



Management Perspectives of Data

Added Information Value Through
Computer Matching of Data

Presented by:

March 27, 1985

Morey J. Chick,
Information Management and Technology Division
General Accounting Office

at NBS Workshop on Data Administration

ABSTRACT:

Computer matching is defined here as a comparison of data that exists in different files, for the purpose of creating new information. The new information that is created by a computer match is a factor that is measurable and that represents a value which may be added to intrinsic value of the information contained in the files that were matched. In an information resources management context, information value must be maximized and information costs must be minimized. In management, these factors, i.e., value vs. cost, are often confused. Nonetheless, they must be measured; the question arises as to whether the value of information can be measured in terms of dollars. Results of some examples of computer matches cited in this presentation appear to indicate that this question can, in some cases, be answered in the affirmative.

BIOGRAPHICAL SKETCH:

Mr. Chick is a manager in the Government-wide Information Technology Studies Group of the Information Management and Technology Division, General Accounting Office (GAO). He is a graduate of Pennsylvania State University, a Certified Public Accountant, and has 24 years of experience with the GAO. Mr. Chick specializes in studies of the problems associated with managing information, automatic data processing, individual information systems, and information policy issues in the Federal government. He has worked on a GAO task force to define and develop methodologies for studying Information Resources Management (IRM) in Federal agencies. He has also instructed GAO professional staff in IRM. Among other things, Mr. Chick has identified and reported to Congress significant problems in automated decision-making, computer security, data standardization, and the implications of automation on employment. Mr. Chick has also prepared Congressional testimony on computer matching as used to detect fraud and error in Federal

032386



benefits. He has written extensively on the subject of the economics of information.

The presentation is a distillation of the views of the General Accounting Office (GAO), the author, and other sources, on computer matching as a tool for the management of information. The views of the General Accounting Office are documented in their report HRD-85-22 entitled, "Eligibility Verification and Privacy in Federal Benefit Programs: A Delicate Balance." The author's views are partially reported in his article, "Information Value and Cost Measures for Use as Management Tools," published in Information Executive, Volume 1, Number 2, 1984. A copy of this article is part of this record of the presentation.

Computer matching is defined here as a comparison of data that exists in different files, for the purpose of creating new information. The files may belong to a single agency, to several agencies at various Federal, state, or local government levels, and/or the files may belong to non-government organizations. The new information that is created by a computer match is a factor that is measurable and that represents a value which may be added to the intrinsic value of the information contained in the files that were matched (Figure 1).

Computer matching is really a type of data analysis. In the "old" technology the process involves a simple match of files from database B against the files from database A on data elements that are common to both files. A match on these data elements generates new information which adds value to the value intrinsic in databases A and B (Figure 2). The purpose of the new information is to detect errors, fraud, and/or internal control problems associated with the management of benefit programs in the Federal government. Dollar values, here, can be measured by the savings resulting from the new information created by the match.

Figure 3 illustrates current technology as moving towards direct linkages of files via telecommunications lines. Location C on this figure represents non-government organizations, such as a credit bureau, a bank, or a school. What we have basically is a de facto centralization of data. Figure 4 represents a hypothetical link comprised of real providers of data. At present, there is no central information on all current linkages.

The concept of computer matching is not a new phenomenon; it has been in existence since approximately 1976. In the time that has elapsed since then, some 126 matches have been performed at the Federal level and some 1200 more at the state level. These matches were made on files that store information on a minimum of 136 Federal programs which benefit three out of ten Americans. The Federal share of total expenditures represented by these programs amounts to approximately \$400 billion a year or 45 percent of the national budget. It is estimated that several billion dollars are overpaid annually because of abuse, fraud,

error, and inadequate verification of applications for benefits. GAO historically supports matching when the benefits exceed costs and the rights of individuals are protected.

Figure 5 presents three examples of major Federal matches of data on income tested programs. The agencies involved were the Veterans Administration (VA) and the Social Security Administration (SSA). The VA pension program files were matched against earnings reports of state unemployment security agency files on at least four data elements: wages, Social Security number, name and employer. This match resulted in the detection of overpayments totaling an estimated \$100 to \$300 million. Benefits realized from two matches of Social Security files are reported in the form of reduction of payments of approximately \$110 million per year in one case, and expected recoveries of \$100 million in the other. Some \$20 million of the latter figure have been recovered to date.

Figure 6 presents examples of three state matches. In the first example, New York City identified companies paying business taxes, but not rent taxes. The City matched the files from several of its own departments and collected \$24.8 million in additional commercial rent payments.

What are the concerns related to computer matching? Some of them are:

- cost vs. benefit ("added value");
- technology and centralization;
- privacy;
- security;
- other concerns.

