EDP ANALYZER

DECEMBER, 1975
VOL. 13, NO. 12

THE DEBATE ON INFORMATION PRIVACY: PART 2

This is the second of two reports on existing and pending "information privacy" legislation in the U.S. The legislation includes the Privacy Act of 1974, which took effect on September 27, 1975, and which applies to agencies of the federal government. Also, at least six states have enacted specific privacy legislation. H.R. 1984, under consideration by Congress, would extend privacy regulations to state and local government units and to all other types of organizations. In addition, many state and local governments are drafting privacy regulations. So privacy regulations are coming—and they will affect many data processing functions. In these two reports, we are considering key issues and controversies in privacy legislation, particularly those that affect the data processing function. Last month, we discussed the changes that will be needed in collecting and maintaining personal information on individuals. In this report, we describe how the handling of personal information will be changed.

What effect will the new privacy legislation have on the operations of organizations covered by that legislation? In one sense, it is too early to say. The Privacy Act of 1974 became effective only at the end of September, so that federal agencies are really still in a period of conversion to comply with the Act.

But in another sense, there is some experience that is relevant to the question. This is the experience of organizations subject to earlier privacy legislation. It is true that the Privacy Act of 1974 and the proposed H.R. 1984 are much more sweeping in their coverage and impact than the previous legislation. But still, the experiences under prior legislation are of interest.

We will discuss briefly the experiences of the U.S. Social Security Administration and of Credit Bureaus, Inc., of Atlanta, Georgia. This material was presented at the "Privacy Mandate" Conference, held in April 1975 and co-sponsored by the U.S. National Bureau of Standards and the Mitre Corporation. A summary of the conference results will be found in Reference 4.

Social Security Administration

The U.S. Social Security Act was passed in August 1935. It required the federal government to collect and maintain personal information about individuals to a degree that U.S. citizens were unaccustomed to at that time. So, from the very outset of the Act, there were strong fears that the information would be used in ways that would affect civil liberties.

One measure of the potential for harm is the number of people on whom social security records are maintained. Initially, less than 20 mil-

lion people were covered. Now the files contain over 230 million records on individuals (150 million of whom are still living) and some 20 million employers. Social security now covers more than just payments to retired persons. It also covers payments to spouses of retired persons, children, disabled persons, Medicare benefits, and supplemental security income benefits. So the files contain information about a very large percentage of all people living in the U.S.

Even in the early days of this new Act, though, fears were expressed about how the information might be used. Since the collection of the personal information was vital to the administration of the Act, the Social Security Board issued a public pledge in November, 1936 which later (in June, 1937) became the famous Regulation 1 of the Social Security Administration.

Regulation 1 said, in effect, that the personal information collected under the Act would be regarded as confidential. It would be disclosed only to those who have a legitimate interest in the administration of the Social Security Act, and to the individual (or business) to whom it pertained. Further, anyone connected with the administration of the Act would respectfully decline to furnish any information forbidden to be disclosed by Regulation 1, in response to a subpoena or otherwise.

It would appear from the early fears and from Regulation 1 that the social security records have a wealth of personal information about individuals. Such is not the case. Only three main types of information are collected and maintained. One is identification information in sufficient detail to conclusively establish the identity of a person. Recognizing that between 2000 and 3000 persons can have exactly the same name, it is clear that other types of identifying information are needed. Another type if the annual contributions of the employee. Finally, the most recent report of earnings includes the employer identification and hence the recent whereabouts of the person. (Other types of information would be required, of course, for persons receiving benefits.)

Even with this rather limited amount of personal information, the Social Security Administration gets many requests and pressures to release information that fall outside of Regulation 1. One type of request, of course, is for the purpose of locating persons: missing persons, persons suspected of crimes, persons wanted in connection with lawsuits, absconding debtors, etc. In some cases, the purposes are humanitarian: persons who are missing heirs, or close relatives of very sick persons. (In humanitarian cases, the administrators do not disclose the locations but will forward letters via the last known employer.) Another type of request is from business: what are the wage patterns of competitors, or how many employees does a particular firm have? Some requests are from political organizations, for political purposes. And during World War II, another agency of the federal government wanted the Social Security Administration to give it a list of all persons with specified foreign backgrounds. There was even one rather recent case where a function within the Social Security Administration itself wanted information from the records of SSA employees, but where the purpose had nothing to do with the administration of the Act. The people in charge of the records are proud of the fact that, to a very great extent, they have been able to deny such requests and pressures. They have tried to conscientiously enforce Regulation 1.

However, Congress has enacted other legislation that has eroded some of the privacy aspects of Regulation 1. Social Security record information may now be released in cases where national security and the safety of the President are involved, as determined by the FBI and the Secret Service. It may also be released in cases involving federal income tax laws, as determined by the IRS. It may also be released in certain cases involving aliens illegally residing in the U.S., for the purpose of imposing penalties on employers who hire such aliens. And it may be released in certain cases involving deserting parents having children on public assistance. Also, the Pension Reform Act of 1973 requires that pension fund administrators notify members periodically about their pension rights. These administrators feel that they need social security information in order to answer such questions completely, and they are requesting such information. It has been pointed out that there are some dangers in releasing social security information for this purpose. For instance, some unions might impose sanctions on members whose records show employment for nonunion employers.

