4.1 INFORMATION IN THE SYSTEM

a. CONTENTS AND SCOPE

ISSUE: DO DESCRIPTIONS OF THE TAS PROPOSAL SUFFICIENTLY IDENTIFY THE INFORMATION TO BE STORED IN THE SYSTEM AND THE SCOPE OF TAXPAYER FILES TO PERMIT CONSIDERATION OF THE POSSIBLE EFFECTS OF THE TAS ON PRIVACY AND OTHER RIGHTS?

SUMMARY

Until specific scope of files and contents of TAS are spelled out, operations cannot be effectively monitored or dealt with by the Executive Branch, Congress, judges, or parties in data-connected tax law disputes. Without the basic ground rules which include some meaningful specifics about the data contents in advance of installation of the system, in the future it may be impossible to determine the extent of the adherence of TAS and its users to public expectations of government performance, and to the demands of new laws such as the privacy Act of 1974, the Freedom of Information laws, the tax return confidentiality provisions of the New Tax Reform law, and new information requirements of tax programs. Similarly, it will be difficult to evaluate the adequacy of administrative and technical barriers designed to protect the confidentiality and integrity of different types of data.

Such information would help provide the basis for considering the possible relationship between TAS and the impact on the individual of future IRS and governmental information collection policies.

QUESTIONS

1. Does the TAS proposal contain any criteria for contents of TAS which may need elaboration?
2. Are available descriptions of TAS files sufficient to afford the full knowledge about IRS records to which law and regulations entitle the individual?

3. Is the type of description in the TAS proposal and the current reporting in the Federal Register of "categories" of personal information in IRS records systems sufficient to inform Congress and the public about contents of TAS so that decisions can be made about the system's potential impact on privacy and on information policies generally?
4. Are descriptions of TAS and reports filed under the Privacy Act on IRS personal information systems sufficient to cover the rights of corporations, businesses, firms and organizations in the new TAS?
5. Is the available information about TAS contents sufficient to permit judgments about the need, if any, for new rules for partitioning files and for requiring administrative and technical safeguards for categories of data?
6. Is there a need for legislation or regulations establishing the contents of TAS?
7. Is there a need for some kind of specific reporting mechanism on TAS contents to allow Congress to monitor the system? How can an effective audit be made if present rules and statutes governing IRS information are unchanged?
8. What specific privacy considerations governed the resizing of the TAS from the original design?
9. Has a review been made recently of standards for the collection and maintenance of IRS records? If so, by what standards? Was it made in connection with requirements of the TAS design or was it in connection with the administration of the Privacy Act and the Freedom of Information Act, or was it in connection with the concerns of the Federal Paperwork Commission for cutting back and simplifying Federal forms?
10. Do the 1976 and 1977 tax reform laws affect collection and maintenance of IRS information in such a way as to alter any plans for the size of the data base of TAS?
11. Is all of the information proposed for the TAS data base required for purposes of administering the Internal Revenue laws and collecting taxes? What other criteria will govern characteristics of the data base?
12. Does the TAS proposal permit modification of the data base should the Freedom of Information Act or Privacy Act be amended, modified, or changed in interpretation by courts? By what means? With what effect on rights of taxpayers?

13. Is there sufficient information to permit consideration of the extent, if any, to which TAS may have an impact on future governmental programs for collecting information from and about citizens?

BACKGROUND

According to a 1972 Academy of Sciences report, *Data Banks in A Free Society,* " 'Privacy' is independent of technological safeguards; it involves the social policies of what information should be assembled in one information system. "

In the Privacy Act of 1974, Congress stated its finding that "the privacy of an individual is directly affected by the collection, maintenance, use, and dissemination of personal information by Federal agencies. "[9]

Despite the importance of privacy in such a sensitive information system, there appears to be an element of secrecy about important aspects of the TAS which affect privacy. Nowhere in the testimony and materials given Congress did the IRS spell out the contents of the files to be consolidated in the new system; nor did it indicate how much of the specific information supplied by taxpayers on tax returns will be in an account in the new sysem.

