation for fifteen years or more. It does not seem credible that the lack of abuse is due largely to "fortuitous accident."

Knowledgeable persons have asserted that the computer is *not* the cause of new threats to privacy. However, the computer is an amplifier of abuses. It has triggered an interest in personal privacy that really should have occurred before the computer era.

What is proposed?

In brief, the privacy legislation that we have seen has two main characteristics: (1) it is omnibus legislation, and (2) it aims at implementing the Code of Fair Information Practices, as spelled out in the HEW report.

Fenwick, in Reference 7, discusses omnibus legislation versus specific legislation. The Fair Credit Reporting Act of 1970, for instance, was specific legislation aimed at correcting problems within the credit industry. But the Privacy Act of 1974, and proposed federal legislation for the private sector (H.R. 1984), attempt to prevent misuse of personal information by regulating *all* uses of such information. For example, H.R. 1984 would have its regulations apply to manual and automated record keeping systems operated by any unit of federal, state, or local government or any public or private entity (which would cover individuals, too).

The Code of Fair Information Practices, as proposed by the HEW committee, is as follows:

CODE OF FAIR INFORMATION PRACTICES

1. There must be no personal information record keeping systems whose very existence is secret.
2. There must be a way for an individual to find out what information about him is on record and how it is used.
3. There must be a way for an individual to correct or amend a record of identifiable information about him.
4. There must be a way for an individual to prevent information about him that was obtained for one purpose from being used or made available for other purposes without his consent.
5. Any organization creating, maintaining, using, or disseminating records of identifiable personal data must guarantee the reliability of the data for their intended use and must take precautions to prevent misuse of the data.

So here is the heart of the debate on information privacy. A new concept of privacy is proposed which says that an individual owns the personal data about himself, limited only by the needs for common security and welfare. At the other extreme is the concept that the file owner owns the data that has been collected. In between is a whole spectrum of ways for dividing the property rights to the information, between the individual and the file owner. This point, we beleive, is central to the debate.

Another aspect of the debate is the constraining phrase "the needs for common security and welfare." What personal information should be held in confidence and what should be publicly available? For example, will it really be good for our elective process, asks Fenwick (Reference 2b), if no personal information about candidates can be collected and divulged without their express permission?

And still another aspect of the debate is that of omnibus legislation versus specific legislation. The Fair Credit Reporting Act was specific to an industry; the Privacy Act of 1974 covers all agencies of the federal government; H.R. 1984 proposes blanket coverage of both the public and private sectors. Which is the best way to go?

Before getting into the discussion of the approaches taken in existing and proposed legislation, it might be well to briefly review the current status of that legislation in the U.S.

The status of privacy legislation

The Privacy Act of 1974 is now a law. It was signed on January 1 and went into effect on September 27 last. Primarily, it applies to agencies of the federal government and to private con-

tractors acting as agents for the government. The Act will also have some application to state and local government agencies, particularly in cases where federal funds are being administered.

House Bill 1984, the Koch-Goldwater bill, is perhaps the foremost piece of proposed privacy legislation being considered by Congress. It proposes to extend the Privacy Act of 1974 in several ways. For one thing, it would apply to "the government of the District of Columbia, or territories or possessions; any State or local government, or any unit of any State or local government or other jurisdiction; any public or private entity engaged in industrial, commercial, or other similar business, as relates to that business." That would seem to cover just about all of us. Fenwick (Reference 7) points out that many of the refinements made in the Privacy Act of 1974, by way of amendments during the final debates, have not been made yet in H.R. 1984.

State bills are under debate in many states. Fenwick sent us a tabulation of bills that was current as of mid-July. At that time, there were 42 bills pending in 22 states. Six bills had been enacted into law, and 23 bills had been defeated. Three of these six (Arkansas, North Carolina, and Utah) regulate state information practices. One (Minnesota) both regulates the state information practices and sets up a study commission to study the public and private sectors. The last two (Rhode Island and Virginia) set up study commissions to study the public and private sectors. In general, much of the proposed state legislation has been copied from, or adapted from, the HEW report.

Local government resolutions, ordinances and administrative policy on the subject of privacy are being considered—and, in a few cases, have been adopted.

The National Association of State Information Systems (NASIS) has developed a "model" bill which they are submitting for consideration to numerous state legislatures.