Cost/benefit analysis presents a very difficult problem, one which is under study by GAO at present. Measurable benefits are being identified and are continuing to be reported; recoveries represent real savings. Reductions in future payments present an added difficulty in that there is a lack of information on how long the benefit payments would have been made or even if they would have been made, in any given case.

Intangible benefits identified include the potential inherent in the use of computer matching as an internal control mechanism, as a means of testing of internal controls, and as a deterrent factor. Benefits of such intangibles are very difficult to measure in dollar terms.

GAO is just now beginning to study the different kinds of costs involved in computer matching. Some of these are:

- cost of match (software, computer time, etc.)
- manual verification (e.g., employers, manual computations, etc.)

- file acquisition costs (from third parties, e.g., credit bureaus);
- costs of poor data quality;
- cost of reducing or deleting payments;
- cost of denying payments (e.g., litigation and related administrative procedures);
- collection costs for recoveries.

The first of the above costs is the traditional one. The second, manual verification is now required by law for certain major programs. There are hidden costs associated with matching in cases where there is a need for employers to verify information. Poor quality of data is partially a result of the lack of data standards. Further costs are those stemming from data sensitivity and privacy issues, such as litigation and related administrative procedures. Currently, GAO is studying the situation, particularly from the standpoint of much-needed methodology for measuring value versus costs associated with computer matching.

In information management, the terms value and cost, are often confused. The cost of information can be equated almost to the cost of producing a commodity from raw materials. Many accounting functions can be applied here; and information value can be described in terms of worth, merit, importance, etc. However, the question remains: can we measure value in terms of dollars. In his journal article cited above, the author presents some ways to measure information value in dollars in some cases. It should be done, where possible for "effective management."

Computer matching does represent a de facto centralization of data, as Figures 7 and 8 indicate. The figures also identify the many and various sources of information for matching purposes. This de facto centralization is not unconstitutional but does raise increased concern about privacy and security. The privacy issue is a very sensitive issue these days, one that is being hotly debated. The GAO report cited above addresses some of these issues. GAO's conclusion is that there is a delicate balance involved between detection of fraud on the one hand aimed at protection of the U.S. taxpayer and the privacy of the individual on the other hand aimed at protection of the U.S. citizen. In many cases, these are the same people.

The sources of citizens' rights to privacy are basically the Constitution, the Fourth, Fifth, Fourteenth (and perhaps other) Amendments, and Common Law. These are the real sources. The Privacy Act of 1974 (P.L. 93-579) is the legal source for Federal data only. The Privacy Commission provided opinion and clarified the principles. Section 552a of Title V of the Privacy Act defines routine use as, "The use of such record for a purpose which is compatible with the purpose of which it is collected" (Figure 9). This is the part of the Act that provides for no disclosure without written consent of the individual citizen. However, there are 11 exceptions to that, and the routine-use

clause of the Privacy Act is one of the exceptions. Executive interpretation is usually related to this clause and has basically increased and facilitated extensive Federal matching. State matches are not covered by this Act.

At this point, the author separated himself from the GAO and presented the views of some of the opponents of computer matching. Some of these views include the following:

- the real possibility of excessively broad interpretation of the routine-use clause;
- matching presumes crime, therefore it does not constitute reasonable search;
- the category of people is of interest to the government;
- fear of misuse of information (big brother);
- matching involves everyone in the file, including the innocent, and even people not receiving benefits, as in the case of credit bureaus, for example;
- purpose of match is to generate evidence of wrongdoing;
- not every program requires a direct notification of a match;
- notification via the Federal Register as required by the Privacy Act is inadequate notification;
- technology linkages increase security vulnerabilities;
- there is no requirement for central approvals of matching.

The Internal Revenue Service (IRS) has a concern about the confidentiality of tax information, as provided for in the Tax Reform Act (Figure 10). Though opening of actual taxpayer information files (forms 1040 and related schedules) is not in sight at the moment, the IRS is concerned about the impact of opening tax records. The potential losses in voluntary tax collection may be more than what may be saved in the benefit programs through matching.

The last major item of concern in this area has to do with computer security. GAO is currently studying this area, and the author is involved in the study. Figure 11 lists the concerns associated with computer security. One of the items on this list is the personal data and privacy issue. The Privacy Act requires adequate technical, administrative, and physical safeguards for the protection of personal data. The last item concerns human safety considerations. Factors such as speed, error, system design problems, human response to speed, and automated decision making are major personal concerns.

Finally, some other major concerns in computer matching include:

- data quality in automated decision-making and the associated practice of direct notification and elimination of beneficiaries without manual verification;
- the question of when to match;

- the SSN as the national identifier;

- alternative verification techniques, such as telephone contacts.