Unless the rules for such large computerized systems have spelled out sufficiently what will be permitted in the system and how the taxpayer can exercise information rights with respect to specific kinds of information, there may be a lack of public confidence in the entire system. Furthermore, without such specific information, it may also be difficult to assure the enforcement of those statutes which are designed to promote the obsemance of information practices which respect privacy and due process rights.

Knowledge, consent to gather, use and shine information, accountability, oversight, Confrontation of records by means of access and right to challenge records, and specific prohibitions against collecting and maintaining certain information are the key elements in considering the contents of TAS. Congress has used a number of principles and techniques in setting controls and limits for sensitive information systems. Among these me the privacy Act of 1974 which states that "an agency that maintains a system of records shall maintain in its records only such information about an individual as is relevant and necessary to accomplish a purpose of the agency

9. Pubic Law 93-579, 93rd Cong. (5 USC 552a).

required to be accomplished by statute or by executive order of the president. " In addition, an agency must publish an annual notice, less general in nature, of the existence and character of the system of records, including the "categories" of individuals on whom records are maintained and the "categories of records" maintained in the system.

Furthermore, the agency must "maintain all records which are used by an agency in making any determination about any individual with such accuracy, relevance, timeliness, and completeness as is reasonably necessary to assure fairness to the individual in the determination. "

Other provisions of the Privacy Act allow certain information rights which have been partially applied to IRS files. Additionally, the Freedom of Information Act creates rights to obtain identified information.

The public has also demonstrated specific concern for protections against abuses in the governmental collection and maintenance in records systems of information which bears on the exercise of First Amendment rights, that is, on how people speak, write, think, organize and associate for religious, political and civic purposes. As a result, the Congress enacted in the Privacy Act a provision that an agency shall "maintain no record describing how any individual exercises rights guaranteed by the First Amendment unless expressly authorized by statute or by the individual about whom the record is maintained or unless pertinent to and within the scope of an authorized law enforcement activity. "

Given the broad IRS mandates for information-gathering under the tax laws, it would be unrealistic to enforce literally this specific provision. Yet past IRS abuses have been identified which involved use of IRS personnel and tax data for non-tax purposes because of First Amendment activities of taxpayers which offended administration politicians.[10]

In view of previous public concerns, the Congress and taxpayers ought to have a way of assuring themselves that the ability of the IRS to observe the spirit of this provision of the Privacy Act will not be adversely affected by the installation of the new system.

Most of the information supplied by IRS addressed the purposes of the new system and the new ventures which IRS proposed to launch with it, its general characteristics and attributes, and what the internal IRS user needs were.

---

10. An IRS Directive of Sept. 29, 1975, "Exercise of First Amendment Rights," provides guidelines for agency compliance.

In deference to the Privacy Act requirements, the TAS proposal states that the "Service has reviewed and revised its practices and procedures relating to the collection and maintenance of records to assure that only such information that is relevant and necessary is maintained. " Policymakers may want to know when and how this review was made, by whom, and under what terms and standards. For instance, were the internal users of TAS allowed to define their own informational needs, and if so, by what standards? was this review made in response to the Privacy Act, to the Freedom of Information Act, or to the concerns Of the Federal paperwork Commission for cutting back on governmental information~ demands and simplifying forms? Was it made in response to the specific needs and environment of the computerized, decentralized TAS?

The Service states that it has complied With the privacy Act notice requirements by publishing in the Federal Register the indices and notices of existing person~ information systems. These are general and brief, indicating how people may discover whether a system contains information on them, how they may learn what information rights, if any, they have within a system, and what files are exempt, what routine disclosures are made, and the nature of sources.[11]

A question was raised by panelists whether or not the description in public documents and the very general Federal Register reports for records and files on individuals can be substituted for the more detailed inventory needed to evaluate the possible impact of any changes under TAS. Furthermore, these Federal Register reports me designed for the personal information systems of government, and may not be extensive enough in this case to account for the data which may be in the system on corporations, businesses, organizations, and other tax entities.

The issue of contents of the TAS is also joined to the issue of the adequacy of whatever safeguards for technical and physical security me proposed to protect the data under various laws and executive branch standards for records management and computerization, including those established by the National Bureau of standards. While this issue is more appropriately raised in a section on "security," it bears citing here to show the importance of the issue of establishing precise rules for management of TAS information.

