

**REQUEST FOR AUTHORITY
TO DISPOSE OF RECORDS**

(See Instructions on Reverse)

LEAVE BLANK	
DATE RECEIVED MAR 10 1972	JOB NO. 172-21
NOTIFICATION TO AGENCY	
In accordance with the provisions of 44 U.S.C. 3303a the disposal request, including amendments, is approved except for items that may be stamped "disposal not approved" or "with-drawn" in column 10.	
4-28-72 Date	<i>James B. Rhoads</i> Archivist of the United States

TO: **GENERAL SERVICES ADMINISTRATION,
NATIONAL ARCHIVES AND RECORDS SERVICE, WASHINGTON, D.C. 20408**

1. FROM (AGENCY OR ESTABLISHMENT)
National Archives & Records Service

2. MAJOR SUBDIVISION
Office of the National Archives

3. MINOR SUBDIVISION
Data Archives Branch (NNPD)

4. NAME OF PERSON WITH WHOM TO CONFER
Gerald J. Rosenkrantz

5. TEL. EXT.
13-21932

6. CERTIFICATE OF AGENCY REPRESENTATIVE:

I hereby certify that I am authorized to act for this agency in matters pertaining to the disposal of the agency's records, that the records proposed for disposal in this Request of 1 page(s) are not now needed for the business of this agency or will not be needed after the retention periods specified

3/10/72
(Date)

Gerald J. Rosenkrantz Chief, NNPD
(Signature of Agency Representative) (Title)

7. ITEM NO.	8. DESCRIPTION OF ITEM (With Inclusive Dates or Retention Periods)	9. SAMPLE OR JOB NO.	10. ACTION TAKEN
	<p>Proposed General Records Schedule 20, describing proposed disposition for machine-readable records, related documentation required for their servicing, and files related to the automatic data processing (ADP) procurement, operations, and management functions.</p> <p>Proposed schedule is submitted in accordance with the authority contained in 44 U.S. Code, Section 3303, paragraph d.</p> <p>Attached xeroxed comments from the Central Intelligence Agency, the Atomic Energy Commission, the General Accounting Office, Bureau of the Census, and the Internal Revenue Service regarding proposed general schedule. Formal final GAO approval has been sought this date, and will be forwarded when received.</p> <p>Attached proposed general schedule in two copies consists of forty pages.</p>		

General Records Schedule, Automatic Data
Processing Records (Disposal Job No. 172-20)

General Schedule of Records authorizing the disposal, after the lapse of specified periods of time, of records of a specified form or character common to several or all agencies of the United States Government that have accumulated and will accumulate in such agencies and that apparently will not, after the lapse of the periods specified, have sufficient administrative, legal, research, or other value to warrant their further preservation by the United States Government. This General Schedule is approved in accordance with Section 3303a, Title 44, U.S. Code.

James B. Rhoads
Archivist of the United States

April 28, 1972
Date

DRAFT
RLMayer/sjd
3/9/71

DATA SELECTION HANDBOOK
FCR
ARCHIVES ANALYSTS

Introduction

This handbook is a guide to the Data Archives Analysts for determining the nature of data files (on magnetic tape) generated by Automatic Data Processing Equipment (ADPE). Many factors which influence the selection of data files for retention on magnetic tape in the National Archives are indicated and explained in this handbook.

General ADP System (ADPS)

1. Most ADPS's are operated in a sequential manner; whether they are primarily transaction (one at a time), or batched record (more than one at a time) systems. In addition to the controlling software, and the operating software, there are master files (old and new), maintenance files (updates to the master files), and reports.

1.1 Master files are new (updated), and old (sources for the new or current) which are updated by data from the maintenance files. Transaction records from the maintenance are sometimes appended to the old master record, to justify the update of the base record in the master file. Retention of the old master records, the old master records with the appended transaction records, or only the new (updated) master records, are the choices available for the selection.

1.2 Maintenance files (transaction records for updating the master files) may be retained as input files, as applied or unapplied update files, or individually (record by record) as the appended records to the master records to which they were applied.

1.3 Unapplied maintenance records (no update was performed) may be put in an error file as output of the ADPS, with an error code indicating why no update was made. Manual adjustment must follow analysis.

1.4 Obsolete, deleted, and incompatible master records, may be output for review, analysis or archival storage. Depending on the system, these may be output immediately (as ascertained), on a cyclical basis, or on a scheduled cleanup (quarterly, annually or other).

2. Within the ADPE there are three distinctly defined classes of operation: input, output, and processing.

2.1 Input includes everything fed into the ADPE. Such files, records or blocks of data, for control or operation by the ADPE, may require manual or automatic preparation and verification before or after, but as they are read into the machine, they are input.

2.2 Output includes everything produced by the ADPE. Master files, reports, error files, summary information, run statistics, and

console typewriter messages are included in the class of Output, whether manual or automatic handling follows, or not.

2.3 Processing by ADPE includes everything that happens to the data from the time it is input, until the time it is ready to be output from the processor. In the processing, determinations are made relative to the application of maintenance records in updating master records, sorting and merging records for output as information or reports, and selection of records or files for retention in active or archival storage, whether on cards, disc, drum, tape or other device. All above actions are under the control of software, which is the heart of an automated system.

3. Preparation of data for an ADPS may be simply the representation of an event (or a condition) as recorded by an automatic sensor, or more complex, as transcription by manual means, through several stages or copying and translating until finally the data is recorded in machine readable form. Through the manual preparation steps, verification, validation, screening and editing must often be included, and each is an opportunity to introduce errors as well as to eliminate previous recording errors.

GENERAL RECORDS SCHEDULE 20

Introduction

This schedule covers machine readable records, related documentation required for their servicing, and files related to the Automatic data processing (ADP) procurement, operations, and management functions.

The decision table format, rather than the columnar format, is used for two reasons: 1) footnote requirements are greatly reduced with this format as compared to the columnar format of the first nineteen schedules; and, 2) the number of times a given file of logical records has been processed is often more important than the name assigned to it. For example, in an update system, the last created version of an interim master file becomes a final master file after the sponsor declares it error free. The only difference between it and its predecessors is the version number. There may be many versions of a given file created during the course of a processing cycle. Failure to promptly return unneeded tapes to the inventory will lead to excessive requirements for tape. For this reason, it is imperative that the creator of machine readable records assign file retention times at the outset--that is to say--at the time of the original system design effort.

The principle machine readable and supporting records common to more than one agency have been divided into four categories. These classes of records correspond roughly to the typical organizational and functional structure found in most ADP installations and their parent organizations.

Data Automation planning and operational records, (Part I), are normally those created during the life cycle of individual computer installations. They deal with planning for, managing, procuring, selecting, utilizing, and accounting for the physical facility investment of the ADP installation and supporting activities.

Documentation required for servicing machine readable records, (Part II), is defined as the organized series of descriptive documents required to initiate, develop, operate, and maintain specific applications of ADP systems. These include project documentation, system specifications, test data and procedures, file and user documentation, and the various installation procedures and standards used in daily operations.

Erasable Media, (Part III), covers all devices which store machine readable records in an erasable mode. At present, only magnetic media are commonly used for such purposes. However, future technological developments may provide the same characteristics (nonvolatility and easy reusability) now found on magnetic tape.

Since magnetic records may be destroyed by overwriting, a variety of protective devices and techniques have been developed over the years to preclude inadvertent erasure of records. The earliest technique, still in use, consists of a mechanical interlock device known as a "write protection ring," inserted or left out of a reel of tape. With the later development of computer manufacturer supplied "operating systems," an additional safeguard was inserted into the software. It consists of writing file identification and expiration dates on a label record at the beginning of each reel of tape. Other magnetic media such as disks depend almost solely on such software devices.

Nonerasable Media, (Part IV), covers such media as punched cards and paper tape. Most ADP installations use media other than magnetic for a variety of roles and functions, but for the most part they are temporary. However, punched cards are sometimes used as documents, such as checks, savings bonds, and requisition forms. In such cases, the functional retention period, developed in other records schedules will apply.

Procedural Analysis of data processing systems (Part V) is a guide for archivists, records officers, and auditors in determining secondary uses for data files. Unlike paper, computers create more working copies, which should be erased promptly. But the secondary value, such as furnishing data for audit trails and statistical analysis must be recognized when appraising machine readable records. Many systems, in becoming more automated upon procurement of newer ADP equipment, drop certain manual controls. Since many systems are dynamic, they change due to corresponding changes in legislation and other factors. Thus, nonoperational programs may have to be kept for site-audit records.

DRAFT JULY 1971

DATA AUTOMATION PROGRAM RECORDS - GENERAL RECORDS SCHEDULE NO. 20

Part I Data Automation Planning and Operational Records

Covering documentation relating to objectives, concepts, policies, and plans providing overall aspects of data automation data needs and systems design of management supporting systems and operational supporting systems, including equipment selection and statistics.

File Designation	Consisting of	Which are	Then
1 Planning documents	master plan, feasibility studies with associated charts and diagrams, supporting data that reflect on the characteristics of the data automation activity	graphic, narrative and tabular information relating to the present and/or planned ADP composition and requirements of the data automation activity	disposal not authorized by this schedule.
2 Program management	development of plans, policy, and procedures governing the conversion to electrical machine operations and the supervision, control, coordination, and operation of the mechanization program	maintained at policy determination level	disposal not authorized by this schedule.
3 Hardware selection	agency requirements, specifications for hardware, software, and support capabilities of vendors of complete installations or of major peripheral equipment	selection criteria for procurements in the establishment or modification of an ADP installation	dispose 2 years after specific configuration of equipment is discontinued
4 Standardization	data elements and codes, standardization requests and justification for all data systems	promulgated Federal or national (except record copies at National Bureau of Standards)	dispose when superseded or obsolete
5		other standards eg. developed by agency	disposal not authorized by this schedule.

MAR 1972

	File Designation	Consisting of	Which are	Then
6	Utilization and Maintenance	forms or cards which equipment operators complete relative to machine use, nonuse, or maintenance	used for daily management of operations	dispose after 3 years
7		daily detail cards, intermediate summary decks, related magnetic tape files and machine listings	used for daily management of operations	dispose after 90 days
8		monthly summary of cost and utilization reports	card decks, magnetic tape files and machine listings	dispose after 3 years
9	Accountability	documents concerning the management of ADPE equipment	original records maintained at data processing installation	dispose 2 years following the date equipment is discontinued
10		requirements for cards, paper and magnetic tape reels and inventory of ADPE supplies		dispose after 1 year
11		contractor's invoices for rental and other charges incurred for use of ADPE		dispose after 3 years
12	Magnetic tape library control records	Library transaction records	card decks and magnetic tape files	dispose when the 4th update cycle is created
13			machine listings	dispose after 90 days
14			transaction slips	dispose after 90 days or when no longer needed

PART I

NOTES:

Item 13: Machine listings of library transactions are often produced daily. Quite often, the transaction listings provide audit trails of the last recording made on a specific reel and may be useful in retrieving a lost file or in determining how a file may have been inadvertently scratched. Accordingly, some installations keep some copies of these listings for as long as one year.

Item 14: Transaction slips for military classified or other sensitive records have longer retention periods. These retention periods are generally specified as a matter of agency policy or regulation.

Part II Documentation Required for Servicing Machine Readable Records

Documentation covering the organized series of descriptive documents relating to all aspects of system development and operation. These include system planning documents, ADP systems specifications, application program manuals, systems operating instructions and various management aids.