We understand that the Secretary of the De-

partment of Health, Education, and Welfare, under which the Social Security Administration operates, has instituted a further policy on disclosure of the information. Should any federal agency have a need to know social security information for a new purpose, the executive of the requesting agency must sign a form saying that he recognizes that he is performing a possible violation of the law. This is perhaps the ultimate in assuring a legitimate "need to know." We suspect that it will result in a curtailment of transfers of social security information to other agencies. (Note that confidentiality does not apply to statistical or other information not relating to any particular person.)

Thus the Social Security Administration has operated under strict restraints on the disclosure of personal information collected for social security purposes.

Credit Bureaus, Inc.

Credit Bureaus, Inc., with headquarters in Atlanta, Georgia, is the second largest credit bureau in the U.S. It has computer-based records on some 24 million people. It has a total of about 1200 terminals, both in customer (usually retailer) establishments and in-house, and about 30,000 miles of data communications lines.

CBI is a credit bureau only. It collects data from credit grantors and then sells that data back to the credit grantors. CBI is a subsidiary of Retail Credit Inc., an investigative reporting agency, but CBI itself does not release information for investigatory purposes.

CBI falls under the regulations of the Fair Credit Reporting Act of 1970 (FCRA). This Act addresses the subject of "information privacy" specifically in the credit industry. Under the Act, if a consumer is denied credit due to a report received from a credit bureau, the consumer has a right to learn the contents of his or her credit record. Moreover, if the consumer disputes some of the information in the record and the credit bureau is unwilling to change the record as the customer requests, the consumer may include a statement of up to 100 words in the record.

What has been CBI's experience under the FCRA of 1970? Back in 1968, two years before the Act, CBI set up a new data retention schedule, with a maximum of five years, to reduce storage costs and increase relevancy. The FCRA set a maximum of seven years, so CBI met this requirement. In 1969, CBI adopted a set of ground rules for broadened privacy protection. These included the data subject's right to know the contents of his record and the purging of outdated information. CBI requires that all employees sign an affidavit telling the penalties if they improperly disclose personal data; the penalties include liabilities for damages and immediate dismissal.

So CBI had taken a number of steps even before the Act became law. After the Act was law, CBI instituted the necessary training program for employees, the development of instructional material and manuals, and the setting up of logs of interviews. During 1974, for instance, over 60,000 man-hours were spent in interviews with consumers, plus additional time for rechecking information. The major cost item resulting from the Act was the training of staff members. Operating costs have increased, due to the interviews with consumers and from carrying some 100-word statements. But these costs have been much less than CBI originally expected; less than 1% of the consumers have asked to see their records and only a fraction of these have had disputes that were not easily settled.

Goldstein (Reference 6) has pointed out two shortcomings of the Act, as far as privacy is concerned. For one thing, the Act only requires that consumers get oral summaries of their credit records; they cannot get printed copies or even examine printed copies. So the consumer may wonder if he or she is hearing *all* of his or her record. Another shortcoming is that the consumer may not even know about the existence of a credit record unless credit (or employment or insurance) is denied *because of a credit report*. The credit grantor may deny credit for a claimed other reason.

CBI goes further than the Act requires in that it shows the consumer a copy of his or her credit record. However, the consumer may not carry away a copy because CBI wants credit grantors to obtain the latest credit information, and not use some previously prepared (and perhaps tampered with) report.

So far, CBI has had no real problem operating under the FCRA. However, proposed privacy legislation (such as H.R. 1984) goes quite a bit further in its regulation of the collection, maintenance and use of personal information than did the

FCRA, and thus may impose more of a burden on credit bureaus.

Mechanisms for handling personal information

In our report last month, we discussed requirements that are (or are likely to be) imposed on the collection and maintenance of personal information. We pointed out that a set of data definitions will be required, for defining privacy aspects of personal data. And we discussed the need for making an inventory of the types of personal information that an organization now has in its files, where that information is located, who has access to it, and what it is used for.

In addition, we pointed out the main mechanisms proposed in pending legislation for the handling of personal information. Most of these mechanisms are included in the Privacy Act of 1974. The mechanisms are:

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 discussed the mechanism of notices of existence last month. We pick up our discussion with rights, records, and reports of access.

Rights, records, and reports of access

The privacy legislation that we have seen includes the provision for the data subject's *right of access* to his or her personal information records upon request. Goldstein (Reference 6) says that this idea has replaced the idea of mailing individual notices to data subjects telling them that their records are in the files.

It should be noted that there are three categories of files of interest here. Active files, which are used to produce output documents such as invoices or reports, can be sub-divided into direct and secondary types. With the direct files, the outputs go to the data subjects themselves—payroll files, billing files, vendor files, customer files, etc. Generally, data subjects are aware of records about them in such files.

With secondary files, the outputs do *not* go to the data subjects. Examples are credit bureau files and medical information bureau files. Data subjects may not even know such files exist, much less know that records about them are in such files. Privacy legislation is attempting to make the existence of such files known to the data subjects and to provide the right of access to the data subjects' own records.

Passive files are ones for which there is usually no direct output. One example is a correspondence file. If the correspondence contains information on personal attributes and activities, the file will probably qualify under the regulations.