The Privacy Act, for instance, requires agencies to "establish appropriate administrative, technical and physical safeguards to insure the security and confidentiality of records and to protect against any anticipated threats or hazards to their security or integrity which could result in

11. U.S. Dept. of the Treasury, Internal Revenue Service. Privacy ***Act of 1974: Resource Material.*** Document 6372 (11-75).

substantial harm, embarrassment, inconvenience, or unfairness to any individual on whom information is maintained. " Unless criteria for the permissible contents are known and established beforehand, it is hard to tell how the appropriateness or adequacy of these safeguards can be evaluated for different kinds of information.

It is helpful to review the degree of specificity already accorded the TAS contents in the public documents sent to Congress. The description in the report filed with several Committees pursuant to the Privacy Act is vague and brief on this point. It does not draw a clear line between what is now gathered and stored and what will be included in the new computerized system. The personal information now received by IRS is characterized in the following ways. It notes the sources of information received, stating that it receives most of its data from "tax returns and related documents required by the IR Code and regulations or forms authorized by them;" that relevant data is also obtained from records "required to be kept," and taxpayers or other sources, as necessary to ascertain the correctness of returns received, or to secure or prepare delinquent returns. The primary source of data is described as the individual income tax return, Forms 1040 and 104OA submitted by the taxpayer and containing personal and financial information. Tax data is also received from third parties via income information documents such as Forms W-2 and 1099 reporting wages, interest, dividends and other taxable income; and from related returns such as partners and beneficiary's income on Forms 1065 and 1041.

The report states that information is received from other government agencies, such as the Agriculture Department reports on taxable farm subsidy payments; that it is also obtained from public and other records; from the taxpayer's own records, financial and other statements; from correspondence and information furnished by the public.

It states that the information the Service receives is "prescribed by the IR Code or supporting regulations, " and briefly cites 28 different provisions of Title 26 of the United States Code referring to required returns and statements.

In a sub-section entitled "Types of Data Retained," the report states 'The Service retains several broad types of data — identification, accounting, status, assignment, cross-reference to related accounts, history, statistical, and data and system control and security. " After all of the various statutory retention policies are cited, the contents of TAS are summarized as follows: "The redesigned system will retain five years' data for all tax accounts in a readily accessible file, and additional years for only unpaid and otherwise active accounts. "

In another section entitled "Information Retained on TAS Files," the report seeks to illustrate "some of the reasons the Service retains certain data," and indicates the broad types of data retained, and how long IRS deems it necessary to retain, or believes it is authorized to retain, various types of data because of various statutory references to liabilities, rights and duties.

It states that: "In accordance with standard accounting practices, the service maintains records and controls on all tax transactions which affect the revenues; that an account is maintained for each taxpayer to which the related tax liabilities, paymentsj credits and other financial transactions are posted, and from which the necessary bilk and refunds and other accounting activities are generated. In addition, the summary (general ledger) data is maintained and used to produce accounting reports such as the reports on gross collections and refunds paid to taxpayers. " It wants "to verify the correctness of information received and to quickly retrieve the taxpayer's figures and the Service's computation to satisfy and resolve taxpayer's inquiries about bills, refunds and other account settlement matters, and to promptly make corrections and tax adjustments to data in the files. "

Further, it states, "some of the transcribed and retained data is used to select returns having the highest potential for tax change and which may require examination. Retention of data from the tax return reduces the costs to locate, pull, control and refile original documents . . . . Also, by retaining data concerning tax filers, it is possible to identify non-filers through comparisons of present and past data as well as other leads or sources. In addition, the data is used to produce operating and statistical reports for management purposes or as required by law such as the publication of statistics of income. "

As a result of a number of concerns which were raised about lack of specificity in TAS documents on criteria for the contents of the new system, OTA sent a questionnaire to the Service asking them to indicate as precisely as possible what items of information under the new expanded TAS will be placed in the taxpayer's file and thereby linked to the taxpayer's name. The response and some comments by individual panelists are included as Part Of the appendix to this report.