You might be wondering about possible conflicts among these many bills. Congressman Koch, when asked a question about this, replied that he hoped H.R. 1984 would provide a "floor" for privacy practices; that is, it would set the standard unless a state's act was more restrictive in some areas. We have talked to people working for firms that operate in all 50 states; they are quite concerned about the possibility of having to operate under 51 privacy laws—one for each state plus a federal law. Also, some of the provisions of H.R. 1984 are sufficiently controversial that they invite diverse state laws. To complicate matters even further, some states are putting privacy provisions in other bills not related to privacy.

Fenwick has commented to us on two other points of concern. One is that H.R. 1984 proposes to extend its coverage to foreign operations of U.S. multi-national organizations, so it has international ramifications. The other point is that many of the bills propose to restrict the transfer of personal data to jurisdictions or countries which do not have privacy legislation.

The state bills and H.R. 1984 have not yet been enacted into laws. There is still time for interested people to study them and make their thoughts known to legislators. In a bibliography we are preparing, we will tell how to obtain copies of proposed privacy bills.

In discussing the debate on information privacy, we will be dealing mainly with the possible impact *on the data processing function* from the proposed privacy legislation. Further, we will be discussing the impact in terms of the U.S., although a few other countries are equally far along, or even further along, in dealing with the problem.

We will break the privacy problem into two parts:

- What information about individuals should be collected?
- How should that information be handled?

What information should be collected?

Actually, the question is broader than stated. It should be: what information should be collected about individuals; which of it should be retained and for how long; to which audiences should disclosure be limited; and how can the accuracy, relevance, timeliness, and completeness of the information be assured?

Fenwick has pointed out that most of the legislation discusses collection and maintenance together. So each organization will not only have to review what information it should collect but must also review all of the types of personal information it now has in its files.

In short, personal data must be "defined" from a privacy standpoint. We suspect that the data

definition languages (DDLs) being proposed for data base managment systems will have to be enhanced in the not-distant future to incorporate privacy considerations.

Somewhat unfortunately, the CODASYL DDL specifications speak of "privacy locks and keys" instead of "security locks and keys." This terminology may lead some to mistakenly believe that CODASYL has taken personal privacy rights into consideration. *No* data definition language proposals with which we are familiar have considered personal privacy. Security is a necessary, but not a sufficient, condition for privacy. It deals with preventing unauthorized access to data and with insuring the integrity of data. Privacy considerations go well beyond these functions.

Data definitions for privacy

We think it would be time and money well spent for organizations—assuming they are subject to privacy legislation of the type we are discussing—to develop data definition standards that pertain to privacy. There should be no need to repeat the mistakes of ten to fifteen years ago on basic data definitions. At that time, each programmer developed the data definitions for his or her own programs. Now data definitions are being developed, at many organizations, independent of any given application system so as to be usable by multiple application systems.

There may be a tendency, when faced with the need to implement privacy safeguards, to handle such definitional work in an ad hoc manner. But eventually, the need for systematic data definitions for privacy will be apparent. We think these systematic definitions ought to be considered at the outset.

In saying that, we recognize that the concept is still in its embryonic stage. We have seen no discussion of this topic specifically. However, various authors that we have read have made points that apply. In the brief discussion that follows, many of the points were obtained from References 6, 7, and 8. Here, then, are some aspects of personal data that must be "defined" from the standpoint of privacy.

Purpose of system. Each data item collected and retained will have to be related to the stated proper purpose(s) of the system(s) using it. (The system, in turn, must be germane to a stated proper purpose of the organization.) So purposes probably will have to be defined, unless already specified by statute or governmental order, and then the individual data items related to those purposes.

Personal information. Those data items which constitute personal information will have to be so defined. The most likely definition of personal information will be: information that describes, locates, or indexes *anything* about an individual.

Voluntary or mandatory. Each data item will have to be defined as voluntary or mandatory (as to whether the data subject must supply it). If mandatory, the appropriate regulations will have to be at least referenced. Perhaps more likely, some understandable descriptive material will be needed to explain the regulations.

Actions if not provided. If the individual does not provide the information, for either the voluntary or the mandatory type, the actions to which he or she will be subject must be specified.