This apparently simple right of access may prove to be one of the most complex privacy provisions. For instance, how does the organization handle the *mixture* of personal and business data, particularly company confidential data? And how does the organization handle a request of the type, "Which of your files am I in?" This question apparently would require one of three solutions: (1) searching all files where there was any chance of inclusion, (2) setting up one massive personal information data base for serving all applications, or (3) maintaining an active, indexed inventory of all personal information records. Solutions (2) and (3) generally do not exist in organizations today; further, they lead to the type of dossier information that privacy advocates do not want.

Another complication is particularly pertinent for correspondence files. Suppose that a letter in the file gives personal information about a third person. Does that third person then have the right of access to see the letter? Reference 9 reports an interpretation of the Privacy Act of 1974 which is illustrated by an example prepared by the Office of Management and Budget. "If Joan Doe's record is customarily accessed by her name, Joan Doe has a right to access it. However, if Joan Doe's record appears in a contract source evaluation of her employer, Corporation XYZ, and the record is not accessed by her name or identifying particulars, then Joan Doe has no right to see or copy the record."

Records of access

Almost equally complex is the subject of *record of access*. The wording of some of the proposed legislation implies that *all* accesses to personal data records must be recorded and accounted for (including the source of the request and purpose of use), and such accountings saved for at least five years. If this interpretation is valid, it means that every time a file on magnetic tape is processed, an accounting record must be created for each record read—in short, for the whole file. Even for files on direct access devices, some legislation would appear to require accounting for all maintenance and updating accesses. In Reference 8a, the point is raised that such accounting may even have to include accesses to any contested data deemed so bad that the organization decides to purge it.

(As a point of interest, Goldstein, in Reference 6, proposes the use of 20-character records of access which he feels will meet the requirements. And Fenwick, in Reference 2 and in a comment to us, says that such records of access themselves may be classified as personal information records and hence subject to the regulations.)

Gerberick, an author of Reference 3, commented to us on a point that frequently arises in connection with records of access. Much of the legislation, he says, requires such records of access only for *non-routine* accesses. Hence any problems should be small in magnitude.

"Routine use"

This point of routine versus non-routine accesses is important. First of all, not all legislation allows the exemption of routine accesses, as far as a record of access is concerned. But even more important, perhaps, is the fact that even the legislation which does allow this exemption does not rigorously define "routine." The concept sounds so simple that a person might believe (as apparently the drafters of the legislation do) that no rigorous definition is needed. Unfortunately, the term means different things to different people, when applied in the context discussed here. *To date, the concept of "routine use" is ambiguous in the legislation.* Both Dr. Willis Ware (Vice Chairman of the Federal Privacy Protection Study Commission) and Mr. William Fenwick (a lawyer who has made an extensive study of privacy legislation) have pointed out to us that problems can occur from this lack of a clear definition.

For instance, Fenwick says that both civil and criminal risk are involved here, due to the ambiguity. He thinks that a huge amount of money may be spent over the next thirty years or so in having the courts work out just what "routine use" really means.

We suspect that unless a rigorous definition of routine use is included in privacy legislation, organizations will be hailed into court by people whose interpretations of the term differ from those of the organizations. And after a few lawsuits, organizations will decide to take the safe course and record *all* accesses.

Further, if *all* accesses must be recorded, and if the legislation covers both manual and computer files (as in both the Privacy Act of 1974 and H.R. 1984), this means every time your secretary goes to the filing cabinet, a record of that access must be made.

Report of access

We are only about one-half way through the complexities associated with the rights of access—and we are only touching on major points, at that. The next area of complexity is *report of access*. At the very least, as far as some legislation is currently drafted, whenever a data subject requests to see his or her records, he or she must also be shown a listing of all accesses made since the last previous request. Some other versions of legislation, if taken literally, would require the organization to notify the data subject every time his or her record was accessed—remembering that the idea of exemptions for routine accesses may be a mirage.

As we will discuss later in this report, this concept of "report of access" may be one of the most expensive aspects of privacy legislation.

Disclosure rules

And then there are *disclosure rules*. If one reads the definitions of disclosure in the legislation, one concludes that disclosure is equivalent to access; every access results in a "disclosure."

Goldstein (Reference 6) notes that before disclosure can be made, the organization must check whether the inquirer is authorized to access the file as well as to access the individual record. The concepts here include both need to know and authority to know.

Again, there is no clear definition of these concepts given in the legislation. In Reference 7, it is reported that under the Privacy Act of 1974, the disclosure rules be left to the individual federal agencies. The agencies' decisions can be challenged in the courts, but there is no federal review agency that reviews the rules used.

In Reference 8a, quite a bit of discussion centered on the problem of authenticating the identity of a inquirer. How should an organization authenticate the identity for inquiries received through the mails? Will all inquiries have to be notarized? What other personal identification information will have to be supplied, in order to make the identification precise? In very large files, such as those of Social Security, or the IRS, or the Civil Service Commission, there are many duplicate names, so precise identification is mandatory. Moreover, this identification information itself will probably fall under the privacy regulations.

Organizations will have to be careful to authenticate the inquirers for at least two reasons. One reason is the possibility of civil and criminal penalties for improper disclosure. Another reason is that a person may submit derogatory information on a second person by pretending to be that second person. So care is needed.

Not only is the *who* of disclosure important, so is the *what*. Again in Reference 8a, the case of an insurance company was brought up. If the legislation says that only a policyholder himself (or herself) can see the information about the policy, what policy information (if any) can the company send through the mail to the person's home? Does it also mean that the company cannot disclose any personal information to its agent who is dealing with the customer, for a review of coverage and for normal business promotion efforts?