The above concerns comprise basically the GAO report now being circulated. In conclusion, matching does represent a delicate balance.

DEFINITION OF A COMPUTER MATCH

- **DATA FROM DIFFERENT FILES**
- **MAY BE DIFFERENT AGENCIES**
- **FEDERAL, STATE, LOCAL**
- **OTHER FILES (NON-GOVERNMENT)**
- **CREATES NEW INFORMATION**
- **DONE BY COMPUTER**

CONCEPT "ADDED VALUE"

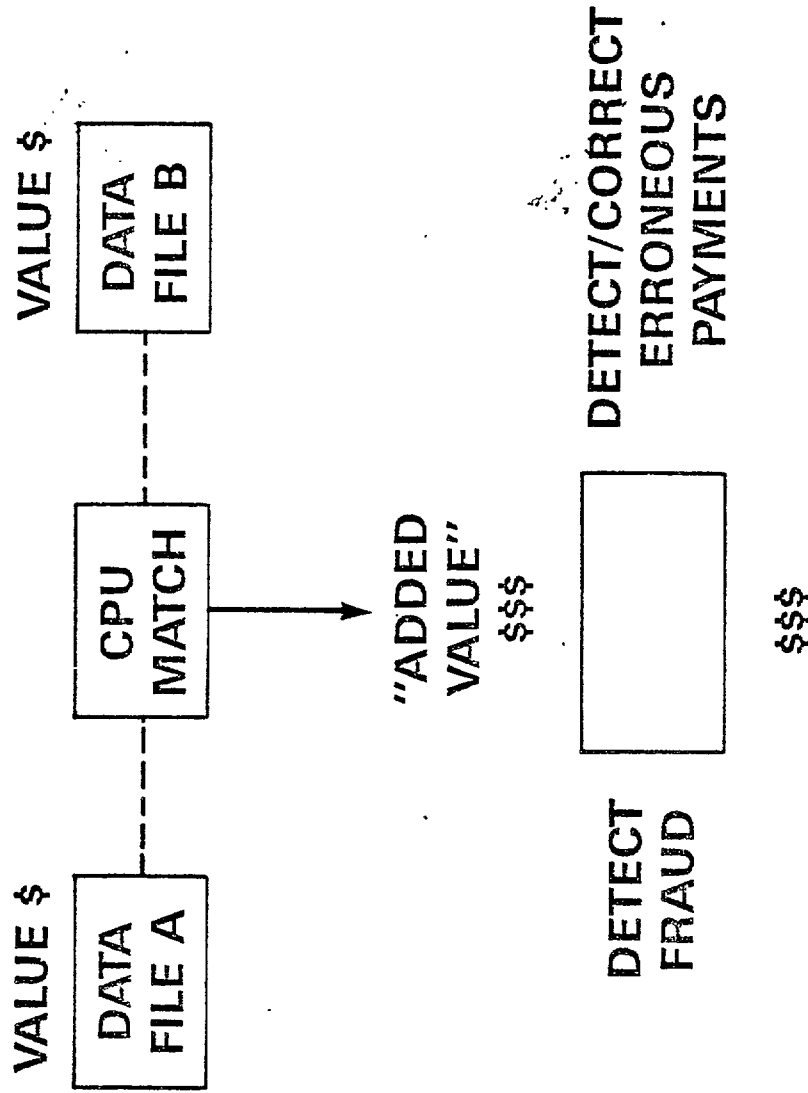


FIGURE 2

CURRENT MATCHING TECHNOLOGY

MOVING TOWARDS DIRECT LINKAGE

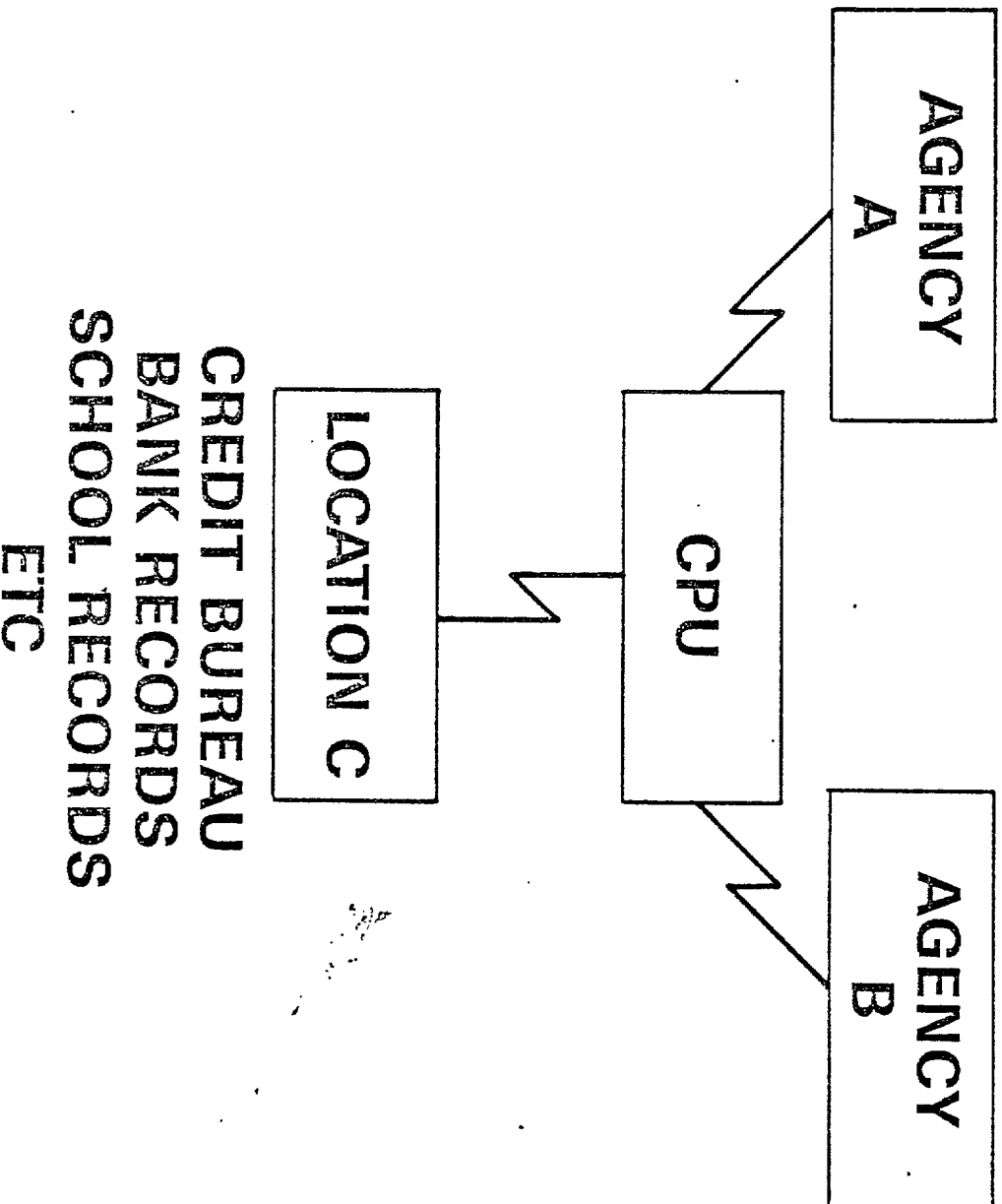
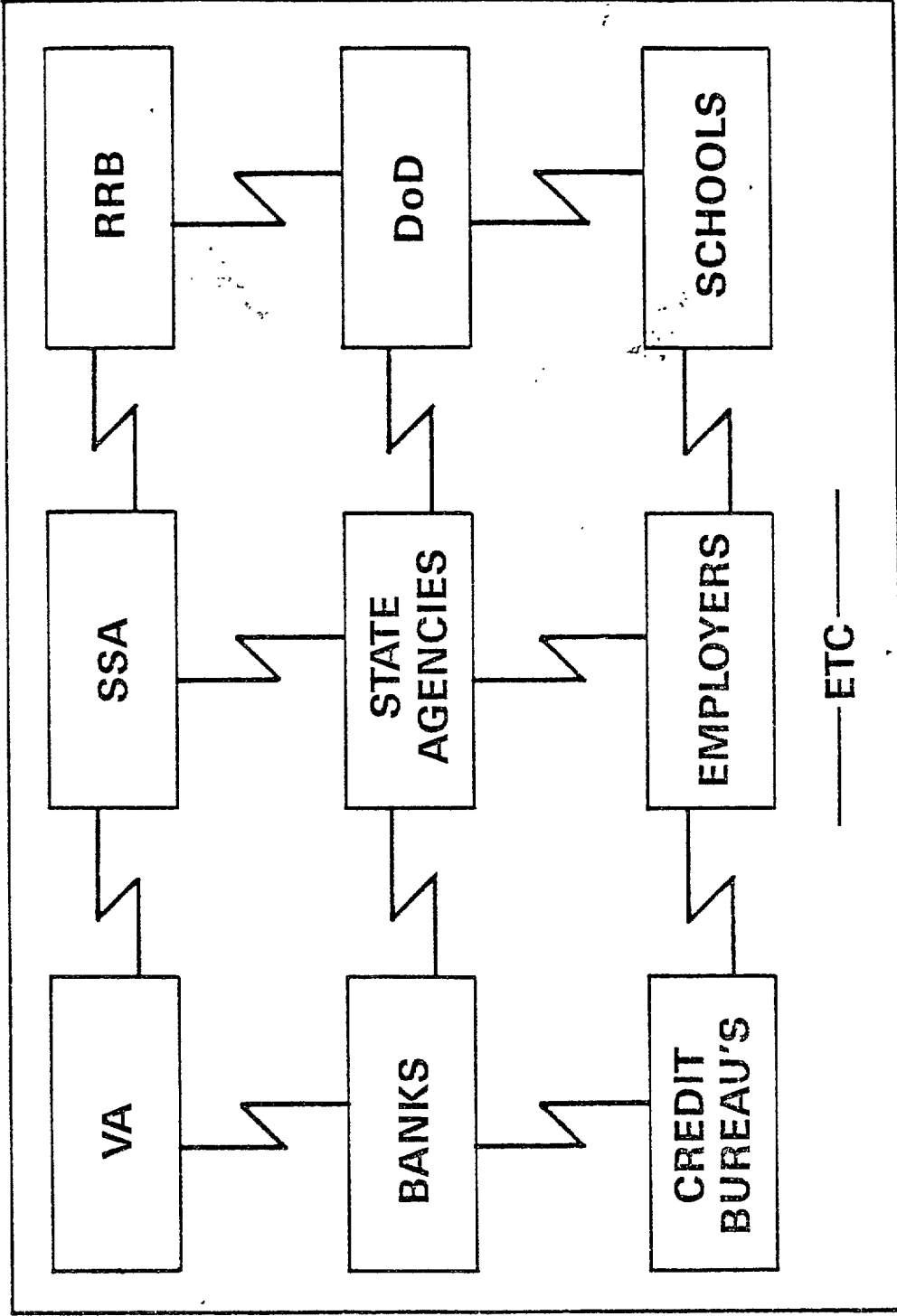