Change in Contents

From the documents supplied on the TAS proposal and discussions about it, it would appear that expense is the major barrier to adding to the contents of the TAS and that even expense may be little deterrent to expansion under certain conditions. TAS officials stated that the estimated

cost for each additional character of data transcribed from all the individual income tax returns and entered into the system is $60,000, and that "for this reason, among others, they strive to capture the minimum amount of data consistent with effective tax administration. " Furthermore, they stated "data requirements created by new legislation are added of necessity. " This statement may deserve elaboration in any review of TAS. For instance, one panelist considering TAS observed that few, if any, agency officials or ADP personnel have proven courageous enough in the past to come forward and protest to Congress in the face of a legislative push for new laws which might overload their ADP or telecommunications systems or which might create a data base difficult to manage from a due process or privacy standpoint. In such event, it is not clear what, if any initiatives IRS or Treasury officials could or would take to alert policy makers or Executive Branch managers of potential problems of privacy, due process, confidentiality, or overload of the system.

Further inquiries may be appropriate on this issue. It may be, given the attractiveness of the technical capacities of TAS for manipulating data, for programming and for retrieval, that some special attention should be directed to the need for installing an "early-warning" system so that the attention of IRS and of appropriate committees of Congress may be alerted when there is an effort which would result in altering the size of the TAS data base and, the scope of individual taxpayer files.

b. RETENTION OF TAX INFORMATION

ISSUE: DO THE TAS DOCUMENTS DESCRIBE RETENTION TIME POLICIES SUFFICIENTLY TO PERMIT A DETERMINATION OF THEIR CONSISTENCY WITH SOUND SOCIAL POLICY, FAIRNESS TO TAXAYERS, AND WITH STATUTORY REQUIREMENTS?

SUMMARY

Another major management benefit planned under TAS is availability of a longer tax history through increased storage capacity. In light of what is known or perceived about the threats from other large computerized person~ information systems containing financial data, and in light of recent public concerns about the IRS and other government information practices, it is important to consider to what extent the longer retention time afforded by TAS might contribute to a public view of it as unfairly inhibiting people from starting anew in society. There is a need to assure that, as programmed and operated, TAS will not stigmatize taxpayers long after thek difficulties with IRS have been resolved in a satisfactory fashion.

In order to evaluate the potential policy impact of the system, it may be important to define for the public record the retention time policies governing the data to be stored in the system together with whatever administative and technical standards and devices might be planned for enforcing those policies.

QUESTIONS

1. h light of what is known or feared about large computerized financial data systems, might there be an undesirable impact on civil liberties and due Process ~terests as a result of the change from the *present 3-years* storage capacity (and from none at all in some cases) to computerizing 5 years of tax history of a tax account, with potential for storing much more, and making it available to users of IRS tax data?

2. To what extent might the TAS lead to difficulties similar to those which have been widely discussed in the administration of justice field over outdated computerized "rap sheets" or in the commercial field over computerized consumer credit files?
3. In light of what is known or feared about large financial data systems, could TAS inadvertently become an instrument for promoting an unsanctioned social policy of stigmatizing taxpayers long after their difficulties with IRS have been resolved in a satisfactory fashion?
4. To what extent can taxpayers be informed under TAS about the full scope of the information potentially available to them? How can they challenge the accuracy or completeness of such information?
5. How might retention and storage policies affect the information rights of the individual taxpayer under IRS rules, under the Privacy Act, the Freedom of Information Act, and other Internal Revenue statutes?
6. What provisions have been made administratively and technically for systematically identifying and purging outdated information in the TAS? For updating it?
7. What provisions should be made by statute? By regulation?
8. What reporting or other accounting method should be installed to assure the taxpayer and Congress that any purging and updating program for TAS is enforced?
9. Should there be further legislation or regulations specifically establishing retention policies for different categories of information?
10. Without specific findings on current and proposed retention policies of user components in IRS and in programs of other users and producers of TAS data, without analysis of the effects on these users of changes in IRS retention policies, is it possible that the technological momentum of the new system may initiate or influence changes in public policy without the input of assigned policy makers?
11. Could IRS retention policies for TAS data affect the vulnerability of taxpayers to unauthorized surveillance and to harassment by IRS or other governmental users of the system?
12. How will the new IRS retention policy affect the information retention policies of other users of income tax return information? Of users of information derived from IRS data? Of taxpayers, employers and corporations who must supply information on Taxpayers under various programs to IRS and to other users of tax data?