Routine uses. The regular, routine uses to which the data is expected to be put must be specified, and in a manner that is understandable to the individual. As we will discuss later in these reports, this "routine use" concept may prove to be very difficult to define.

Permission to use. All items of personal information may well require a (signed?) permission to use for the defined routine uses; specific permission for any new uses will have to be obtained. Third party agencies, such as service bureaus, may well need to have copies of such signed permissions. If personal information is collected orally, or over the telephone, or via computer terminal, one wonders how the "permission to use" problem will be solved.

Timeliness. Personal information that attempts to represent the current situation must be timely. For instance, "number of children" is a quantity that would have to be verified from time to time, particularly for a young married person. The problem is, on what time schedule should such verification be sought? What happens if the individual cannot or will not provide the information?

Probably a "date of last verification" field must be added to each personal data record. Hopefully, it will not be necessary to use a "date of last verification" field at the data *field* level.

Purge rules. Privacy legislation makes a big point of the requirement that personal informa-

tion which is no longer relevant, accurate, timely, or complete be purged. So the rules purging must be specified, perhaps at the record level.

The rules of purging probably will have to cover transient, permanent, and archival data; at least, H.R. 1984 would require that. This requirement probably applies to areas of main memory that held personal data, disk areas and magnetic tapes that held personal data, and so on.

Rules for purging of archival personal data may be complicated. One company mentioned in Reference 8 conducts periodic performance reviews of its employees. It would like to save such information as evidence for years, in case charges of discrimination are brought against it.

Accuracy. Privacy legislation aims, among other things, at protecting individuals from harm due to erroneous or false information in their records. Isolated data errors are just as important to the individuals whose records are in error as are systematic errors to the organization maintaining the records. So rules for enhancing the accuracy of all personal data fields will be required, as will rules for quickly and efficiently correcting errors.

Completeness. One concern of privacy legislation is that a personal data record should give a fair picture of a situation. One frequently cited example of abuse is that of arrest records that do not show the disposition of the cases. Another example is that of billing systems where a dispute exists between the parties as to payments, receipt of merchandise, and so on. So the data definitions probably will have to allow for explanatory remarks that are related to some of the data fields.

Disclosure. H.R. 1984 defines disclosure as the release, transfer, or otherwise communication of information—orally, in writing, by electronic means, or by any other means. Such a definition goes well beyond disclosure by printed output. So the rules of disclosure, for various types of personal data, may well have to cover all of these cases.

Another possible aspect of disclosure is that of sensitivity levels. National security information has three levels—top secret, secret, and confidential. Personal information might also have multiple levels of sensitivity. Name and address information might be the least sensitive, while medical history might be the most sensitive. Fenwick, in Reference 7, points out that such sensitivity is dynamic. Telling one's age to one's doctor is one thing; telling it to a prospective employer can be something else.

The rules of disclosure will have to cover routine, authorized need-to-know cases on one hand versus all other cases on the other hand. As mentioned earlier, "routine disclosure" may prove diffcult to define. And rules will be needed for handling the "all other" cases.

Ownership. While some privacy advocates claim that an individual "owns" all of the personal information about him or her, this concept is not yet embodied in the law. One wonders about personal information that is created as a part of a person's employment—the skills he has developed on the job and is qualified to practice, the training he has received at the organization's expense and the grades he received, the dates on which he was promoted, the reviews of his job performance, and so on. It would seem to be necessary to define, at the very least, the property rights for each data field. These property rights, of course, may be determined by legislation.

Returning to the question of what personal information should be collected, the views of Aryeh Neier, executive director of the American Civil Liberties Union, are of interest (Reference 11). "Don't collect personal information in the first place," he says, "and if you have collected it, destroy it as soon as you no longer need it for the purpose for which it was originally collected." This is an extreme position that probably will be challenged by managers, researchers, lawyers, historians, and others. But it seems to us that privacy legislation, with all of the constraints that it imposes on the handling of personal information, is leading in the direction that Neier advocates.

What personal information do you have?

Privacy legislation for the private sector probably will be written to cover *all* personal information in the possession of an organization—manual as well as computerized records, in transient, permanent, or archival form.