FIGURE 3

DE FACTO CENTRALIZATION OF DATA



NO CENTRAL INFORMATION ON ALL CURRENT LINKAGES

FIGURE 4

EXAMPLES OF MAJOR FEDERAL MATCHES (INCOME TESTED PROGRAMS)

YEAR INITIATED (APPROXIMATE)	PRIMARY AGENCY	PROGRAMS	MATCHED AGENCIES	DATA ELEMENTS MATCHED (EXAMPLES)	BENEFITS
1983	VA	VA PENSIONS	STATE UNEMPLOYMENT SECURITY AGENCIES	(EARNINGS REPORTS) WAGES, SSN, NAME, EMPLOYER.	FROM \$100 MILLION TO \$300 MILLION (TOTAL OVERPAYMENTS ESTIMATED)
1976	SSA	SSI (AGED, BLIND DISABLED)	VA (COMP./PEN) RRB (PENSION) OPM (PENSION) DoD (RETIREMENT)	UNEARNED INCOME (PENSIONS) SSN, NAME.	\$110 MILLION A YEAR (REDUCTIONS)
1979	SSA	OLD AGE SURVIVORS AND DISABILITY	STATE DEATH DATA (30 STATES SO FAR IN CENTRAL FILE IN HHS SSA ACT AMENDMENT OF 1983)	DATE DECEASED, SSN, NAME, DATE BORN.	EXPECT \$100 MILLION (RECOVERIES) \$20 MILLION TO DATE