Privacy legislation says that when a data subject asks to see his or her records, that the records be presented in a form comprehensible to the data subject. At the very least, any coding used in the computer version of the record must be translated. But, asks Fenwick, will there be a need to explain the meaning of the terms, as used by the organization? If so, this can mean a substantial staff training burden imposed on the organization.

The Privacy Act of 1974 specifies that a data subject who wants to see his own record may be accompanied by a third party of his choice. In Reference 8a, it was pointed out that the data subject may be pressured by an organization (say, his union) to have someone from the organization be that third party. In such a case, a privacy *loss* may well result.

Dr. Dorothy Denning, on a project conducted at Purdue University, has developed a mechanized method for certifying computer programs, as far as the access to and the dissemination of personal data are concerned. Classes of access and dissemination rights are defined at the file level, record level, or if desired, at the data field level. The certification program analyzes the program being certified, for both explicit and implicit data transfers between input and output, and checks the legitimacy of these transfers. Concerning implicit transfers, one or more tests of the value of a field in record X can be used to implicitly transfer information about that value to record Y; the certification programs checks for such implicit transfers. There are other types of implicit transfers that cannot yet be detected. For more information about this project, see Reference 11.

As we say, the whole idea of rights of access—while it sounds simple at first—has a wealth of complexities. We would hope that the debate on information privacy will sharpen and refine the concepts so that they will protect privacy without being unduly oppressive to legitimate information handling practices.

Constraints on the transfer of data

The concept here is to limit the merging/matching of personal data records from two or more files, as well as to prevent new, unauthorized uses of the data. Control is exercised by restricting the right to "transfer" data between systems, as a substitute for the linking of the files.

An example of the transfer of data would be the exchange of data among federal agencies, such as the FBI telling the IRS who to examine for potential income tax evasion. This would be a new use of data originally collected for another use.

The way that the provision seems to be emerging, the constraints will apply to the transfer between *systems* of records. As mentioned last month, it is not yet clear just what is meant by "system." Does one organization at one geographic location have just one system? If it has a data base serving all applications, is it one system?

Or will the regulation be interpreted to mean that you will be constrained from transferring data between your payroll system and your personnel system?

Once again, here is an area of ambiguity that probably will cause trouble. There may be a number of lawsuits to determine just what is meant by the term "system" unless the legislation clearly defines it.

Consider another legitimate business use of personal data—doing operational planning, perhaps using a simulation model. For instance, if the function were job shop scheduling, records representing specific machine operators and their skills would be needed. The logical place to get this information would be the personnel file, perhaps leaving off the employee name. It is not yet clear whether this transfer of data could take place without the consent of the employees.

Note that removing the name of the employees does not make the new file any less of a personal data file. If the records give any personal attributes at all and can be accessed by any means of identification (employee number, name, voice print, finger print, or whatever), it is a personal data file and hence subject to the regulations.

Researchers very much want to retain "statistical" files on human subjects, over a period of years. In order to update the information, records must be set up on individuals and accessed by means of personal identifiers. As such, a file of this type would probably qualify as a personal information file. Privacy legislation, such as the Privacy Act of 1974, allows the use of statistical records as long as the records are "in a system of records maintained for statistical research or reporting purposes only and not used in whole or in part in making any determination about an identifiable individual . . ."

At the NBS/Mitre Conference (References 4 and 5), Ms. Naomi Seligman made an interesting point in connection with the restrictions on the transfer of data. As some of the proposed legislation is currently written, it would seriously hamper the use of third party service bureaus. It is not always feasible to get authorization from all data subjects that would cover the use of service bureaus. The legislation should specifically recognize this legitimate type of activity.

Gerberick stressed to us that the intent of the constraint on the transfer of data is to avoid new uses of data, not to restrict legitimate business operations. We believe that the wording of the legislation must be carefully refined in order to achieve this end.

A subject related to the transfer of data is that of universal identification numbers. (We have seen these referred to as UIDS, UINS, and SUIS; we'll refer to them as UIDS.) Many privacy advocates have deplored the existence of a UID since it would make the merging/matching of records easier and thus facilitate building dossiers. They argue that people should resist giving their social security numbers for other than social security purposes, so as to prevent the social security number from becoming a UID.

The main thrust of the "social security number" argument sometimes is lost sight of. The Council of the Association for Computing Machinery recently adopted a resolution expressing its concern that the social security number might become a UID. But Dr. Peter Denning (Reference 10, Letters to the Editor), as a member of the Council, feels that the Council acted too hastily. It is not the matter of a UID that is of concern, he says, but rather it is the merging/matching of data via a UID that should be controlled. In fact, if merging/matching were allowed to continue in the absence of a UID, the chances of erroneous matching are greater than if a good UID were in use. This, in turn, would lead to incorrect data in some records, which is one of the trouble spots that privacy advocates are trying to rectify.

As a matter of fact, the social security number should *not* be used as a UID, we understand—not for any ideological reasons but because it is a very poor UID. Literally millions of people have multiple or duplicate numbers or erroneous numbers or such. The social security number does not have a self-check digit attached, to help guard against transpositions, etc. When people are eligible to receive social security benefits, the Social Security Administration people work out such problems as best they can. But for everyday use as a UID, the social security number is inappropriate.