	File Designation	Consisting of	Which are	Then
1	specific data systems planning records	documents containing definition of the system including the system objectives, request for the system, authorizing directives, source data, detailed studies reflecting advantages and disadvantages of alternate solutions; equipment requirements, tangible benefits; output requirements and schedule for completion	at departmental level headquarters	Disposal not authorized by this schedule. Review after 5 years.
2			supplementary files at ADP unit level	Dispose 5 years after final action
3	system test documentation	system test specifications; test runs, machine listings of test data, test results	approved system	Dispose 1 year after discontinuance of the system
4			disapproved proposed system	Dispose 1 year after final action
6	systems design specifications	documents which contain operating procedures for implementation of a specific data system, including policies, instructions, details of computer technique, logic charts, input/output document flow data	for systems for which related magnetic tape data is authorized for blanking	Dispose at time final mag. tape records produced by system have been
			for systems for which the related magnetic tape data is not authorized for blanking	retain with the related magnetic tape

	File Designation	Consisting of	Which are	Then
7	File(s) specifications	narrative description of the source and functional characteristics of the file(s), a definition of the content of each record	for system the related magnetic tape data is authorized for blanking	Dispose of at time final mag. tape records produced by system have been blanked.
8		in terms of the relative position name, length, and type of each data element in a field (run layout) explanation of the coding system and a cross reference code manual of every code used together with all their values	for system the related magnetic tape data is not authorized for blanking	retain with the related magnetic tape
9	input specifications	detailed description of each transaction that generated some activity in the system in the form they appear at the time they enter the computer system; identification	for system the related magnetic tape data is authorized for blanking	Dispose at time final mag. tape records produced by system have been scratched
10		title, recording media, purpose, frequency, volume and source; detailed description of the contents of each input to the basic record file and a graphic illustration of each.	for system the related magnetic tape data is not authorized for blanking	retain with the related magnetic tape

	File Designation	Consisting of	Which are	Then
11	Output (Report Forms) specifications	detailed description of products of the system that are to be used outside the computer center	a listing of the outputs by sequence, name, media, purpose, frequency, volume and distribution; a detailed record description; samples of output in the form of layouts or copies, keyed to names and numbers in the output listings	dispose on termination of system by either obsolescence, update or discontinuance
12	Application program manual	documents reflecting the latest information for a general description of the function, use and methodology of the program	a description of input, files, and output; source and object code listings, flow diagrams showing the logic of the program; description of program output messages; coding information, test plan, program test, and operating instructions	dispose on termination of system by either obsolescence, update or discontinuance
13	User guides	information used in training or explaining overall system	handbooks, guides to data availability, and procedures for querying files	retain with systems specification

	File Designation	Consisting of	Which are	Then
14	System operating procedures	user oriented instructions: 1) to prepare input data; 2) for control and interpretation of output reports; and 3) for processing work on the computer	for systems the related magnetic tape data is authorized for disposi- tion	Dispose at the time magnetic tape reels are scratched.
15			for systems which require retention of related magnetic tape data.	retain with file (systems) specifications
16	Report	printed final report containing the statistical tabulation and an analysis of the findings of a study or survey including a narrative description of methodology employed	for systems which re- quire retention of related magnetic tape data.	retain 1 copy of the printed report with related file specification

Part III Erasable Media

The term erasable media refers to tape (analog, digital) drums, disks, disk packs, data cells and other devices which store data in an erasable mode. The term "dispose" in column 4 is synonymous with the terms "scratch", "erase," and "blank!"

	File designation	Consisting of	Which are	Then
1	Scratch tape (blank tape)	temporary magnetic tape used by the console operators or tape handlers to facilitate general computer runs such as sort and merge runs	not included in a tape library control or files whose retention dates have expired, or new tape	available for immediate use or reuse
2	Test tape	magnetic tape used in testing a proposed system	used by programmer for individual run testing and not under library control	Dispose after system has been accepted or discontinued, whichever is sooner
3			system debugging test data	Retain until related program is discontinued.
4			system acceptance test data	
5	Program tape or disk pack	tapes (disk packs) containing sequence of instructions required to accomplish the processing of data or solving a problem	Updated	Dispose after 3rd update cycle.
6			the last update of specific EDP application used in a terminated system	Dispose after agency has exhausted its use of the tape
7			required in Audit Trail	Dispose in accordance with functional guidelines provided by GAO.

	File designation	Consisting of	Which are	Then
8	Raw data input	magnetic tapes containing data abstracted from source documents or other media and entered into the system for the first time	used for updating with existing program and which are required to support reconstruction of master files	Dispose 1st generation data upon successful completion of 4th processing machine pass
9			not required to support reconstruction of master file and/or used as input for a one-time study or survey	Dispose after raw data is processed into final data and proven to be satisfactory
10			officially designated to replace or serve as the basic source data in lieu of the "hard copy" or other input source document	Dispose in accordance with instructions applicable to the "hard copy" or other files documenting the same process, transaction or case

	File Designation	Consisting of	Which are	Then
11	Working tape Input/Output	magnetic tape containing output or control within or from one run to a subsequent run which manipulate, sort and/or move data thru the systems. Includes checkpoint, edit, correction, reject list, unmatched data eliminating error, rerun tapes	used in an updated system	Dispose after subsequent magnetic tapes which contain the accepted detail data have been created and proven to be satisfactory
12			used in a one time study or survey	Dispose after master data tape has been proven to be satisfactory
13	Valid Transaction	magnetic tapes containing valid file of items used with a Master data tape input file for creation of Master data tape output file	partially valid transaction after all outstanding items are liquidated from current status tapes	Dispose after 3rd update cycle.
14			valid transaction after cumulative final Master tape is prepared and determined successful and there is no necessity for statistical analysis	Dispose after 3rd update cycle.
5			used in additional statistical analysis	Disposal not authorized by this schedule.

	File Designation	Consisting of	Which are	Then
16	Information retrieval system master reference	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file (including the security copy tape of data on disk packs)	a cumulative index to scientific and technical publications; bibliographic and other non record material	Dispose after 3rd update cycle.
17			an index to record material such as correspondence; legal hearings and decisions; patents, trademarks; and record copy of publications	Disposal not authorized by this schedule.
18	Federal loan and grant program master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file (initial data includes excerpts from forms placed in case files)	cumulative data of funds made available through Federally supported loan and grant programs	Dispose after 3rd update cycle.
19			non cumulative periodic file of status of Federal loan and grant activity	Disposal not authorized by this schedule.
20	<i>Administrative</i> "housekeeping" systems: master data file	magnetic media containing data for such <i>adm. systems</i> "housekeeping systems" as fiscal accountability, supply management, payroll administration	not required for General Accounting Office site audit	Dispose in accordance with instructions applicable to the hard copy or other files documenting the same process, transaction or case.
21			required for General Accounting Office site audit	Dispose in accordance with functional guidelines provided by GAO

	File Designation	Consisting of	Which are	Then
22	Economic statistics Master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative data such as status of banks and insurance institutions; production, consumption and monetary status of industry and agriculture; value of foreign commerce and other economic indicators construction of houses and buildings; motor, rail and air travel; communications including broadcasting, telephone and telegraph	Dispose after 3rd update cycle.
23			noncumulative; used to prepare reports covering a limited period of time	Disposal not authorized by this schedule.
24			noncumulative recurring periodic surveys including wholesale and consumer price indexes; annual industry; housing vacancy and other economic indicators	Disposal not authorized by this schedule.
25			noncumulative economic census taken during five year intervals	Disposal not authorized by this schedule.

	File Designation	Consisting of	Which are	Then
26	Social statistics Master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative social and demographic data such as births, deaths, and marriages; income taxes paid; social security accounts; employment information; law enforcement, crime and civil disturbance, and other social indicators	Dispose after 3rd update cycle.
27			noncumulative; used to prepare reports covering a limited period of time.	Disposal not authorized by this schedule.
28			noncumulative recurring periodic surveys including current population statistics; annual industry; housing vacancy; voter participation; statistics of income sample.	Disposal not authorized by this schedule.
			noncumulative demographic censuses	Disposal not authorized by this schedule.

	File Designation	Consisting of	Which are	Then
30	Natural Resources Master file	continuously updated magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative data on characteristics, use and ownership of natural resources such as land titles, water, mineral and timber resources	Dispose after 3rd update cycle.
31			noncumulative; used to prepare reports covering a limited period of time	Disposal not authorized by this schedule.
32	Longitudinal studies Master data file	magnetic tape containing data recorded over time from one or more sources	a series of observations relating to individual units (persons, places, things)	Disposal not authorized by this schedule.

	File Designation	Consisting of	Which are	Then
33	Scientific data files	magnetic media source data recordings received from experimental sensor instruments for scientific measurements such as outer space orbiting spacecraft, oceanographic and geophysical phenomena and medical research (including analog tape)	converted to raw data digital magnetic tape media	Disposal after meaningful data has been analyzed
34			not converted or converted only in part to raw data digital magnetic tape media	Dispose after determination has been made that the data will not be converted to raw data digital magnetic tape media
35		magnetic media containing data created either from analog magnetic tape or recorded directly on magnetic digital tape for scientific measurements of astronomic, outer space, oceanographic phenomena; air and water quality, and medical research measurements	held in national data centers	Disposal not authorized by this schedule.
36			not duplicated in national data centers	Disposal not authorized by this schedule.
37			duplicated in national data centers	Dispose after determination is made that data is not required outside of the data centers
38			not calibrated or validated	Dispose after subsequent magnetic tapes containing the accepted data have been created and proven to be satisfactory

	File Designation	Consisting of	Which are	Then
39	Summary data file	magnetic tape containing aggregates of individual observations from valid transaction or master data file which are disaggregates of published data	substantially unpublished such as tapes containing data that are disclosure free	Disposal not authorized by this schedule.
40	Publication tape	magnetic tape containing source output data extracted from the system (without destroying the source tapes)	reproduced and disseminated as a publication or used for reproducing a printed publication	record copy not authorized for disposal by this schedule.
41	Print tape		used for producing required printouts of tabulations, ledgers, tables, registers and reports	Dispose after output has been released and approved
42	Reformatted data file	magnetic tape containing essentially duplicate data from the master data file but which is created for use with other computer hardware systems	created for the specific purpose of information interchange	Dispose as provided for master data tape
43			of specific application for agency computer hardware systems	Dispose when determination is made that such format is unnecessary
44	Sample and sub-sample data files	magnetic tape containing individual observations selected from a larger census or survey file such as stratified or pure random sample files with or without weighting factors	disclosure free or useful in statistical analysis or policy formulation models and simulation studies	Disposal not authorized by this schedule.

	File Designation	Consisting of	Which are	Then
45	Security back up file	magnetic tape which is identical in format to master tape retained as security in case master tape is damaged or inadvertently erased	updated	Dispose after 3rd update cycle.
46			a one time study or survey	Dispose or retain in accordance with standards for scratching of corresponding master file
47	Other agency files	magnetic tape created by other agencies	not altered substantially by the receiving agency	Dispose when no longer needed.

PART III

NOTES:

Items 3 and 4. This type of data is differentiated from simple debugging test data in that the data set is used to exercise all possible data system options within the complete set of programs. System debugging test data means data used to debug individual programs or groups of programs prior to final acceptance testing. It must be retained until the related program is discontinued.

Acceptance test data may also be a contractually defined specification or item in software systems being procured and it or a listing of it may have to be kept with the contract file. For details in this case, see General Records Schedule 3, item 4.

In other cases, particularly in systems where accounting for funds is involved, the files may be required to be kept until a particular version of a system has been audited and approved by the General Accounting Office. Retention periods in this case will be in accordance with the specific functional file in one of the other General Records Schedules. This means that specific acceptance test data sets might have to be kept for the life of the particular version of a software system or until all records produced under that system have been disposed of.

Item 7. Just as the acceptance test data may be required to be kept beyond its useful life for auditing purposes, programs which processed that data may also be kept for audit purposes beyond the operational life of the particular system. Disk packs are relatively expensive for long term storage and there is usually a back-up copy of the system on magnetic tape. In these cases, the tape copy of the program together with all relevant documentation may be used in lieu of the disk pack version. Either source or object versions of the system may be used for this purpose.

Items 16, 18, 22, 26, and 30. "Cumulative date" implies no earlier data is deleted in present pass.

Part IV. Nonerasable Media

Nonerasable media refers to ADP punched cards, paper tape, and other nonerasable, machine-readable media.