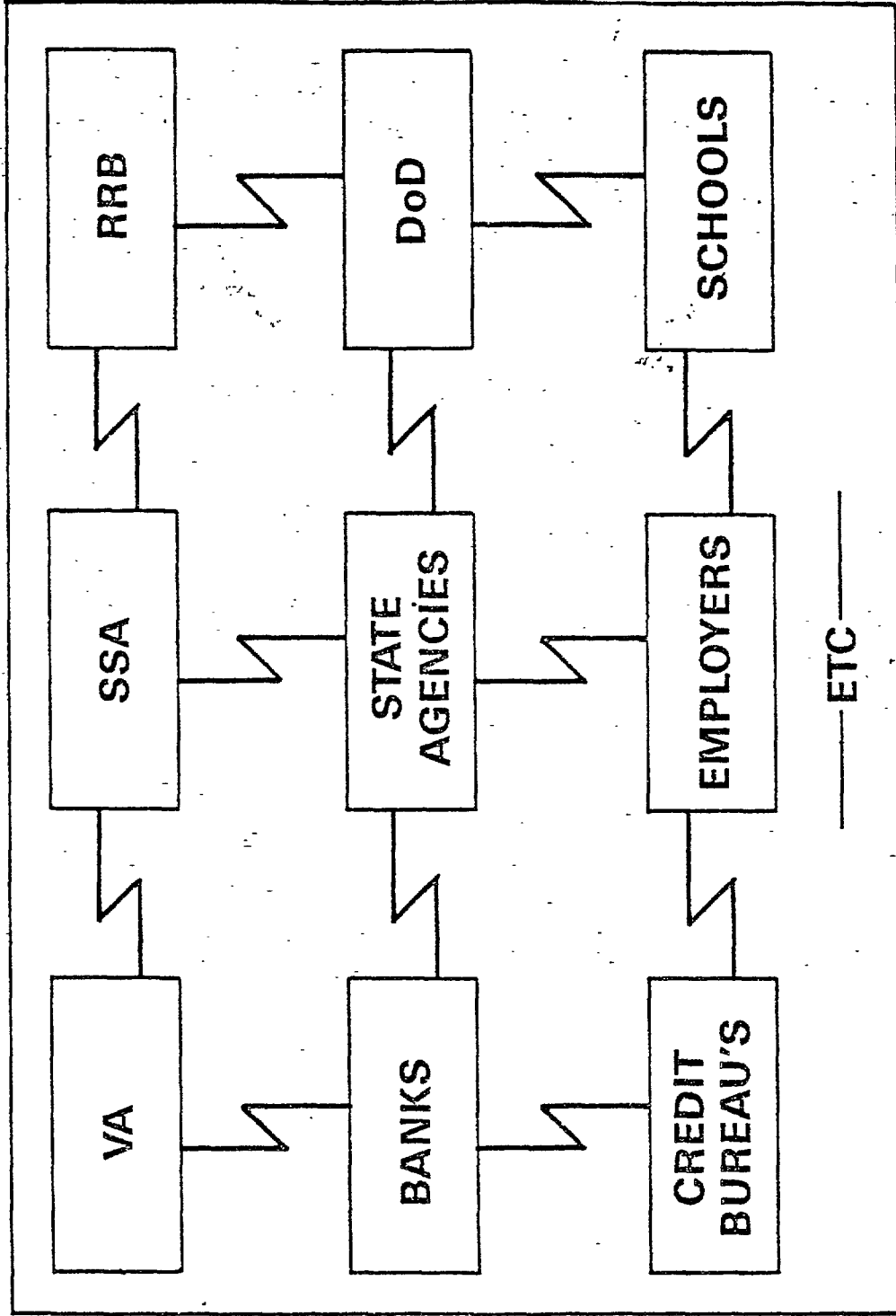
FIGURE 5

EXAMPLES OF MAJOR STATE MATCHES

<u>YEAR INITIATED</u>	<u>PRIMARY AGENCY</u>	<u>PROGRAM</u>	<u>MATCHED AGENCIES</u>	<u>DATA ELEMENTS MATCHED (EXAMPLES)</u>	<u>BENEFITS</u>
1981	NEW YORK CITY DEPARTMENT OF FINANCE (COMMERCIAL RENT FILES)	COMMERCIAL RENT TAX (OCCUPANCY TAX ON COMMERCIAL TENANTS)	SAME (BUSINESS TAX FILES)	CONTROL NUMBERS BUSINESS NAME ETC	\$24.8 MILLION IN ADDITIONAL COMMERCIAL RENT TAX PAID
1981	NEW YORK CITY DEPARTMENT OF FINANCE	BUSINESS TAX FILERS IN 1976 (NOT FILING IN LATER YEARS)	SAME (SAME FILE, PREVIOUS YEARS)	CONTROL NUMBERS BUSINESS NAME	\$20.2 MILLION IN ADDITIONAL BUSINESS TAX PAYMENTS
1984	MICH. DEPT OF SOCIAL SERVICES (WELFARE FILES)	STATE WELFARE STATE: AFDC FOOD STAMPS MEDICAID	SSA BENEFICIARY EARNINGS INDEX (BENDEX)	EARNINGS, BENEFITS, RETIREMENT AND DISABILITY SSN	550 CASES OF FRAUD IN ONE COUNTY ALONE (WAYNE) ESTIMATE: \$6.3 MILLION IN FRAUD