But in fact, will it cover *all* personal information? Some types of records, such as personnel and payroll records, surely will fall under the regulations. But there are other types of records that are more questionable. Goldstein assumes that Christmas card lists and personal telephone lists will not be covered. (As will be discussed, organizations will probably have to publish an-

nually lists of their personal information files and provide access to these files by the individuals whose records are in them, so the definition of "personal information files" is important.) Parsons, in Reference 7, believes that an attorney's files incident to litigation in which the government is engaged will not be classified as a system of records under the Privacy Act of 1974.

But how about data records that have only one field in them that could be classified as personal information, such as employee number? Would that one field make the record a personal information record?

And how about correspondence files? The names and addresses of writer and recipient of each letter would seem to make the file personal information. And anyone else even identified in a letter might then become a data subject, under the legislation, and have the rights and privileges available to data subjects. If an individual asks an organization (as he is entitled to do under much of the proposed privacy legislation), "Do you have records about me in any of your files?" it may be necessary to include the correspondence files in the search.

Probably one of the first steps that an organization will have to take, to comply with new privacy legislation, is to make an inventory of the personal information it already has.

Fenwick, in Reference 7, outlines some of the steps that may be involved in making such an inventory. Review all files, to determine which ones have personal information in them he says. Determine who is using them, why, and who has access to them. Determine how the files relate to the purposes of the organization. Determine which information is necessary to keep, which is necessary to use, which is necessary to disseminate, and which should be purged.

The organizations which have made such studies have found them to be more expensive and time consuming than originally expected, we are told. One organization, cited in Reference 8a, found many records on the same individuals scattered over a number of geographical locations and under the control of different people.

One privacy legislation advocate stated his opinion on this subject: if an organization does not know what personal data it is keeping and on whom it is being kept, then the organization has no real need for that data and the data should be destroyed. This argument is specious, in our opinion. Most people do not have inventories of *all* of their personal possessions but that does not mean they don't need those possessions. Most organizations do not yet have centrally maintained inventories of all of their personal data items, identified by location, use, purpose, etc., but that does not mean all those data items are not needed.

What information should be collected and maintained?

The process of identifying, defining, and making an inventory of all personal information and the uses to which it is put will, in itself, go a long way toward enhancing personal privacy. Organizations will stop collecting some data they currently collect. Some data will be purged from files. Retention schedules for personal data will cause the purging of obsolete data.

Developing data definitions for privacy and developing an inventory of all personal information can be costly. But many of the other steps involved in this process will be relatively easy to take. Relatively easy, that is, if the legislation gives an adequate grace period for instituting new procedures and getting rid of old forms, records, and such.

It is when the question of "how should personal information be handled?" is considered that the problems begin to get complicated.

How should personal information be handled?

The main mechanisms proposed by pending legislation, for implementing the Code of Fair Information Practice, are the following. Most of these are included in the Privacy Act of 1974.

HANDLING OF PERSONAL INFORMATION

1. Published public notices of the existence and the detailed characteristics of all files containing personal information.
2. Records of accesses to personal data records, indicating source of request, purpose of use, and which data records were accessed.
3. Constraints upon the "transfer of data" between files and/or systems, to control the merging/matching of personal information from multiple files.
4. Constraints upon the new uses of personal data, requiring that an individual consent to a new use of information about him- or herself that has not been previously authorized.
5. Procedures for handling disputes about personal data, between the data subject and the file owner.
6. Improvements in data validation and data security.

We will discuss each of these mechanisms, in the remainder of this report and in our report next month, to give some idea of the intent of the mechanisms and some of the complications that we have heard discussed.

Notice of existence of files

Goldstein (Reference 6) discusses some of the early thinking, in developing legislative bills, for letting individuals know that data about them is stored in specified files. The first thoughts were to send annual notices to these "data subjects," to let them know which files they were in. But upon study, this approach was dropped as ineffective and too expensive. Individuals would be inundated with notices of inclusion, to the point where they might well pay no attention to the notices. In fact, the main reaction might be, "Why are you wasting so much money sending me those notices."

Instead, what is being proposed today is either publication of the existence of files or registration of the files at an appropriate state office.