	File Designation	Consisting of	Which are	Then
1	ADP program card files	punched cards containing common language source program data (source deck)	processed with a processor or utility program to produce a machine coded object program	Dispose of individual cards when replaced ^d by new ones.. Dispose of program deck after program has been removed from system. See Note Part. III Item 4, Item 7
2		machine punched cards containing coded machine language instructions arranged in proper sequence (object deck)	read into computer memory before running a program to cause the computer to perform data processing functions	Dispose after successful completion of a program revision or after related program has been removed from system See Note Part III Item 7
3		prepunched utility or processor program card decks	Used to update installation systems software.	Dispose after receipt and successful use of new cards from the manufacturer or programmer, or 1 year after discontinuance of program or system
4		job stream (job stack, job control) card decks	used to activate program processing modules performing a data processing job	Dispose of individual cards or sets of cards when replaced by new cards and when necessary changes (if any) have been made to appropriate data processing manual

	File Designation	Consisting of	Which are	Then
5	ADP program control cards	punched cards containing data for program control generated by the producer or user	pertinent to a specific run or cycle	Dispose individual cards or sets of cards when replaced by new cards and when necessary changes (if any) have been made to appropriate data processing manual
6			for repetitive use and are updated, either by ADP or user.	Dispose of individual cards after replacement by new cards destroy control deck 1 year after program has been removed from system, or after system has been discontinued
7	ADP source data cards (or paper tape as applicable)	punched cards or paper tape containing data abstracted from source documents and used for conversion to magnetic tape or processing on (EAM) electric accounting machine equipment created after January 1, 1970	retained by ADP operational elements as backup to magnetic tape or disk	Dispose when related magnetic file has been proven to be satisfactory and has grandfather back up
8			EAM output listings and reports	dispose after 180 days if used in processing without being converted to magnetic tape
9			on magnetic tape	Dispose after verification of data on related magnetic tape
10			punched cards that contain original entry data with film or written inserts.	source documents

PART IV. ADP Punched Card and Paper Tape

NOTES:

Items 5 and 6. These items refer to parameter cards associated with the execution of various options of operational programs. These things include date cards, periodic (monthly or quarterly) options executed only occasionally and queries to information retrieval systems. They do not include card decks for generalized interpreter systems used with computer simulation software packages such as SIMSCRIPT, GPSS, DYNAMO, and similar systems. These decks have the status of program source decks. Similarly, all except report generation decks in file management systems are considered to be source program decks and should be retained or destroyed in accordance with the criteria of Items 5 through 7 of Part III.

PART V. PROCEDURAL ANALYSIS OF DATA PROCESSING SYSTEMS -- GUIDELINES
FOR APPRAISING FILES AND DATA SETS FOR PERMANENT RETENTION

1. Introduction

This section is a guide to ADP systems analysts, records officers and archivists for determining the nature of data files (also called data sets) generated by computers. Factors which influence the selection of specific data files for permanent retention in machine readable form (chiefly on magnetic tape) are indicated and explained here.

In examining a variety of documentation files for different ADP systems, substantial differences were found in the usage of technical terms between agencies, and in some cases, within agencies. These differences are being resolved by several vocabulary standardization groups, among them Federal Information Processing (FIP) task Group 5 and its successors and the American National Standards Institute (ANSI) X.3.5 Committee on Vocabulary. However, the definitions in the vocabulary have not been standardized to the extent that flow chart symbols have been in ANSI Standard X3.12-1968, Flowchart Symbols and Their Usage in Information Processing.

Accordingly, better guidance for appraising data and documentation files can usually be achieved by studying the high-level system flow charts in addition to the narrative description found in the system documentation files. The system files are enumerated and described in Part II of the schedule. This section has been written based on the fact that virtually all Automatic Data Processing systems are composed of a small number of basic procedure types connected in sequences that can be called modules. The text and charts in the following sections are organized around this concept. Almost all existing ADP systems can be analyzed into portions or groupings of these charts.

2. The elements of Data Processing Systems

Data processing systems are composed of four basic classes: hardware, software, peopleware, and data files. The hardware consists of the central processing unit and all of its peripheral devices and recording media. The software consists of the machine instructions which direct the hardware to perform the processing. Peopleware is listed in Parts I and II and consists of specifications, hard-copy documentation and user manuals for all personnel involved in running a system. The data files themselves are listed and described in Part III. Appraisal criteria for them will constitute the bulk of sections 3 and 4, below.

2.1 Hardware

Computer hardware and recording media are still undergoing relatively rapid evolution and this presents a problem in attempting to find equipment which can successfully read some older machine readable file. Files to be retained permanently may have to be recopied periodically onto newer media or totally converted in format and most other physical characteristics. Since costs for this type of work are declining, this situation presents no undue burden to the holder of this data. In general, the property value and conversion costs of machine readable records are less than one-tenth of one percent of the data collection and editing costs of the information recorded on it. Upon consultation, the Office of the National Archives, National Archives and Records Service will recommend procedures and techniques needed for the physical preservation of the record content beyond the life of the recording medium.

2.2 Software

This is divided into two main types, systems and application software. System software is furnished by the computer manufacturer and is designed primarily to manage the available resources of the computer complex in an efficient manner. The computer complex consists of the central processor and its attached peripheral devices such as card readers, magnetic tape drives, high speed printers and others. In general, this type of software is not related to any specific file or record maintained in an installation. It is, therefore, of no permanent value except as evidence of history of the development of computer science. Selected portions of systems software specifications are useful for reading files produced on one computer with another equipment configuration. However, this information may be documented in less than one page and does not require extensive documentation. Subclasses of system software include utility, operating system, sorts, merge and compiler software.

An exception is application software written in one of the standardized machine independent programming languages. COBOL, FORTRAN, and PL-1 are the three most widely used such languages. In most cases, application software written in these languages may be considered for retention with the related files. However, only a small portion of the total software written for an application need be retained permanently. For example, a file that has been closed off and covers a specific period of time will not be updated. Therefore, the update software is unlikely to be ever required again and is disposable.

If the file is a large complex data base designed to service many inquiries, retention of the query software may be warranted. However, much good query software is commercially available to handle the problems of file inquiry. Therefore, retention of this software is less important than retention of user documentation described in section 2.3, below.

The final class of software which may have permanent value is that used in computer simulation work. There are several software systems which have been used in policy formulation and evaluation work for high level management in agencies. The three best known such software systems are SIMSCRIPT, DYNAMO and GPSS (General Purpose Systems Simulator. Like COBOL and FORTRAN, these systems are available for most computers on the market. It is also likely that they will continue to be available for the foreseeable future. What is important to save in such applications are the source programs decks. The policy alternatives and much of the information on a project is contained in these decks and they often constitute records of intrinsic value for historical purposes. Economic and financial projection models and war game software are typical examples.

2.3 People-ware

A wide variety of hard-copy documentation is produced in data processing systems. People-ware is that documentation required by the personnel involved in the design, development, operation and maintenance of ADP systems. The files are listed in parts I and II of this schedule. Of interest in this section are primarily those files required for the direct serving of files declared permanent.

The basic concept to grasp in data processing is that the record constitutes a representation of an event and not the event itself. As such, the representation or record may have been recorded by a sensor (as in scientific measurement) or may have been transcribed and encoded from some other document or document group as in all transaction reporting. In either case, a researcher needs to know what kind of transformations occurred between the actual event and its representation on magnetic tape. This knowledge is in the documentation described in part II of this schedule.

For example, most housekeeping systems usually encode events using elaborate code tables rather than narrative fields on the record. A payroll system may have dozens of deduction code possibilities as well as an equal number of pay plans. Typical codes would represent bond deductions, local tax rates for states and municipalities, bond and charity deductions, overtime and premium shift differential rates, etc. In scientific work, instrument readings represent observations of physical phenomena and other occurrences.

Each time a transaction is encoded or instrument reading is made, there is a possibility of an error or distortion taking place in the process. The errors may be simple random occurrences such as digit transposition by position by keypunch operators or transcribers or systematic because of some bias in the recording instrument or observer. In general, the scientist

attempts to calibrate his instruments and adjusts instrument readings for other known factors and the systems accountant devises consistency checks, batch totals, and clerical training programs to assure "accurate" recording of his data.

Permanent records of this class include the file and input specifications (items 7 through 10 of part II) along with the final version of the related tape file. They tell a future user of the probable quality and coverage of the file and for those with much encoding, such as accounting files, the meaning of all of the descriptive data fields along with the bias and judgement that went into transcribing a record of an event into a coded element. Some portions of system operating procedures and user guides (items 13 to 15 of part II) are also useful for later reference work. These records are essential for determining how the related data files were used for operations and research and must be retained even if the related software is disposed.

3. Data Processing Systems Flow Charts and Their Use in File Appraisal

Data records in ADP systems are processed both manually and mechanically before finally residing in a file as a correct record. This section presents typical systems charts found in the high level documentation of most such applications. These charts should be used by an appraiser for determining which files among many are most useful for permanent retention.

Data processing systems are categorized by two sets of terms. One breakdown is between continuing and one-time systems; the second is between real-time and batch processing systems. Real-time systems handle one transaction at a time and complete the function of posting and validation before going on to the next transaction. These operations occur at the time the actual real-world event occurs or at most, soon afterward. Batch processing systems perform one stage of processing for a group (or batch) of transactions. These operations occur after the real world event took place. The delays may range from hours in some cases to months in others.

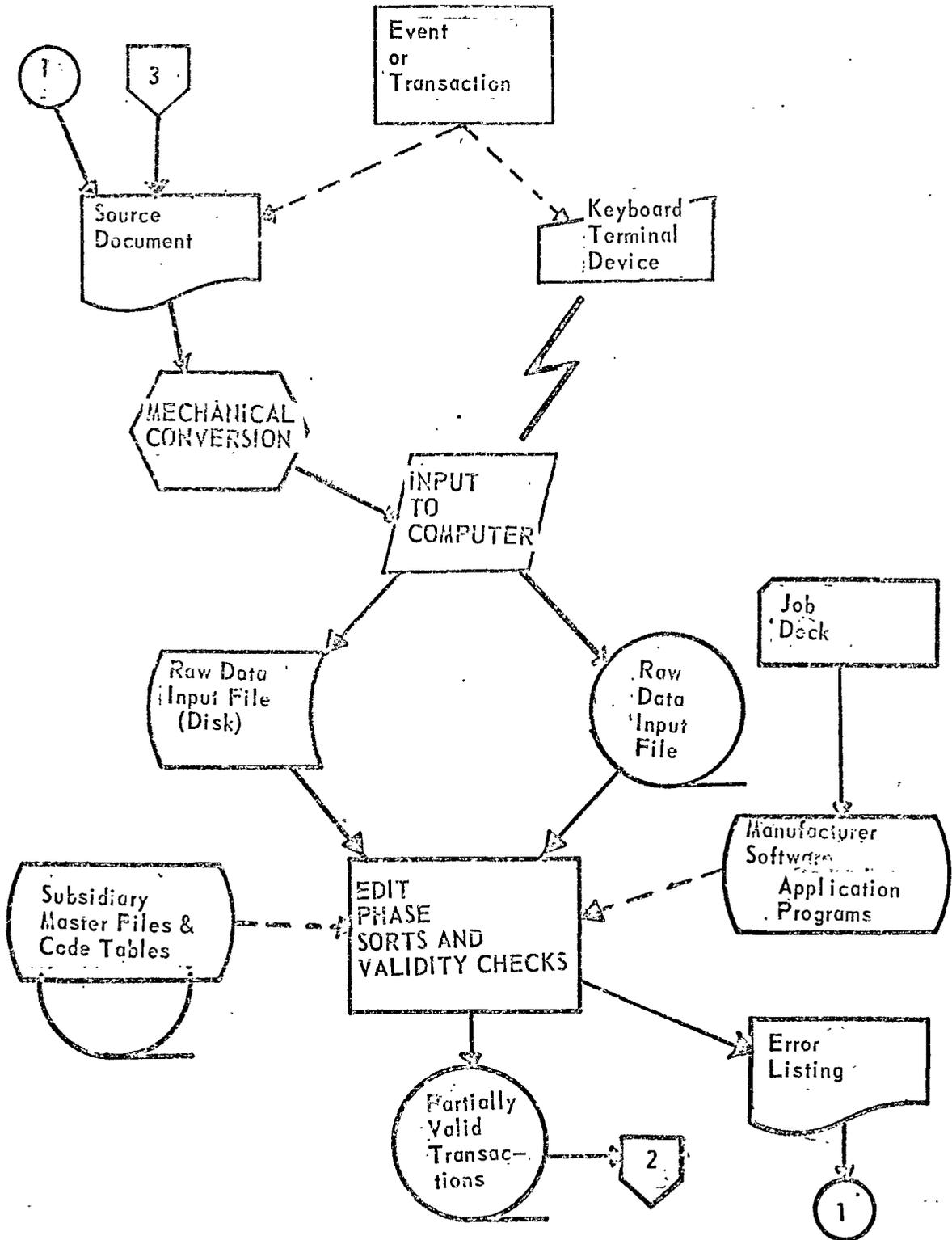
Continuing systems are those which are run periodically with a repetition rate ranging between hours and annually. Most housekeeping systems are of this type. The most familiar applications are payroll, inventory control and financial management. Although the file contents are continually changing, such systems have high continuity from one period to the next and are well documented for auditing and operational purposes. One-time systems are less well documented than continuing systems. There is usually pressure to deliver the results in the form of reports within a time constraint. Many undocumented ad-hoc decisions are made during the course of these projects to meet project deadlines resulting in files containing data errors which may not reflect the contents of published results. Surveys, simulation projects and censuses fall into this category.