FIGURE 6

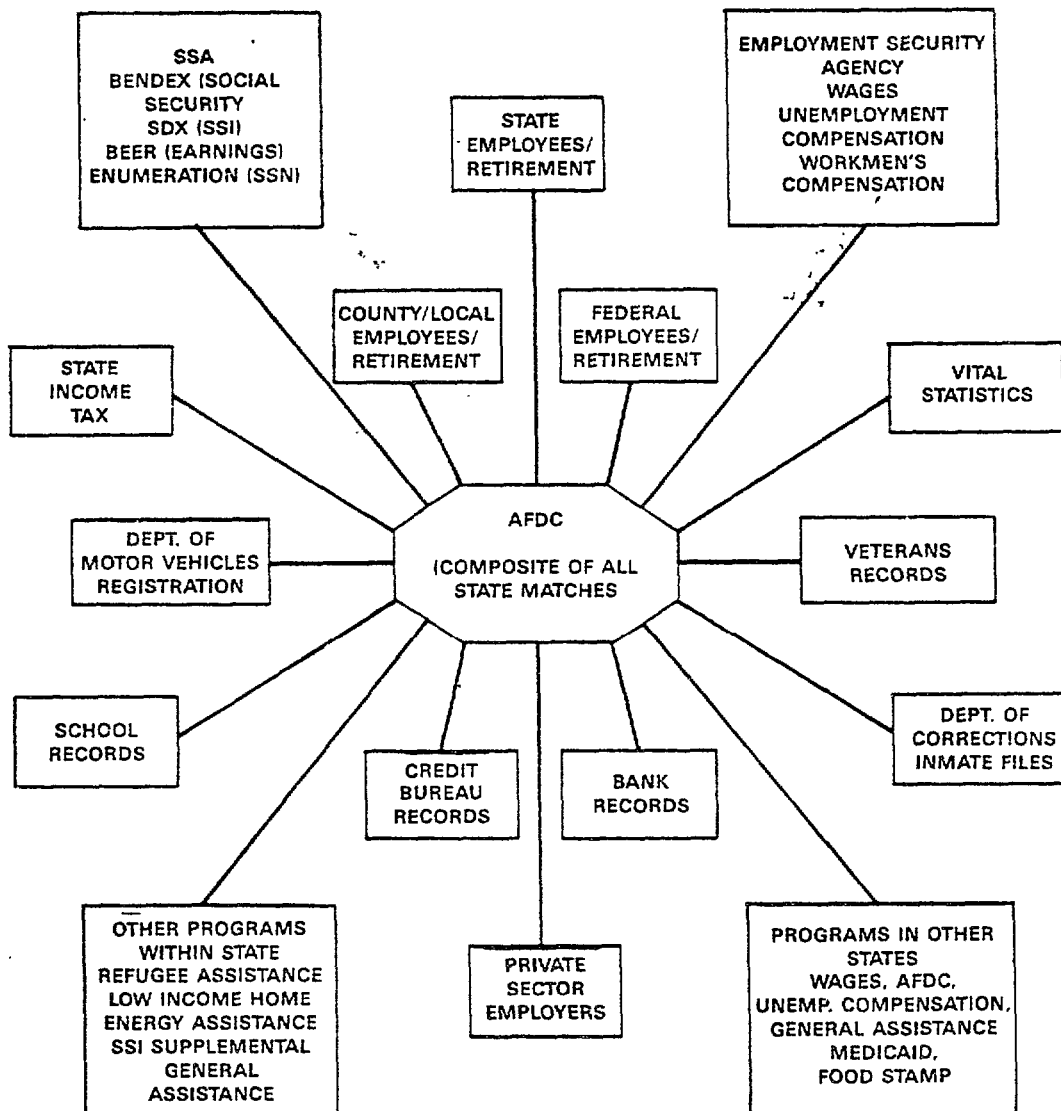
DE FACTO CENTRALIZATION OF DATA



NO CENTRAL INFORMATION ON ALL CURRENT LINKAGES

Figure 7

**COMPOSITE OF DATA LINKAGES THROUGH COMPUTER MATCHES
BY AFDC PROGRAMS IN VARIOUS STATES**



NOTE: NO SINGLE STATE HAS ALL OF THESE LINKS, BUT EACH LINK OCCURS IN AT LEAST ONE STATE. WITH A FEW EXCEPTIONS, HOWEVER, THESE TYPES OF SOURCES COULD BE AVAILABLE IN EVERY STATE.