Fenwick, in a comment to us, pointed out that there are regulations already in force on some types of data transfers. Unless abuses or potential abuses can be identified, he says, regulating the transfer of data generally is not a good idea, since the transfer of information is one of the most important aspects of freedom of speech.

Constraints on new uses of data

One of the main concerns of advocates for privacy legislation is that an individual be made aware of, and consent to, the uses to which his personal data will be put at the time he first provides that data. Further, they say there should be no new uses made of that data without the data subject's explicit consent.

We pointed out above that the intent of the constraints on the transfer of data is to inhibit new uses of data without the consent of the individuals involved.

We suspect that few people would argue against this principle. Trying to implement it, however, raises a number of challenging problems.

In general, the proposed legislation requires explicit actions on the part of a file owner if personal data within that file is to be used for a new purpose. The file owner must get the (written) authorization of the data subjects involved.

Unless carefully worded in the legislation, this constraint can impose serious problems on legitimate business operations—and, in fact, may sometimes act to the detriment of the data subjects it is trying to protect. It will be very difficult to define at any one point in time all of the truly legitimate uses for personal data. Some of these uses will be of a planning nature, such as how some contemplated future activities will be staffed. Some will be of a statistical nature, such as how many employees of the company fit a given set of criteria. Some will be ad hoc queries, such as who in the engineering department is qualified to be sent on a new assignment. What we are saying is that legitimate "new" uses, of which these might be examples, are almost sure to arise.

What is the problem, you might ask? Why not get an employee's authorization to use personal data for "planning, statistical, and normal business query purposes?" The trouble is, it is not yet clear just how precisely the uses will have to be defined—planning for *what*, statistics for *what*, queries for *what* reasons? The less precise the definitions, the less understanding the data subject has of how his data will be used. The more precise the definitions, the more likely it is that the file owner will continue to find legitimate needs not covered by the authorizations. If the file owner continues to seek new authorizations, it is easy to visualize a number of data subjects refusing to give the desired authorizations. In the extreme case, each record might end up with a table showing which uses it is authorized for and which ones not.

What about a blanket permission covering all "routine, normal business" uses of personal data? Fenwick, in a comment to us, says that this probably will not suffice. Most courts will require that the consent be an informed consent, he says. The solution may well be to draft a general consent, then specify all of the known uses, and finishing with a provision that a listing of the specific is not intended to limit the general. And in Reference 8a, it was noted that the definition of routine uses (in a given organization) might be written in a sufficiently broad manner to cover a large number of possible uses. The definitions may even be rewritten and republished, if the need should occur.

Goldstein (Reference 6) mentions several possible problems in connection with this constraint. It may be necessary to get permission to use personal data from the data subject, even if the data was not collected from the data subject originally. (For instance, the data may have come from investigative agencies, schools, etc.) Further, he says, it may be difficult to obtain permission from the data subject (1) to use data obtained from third parties, or (2) to make new uses of data collected from the data subject himself, unless such uses are clearly beneficial to the data subject. Goldstein mentions two possible solutions to the problem (and recognizes the shortcomings of both): assume that consent is given if the data subject does not reply to a request for permission, or publish a public notice of the intended uses and assume that consent is given if no communication is received from the data subjects. (Elsewhere in the book, Goldstein points out that the safe approach probably will be to treat "no reply" as "no permission.") Still another problem area that he mentions is the mailing expense associated with mailing out requests for permission for very large files. Two-way postage costs, clerical and handling costs, follow-up costs when no replies are received—all of these may make new uses of personal information very expensive.

Gerberick, in a comment on the draft of this report, takes us to task for being too concerned about the rights of the organization in this regard, as opposed to the rights of the individuals in controlling the uses of their personal information.

This issue lies at the heart of the debate on information privacy.

Most of what we have discussed has been in terms of employee personal data. But it should be evident that similar questions arise for personal data on customers, suppliers, advisors, competitors, and so on. Getting permission to use such personal data from people outside of the organization will be even more difficult than for employees.

The Privacy Act of 1974 says that agencies of the federal government may not sell or rent lists of names and addresses of individuals, for direct mail or other purposes, unless specifically authorized by law.

As we said earlier, few people would argue with the principle behind the constraints on new uses of data. But the legislation must be carefully drafted so that this principle can be implemented without severe disruptions in our business life and with due concern for the individual. As Dr. Willis Ware says, it is *balance* that the privacy advocates are struggling to attain.

Procedures for handling disputes about data

Most proposed privacy legislation for the private sector, as well as the Privacy Act of 1974, allow a data subject to correct or amend his personal data record. In many instances, the needed changes will be due to errors of input and the file owner will be willing to make the changes.

Fenwick, in a comment, points out that the requirements for "complete, relevant, and timely" data may cause most of the disputes. These terms are meaningless except in context, he says, and he feels they have caused the most problems with credit bureaus in complying with the Fair Credit Reporting Act.

Whatever the cause, there are, and will continue to be, cases where a data subject is deprived of some right or benefit due to information in a record about him or her. Further, the data subject may feel that the data is incorrect or a half-truth or misleading, and will want to either correct it or amplify it. The file owner may or may not agree with the data subject's position. If the file owner does not agree, then a dispute arises.

The types of files where such disputes are most likely to arise are those involved in the control of rights and benefits. These rights and benefits pertain to applications for credit, insurance, licenses, security clearances, disability benefits, and such.