TITLE(S)

GENERALIZED INPUT UPDATE (I)

TYPE OF CHART

CODE NO.



Various combinations of these attributes are found in ADP systems. Real-time systems are used in continuous housekeeping systems, the most notable being military command and control and airline reservation applications. Batch processing is used for both one-time and continuous systems and constitutes the bulk of existing data processing applications.

The files in all types of systems are increasingly being put on mass storage devices such as disks rather than tape. However, even in real-time systems, backup and recovery procedures dictate that magnetic tape copies of the file be created. These are usually called either "file dumps" or "safe data dumps." They are also created for running off summary reports since total file scanning of disk files is inefficient. These files can be appraised in the same manner as tape resident files.

3.1 Input and Update Subsystem Phases

These two phases are common to all data processing systems which involve file maintenance. Two typical flow charts are shown and labelled "Generalized Input Update." They show the processing steps taken to record, convert, check, edit and post a record to a file for later use.

3.1.1 Source Data Conversion Phase

Data can be converted to machine-readable form by several methods. Formerly, data were transcribed from source documents onto transcript sheets. They were keypunched, converted to magnetic tape and then processed. More recent methods either record machine-readable data onto source documents (turnaround documents) or accept input directly into computers through keyboard driven terminals (source data automation).

The data control function is closely interwoven with the mechanical conversion process. One part of data control consists of keeping count of the documents in each batch to be processed and control totals of one or more quantity fields. Examples would be dollar totals and counts of checks or invoices. The other part of data control is manually editing source documents. Such editing consists of checking codes and resolving errors.

The typical sequence is shown on the chart entitled "Generalized Input Update I." A manual handling phase is followed by a media conversion step. The machine-readable transaction is further validated by a series of computer runs. Errors may be introduced and detected during each stage of the process. Correction procedures depend on the stage at which the error is detected, the type of error, and the conversion hardware used. For example, in keypunch oriented systems, verification is used to minimize conversion errors while the computer passes are used to catch logical and transcription errors. Some systems combine error

detection and correction processes by attaching the conversion equipment to a computer.

However many steps and cycles occur in the process, the end product is usually called a raw data tape. Other names used are unsorted transactions, partially edited transactions, sorted preliminary update file, etc.

Raw data tapes are seldom of any permanent value since they contain erroneous and duplicate records. A possible exception is the case of real-time systems where the tape may be named "logging file" in such applications as message switching or production control systems. Usually, such tapes are kept for a short period of time as backup to recreate a real-time file. The sole other usage is for system test data or some simple transaction counting for real-time system workload studies.

3.1.2 Edit Phase-Sorts and Validity Checks

Many tapes with records of temporary value are produced at this stage. Another common designation for these tapes is work tapes. This phase or module processes transaction files against various editing and validation criteria. These criteria may be found in a computer program, such as a table of valid transaction codes or in a subsidiary master file, such as valid account numbers or a name and address file. Other common checks performed here are for numeric characters in quantity fields, transaction batch totals, transaction counts and consistency checks.

The output of such a phase is a file of partially validated transactions. Two methods are used in handling errors at this time. In one invalid

transaction items are listed on a printer along with the error indicators for immediate correction on a batch basis. In other systems, erroneous records are coded to indicate the presence of certain errors, but are not deleted from the transaction file. Instead, they are kept in the transaction file for still further checks in the update program itself. This gives a consolidated error listing at one time for a given batch of transactions. The most common additional tests performed on the data would be tests against the master file key itself. Examples are transactions which attempt to delete a nonexistent record or insert a duplicate record into a file; others may be quantities which are checked for "reasonableness".

3.1.3 Update Master File or Data Base Phase

The edit phase's output is the updating run's input. This is shown in "Generalized Input Update II." If the file contains only one application, such as accounts payable or receivable, it is normally called a "master file." If it contains data from a series of applications, or summary data from a variety of sources, then it is a "data base." Under certain circumstances, individual transaction and status report is the direct concern of an organization's top management. Examples are status reports for an important research and development project, construction job, or loan -- all of which might have enduring value.

There are several types of master files. In some systems, they include only currently active records. Purged master file records are periodically transferred to dormant account (history) files. Personnel and payroll files are typical with the purge generally occurring at year's end. Many other files are cumulative and continue to grow in size depending upon the application. As in the case of paper records, the frequency of use is the major criterion when deciding upon the length of time "detailed transactions" will remain on the master file.

In the case of periodically updated files, where transactions are deleted, purged records are often merged into historical files. These files are valuable. However, they may lack data found on the master record file. Therefore, both the merged periodic transaction file and the master file should be retained. Items 13 through 15 of Part III denote the types of transaction files created. Items 16 through 32 of Part III furnish disposition criteria for master files.

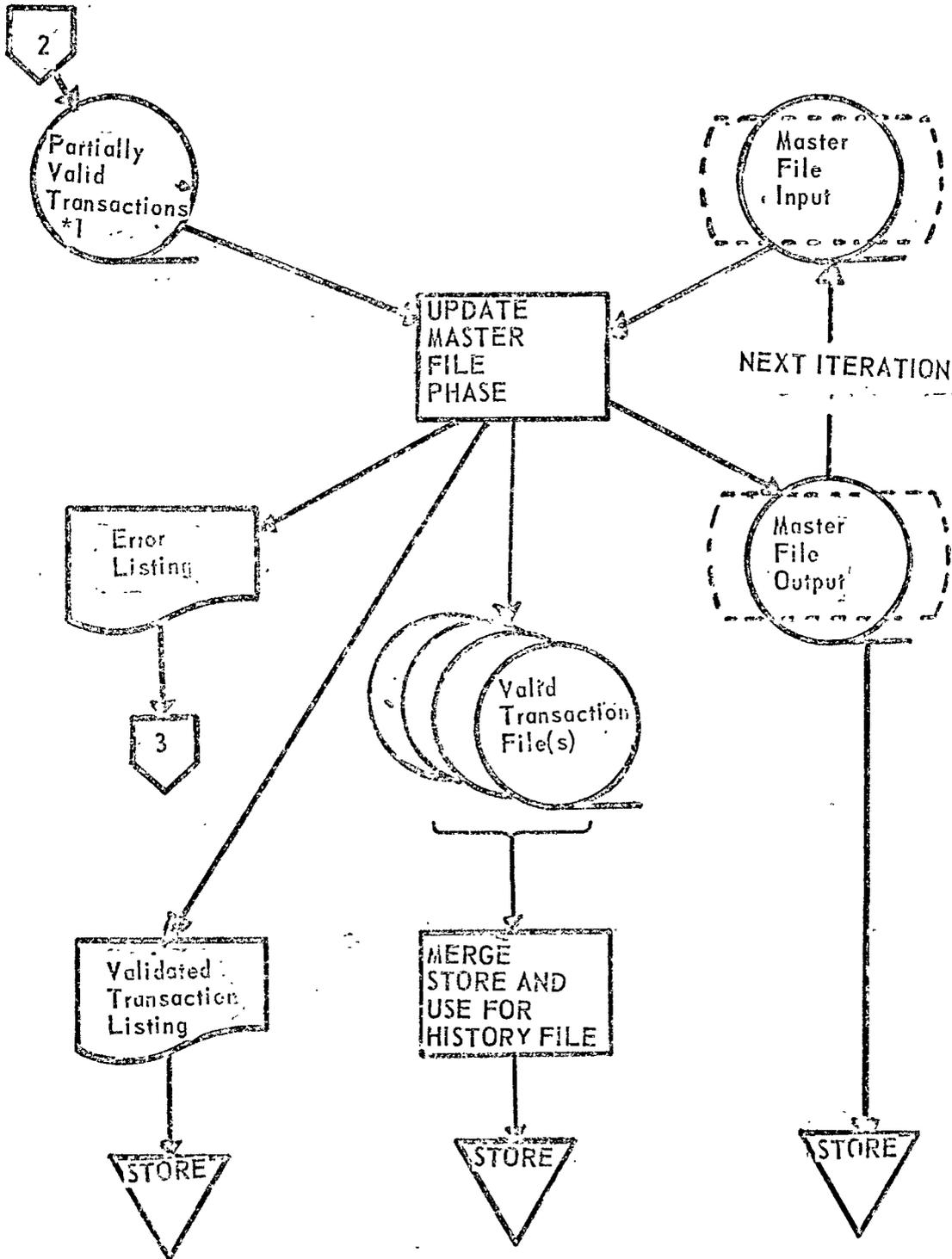
Master files are seldom updated for a given period in a single update pass. Some errors cannot be detected until the actual posting attempt is made. This creates a series of "interim" master files. The only valid file would be the one from which the periodic output was run. Usually, processing deadlines determine which version is "valid." Interim master files are usually retained for short periods as backup tapes for the final master file. (See items 44 and 45 of Part III). This retention plan is called the "grandfather system." In appraising master files for permanent retention, it is preferable to retain the

TITLE(S)

GENERALIZED INPUT UPDATE (II)

TYPE OF CHART

CODE NO.



*1 or work tape

*3 May be cumulative or noncumulative file.

*2 A series of these files may be merged into a continuous history file.

*4 Until final version is approved, all prior versions are interim master files.

"as-of" date from the official files. While this is usually possible, there are many cycle billing systems in which the master file is never completely purged of detail transactions -- thus, never complete. In such cases, it may be preferable to retain extracts of the master file made for reporting purposes and audit trails rather than the master file itself. Items 39 through 43 of Part III describe alternative selections to master record tapes.

3.2 Report Generation Phase

This section describes the files, processing and software used to produce output from ADP systems. The chart labeled "Report Generation Phase" shows the usual processing sequence in such modules from the machine-readable record to the final printed report or listing. Since mass, random-access storage is increasingly used, the chart shows tape and disk files used interchangeably, although in practice, one or the other medium will dominate.

As indicated previously, unless its usefulness or transferability to future computing systems is assured, it is unnecessary to retain output-oriented software. This evaluation should be ad hoc.

3.2.1 Report Data Extract and Format Phase

If the printed report and the master file are in identical sequence, the data selection, tabulation, and printing phase may occur in one program. This is characteristic of billing, of payroll, and of most housekeeping systems. However, often it is necessary to print the report in a sequence different from that of the master file. The use of a sort will resequence the selected records as desired, creating a series of intermediate work files between the master file and the printed output. The flowchart labeled "Report Generation Phase" shows both tapes and disk files in the processing sequence.

Newer computers and operating systems seldom produce work tapes except for the largest multi-reel files. The intermediate files reside on disk as transient files within the "job stream." (See item 4 Part IV). The input file, the job control deck and the final printed output only are visible to the uninformed. Thus, there is seldom need to retain intermediate files, because they can always be recreated from the master file.