13. How do the retention policies comport with the work of the Federal Paperwork Commission to cut back on the amount of information collected and retained by federal agencies, and to simplify income tax returns?
14. How does TAS retention policy comport with the policies of the Treasury Department for retention of records of that Department?
15. Does the Privacy Act of 1974 authorize some retention of data as described by the IRS?
16. Has there been a review of the policies governing retention time for the data to be stored in the system and is it available?
17. What provision has been, or will be made, to assure that data originally collected for one purpose (and the taxpayer so informed) will not be retained for another purpose in another location, longer than permitted by the policy for that kind of data?

BACKGROUND

There may be a need to consider to what extent the programmatic or operational aspects of TAS may prevent taxpayers from ever redeeming themselves from the adverse effects of previous infractions, misunderstandings of tax rules, investigations, audits, debts, petty transgressions and records of old tax events.

Public apprehension about large computerized personal data systems, and this reflects current attitudes toward government decision-making generally, is that they may facilitate the storage and the use of personal information which is irrelevant or outdated for making decisions on the merits of a case. For this reason, a special interest in privacy and due process in recent years has been to prevent certain kinds of sensitive information from ever being collected or stored in a system in the first place. Another important aspect of the privacy issue has been the setting of precise reasonable rules for the length of time information is kept and for assuring that it is eliminated from a data system at the end of that time, unless new judgments are made as to further use.

The 1972 report by the National Academy of Science explained the civil liberties issues as follows :

> [6]'Not only should the need for and relevance of specific items of personal data have to be established in positive terms but serious consideration should be given to wlzether some entire record-keeping programs deserve to be continued at all. . . . A further consideration where the need for collecting data is at issue is whether records should be retained beyond their period of likely use for the purposes for which they were originally collected. A related but more complicated question concerns the continued existence in files of information which is no longer supposed to be used for making decisions about

> individuals. Many cumulative records about individuals in various sectors of the organizational world are filled with facts and evaluations set down in earlier time, under a different sociopolitical ethos. In this setting, it is not enough to say 'from now on we will not . . . . .' Steps need to be taken to remove from historical records in high schools, colleges, commercial reporting agencies, law-enforcement files and other organizations the personal information previously gathered about political, racial, cultural, and sexual matters that would not be put in the files under present rules. To the extent that evaluators today have such record to consult, especially for decisions that are not visible to the individual, the presence of such information represents a dead (and improper) hand from the past."12

Thus, a major concern about computer systems in both private industry and government is that the existence of the technological capacity to store data and to have large portions of it available in real time, will lead to searches for additional kinds of data. The costs of computer systems can more easily be justified when they are used to maximum or new maximum capacity. This leads to strong incentives for maintaining and storing data which may be unnecessary, outdated, or even malicious.

There has been a concern that information kept too long, whether or not it used as a matter of official policy, will, by its very existence and its potential for misuse, have an intimidating or "chilling" effect on the taxpayer. On the other hand, in a society in which people are judged on their merits, standards of administrative due process demand that all relevant information be considered in making a fair decision. There has been) therefore) a countering trend of concern to the privacy one that terminal users and other decisionmakers in mp systems may not have enough accessible information on individuals to make fair decisions. For instance, availability of a benign tax record going back some years could obviously be helpful to a taxpayer who suddenly has problems.

If extensive information in a case is denied the IRS employee, it may affect his ability to set priorities with which duties are carried out with regard to that case and others as well.

It is difficult to separate the policy from the technical considerations on this subject. Issues of retention time and storage capacity are intertwined with issues of contents of the system and the relevancy of the information collected, stored, and used in it. Retention is related to questions of when it becomes outdated, how it is purged, and how these decisions can be enforced administratively and technically.

---

12. DattJ ***Banks in a Free Society,*** by Alan F. Westin and Michael A. Baker, Report of the project on Computer Databanks of the Computer Science and Engineering Board, National Academy of Science, 1972.