Ms. Naomi Seligman, at the NBS/Mitre conference and in Reference 5, has pointed out that there are large variations in what the different bills propose should be listed. In the HEW report, it is recommended that any governmental agency with personal data files publish the following information about each such file:

HEW: INFORMATION ABOUT FILES

1. The name of the system (or file).
2. The nature and purpose(s) of the system (or file).
3. The categories and number of persons on whom data are (to be) maintained.
4. The categories of data (to be) maintained, indicating which categories are (to be) stored in computer-addressable files.
5. The organization's policies and practices regarding data storage, duration of retention of data and disposal thereof.
6. The categories of data sources.
7. A description of all types of use (to be) made of the data, indicating those involving computer-accessible files, and including all classes of users and the organizational relationships among them.
8. The procedures whereby an individual can be informed if he is the subject of data in the system (file), can gain access to such data, and can contest their accuracy, completeness, pertinence, and the necessity for retaining them.
9. The title, name, and address of the person immediately responsible for the system (file).

Ms. Seligman pointed out that item (7) by itself could be a large undertaking, if interpreted literally.

If provisions of this magnitude are incorporated in the legislation, many people hope that some exemption will be offered for "routine use." But, as we mentioned earlier and will discuss next month, the concept of "routine use" is a difficult one. It sounds very simple, but it actually means different things to different people. Coming up with an acceptable, common definition may be a complex matter. It was pointed out to us that the Privacy Act of 1974 is vague on its definition of routine use and that this vagueness is very likely to cause difficulties as federal agencies try to administer the privacy regulations in the Act.

At the other extreme is the Cullen Bill in California (Assembly Bill 150). Assuming that the bill passes, the only notice of existence required would be once—via registering each system (file) with the Office of the Secretary of State in California by January 31, 1977. The notice would become a permanent public record. Thereafter, any business entity or agency proposing to establish or terminate such a system, or to change the type, use, or categories of users in such a system, would have to file such a notice within 90 days of such action. Further, the Cullen Bill would require that the following information be included in the notice:

CULLEN: INFORMATION ABOUT FILES

1. The name of the system and the organization, plus location and person to contact.
2. The purposes of the system and the uses made of personal information.
3. The categories of data subjects on whom personal information is maintained.
4. The categories of personal information to be maintained.
5. The categories of routine or usual business entities, agencies, or individuals who may receive or use the personal information.
6. Whether the system or any portion thereof is exempted from other provisions of the bill (such as some criminal justice information, some medical history information, etc.).

Ms. Seligman pointed out three different objectives that her company's study observed in these proposed public notices or registrations. One objective was to involve the public in system evaluation, design, and/or use of information, as exemplified by the HEW report. A different objective was to certify compliance with the law. A still different objective was to notify the public

that a data base exists with a specified purpose and population.

The original intent would seem to be that of eliminating the chance of secret files. But the HEW report goes well beyond this. Before a proposed new data file is set up with personal information, the HEW report would like the public to know about it and be able to comment on it. As mentioned earlier, the HEW report was developed with *government* data systems in mind, and this attitude might be appropriate. But is is debatable whether a company must advertise that it is planning to set up a retirement system and would be setting up a new data file—particularly in view of all of the other checks and balances that the government imposes on retirement systems. In fact, there might well be legitimate reasons why organizations would not want to *publicly* announce the creation of new files involving, say, only their management personnel.

Fenwick has pointed out, in a comment to us, that there are some types of files which, if subjected to such privacy legislation, will simply not be kept. An example might be a corporate plan of succession to be used in the event of death or some other disaster affecting corporate executives. If privacy legislation makes such a plan available for access by anyone named in it, this earlier-than-expected release of the information might have a disastrous impact on the organization.

Long, in Reference 8c, argues against the whole idea of public notice of personal data files. Such publication will do little good because it will flood the newspapers with 15 to 20 million data file announcements, overwhelming most readers. Such public announcement tends to penalize the file owners for what they *can* do and not for what they have actually done; this is a presumption of guilt not justified by the evidence. Also, he says, public notice amounts to registration which in turn can lead to file licensing, and that in turn can lead to file taxation. Public *registration*, as proposed in the Cullen Bill, would avoid overwhelming the public with announcements about files, of course.