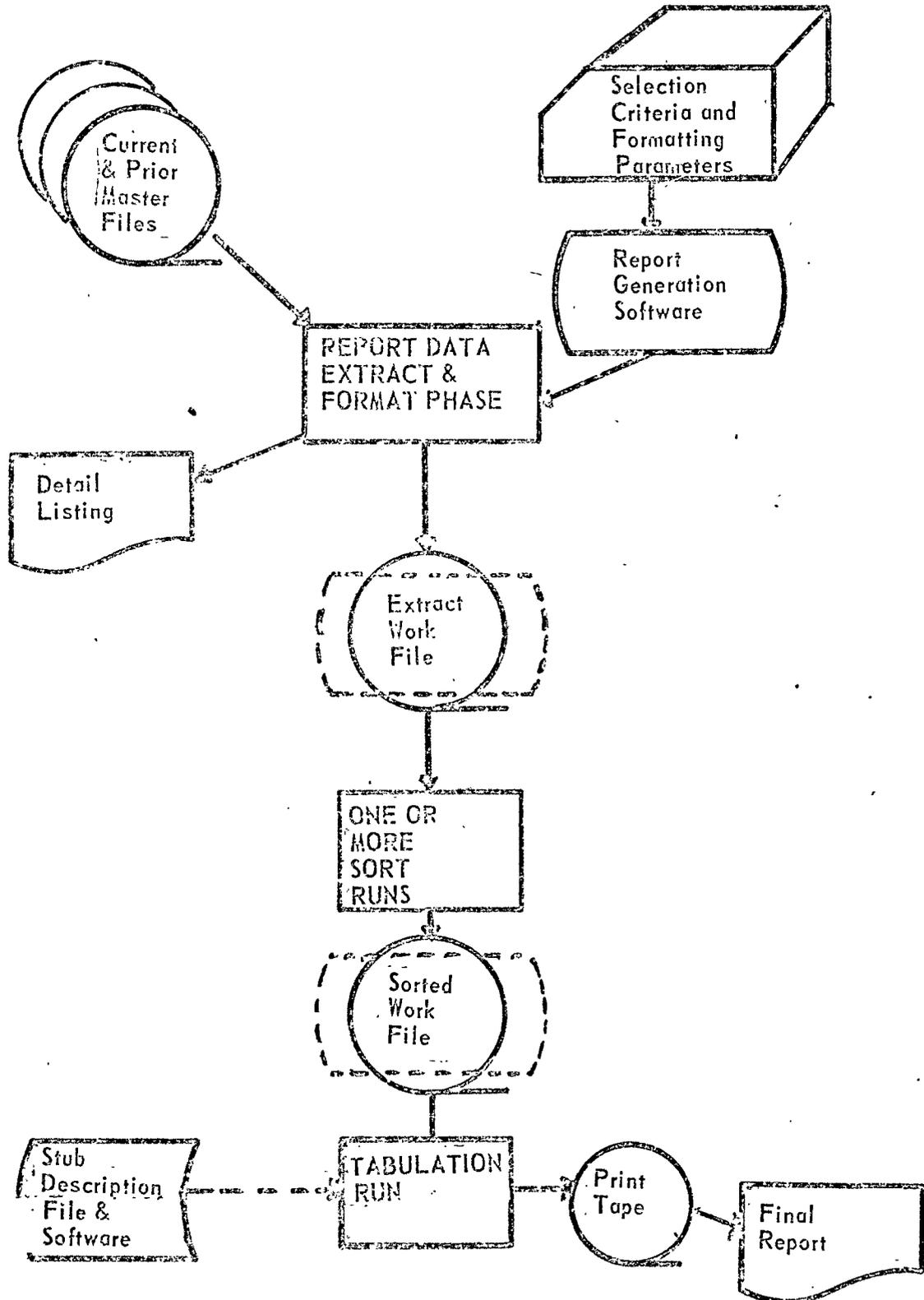
In one case, extract files are useful and should be retained. Files that contain "statistical samples of the entire body of data often have long term value when the methodology is documented. These sample files, along with appropriate weighting factors and stripped of identifying information disclosing individual persons or establishments are immediately releasable to public researchers.

TITLE(S)

REPORT GENERATION PHASE

TYPE OF CHART

CODE NO.



The next problem is to determine which of several work files to retain. In general, this depends upon the degree of decoding stub descriptors required to interpret the file. Heavily encoded files with little or no narrative descriptor are suitable provided that the stub descriptor files and tables required for human reading and interpretation are of reasonable length. When the code is a Federal Information Processing Standard, the length of the code table is unimportant. An example is the table of State and county codes of the United States with over 3,000 entries. Tables of less than 200 entries developed for individual agencies or one-time studies are reasonable to leave encoded as they can be decoded by simple computer programs. For large code tables stub descriptions are preferable for long-term preservation.

3.2.2 One or More Sorting Runs

Extract files are often in the wrong sequence for producing reports or tabulations required. In fact, the same file may be sorted into as many as ten different sequences for different types of analysis and tabulation. The criteria for retention of sorted work tapes are the same as for extract tapes described above. The output of this phase is a sorted work tape or file ready for tabulation, summarization and editing.

This type of file on tape is often a useful research file, particularly if there has been some editing and interpreting has been performed. In general, continuing administrative systems have relatively few processing steps between the first extract run and the final output pass. This is different from one-time reports as described in Section 4, below.

3.2.3 Tabulation Run

The inputs to the tabulation run are the sorted work files, and usually with a stub descriptor file. The stub descriptor file is invariably used when a very large list of codes must be displayed in plain text. If it is on magnetic tape, one or more sorting runs are typically required of the extract tape in order to apply the descriptors. If on disk, most of this decoding can be performed during the tabulation run. When files are considered for retention, the information necessary to decode such data elements must be retained. This may either be a hard copy document as described in Section 2.3 above or a machine-readable file.

The final output of a tabulation run may consist of either summary data files or a print tape (Items 39 and 41 of Part III). Summary data files may also serve as publication tapes (Item 40 of Part III) when they are reproduced and disseminated to the public and/or other Federal agencies. Summary data files are occasionally used as input for published and widely-disseminated printed reports.

Many installations do not use print tapes when they produce a computer listing or report. Instead, the data are temporarily transcribed on disks until the printed report is complete. However, tapes can be created upon specific request if there is a known demand and further use for the same information in machine-readable form. This procedure is often followed by producers of general proposal statistics.

When a tape file may be classified under more than one of these three categories, disposal is not authorized by this schedule.

4. One-Time Surveys and Report Generation Systems

The sequence of operations in one-time surveys, censuses and tabulations is shown in the following two charts. When the flow process charts are compared to the typical continuous file maintenance system, the similarities are evident. The basic difference between continuously running systems and one-time jobs is the much higher amount of manual editing and encoding required. Unless the job is a very large effort with many thousands of observations, the forms used allow somewhat more variability in field entries than accounting type documents. Since line respondents to these surveys rarely have opportunity to correct the inputs, much more manual editing and encoding is required to correct (clean up) a file prior to its use in tabulations.

As the second sheet shows, there is a file buildup process which occurs with no changes occurring to individual records after they have entered the file. Where a multiplicity of systems and sources feed the file, the individual records are usually of variable length to minimize storage requirements.

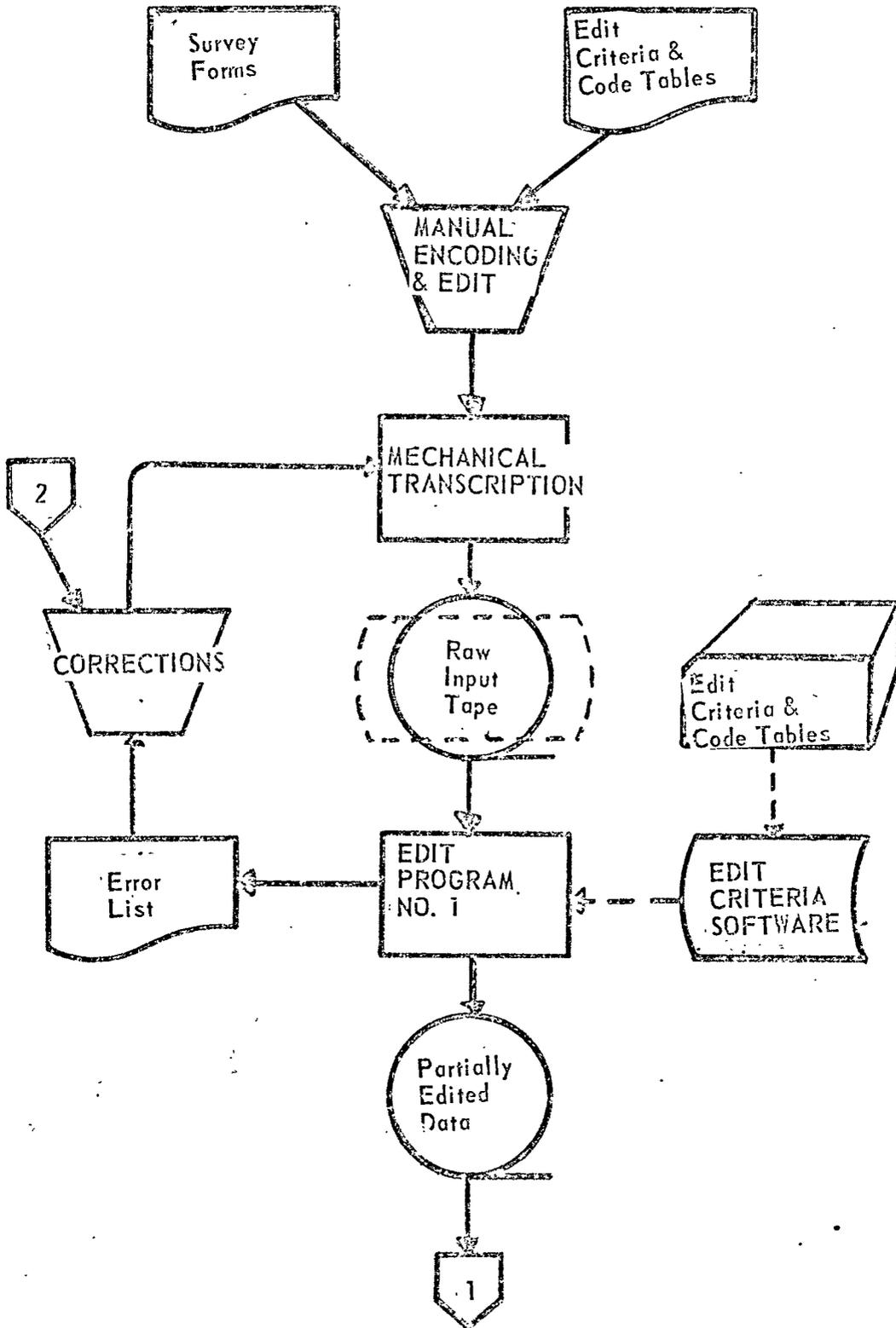
Documentation for such systems contains complex record formats but include few of the elaborate codes found in administrative systems. When these files are retained, it is important to identify the original sources of information, the instructions to respondents for filling out the forms, together with sample forms, and the directions given to the response form editors for proper interpretation and secondary usage of these files. Most such files are described in part II of this schedule.

TITLE(S)

TYPE OF CHART

CODE NO.

TYPICAL ONE-TIME PROCESSING SEQUENCE (I)



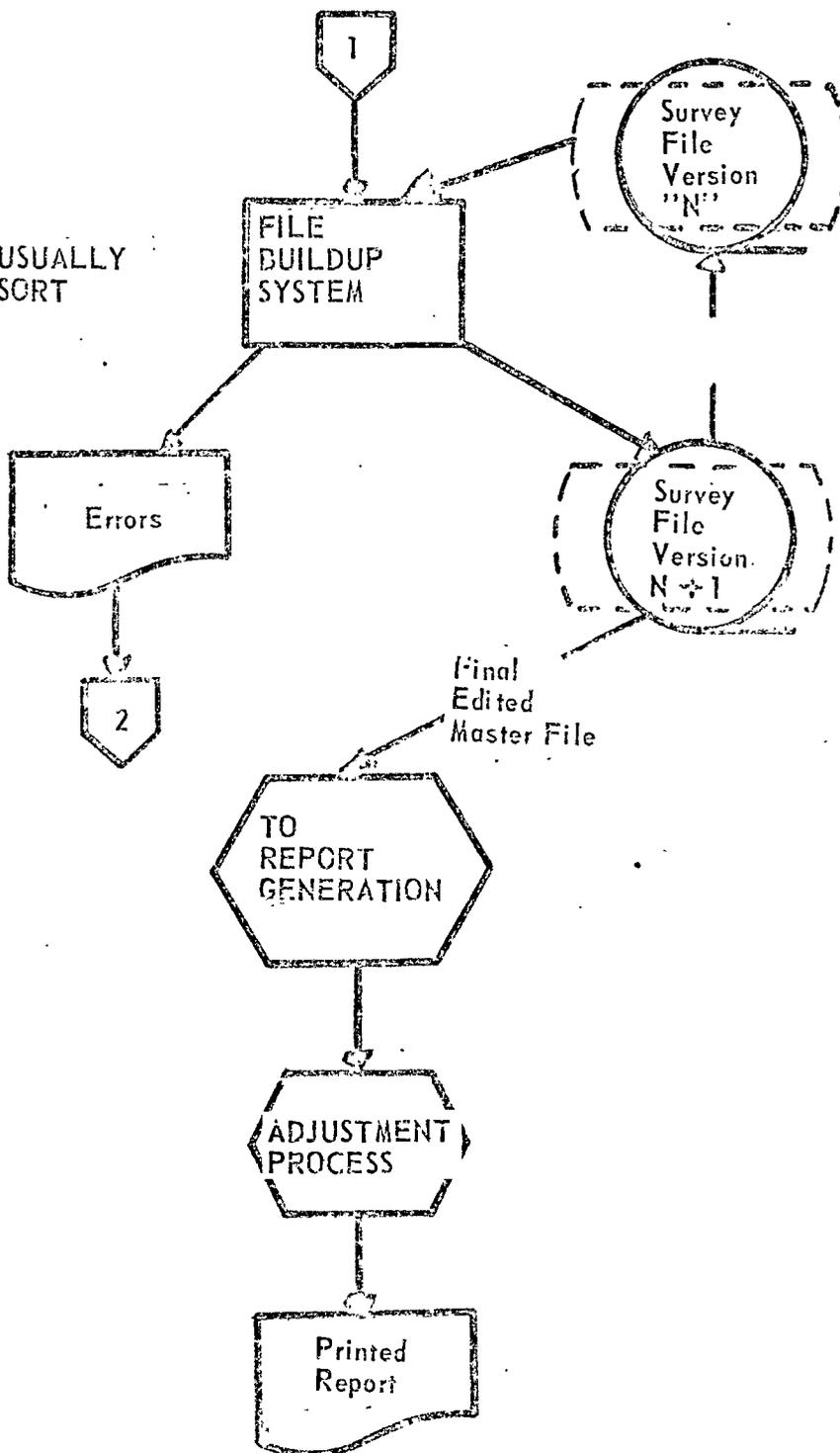
TITLE(S)