In many agencies, policy on such issues has tended to evolve from a series of management housekeeping decisions made incrementally over many years to meet the administrative and political needs of the moment, the changing capacity of the equipment, or cost/benefit concerns. Consequently, there has been little comprehensive review Of the PUBLIC POLICY implications of retention time of file data.

With all of these developments, the chance to start anew is now seen not as a mere concept, but as a right to be respected in the administrative process of government ~d organizations. In recent years, as new or expanded government computerized systems have been proposed, interest GROUPS and others feeling the pressures of certain policy aspects of Federal data programs have urged the evaluation of retention time of data as a policy issue.

This public concern has been expressed in numerous acts of state and federal legislatures for various kinds of records. Many states have adopted policies of sealing old records. Legislation and regulations governing arrest records and other information used in the administration of justice have sought to impose time limits for information kept in the system and to devise detailed administrative and technical methods for challenging and purging irrelevant and outiated information. Courts have also sought to formulate judicial standards for purging information and assuring that outdated information is not communicated to other systems.

In the Fair Credit Reporting Act, Congress recognized this right in a modified way by setting time limits on the use of credit information and allowing the consumer to start anew with a clean slate.

Again in the Privacy Act of 1974, Congress indicated its concern for obsolete, irrelevant data in government files, including those of the Internal Revenue Service, by requiring all Federal agencies to "maintain all records which are used by the agency in making any determination about any individual with such accuracy, relevance, timeliness, and completeness as is reasonably necessary to assure fairness to the individual in the determination," and it allows the individual to seek correction or purging of the records.

Computer users and managers have examined the retention ISSUE as an economic or a technical problem, but not as a social or political one. In some organizations and agencies acquiring new electronic data processing means, one response to such concern has been not to purge, but to reduce the data for storage still further, and to argue that the economics of the situation make it easier to retain the data than to purge it of irrelevancies and outdated matter.

As a result of all of these trends, decisions on need and retention time of specific data elements can be key public policy issues for new systems. Consequently, one technological attribute being asked by the public, by buyers and users of these systems is the ability to purge outdated information and the requirement to purge it in those systems where there appears to be provable evidence of potential detriment to the individual of outdated or irrelevant information.

The TAS Proposal

According to IRS sources, the present Master File tapes keep only three years of tax history. After the 4th year, the oldest year is put on a retention register either on tape or microfilm. These records may be kept forever. Additional research for previous years is done by requisitioning stored tax returns or searching microfilm records. In addition, a separate system, the Integrated Data Retrieval System (which will be replaced by TAS) puts on-line, accessible on terminals in the regions, tax history for those taxpayer accounts with problems or where activity is expected. (This generally amounts to about ten percent of taxpayers' accounts.) According to one source, this usually may cover as many years as necess~ to deal with the account. According to others, it is usually for one year.

The TAS will provide means for retaining 5-years data for all tax accounts. When it is outdated, according to IRS, the information will then be taken off the system and microfilmed or otherwise stored for at least 10 years or, in some cases, possible forever, since there is no destruction period for basic data.

The TAS storage approach has three different storage levels according to TAS documents; it permits record migration or movement to less expensive and less responsive on-line storage devices unless subsequent events, such as inquiry or analysis needs, demonstrate a need for extended retention and for frequent access. The first level offers immediate accessibility. Records in the second level would be available immediately, most of the time, but usually overnight. Records in the third level of storage would be on disc or tape and available probably within a week. For instance, if the audit division is auditing all five years, then that would be in the immediately accessible storage level.

Thus, several records comprising one taxpayer's account may reside on several devices with differing access characteristics and times.

This range of retention time and levels of storage, in the opinion of some people concerned with civil liberties, may affect the ability of the taxpayer to understand the system and to exercise information rights in the programs of the various usem of the TAS. Yet, it is exactly the ability to understand the system which leads to a belief in its fairness.

In the proposal for any such computerized system, especially since the passage of the Privacy Act, it ought to be very clear whether or not the subjects of the files will be informed about the full scope of the information potentially available to them. If not, the~ ability to challenge the accuracy or completeness of the information will be severely limited with such a range of storage. It ought to be clear how important an item of information has to be before a search of relatively inaccessible storage devices is instituted.