Fenwick, in Reference 7, points out a further complication due to the way H.R. 1984 proposes that such annual notices be published. Organizations in the private sector would be required to publish such notices "in local or regional printed media *likely to bring attention to the existence of the records to data subjects.*" (Italics ours.) This means, says Fenwick, that the organizations must determine the location of all data subjects and then determine what publications are available which are likely to bring attention to the data subjects of the existence of the records. Since some personal information files have records about people at widely scattered geographic locations—including locations in other countries, as in the case of subscriber lists—this will be no small task.

Next Month

Next month, we will continue our discussion of the proposed privacy legislation. The discussion will cover the other mechanisms that are planned for the handling of personal information—records of usage, constraints on the transfer of data, constraints on new uses of personal information, procedures for handling disputes, and so on.

In addition, we will give a discussion of the possible costs of privacy legislation, both conversion costs and operating costs, as they might impact an organization.

From the above discussion, plus the fact that privacy legislation is being enacted at federal, state, and local levels, it should be clear that this is a subject of prime importance to data processing management. Privacy legislation is coming; it is not going to go away. It is to be hoped that the legislation is thoroughly debated before it is made into law.

At the heart of such a debate are some very important principles. Who owns personal information? What kinds and how much personal information should be publicly available? What kinds and how much should be completely out of the control of the individual? Should the legislation be specific or omnibus?

Then there are the mechanisms that have been proposed for protecting an individual's information privacy. Each should be debated and refined, so it can do an effective job of protecting privacy without at the same time being too oppressive or too costly for legitimate business or governmental purposes.

We urge you to study the subject and make your views known to legislators.

REFERENCES

1. "Records, Computers, and the Rights of Citizens," report of the Secretary's Advisory Committee on Automated Personal Data Systems, Department of Health, Education, and Welfare, July 1973. Order from Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402; Cat. No. 1700-00116. It is also available in paperback form through M.I.T. Press, Cambridge, Mass. 02139.
2. Fenwick, Wm. A. (Davis, Stafford, Kellman & Fenwick, Two Palo Alto Square, Palo Alto, California 94304), has spoken and written extensively on privacy legislation. Two of his writings are:
   a) "Privacy," *Data Management* (DPMA, 505 Busse Highway, Park Ridge, Illinois 60068), May 1975, P. 18-21; price $1.50.
   b) "Privacy Legislation," address to American Library Association, June 1975, scheduled for publication in one of the Association's publications (50 East Huron Street, Chicago, Illinois 60611).
3. "Privacy, Security, and the Information Processing Industry," report by Ombudsman Committee on Privacy and Security, Los Angeles Chapter of ACM, Dahl Gerberick, Chairman, Fall 1975; to be published by Association for Computing Machinery (1133 Avenue of the Americas, New York, N.Y. 10036) in late 1975.
4. "The Privacy Mandate," summary report of a symposium/workshop held in April 1975, co-sponsored by U. S. National Bureau of Standards and The Mitre Corporation. Order from: The Mitre Corporation, Westgate Research Park, 1820 Dolly Madison Blvd., McLean, Virginia 22101, Attn: Mr. Gene Raichelson; price $4.
5. "Analysis of Alternatives," a cross analysis report of state and federal privacy bills, prepared by McCaffery, Seligman & von Simson, Inc., for their clients. Summaries of this report are available to state and federal legislators. Write to the company at 251 East 61st Street, New York, N.Y. 10021. A very brief summary is given in Reference 4.
6. Goldstein, Robert C., *The Cost of Privacy*, published by Honeywell Information Systems, Inc. (Distribution Services, HIS, 40 Guest Street, Brighton, Mass. 02135), 1975, price $10.
7. "A Briefing on the Impact of Privacy Legislation," proceedings of a briefing session held in May 1975, published by Data Processing Management Association (DPMA, address above), 1975, price $15.95.
8. Publications of U.S. National Bureau of Standards; order from Superintendent of Documents (address above):
   a) Berg, J. L. (Ed), "Exploring Privacy and Security Costs—A Summary of a Workshop," 1975, Cat. C13.46:876; price 85¢.
   b) Renninger, C. R. and D. K. Branstad (Eds), "Government looks at Privacy and Security in Computer Systems," 1973; Cat. C13.46:809; price 85¢.
   c) Renninger, C. R. (Ed), "Approaches to Privacy and Security in Computer Systems," 1974; Cat. C13.10:404; price $1.20.
9. Two issues of IBM's employee magazine *Think* (March 1974 and June 1974) present some of the company's views on privacy. For copies, write *Think*, IBM Corporation, Armonk, New York 10504.
10. Two papers prepared for publication by members of the Rand Corporation (1700 Main Street, Santa Monica, Calif. 90406):
    a) Turn, R. and W. H. Ware, "Privacy and Security in Computer Systems," P-5361, January 1975.
    b) Turn, R., N. Shapiro, and M. Juncosa, "Privacy and Security in Centralized vs Decentralized Databank Systems," February 1975, P-5346.
11. Flato, L., "Privacy: the ACLU takes a stand," *Computer Decisions* (50 Essex Street, Rochelle Park, New Jersey 07662), April 1975, p. 32-34; price $3.
12. We are preparing a more extensive bibliography on information privacy which will include, among other things, methods for obtaining copies of state and federal privacy bills. It will be available after next month's report has been published.