TYPE OF CHART

CODE NO.

TYPICAL ONE-TIME PROCESSING SEQUENCE (II)

NOTE; THIS USUALLY
 CONTAINS A SORT
 PROGRAM.



5. Adjustments in One-Time Jobs

A variety of additions, changes, and deletions can occur to either individual or groups of entries within a file at any stage of processing. They can occur for a variety of reasons and lead to magnetic tape files of different accessibility and validity. If the files are to be retained permanently, it is important to document the corrections and adjustments. This record of changes constitutes the equivalent of the accountant's audit trail for evaluating the accuracy of a financial file.

If a payroll record contains an error, the originating office usually hears about it in short order, particularly when an employee is short changed. In sample surveys, respondents seldom correct reporting and transcription errors unless elaborate procedures have been established for a review of the machine prepared record. This correction and review process almost always occurs in accounting systems while cost usually precludes this process in most one-time jobs.

Therefore, such files of recorded observations contain a variety of errors which in summary tabulations are nonsense entries. For example, male widows may appear in a tabulation. Such errors arise from a number of causes. Correction of the tabulations can be made at any step in the process between the final survey file and the printed report. The accompanying flow chart shows the points in the procedure where this is usually made.

If the error is thought to be a random event, the illogical counts are generally distributed to all other possible categories and deleted from the tabulation array. This would lead to a discrepancy between the published table and the final master file. The illogical records would remain in the file uncorrected.

Systematic final edited master errors also occur frequently in encoding and processing. In these cases, the summary file may be corrected by moving the entire nonsense count to the correct location in the table. These errors can also be corrected in the final master file using the computer.

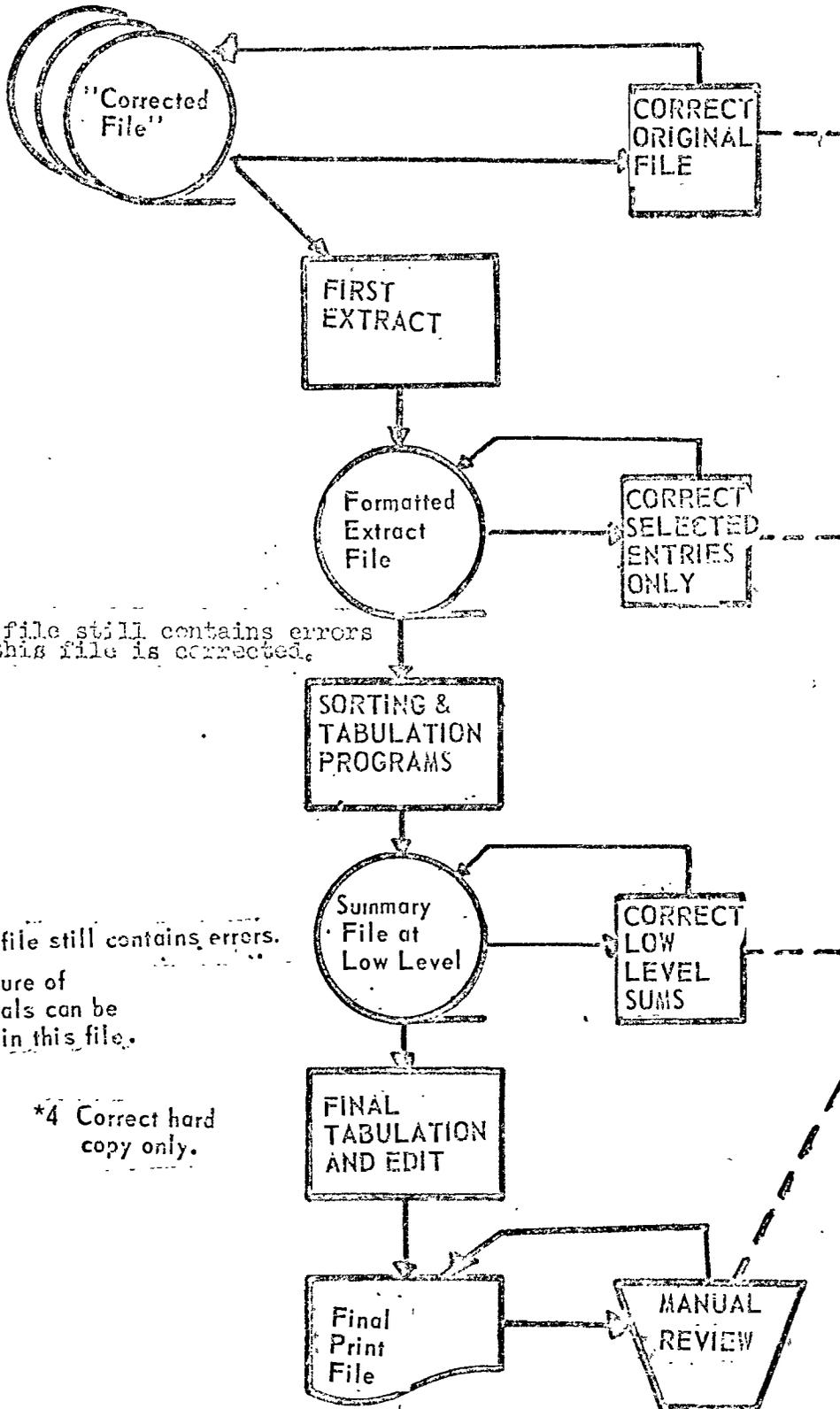
Another common adjustment operation occurs when a tabulation discloses individual confidential information. Confidentiality is protected in one of three ways: 1) by deleting the entry on the summary file and combining with enough other tabular entries to eliminate individual disclosure; 2) by correcting the print tape and 3) by correcting only the printed report. In the first method, the summary file is releasable to the public. In the second, the summary file is not releasable to the public, but the print file is.

TITLE(S)

WHERE OBSERVATIONS AND SUMMARIES
 GET ADJUSTED IN ONE TIME JOBS

TYPE OF CHART

CODE NO.



* Original file still contains errors if only this file is corrected.

*2 Extract file still contains errors.

*3 Disclosure of individuals can be deleted in this file.

*4 Correct hard copy only.

~~SECRET~~
BY A D Bennett
DRAFT JULY 1971

GENERAL RECORDS SCHEDULE 20

DATA AUTOMATION PROGRAM RECORDS

This schedule covers machine readable records recorded on erasable media, related documentation required for their servicing, and files related to the ADP operations, procurement and management function.

The principal machine readable and supporting records that are common to several or all agencies have been divided into four categories. These roughly correspond to the typical organization and functional structure found in most ADP installations and their parent organizations.

In the development of this schedule, the decision was made to use the decision table format rather than the columnar format used in the other General Records Schedules. ~~It was based on two reasons. These are~~ 1.) footnote requirements are greatly reduced with this format as compared with the first nineteen schedules; and 2.) the number of times a given file of logical records has been processed is often more important than the name assigned to it. For example, in an update system, an interim master file becomes a final master file after the sponsor declares it error free. The only difference between these two files would be the version or generation number on the file label record.

Data automation planning and operational records are normally created during the life cycle of individual computer installations. They deal with planning, managing, procurement, selection, utilization and accountability for the physical facility investment in ADP systems and supporting activities.

Documentation required for servicing machine readable records covers the organized series of descriptive documents required to initiate, develop, operate and maintain specific applications on ADP systems. These include project documentation, system specifications, test data and procedures, file and user documentation and the various installation procedures and standards used in daily operations.

The term magnetic media refers to all devices which store data in an erasable mode. At present, only magnetic media are commonly used for such purposes. However, other technologies may in the future have the same characteristics now present on magnetic tape. These are nonvolatility combined with the characteristic of easy reusability.

Since magnetic media can be overwritten, a variety of protective devices and techniques have been developed over the years to insure that inadvertent erasure of records takes place. The earliest technique, is still in use, and consists of a write protection ring inserted or left out of a reel of tape. This is a mechanical interlock device and is now a standard

PART I

NOTES:

replaced by blank

Item 13: Machine listings of library transactions are often produced daily. Quite often, the transaction listings provide audit trails of the last recording made on a specific reel and may be useful in retrieving a lost file or in determining how a file may have been inadvertently scratched. Accordingly, some installations keep some copies of these listings for as long as one year.

Item 14: Transaction slips for military classified or other sensitive records have longer retention periods. These retention periods are generally specified as a matter of agency policy or regulation.

Part III, column 4

Nos. 15, 17, 19, 23, 27, etc: "disposition by individual agency authorization," Why not, since this is ~~permitted~~ permissive, put in a suggested disposal date, such as 5 yrs, 10 yrs, etc, as is done in Parts I & II?

DRAFT JULY 1971

DATA AUTOMATION PROGRAM RECORDS - GENERAL RECORDS SCHEDULE NO. 20

Part I Data Automation Planning and Operational Records

Covering documentation relating to objectives, concepts, policies, and plans providing overall aspects of data automation data needs and systems design of management supporting systems and operational supporting systems, including equipment selection and statistics.

	File Designation	Consisting of	Which are	Then
1	Planning documents	master plan, feasibility studies with associated charts and diagrams, supporting data that reflect on the characteristics of the data automation activity	graphic, narrative and tabular information relating to the present and/or planned ADP composition and requirements of the data automation activity	disposal not authorized
2	Program management	development of plans, policy, and procedures governing the conversion to electrical machine operations and the supervision, control, coordination, and operation of the mechanization program	maintained at policy determination level	disposal not authorized
3	Hardware selection	agency requirements, specifications for hardware, software, and support capabilities of vendors of complete installations or of major peripheral equipment	selection criteria for procurements in the establishment or modification of an ADP installation	dispose 2 years after specific configuration of equipment is discontinued
4	Standardization	data elements and codes, standardization requests and justification for all data systems	promulgated Federal or national (except record copies at National Bureau of Standards)	dispose when superseded or obsolete
5			other standards eg. developed by agency	disposal not authorized

	File Designation	Consisting of	Which are	Then
6	Utilization and Maintenance	forms or cards which equipment operators complete relative to machine use, nonuse, or maintenance	used for daily management of operations	dispose after 1 year
7		daily detail cards, intermediate summary decks, related magnetic tape files and machine listings	used for daily management of operations	dispose after 90 days
8		monthly summary of cost and utilization reports	card decks, magnetic tape files and machine listings	dispose after 1 year
9	Accountability	documents concerning the management of ADPE equipment	original records maintained at data processing installation	dispose 2 years following the date equipment is discontinued
10		requirements for cards, paper and magnetic tape reels and inventory of ADPE supplies		dispose after 1 year
11		contractor's invoices for rental and other charges incurred for use of ADPE		dispose after 3 years
12	Magnetic tape library control records	library transaction records	card decks and magnetic tape files	dispose when the 4th update cycle is created
13			machine listings	dispose after 90 days
14			transaction slips	dispose after 90 days or when no longer needed

Part II Documentation Required for Servicing Machine Readable Records

Documentation covering the organized series of descriptive documents relating to all aspects of system development and operation. These include system planning documents, ADP systems specifications, application program manuals, systems operating instructions and various management aids.