The existing and proposed legislation calls for the right of the data subject to *require* the file owner to indicate that the data is in dispute. Further, the data subject has the right to append a statement (generally of not over 100 words) to the record, stating his or her position on the disputed data. The file owner, in turn, may append a concise statement of the reasons for not making the requested amendments. The Cullen Bill, in California, proposes another solution in the case of a dispute. Within a reasonable period after the receipt of a written statement of disagreement from the data subject, the business entity may bring an action for declaratory judgment as to the dispute, in a court. Or the business entity may use the statements, just mentioned.

A number of points should be considered. For one thing, should disputed data be released at all? Some proposed legislation says that it should be released only in cases of demonstrated necessity, and then only with the data subject's statement attached. Other proposed legislation would allow the release of the data for any legitimate requests, but with the appended statement included. Still other proposed legislation, we understand, would allow the release of the disputed data if properly marked as the subject of a dispute—but would require the release of the statements only upon request of the data subject or the recipient of the data.

Some legislation also requires that such notices of dispute and the dispute statements be sent to anyone who has received the personal information for some period in the past.

If a substantial portion of the records in a file required dispute statements, the additional costs could be significant. Experience to date, however, indicates that the percentage of such records in a file is low, less than 1%. Of course, harassment by data subjects could occur, and possibly some protection against this eventuality is needed.

One proposal in some privacy legislation is that the *sources* of personal information be shown in the records. In Reference 8a, it is mentioned that the U.S. Civil Service Commission plans to protect the sources of confidential information, but individuals may see all of the data in their files. The privacy point under debate here is whether the individual has the right to know who supplied

(possibly derogatory) information. Some people feel that evaluators may be less candid in their evaluations—but others feel that this provision will make evaluators more careful in their statements.

Improvements in data validation, security, and audit

As we discussed last month, organizations may have to take even more steps to assure the accuracy of personal information in their record keeping systems than they now take, when privacy legislation is in force. We see several types of steps that may be taken.

One is to define the accuracy requirements for each data field, as part of the overall privacy data definitions, and then design the data collection procedures so as to meet those requirements. For instance, one possible procedure is to have someone verbally verify with the data subject the entries that the data subject has made on a form, whenever personal information is collected. This may be time-consuming and annoying in some instances, but it may prove to be necessary to safeguard against law suits.

Another step is to use more extensive logic in data validation programs, in an attempt to detect errors at the input stage.

Still another step would be to verify personal data annually. For information in direct files, outputs of which go to the data subjects, the organization might ask for verification as a part of normal contacts. For information in secondary files, outputs of which normally do not go to the data subjects, the organization would probably just pass the responsibility along to the owners of the direct files dealing with those data subjects. This procedure might operate on a negative basis, such as "Here is the information we have on you; let us know if there is anything that is *not* correct," but at some risk of inaccuracy.

Security requirements

The security requirements imposed by privacy legislation, such as the Privacy Act of 1974, are discussed in References 6, 7, and 8. The Act says that "the organization shall establish appropriate administrative, technical, and physical safeguards to insure the security and confidentiality of records." What does this mean? Must the operating systems be certified? Must the data communications network be secure? If so, who can perform such certifications? Can sensitive and non-sensitive data be mixed in data transmissions? Must encryption be used? If so, how can the security of the encryption procedure be assured? Must the identity of terminal users be authenticated? Must access controls be at the record level, or even at the data field level? What are the "reasonable precautions" that an organization must take for physical security?

Parsons, in Reference 7, gives her interpretation of security requirements. By and large, she says, the conditions of the disclosure section of the Act ratify existing practice. It is not necessary to certify operating systems under the Act, she says; no one is stealing data now, so deliberate penetrations are not a worry. The things to worry about are accidental and inadvertent disclosures. Our reaction is that what she says may well be the intent of the legislation—but it is not what the legislation says. Today's on-line files *can* be penetrated, as we have mentioned many times in the past. If someone's personal data is stolen and some sort of harm comes from that disclosure, it might very well end up in a law suit. *That* is when the interpretation of the security provisions would really start to occur.

Audit considerations

Some privacy legislation specifies that transfers of personal information may not take place from an organization that meets the privacy requirements to one that does not.

The question is: how does one organization know whether the other organization is meeting the privacy requirements? One way is for the first organization to perform an inspection of the other. Another approach is to use independent third party auditors who will inspect and certify that organizations are meeting privacy requirements. Since both civil and criminal penalties can be involved, for the managers who make the decisions to transfer data, it is likely that they will not just accept someone's word that another organization meets the privacy requirements.

So this area, too, needs debate and refinement so that the mechanisms will safeguard the rights of the individuals and at the same time not be oppressive to legitimate organization activities.

The costs of implementing privacy legislation

When considering the existing and proposed privacy legislation, a question that immediately comes to mind is: what will be the cost of implementing the provisions? The only answer that we know of at this time is: it is too early to tell. There has been experience in implementing the Fair Credit Reporting Act, but that Act is much more limited in the regulations it imposes than is most of the privacy legislation.

Within the next year, the costs of implementing the Privacy Act of 1974 will become clearer. Federal agencies are required to report their annual costs of privacy, for instance.