Although an essential element of the redesigned systim is reported to be "quicker access to more current data' by those IRS employees who need the data to resolve a specific inquiry or process a case, the other major function of the system will be to afford them access to older information on the taxpayer.

There may be a need to acquire information to determine whether there are, or should be, policy guidelines backed by administrative and technical controls on retention of specific data for each IRS component office to assure that outdated information does not work unfairly to the detriment of taxpayers and businesses or hamper the Service in the effectiveness of its work.

Originally, the TAS proposal sent to Congress called for putting five tax years of history on-line, but in the budget review process at the Office of Management and Budget, the system was cut back to three years on-line, with two years history in slower storage for most files except where specific problems existed. There is no guarantee, however, that this policy will not change with a lessening of budget restraints and that the retention time till not be extended pursuant to internal managemenent decisions and without any Congressional review of its possible impact.

A cost-benefit analysis made for IRS by a private contractor, an internal document, merely specifies the various offices within the IRS who expect to make me of the data, but does not specify which data they need for what length of time. The public documents on TAS describe in very general terms the type of information now in the IRS manual and other files which might be included in the computerized files. They do not specify how 1ong each type of information is presently maintained.

The report sent to Congress under the privacy Act refers generally to statutory requirements and discretion to examine records or to carry out IRS duties but does not describe what specific policies will govern retention under the new System. Without such information, it may be difficult to determine to what extent tax programs may be altered by expansion of the retention time for those records and documents to be computerized or for those records already in the Master Files. The report notes that the IRS Code provides the basic retention rules followed by the Service. The period for assessing an additional tax liability is 3 years from due date or date the return is filed, whichever is later. There is a general 3-year rule for taxpayers to file a claim for credit or refund. There is a 6-year statutory period to collect assessed tax liabilities. Income averaging involves the current year plus the past 4 years. Net operating loss carryback and forward claims may pertain to more than 3 years. There is a 6-year statute of limitation where there has been a substantial understatement of gross income. There are exceptions to the general rules, which cause the Service to receive claims and other transactions concerning accounts which have been inactive for more than 3 years; normally 200,000 to 250,000 such items are received each year.

According to the IRS, retention of the additional 2 addition~ tax years of data will, it is stated, satisfy almost all research requirements and reduce the need to requisition tax returns or to maintain a microfilm system.

c. CONSOLIDATION AND LINKAGE OF INFORMATION

ISSUE: WHAT CONSOLIDATION AND WHAT LINKAGE OF TAX DATA IS PLANNED AND WHAT UNINTENDED EFFECTS FROM THESE SHOULD BE GUARDED AGAINST IN THE DEVELOPMENT AND OPERATION OF **TAS?**

SUMMARY

In order to determine if there may be opportunities for accidental or intentional misuse of information and in order to evaluate such issues as privacy and organizational change, a more precise description is needed of the extent of consolidation, or association, of IRS files on data and of the linkage of data.

QUESTIONS

1. Exactly what consolidation of records and files is planned under the TAS?
2. Exactly what linkage of data elements is planned?
3. What are the implications of the consolidation of records for threats of surveillance and harassment of the taxpayer?

BACKGROUND

There may be a need in planning for TAS to identify technical or administrative linkage and consolidation of information, whether intended and unintended to assess possible consequences for decisionmaking when information is disclosed in a new consolidated form, and to determine whether particular linkages or consolidations should be authorized or prohibited.

Where linkages and consolidation are approved, there is a need **to determine** whether proposed technology and safeguards permit sufficient social, administrative, and statutory control.

If TAS is found to be a more efficient process of consolidation and linkage actually required by statute, there is a need to determine whether changes in efficiency and effectiveness may have

negative impacts to be weighed by IRS and Congress. The public documents and testimony on TAS are not sufficiently informative to permit judgments on these issues.

The present inability to associate related returns and increasing paper and storage burdens are major reasons cited in advocacy for the new system. IRS officials testified, "We can consolidate and link the taxpayer accounts in our Master Files, cross-relating one to the other. This can be accomplished in a data base system of this kind much easier and more efficiently that it can be in a serial ordered tape system. " (IRS testimony before the House Appropriations Committee in 1976.)