	File Designation	Consisting of	Which are	Then
1	specific data systems planning records	documents containing definition of the system including the system objectives, system establishment request, authorizing directives,	at departmental level headquarters	Disposal not authorized, review after 5 years
2	showing	source data, detailed studies reflecting advantages and disadvantages of alternate solutions; equipment requirements, tangible benefits; output requirements and schedule for completion	supplementary files at ADP unit level	destroy 5 years after final action
3	system test documentation	system test specifications; test runs, machine listings of test data, test results	approved system	destroy 1 year after discontinuance of the system
4			disapproved proposed system	destroy 1 year after final action
5	systems design specifications	documents ^{that} contain operating procedures for implementation of a specific data system, including policies, instructions, details of computer technique, logic charts,	for systems for which related magnetic tape data is authorized for blanking	dispose of at the time magnetic tape reels are <u>blanked</u>
6		input/output document flow data	for systems for which the related magnetic tape data is not authorized for <u>blanking</u>	retain with the related magnetic tape

Erasing?

Erased

	File Designation	Consisting of	Which are	Then
7	File(s) specifications	narrative description of the source and functional characteristics of the file(s), a definition of the content of each record in terms of the relative position name, length, and type of each data element in a field (run layout) explanation of the coding system and a cross reference code manual of every code used together with all their values	for system the related magnetic tape data is authorized for blanking	dispose of when all the related magnetic tape reels are blanked
8			for system the related magnetic tape data is not authorized for blanking	retain with the related magnetic tape
9	input specifications	detailed description of each transaction that generated some activity in the system in the form they appear at the time they enter the computer system; identification title, recording media, purpose, frequency, volume and source; detailed description of the contents of each input to the basic record file and a graphic illustration of each.	for system the related magnetic tape data is authorized for blanking	dispose of when all the related magnetic tape reels are blanked
10			for system the related magnetic tape data is not authorized for blanking	retain with the related magnetic tape

	File Designation	Consisting of	Which are	Then
11	Output (Report Forms) specifications	detailed description of products of the system that are to be used outside the computer center	a listing of the outputs by sequence, name, media, purpose, frequency, volume and distribution; a detailed record description; samples of output in the form of layouts or copies, keyed to names numbers in the output listings	dispose on termination of system by either obsolescence, update or discontinuance
12	Application program manual	documents reflecting the latest information for a general description of the function, use and methodology of the program	a description of input, files, and output; source and object ^{code} listings, flow diagrams showing the logic of the program; description of instructive output messages; coding information; test plan; and program test and operating instructions	dispose on termination of system by either obsolescence, update or discontinuance
13	User guides	information used in training or explaining overall system	handbooks, guides to data availability procedures for querying files	retain with systems specification

	File Designation	Consisting of	Which are	Then
14	System operating procedures	user oriented instructions to prepare input data and for control of output reports and interpretation, and for processing work for the computer	for systems the related magnetic tape data is authorized for disposition	Dispose at the time magnetic tape reels are blanked <i>Draw?</i>
15			for systems the related magnetic tape data is to be retained	retain with file (systems) specifications, <i>Items 7-8</i> ?
16	Report	printed final report containing the statistical tabulation and an analysis of the findings of a study or survey including a narrative description of methodology employed	for system the related magnetic tape data is to be retained	retain 1 copy of the printed report with related file specification same deposition as file specification - Item 7-8

Part III Magnetic Media

The term "magnetic media" refers to tape (analog, digital), drums, disks, disk packs, data cells, and other devices which store data magnetically

ERASE ?

	File designation	Consisting of	Which are	Then
1	Erased: Scratch tape (blank tape)	temporary magnetic tape used by the console operators or tape handlers to facilitate general computer runs such as sort and merge runs	not included in a tape library control or files whose retention dates have expired, or new tape	available for immediate use or reuse
2	Test tape	magnetic tape used in testing a proposed system	used by programmer for individual run testing and not under library control	scratch after system has been accepted or discontinued, whichever is sooner
3			system debugging test data	
4			system acceptance test data	SEE NOTE BELOW.
5	Program tape or disk pack	tapes (disk packs) containing sequence of instructions required to accomplish the processing of data or solving a problem	Updated	scratch after 3rd update
6			the last update of specific EDP application used in a terminated system	scratch after agency has exhausted its use of the tape
7			required in Audit Trail	scratch in accordance with GAO guide lines for specific functional application (SEE NOTE)

	File designation	Consisting of	Which are	Then
8	Raw data input	magnetic tapes containing data abstracted from source documents or other media and entered into the system for the first time	used for general input for update with existing program and are required to support reconstruction of master file	scratch-1st generation data upon successful completion of 4th processing machine pass
9			not required to support reconstruction of master file and/or used as input for a one-time study or survey	scratch after raw data is processed into final data and proved to be satisfactory
10			officially designated to replace or serve as the basic source data in lieu of the "hard copy" or other input source document	scratch in accordance with instructions applicable to the "hard copy" or other files documenting the same process, transaction or case

	File Designation	Consisting of	Which are	Then
11	Working tape Input/Output	magnetic tape containing output or control within or from one run to a subsequent run which manipulate, sort and/or move data thru the systems. Includes checkpoint, edit, correction, reject list, unmatched data eliminating error, rerun tapes	used in an updated system	scratch after subsequent magnetic tapes which contain the accepted detail data have been created and proven to be satisfactory
12			used in a one time study or survey	scratch after master data tape has been proved satisfactory
13	Valid Transaction	magnetic tapes containing valid file of items used with a Master data tape input file for creation of Master data tape output file	partially valid transaction after all outstanding items are liquidated from current status tapes	scratch after creation of 4th cycle
14			valid transaction after cumulative final Master tape is prepared and determined successful and there is no necessity for statistical analysis	scratch after creation of 4th cycle
15			used in additional statistical analysis	disposition by individual agency authorizat ^o n *

* Standard Form 115

	File Designation	Consisting of	Which are	Then
16	Information retrieval system master reference	magnetic media ⁽¹²⁾ containing data created by the merging of prior master file with valid transaction data to create a new master file (including the security copy tape of data on disk packs)	a cumulative index to scientific and technical publications; bibliographic and other non record material	(scratch after creation of 4th cycle
17			an index to record material such as correspondence; legal hearings and decisions; patents, trademarks; and record copy of publications	disposition by individual agency authorization *
18	Federal loan and grant program master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file (initial data includes excerpts from forms placed in case files)	cumulative data of funds made available through Federally supported loan and grant programs	scratch after creation of 4th cycle
19			non cumulative periodic file of status of Federal loan and grant activity	disposition by individual agency authorization * <i>not detailed</i>
20	"Housekeeping systems: master data file	magnetic media containing data for such "Housekeeping systems" as fiscal accountability, supply management, payroll administration	not required for General Accounting Office site audit	(scratch in accordance with standards applicable to the equivalent "hard copy" records
21			required for General Accounting Office site audit	scratch in accordance with General Accounting Office requirements

* Standard Form 115

	File Designation	Consisting of	Which are	Then
22	Economic statistics Master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative data such as status of banks and insurance institutions; production, consumption and monetary status of industry and agriculture; value of foreign commerce and other economic indicators construction of houses and buildings; motor, rail and air travel; communications including broadcasting, telephone and telegraph	(scratch after creation of 4th cycle
23			noncumulative; used to prepare reports covering a limited period of time	disposition by individual agency authorization * <i>had to wait, wait</i>
24			noncumulative recurring periodic surveys including wholesale and consumer price indexes; annual industry; housing vacancy and other economic indicators	disposition not authorized
25			noncumulative economic census taken during five year intervals	disposition not authorized

* Standard Form 115

	File Designation	Consisting of	Which are	Then
26	Social statistics Master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file.	cumulative social and demographic data such as births, deaths, and marriages; income taxes paid; social security accounts; employment information; law enforcement, crime and civil disturbance, and other social indicators	scratch after creation of 4th cycle
27			noncumulative; used to prepare reports covering a limited period of time	<i>request</i> disposition by individual agency authorization *
28			noncumulative recurring periodic surveys including current population statistics; annual industry; housing vacancy; voter participation; statistics of income sample.	<i>al</i> disposition not authorized
29			noncumulative demographic censuses	<i>al</i> disposition not authorized

* Standard Form 115

	File Designation	Consisting of	Which are	Then
30	Natural Resources Master file	continuously updated magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative data on characteristics, use and ownership of natural resources such as land titles, water, mineral and timber resources	scratch after creation of 4th cycle
31			noncumulative; used to prepare reports covering a limited period of time	disposition by individual agency authorization *
32	Longitudinal studies Master data file	magnetic tape containing data recorded over time from one or more sources	a series of observations relating to individual units (persons, places, things)	disposition not authorized

* Standard Form 115

Substantiated?

	File Designation	Consisting of	Which are	Then
33	Scientific data files	magnetic media source data recordings received from experimental sensor instruments for scientific measurements such as outer space orbiting spacecraft, oceanographic and geophysical phenomena and medical research (including analog tape)	converted to raw data digital magnetic tape media	scratch after meaningful data has been analyzed
34			not converted or converted only in part to raw data digital magnetic tape media	scratch after determination has been made that the data will not be converted to raw data digital magnetic tape media
35		magnetic media containing data created either from analog magnetic tape or recorded directly on magnetic digital tape for scientific measurements of astronomic, outer space, oceanographic phenomena; air and water quality, and medical research measurements	held in national data centers	disposition not authorized
36			not duplicated in national data centers	disposition by individual agency authorization *
37			duplicated in national data centers	scratch after determination is made that data is not required outside the data centers
38			not calibrated or validated	scratch after subsequent magnetic tapes containing the accepted data have been created and proven to be satisfactory

	File Designation	Consisting of	Which are	Then
39	Summary data file	magnetic tape containing aggregates of individual observations from valid transaction or master data file which are disaggregates of published data	substantially unpublished such as tapes containing data that are disclosure free	disposition by individual agency authorization *
40	Publication tape	magnetic tape containing source output data extracted from the system (without destroying the source tapes)	reproduced and disseminated as a publication or used for reproducing a printed publication	record copy not authorized for disposal
41	Print tape		used for producing required printouts of tabulations, ledgers, tables, registers and reports	scratch after output has been released and approved
42	Reformatted data file	magnetic tape containing essentially duplicate data from the master data file but which is created for use with other computer hardware systems	created for the specific purpose of information interchange	disposition as provided for master data tape
43			of specific application for agency computer hardware systems	scratch when determination is made that such format is unnecessary

* Standard Form 115

	File Designation	Consisting of	Which are	Then
44	Security back up file	magnetic tape which is identical in format to master tape retained as security in case master tape is damaged or inadvertently erased	updated	scratch after creation of 4th cycle
45			a one time study or survey	scratch or retain in accordance with standards for scratching of corresponding master file
46	Other agency files	magnetic tape created by other agencies	not altered substantially by the receiving agency	scratch after determination is made that retention of the data is no longer necessary

PART III

NOTES:

Item 4. This type of data is differentiated from simple debugging test data in that the data set is used to exercise all possible data system options within the complete set of programs. System debugging test data means data used to debug individual programs or groups of programs prior to final acceptance testing. It does not have to be retained beyond a convenient time period.

Acceptance test data may also be a contractually defined specification or item in software systems being procured and it or a listing of it may have to be kept with the contract file. For details in this case, see General Records Schedule 3, item 4.

In other cases, particularly in systems where accounting for funds is involved, the files may be required to be kept until a particular version of a system has been audited and approved by the General Accounting Office. Retention periods in this case will be in accordance with the specific functional file in one of the other General Records Schedules. This means that specific acceptance test data sets might have to be kept for the life of the particular version of a software system or until all records produced under that system have been disposed of.