But there have been some studies of possible costs of privacy. The conference reported in Reference 8a concerned itself largely with expected costs and expected benefits of privacy regulations. The study reported in Reference 5 gave consideration to the costs of the several privacy mechanisms. The most thorough analysis of possible costs, though, is the one performed by Goldstein (Reference 6). The subject is a large one and we can give only a brief summary here.

First, it should be recognized that there are three aspects of costs to be considered. There are "legitimate" costs of privacy and there are costs that will probably be improperly charged to privacy. Within the legitimate costs, there are conversion costs and operating costs.

What about those "improper" costs? Whenever mandatory change is imposed, some people use the occasion to do things that they have been wanting to do anyway and charge them up to the mandatory change. As far as privacy regulations are concerned, these things might include installing a satisfactory physical security system, or ceasing to collect and maintain more personal information than is really required, or installing a data management system (to aid in answering privacy inquiries), or purging and destroying obsolete data. Some organizations may use privacy as a catch-all for costs of changing application systems, or changing programs, or bringing in new equipment. The General Accounting Office, for instance, will be checking to see that federal agencies do not improperly charge costs to privacy.

As far as the legitimate costs of privacy are concerned, Goldstein's analysis goes the furthest in showing what those costs might be like. (Note that a perhaps oversimplified summary of Goldstein's work appeared in the March-April 1975 issue of the *Harvard Business Review*. We recommend that interested persons read Reference 6 instead.)

Goldstein's analysis of costs

Goldstein developed a cost model of privacy regulations based upon the HEW report (Reference 1). The model was completed and run in mid-1974, prior to the passage of the Privacy Act of 1974. So Goldstein had to assume what the eventual privacy legislation might contain, as applied to both the public and the private sectors. (More recently, Goldstein has been a consultant to the federal government for the development of a cost model based on the Privacy Act of 1974; it has been completed and has been offered for use to agencies of the government.)

Goldstein obtained operating figures from six organizations, for the purpose of estimating the costs of the assumed privacy regulations for those organizations. One was a large network of hospitals, another was a large on-line law enforcement system containing information on people who had been arrested, another was a smaller on-line law enforcement system of outstanding warrants on people and vehicles, another was a large consumer credit system, a fifth was a personnel system in an organization with 10,000 employees, and the sixth was a large casualty insurance system.

Goldstein issued some warnings about the cost figures developed by his model for these six organizations. One warning, of course, was that the model was based on the HEW report, not on actual legislation. But he also pointed out that the study showed costs to be very system-dependent. Some costs were relatively fixed for all systems, such as physical security. Some varied with transaction volume, such as searching the records-of-access files. Some varied with the number of records in the files, such as getting permission for new uses of data. Some varied by the number of inquiries received, such as data subjects wanting to see their records. And some varied with the number of users, such as training costs.

So, said Goldstein, be careful in generalizing from the cost figures presented. For this reason, we will not mention the actual cost figures that he

developed but instead will concentrate on his conclusions about relative costs.

High conversion costs. Goldstein considered costs related to 17 privacy mechanisms. If costs spread evenly, each would have 1/17 of "total" costs, or about 6%. He segregated out those which had at least twice this amount. He identified three high conversion costs.

One of these was the cost of new forms which should include the notification of "rights" that a data subject has when providing information for the forms. Note that the Privacy Act of 1974 was imposed nine months after signing, with not much of a grace period for using up old forms.

Another high conversion cost was installing a physical security system, for those organizations that did not already have a satisfactory system.

The third high cost item was employee training. It will not be sufficient, said Goldstein, just to devise and publish a set of rules. Employees must be trained in the use of the new procedures—and the training must be repeated as new employees are hired.

One conversion cost that we did not see analyzed was that of getting permission from the data subjects to use the personal information already in the files, at the time the legislation takes effect. It is one thing to get personal information when a person is being employed; if he does not provide it, he might not get the job. It is another thing altogether to ask him years later for permission to continue to use the data for a set of specified uses.

High operating costs. The only large computer operating cost that Goldstein uncovered was that related to searching the files of "records of access." He assumed that about 1% of the data subjects would want to see their records, and about one-half of those would want to see a list of people or organizations to whom the records had been disclosed. The searching of the disclosure activity files might cost tens of dollars per inquiry, he found. (In Reference 8a, it was mentioned that making one search in a huge sequential file might cost $1,000; we must wait for experience to develop, to see what actual costs are.)

Another high operating cost was concerned with the executive time needed for handling data subjects' complaints about the accuracy of the data in their records. He pointed out that it would probably require executives, not clerical people, to handle these complaints.

Finally, operating the physical security system and the monitoring of privacy safeguards can be high extra cost items in organizations not already performing them.

Other cost aspects. Goldstein found (under the assumptions he used) that for systems with over one million records, privacy costs were relatively constant and relatively low on a per-record basis. Below one million records, costs were more variable and higher on a per-record basis. If this finding turns out to be true in practice, it may encourage more centralization of personal data records.

In general computer costs associated with privacy were not large, except for searching records-of-access files. Showing data subjects their records was not a large cost. Each data subject might be charged a nominal charge, which probably would cover the cost of copying the records but not the cost of searching for the records. Even carrying 100-word dispute statements was not costly, since less than 1% of the records were assumed to have such statements (based on experience in the credit bureau field). However, sending copies of dispute statements *retroactively*—to people who received a copy of a disputed record before the dispute statement was entered—could be costly.