SOURCE: DEPARTMENT OF HEALTH AND HUMAN SERVICES, OFFICE OF INSPECTOR GENERAL, INVENTORY OF STATE COMPUTER MATCHING TECHNOLOGY; AND GAO OBSERVATION (HRD 85-22)

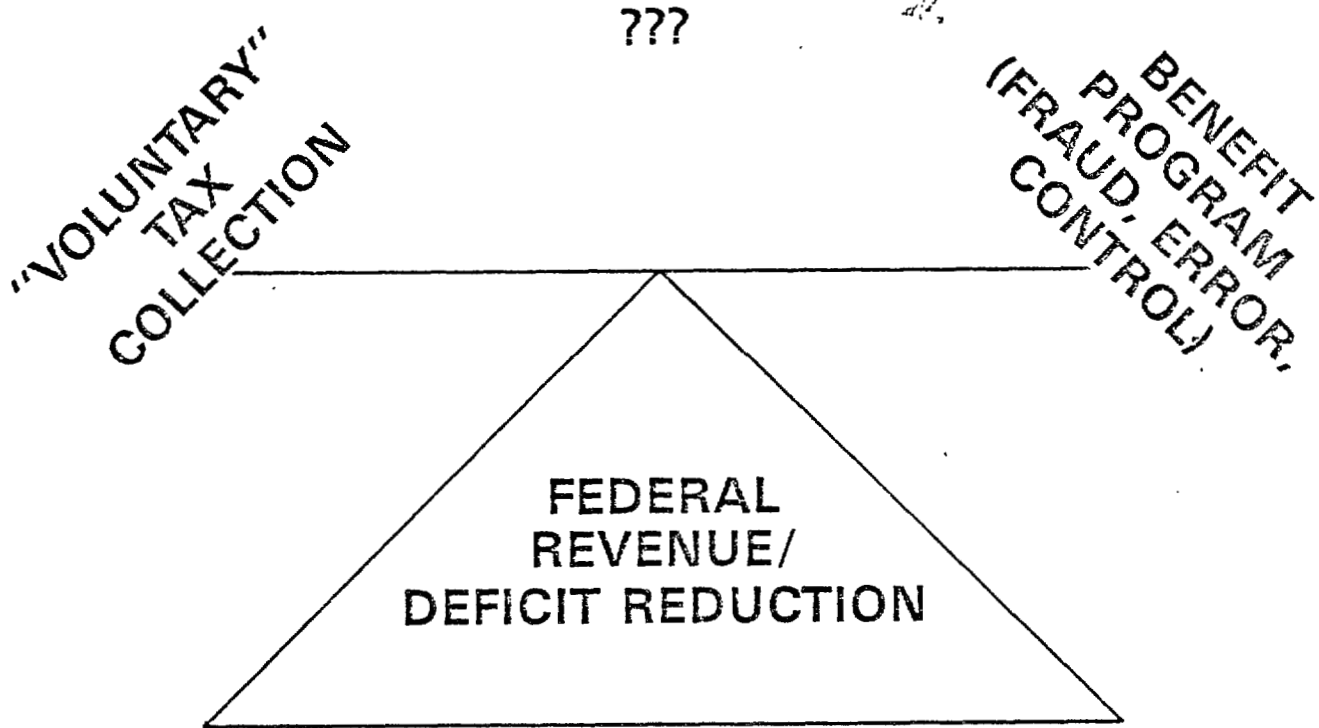
FIGURE 8

**MATCHING AND DISCLOSURE PRINCIPLE
(PRIVACY ACT) - SECTION 552a**

- **"ROUTINE USE" *** THE USE OF SUCH RECORD FOR A PURPOSE WHICH IS COMPATIBLE WITH THE PURPOSE FOR WHICH IT IS COLLECTED."**
- **"CONDITIONS OF DISCLOSURE *** NO AGENCY SHALL DISCLOSE ANY RECORD *** (WITHOUT WRITTEN CONSENT OF THE INDIVIDUAL TO WHOM THE RECORD PERTAINS, UNLESS DISCLOSURE OF THE RECORD WOULD BE *** FOR A ROUTINE USE"**

COMPUTER MATCHING

IRS VIEW ABOUT TAXPAYER RETURNS



WHAT IS THE PROPER BALANCE ???

WHAT IS THE POSSIBLE NET \$ EFFECT ???

REASONS FOR CONTINUING CONCERN FOR ADP/TELECOM. SECURITY

- **DEPENDENCY**
- **HIGH DOLLARS**
- **SENSITIVE DATA**
- **PERSONAL DATA ("PRIVACY")**
- **ADP/TELECOMMUNICATIONS INTEGRATION**
- **RAPID ADVANCES**
- **OLDER SYSTEMS**
- **LEGISLATION**
- **PAST/CURRENT PROBLEMS**
- **AGGRESSIVE HACKERS, CRIMINALS, AND OTHER
UNFRIENDLY SOURCES**
- **HUMAN SAFETY CONSIDERATIONS**