Linkage of tax accounts to other relevant data was a major requirement listed by all TAS users. Association of individual returns with business-related returns is a major area of changed capability under the new system which the Service believes would enable greater compliance with statutory mandates to enforce the tax code and would encourage increased taxpayer compliance in the face of this capacity. At present, according to the IRS, with the exception of sole-proprietorships, such direct association is not possible; partnership returns, individual controlling shareholder-corporate returns, and the link are available for reconciliation only if the business entity is chosen for scrutiny, and related individual returns are then acquired on request for agent analysis.

Other examples of actual and potential linkage of associations could be cited, such as the taxpayer and the names of people and **groups** related to deductions for charities, subscriptions, or business lunches.

Secondary Linkage

Beyond its instant on-line capacity, TAS will facilitate a secondary linkage to other files in storage and in other administrative data systems. A code symbol will flag the account, removing it from the routine processes and alerting the decisionmaker that there is other action pending or that another office or agency may want information or be concerned with the case, and guide the person to additional intelligence or other data within IRS and other agencies. This question is closely tied to the proposed contents of the system and access questions raised elsewhere.

The actual and potential uses of the TAS for secondary linkages so that any possible negative effects on privacy and due process rights have not been sufficiently identified and evaluated in the IRS documents.

d. DERIVED DATA

ISSUE: COULD THE TAS, WITHIN THE PROCESSES IT SERVES, RAISE PROBLEMS OF DERIVED DATA, THAT IS, OF CREATION OF NEW DATA OUT OF SEVERAL PIECES OF PRE-EXISTING INFORMATION, WHICH MAY REQUIRE SPECIAL SAFEGUARDS TO PREVENT THREATS TO PRIVACY OR OTHER RIGHTS?

SUMMARY

The problem of derived data is an implicit one for any large sanitized personal information system or one where personal data may be derived programmatically. It arises where information from other sources is combined with information from the individual file in order to derive other information. The problem is related to the overall problem of linkage in that both assume a matching of sources of information with the object file. Derived data obtained in this fashion are only inferred to be correct because there is not a direct link, only implicit linking. There could be unknown factors which, if known, could prove the derived data wrong, or prove the derivation. Public documents on TAS do not indicate what safeguards are planned for dealing with this problem.

This line of inquiry may be particularly important since the use of the social security number in TAS and other large personal data systems is cited as a means of preserving the anonymity of the individual when the files are used for research, statistical or non-tax purposes.

BACKGROUND

The area of derived data is one which traditionally has concerned segments of the public in census and other statistical information gathering programs. **Lately, with more complex** technology and ingenuity in devising programs, particularly in the intelligence area, it has caused increased concern in computerized personal data systems.

One Commentator describes the problem as follows:

> "The derived data problem is another technological issue not yet clearly understood nor treated in current legislation. The problem takes at least two forms: First, to

> what extent may data not identified with individuals be analytically or statistically associated with them? For example, there may be information in one file about an unidentified individual with a specified salary and other personal details, including the census tract in which he resides. Another file could have information about an identified individual stating his salary and place of residence. By matching the known information common to both of these files, the file of data about the unidentified individual can easily be identified from the name supplied in the other file. This kind of problem often occurs when data about individuals are unique or limited to small numbers of people.
>
> "The second kind of derived data problem is that there are types of personal data that may be represented programmatically rather than directly in the form of stored data. A file can contain names of individuals and limited amounts of data which can be processed by computer programs that contain generic data to produce significant additional information about the individual. Thus, this type of program must also be treated with the same sensitivity as the data that the program produces. Current legislation does not appear to take into account programmatically derived personal information. "[13]

in view of the public concerns and perceptions about threats from surveillance and the technical possibilities for deriving data, the regulatory rules or regulations governing any large personal information system and its data banks planned today should apply certain standards not only to personal information in the files but include language covering all additional personal information derived from it.

Consideration might be given to how extensive this problem could be in TAS to the detriment of rights to privacy and due process of taxpayers not only in IRS programs but in those of other governmental users of TAS data.

13. Dorm B. Parker, ***Crime By Computer,*** Charles Scribners Sons, New York, 1976, p. 250.