SUBJECTS COVERED BY EDP ANALYZER IN PRIOR YEARS

1972 (Volume 10)

Number

1. Computer Security: Backup and Recovery Methods
2. Here Comes Remote Batch
3. The Debate on Data Base Management
4. Intelligent Terminals
5. COBOL Aid Packages
6. On-Line Development of COBOL Programs
7. Modular COBOL Programming
8. New Training in System Analysis/Design
9. Savings from Performance Monitoring
10. That Maintenance "Iceberg"
11. The "Data Administrator" Function
12. The Mini-Computer's Quiet Revolution

1973 (Volume 11)

Number

1. The Emerging Computer Networks
2. Distributed Intelligence in Data Communications
3. Developments in Data Transmission
4. Computer Progress in Japan
5. A Structure for EDP Projects
6. The Cautious Path to a Data Base
7. Long Term Data Retention
8. In Your Future: Distributed Systems?
9. Computer Fraud and Embezzlement
10. The Psychology of Mixed Installations
11. The Effects of Charge-Back Policies
12. Protecting Valuable Data—Part 1

1974 (Volume 12)

Number

1. Protecting Valuable Data—Part 2
2. The Current Status of Data Management
3. Problem Areas in Data Management
4. Issues in Programming Management
5. The Search for Software Reliability
6. The Advent of Structured Programming
7. Charging for Computer Services
8. Structures for Future Systems
9. The Upgrading of Computer Operators
10. What's Happening with CODASYL-type DBMS?
11. The Data Dictionary/Directory Function
12. Improve the System Building Process

1975 (Volume 13)

Number

1. Progress Toward International Data Networks
2. Soon: Public Packet Switched Networks
3. The Internal Auditor and the Computer
4. Improvements in Man/Machine Interfacing
5. "Are We Doing the Right Things?"
6. "Are We Doing Things Right?"
7. "Do We Have the Right Resources?"
8. The Benefits of Standard Practices
9. Progress Toward Easier Programming
10. The New Interactive Search Systems
11. The Debate on Information Privacy: Part 1

(List of subjects prior to 1972 sent upon request)

PRICE SCHEDULE

The annual subscription price for EDP ANALYZER is $48. The two year price is $88 and the three year price is $120; postpaid surface delivery to the U.S., Canada, and Mexico. (Optional air mail delivery to Canada and Mexico available at extra cost.)

Subscriptions to other countries are: One year $60, two years, $112, and three years $156. These prices include AIR MAIL postage. All prices in U.S. dollars.

Attractive binders for holding 12 issues of EDP ANALYZER are available at $4.75. Californians please add 29¢ sales tax.

Because of the continuing demand for back issues, all previous reports are available. Price: $6 each (for U.S., Canada, and Mexico), and $7 elsewhere; includes air mail postage.

Reduced rates are in effect for multiple subscriptions and for multiple copies of back issues. Please write for rates.

Subscription agency orders limited to single copy, one-, two-, and three-year subscriptions only.

Send your order and check to:
EDP ANALYZER
Subscription Office
925 Anza Avenue
Vista, California 92083
Phone: (714) 724-3233

Send editorial correspondence to:
EDP ANALYZER
Editorial Office
925 Anza Avenue
Vista, California 92083
Phone: (714) 724-5900

Name_____

Company_____

Address_____

City, State, ZIP Code_____