He felt that the annual notices of file existence would not be costly. If the file owner need only submit an annual notice to a government agency, his conclusion certainly should be true. Even if the file owner had to publish a notice in one publication, the result still holds. But if the requirements of H.R. 1984 apply, the file owner may have to publish notices in many publications and in multiple countries—and this could be costly.

Then there are legal costs to be considered, for defending against lawsuits related to the privacy regulations.

And finally, there are "shadow costs"—costs of not doing something that was feasible before the privacy legislation.

Can it be assumed that the costs of privacy regulations for the private sector will be reasonable? At this point, we would have to say No, it cannot be so assumed. It is quite possible that the costs can and will be reasonable—*but it depends upon the wording of the legislation.* In these two reports, we have tried to point out some of the areas where the concepts and the wording of the proposed legislation is ambiguous and where the

potential for trouble exists.

The way to avoid unreasonable costs, of course, is to refine the legislation, by debate and by input from the public and private sectors, before the legislation is enacted.

There are a number of other aspects of privacy legislation that we could discuss. But we have covered what seem to us to be the high points. So we will conclude with a brief summary of the experience of the State of Minnesota.

The Minnesota experience

The State of Minnesota's experience with privacy legislation was presented at the NBS/Mitre Conference which is reported very briefly in Reference 4. Minnesota's privacy law, passed in 1974, was the first omnibus bill aimed at state and local government agencies. It was based largely on the HEW report and on Sweden's privacy law. It covered *all* personal data collected, stored, and disseminated by state and local government agencies, not just computer-based data.

Minnesota has some 3600 governmental jurisdictions. Notices were sent out to them, asking who the contact person would be. The better part of a year later, only 60% of the jurisdictions had replied to this simple request. Only two of the 3600 units had reported what files they had.

To get a better idea of the problems involved, a detailed study was made of the types of personal data handled by four county governments and seven city governments. The study found large differences of opinion in the handling of personal data. For instance, government employee salary data was considered private by one of the jurisdictions—and was published annually in newspapers in another jurisdiction. The study found a major need for standardized definitions of public, private, and confidential information.

This past July, the Minnesota legislature passed several amendments to their privacy law, for making the law more workable. A privacy study commission was created, to study the workings of the law and to recommend courses of action. Standard definitions of public, private, and confidential information were established. Public information is open to the public at large. Private information is available to the data subject to whom it pertains and to authorized users but not to the public at large. Confidential information is available only to authorized users, not to the data subjects nor to the general public. Another amendment eliminated the need for state and local governmental units to report the existence of personal information records classified as public information.

Note that Minnesota has not yet really reached the point where the new law has had a major impact. Most of the questionable points discussed in these two reports have not been confronted by the governmental jurisdictions involved. Minnesota is in the process of getting started, and is finding that it must still debate and refine its privacy legislation.

This debating and refining is what we hope happens with *all* privacy legislation. The more it is debated and refined prior to enactment, the more likely it is to be effective, workable, and not oppressive.

Your legislators would like input from you that is pertinent to the debate on information privacy.

REFERENCES

(Note: References 1 to 8 are the same as given last month. We will give only abbreviated citations here. For more detail, see last month's report or the free bibliography, Reference 12.)

1. "Records, Computers, and the Rights of Citizens," the well-known HEW report.
2. Fenwick, W. A.:
   a) "Privacy," *Data Management,* May 1975.
   b) "Privacy Legislation," American Library Association
3. Report of the Ombudsman Committee on Privacy and Security, Dahl Gerberick, Chairman, Los Angeles Chapter of ACM, 1975.
4. "The Privacy Mandate," summary report of NBS/Mitre conference held in April 1975.
5. "Analysis of Alternatives," report prepared by McCaffery, Seligman & von Simson, Inc., 1975.
6. Goldstein, R. C., *The Cost of Privacy,* published by Honeywell Information Systems, Inc., 1975.
7. "A Briefing on the Impact of Privacy Legislation," published by DPMA, 1975.
8. Publications on privacy by U. S. National Bureau of Standards:
   a) Berg, J. L. (Ed), "Exploring Privacy and Security Costs—A Summary of a Workshop"
9. Bigelow, R. P., "The rights of individuals to inspect and correct their files," *Computer Law and Tax Report* (210 South Street, Boston, Massachusetts 02111), August 1975, p. 4–7; price $36 per year.
10. Denning, Peter, "Objections to ACM's Resolution on UIDs," *Communications of the ACM* (1133 Avenue of the Americas, New York, New York 10036), May 1975, p. 303–4; price $5.
11. For more information on the certification of programs for access and dissemination of personal data, write: Dr. Dorothy Denning, Computer Science Department, Purdue University, Lafayette, Indiana 47907.
12. For a free copy of a selective bibliography on the subject of privacy legislation, including methods for obtaining copies of federal and state bills, write EDP ANALYZER.

Next month we begin a discussion of some of the problems associated with the data processing function in multi-national organizations, and solutions that are evolving for those problems. It might seem that managing multi-national data centers is not much different from managing multiple centers within one country. But the situation is *different. Next month, we will concentrate on what the main problems seem to be, based on our discussions with a number of data processing executives in multi-national organizations. And in February, we will concentrate on solutions for one key problem area—staff training in a multi-national environment.*

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
12. The Debate on Information Privacy: Part 2

(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_____