Item 7. Just as the acceptance test data may be required to be kept beyond its useful life for auditing purposes, programs which processed that data may also be kept for audit purposes beyond the operational life of the particular system. Disk packs are relatively expensive for long term storage and there is usually a back-up copy of the system on magnetic tape. In these cases, the tape copy of the program together with all relevant documentation may be used in lieu of the disk pack version. Either source or object versions of the system may be used for this purpose.

Part IV ADP Punched Card and Paper Tape

	File Designation	Consisting of	Which are	Then
1	ADP program card files	punched cards containing common language source program data (source deck)	processed with a processor or utility program to produce a machine coded object program	dispose of individual cards when replacement by new ones. destroy program deck after program has been removed from system See Note Part III Item 1) <i>view 10</i>
2		machine punched cards containing coded machine language instructions arranged in proper sequence (object deck)	read into computer memory before running a program to cause the computer to perform data processing functions	destroy after successful completion of a program revision or after related program has been removed from system See Note Part III Item 1) ?
3		prepunched utility or processor program card decks furnished by computer manufacturers (systems subroutines or supplemental programs written by agency programmers)	provided to assist in computer operation "housekeeping" functions	destroy after receipt and successful use of new cards from the manufacturer or programmer, or 1 year after discontinuance of program or system
4		job stream (job stack, job control) card decks	used to activate program processing modules performing data processing job (a) ?	destroy individual cards or sets of cards when replaced by new cards and when necessary changes (if any) have been made to appropriate data processing manual

	File Designation	Consisting of	Which are	Then
5	ADP program control cards	punched cards containing data for program control generated by the producer and/or user (See Note) 407?	pertinent to a specific run or cycle	destroy individual cards or sets of cards when replaced by new cards and when necessary changes (if any) have been made to appropriate data processing manual
6			for repetitive use and are updated by ADP and/or user	dispose of individual cards after replacement by new cards destroy control deck 1 year after program has been removed from system, or after system has been discontinued
7	ADP source data cards (or paper tape as applicable)	punched cards or paper tape containing data abstracted from source documents and used for conversion to magnetic tape or processing on (EAM) electric accounting machine equipment created after January 1, 1970	retained by ADP operational elements as backup to magnetic tape or disk	destroy when related magnetic file has been proved satisfactory and has grand-father back up 3rd copy?
8			EAM output listings and reports	dispose after 180 days if used in processing without being converted to magnetic tape
9			on tape	destroy after verification of data on related magnetic tape
10			punched cards that have source data entered directly on them; with film inserts; with written information	source documents

PART IV. ADP Punched Card and Paper Tape

NOTES:

Items 5 and 6. These items refer to parameter cards associated with the execution of various options of operational programs. These things include date cards, periodic (monthly or quarterly) options executed only occasionally and queries to information retrieval systems. They do not include card decks for generalized interpreter systems used with computer simulation software packages such as SIMSCRIPT; GPSS, DYNAMO, and similar systems. These decks have the status of program source decks. Similarly, all except report generation decks in file management systems are considered to be source program decks and should be retained or destroyed in accordance with the criteria of Items 5 through 7 of Part III.

DRAFT JULY 1971

DATA AUTOMATION PROGRAM RECORDS - GENERAL RECORDS SCHEDULE NO. 20

Part I Data Automation Planning and Operational Records

Covering documentation relating to objectives, concepts, policies, and plans providing overall aspects of data automation data needs and systems design of management supporting systems and operational supporting systems, including equipment selection and statistics.

File Designation	Consisting of	Which are	Then
1 Planning documents	master plan, feasibility studies with associated charts and diagrams, supporting data that reflect on the characteristics of the data automation activity	graphic, narrative and tabular information relating to the present and/or planned ADP composition and requirements of the data automation activity	disposal not authorized
2 Program management	development of plans, policy, and procedures governing the conversion to electrical machine operations and the supervision, control, coordination, and operation of the mechanization program	maintained at policy determination level	disposal not authorized
3 Hardware selection	agency requirements, specifications for hardware, software, and support capabilities of vendors of complete installations or of major peripheral equipment	selection criteria for procurements in the establishment or modification of an ADP installation	dispose 2 years after specific configuration of equipment is discontinued
4 Standardization	data elements and codes, standardization requests and justification for all data systems	promulgated Federal or national (except record copies at National Bureau of Standards)	dispose when superseded or obsolete
5		other standards eg. developed by agency	disposal not authorized

	File Designation	Consisting of	Which are	Then
6	Utilization and Maintenance	forms or cards which equipment operators complete relative to machine use, nonuse, or maintenance	used for daily management of operations	dispose after 3 years
7		daily detail cards, intermediate summary decks, related magnetic tape files and machine listings	used for daily management of operations	dispose after 90 days
8		monthly summary of cost and utilization reports	card decks, magnetic tape files and machine listings	dispose after 3 years
9	Accountability	documents concerning the management of ADPE equipment	original records maintained at data processing installation	dispose 2 years following the date equipment is discontinued
10		requirements for cards, paper and magnetic tape reels and inventory of ADPE supplies		dispose after 1 year
11		contractor's invoices for rental and other charges incurred for use of ADPE		dispose after 3 years
12	Magnetic tape library control records	library transaction records	card decks and magnetic tape files	dispose ^{after} when the 3rd update cycle is created
13			machine listings	dispose after 90 days
14			transaction slips	dispose after 90 days or when no longer needed (see note)

PART I

NOTES:

Item 13: Machine listings of library transactions are often produced daily. Quite often, the transaction listings provide audit trails of the last recording made on a specific reel and may be useful in retrieving a lost file or in determining how a file may have been inadvertently scratched. Accordingly, some installations keep some copies of these listings for as long as one year.

Item 14: Transaction slips for military classified or other sensitive records have longer retention periods. These retention periods are generally specified as a matter of agency policy or regulation.

Part II Documentation Required for Servicing Machine Readable Records

Documentation covering the organized series of descriptive documents relating to all aspects of system development and operation. These include system planning documents, ADP systems specifications, application program manuals, systems operating instructions and various management aids.

	File Designation	Consisting of	Which are	Then
1	specific data systems planning records	documents containing definition of the system including the system objectives, system establishment request , authorizing directives, source data, detailed studies reflecting advantages and disadvantages of alternate solutions; equipment requirements, tangible benefits; output requirements and schedule for completion	at departmental level headquarters	Disposal not authorized, review after 5 years.
2			supplementary files at ADP unit level	destroy <i>Dispose</i> 5 years after final action.
3	system test documentation	system test specifications; test runs, machine listings of test data, test results	approved system	destroy <i>Dispose</i> 1 year after discontinuance of the system
4			disapproved proposed system	destroy 1 year after final action
5	systems design specifications	documents which contain operating procedures for implementation of a specific data system, including policies, instructions, details of computer technique, logic charts, input/output document flow data	for systems for which related magnetic tape data is authorized for blanking <i>AS per tape</i>	Dispose of at time final mag. tape records produced by system have been blanked <i>scrapped</i> .
6			for systems for which the related magnetic tape data is not authorized for blanking	retain with the related magnetic tape

	File Designation	Consisting of	Which are	Then
7	File(s) specifications	narrative description of the source and functional characteristics of the file(s), a definition of the content of each record	for system the related magnetic tape data is authorized for blanking	Dispose of at time final mag. tape records produced by system have been blanked.
8		in terms of the relative position name, length, and type of each data element in a field (run layout) explanation of the coding system and a cross reference code manual of every code used together with all their values	for system the related magnetic tape data is not authorized for blanking	retain with the related magnetic tape
9	input specifications	detailed description of each transaction that generated some activity in the system in the form they appear at the time they enter the computer system; identification title, recording media, purpose, frequency, volume and source; detailed description of the contents of each input to the basic record file and a graphic illustration of each.	for system the related magnetic tape data is authorized for blanking	Dispose of at time final mag. tape records produced by system have been blanked.
10			for system the related magnetic tape data is not authorized for blanking	retain with the related magnetic tape

when affected like input in the computer system

	File Designation	Consisting of	Which are	Then
11	Output (Report Forms) specifications	detailed description of products of the system that are to be used outside the computer center	a listing of the outputs by sequence, name, media, purpose, frequency, volume and distribution; a detailed record description; samples of output in the form of layouts or copies, keyed to names and numbers in the output listings	dispose on termination of system by either obsolescence, update or discontinuance
12	Application program manual	documents reflecting the latest information for a general description of the function, use and methodology of the program	a description of input, files, and output; source and object code listings, flow diagrams showing the logic of the program; description of instructive output messages; coding information; test plan; and program test; and operating instructions	dispose on termination of system by either obsolescence, update or discontinuance
	User guides	information used in training or explaining overall system	handbooks, guides to data availability, procedures for querying files	retain with systems specification

*to the availability of
data
to access to data
and procedures for
querying files*

interpretation

	File Designation	Consisting of	Which are	Then
14	System operating procedures	user oriented instructions; to prepare input data and for control of output reports; and interpretation , and for processing work for the computer ?	for systems the related magnetic tape data is authorized for disposi- tion	Dispose at the time ⁽²⁰⁰⁾ magnetic tape reels are blanked scratched
15			for systems the related magnetic tape data is to be retained	retain with file (systems) specifications,
16	Report	printed final report containing the statistical tabulation and an analysis of the findings of a study or survey, including a narrative description of methodology employed	for systems ^{of} the related magnetic tape data is to be retained <i>which will require retention of related magnetic tape data</i>	retain 1 copy of the printed report with related file specification <i>see</i> same deposition as file specification - Items 7 & 8

Part III Magnetic Media

The term "magnetic media" refers to tape (analog, digital), drums, disks, disk packs, data cells, and other devices which store data magnetically

	File designation	Consisting of	Which are	Then
1	Scratch tape (blank tape)	temporary magnetic tape used by the console operators or tape handlers to facilitate general computer runs such as sort and merge runs	not included in a tape library control or files whose retention dates have expired, or new tape	available for immediate use or reuse
2	Test tape	magnetic tape used in testing a proposed system	used by programmer for individual run testing and not under library control	scratch after system has been accepted or discontinued, whichever is sooner
3			system debugging test data	
4			system acceptance test data	(SEE NOTE BELOW.)
5	Program tape or disk pack	tapes (disk packs) containing sequence of instructions required to accomplish the processing of data or solving a problem	Updated	scratch after 3rd update
6			the last update of specific EDP application used in a terminated system	scratch after agency has exhausted its use of the tape
7			required in Audit Trail	scratch in accordance with GAO guide lines for specific functional application (SEE NOTE)

	File designation	Consisting of	Which are -	Then
8	Raw data input	magnetic tapes containing data abstracted from source documents or other media and entered into the system for the first time	used for ^{updating} general input for update with existing program and ^{input} required to support reconstruction of master file.	scratch 1st generation data upon successful completion of 4th processing machine pass
9			not required to support reconstruction of master file and/or used as input for a one-time study or survey	scratch after raw data is processed into final data and proved to be satisfactory
10			officially designated to replace or serve as the basic source data in lieu of the "hard copy" or other input source document	scratch in accordance with instructions applicable to the "hard copy" or other files documenting the same process, transaction or case

	File Designation	Consisting of	Which are	Then
11	Working tape Input/Output	magnetic tape containing output or control within or from one run to a subsequent run which manipulate, sort and/or move data thru the systems. Includes checkpoints, edit, correction, reject list, unmatched data eliminating error, rerun tapes	used in an updated system	scratch after subsequent magnetic tapes which contain the accepted detail data have been created and proven to be satisfactory
12			used in a one time study or survey	scratch after master data tape has been proved satisfactory <i>incepted</i>
13	Valid Transaction	magnetic tapes containing valid file of items used with a Master data tape input file for creation of Master data tape output file	partially valid transaction after all outstanding items are liquidated from current status tapes	scratch after creation of 4th cycle <i>see 38</i>
14			valid transaction after cumulative final Master tape is prepared and determined successful and there is no necessity for statistical analysis	scratch after creation of 4th cycle
15			used in additional statistical analysis	disposition by individual agency authorizatn * <i>cb</i>

* Standard Form 115

	File Designation	Consisting of	Which are	Then
16	Information retrieval system master reference	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file (including the security copy tape of data on disk packs)	a cumulative index to scientific and technical publications; bibliographic and other non record material	scratch after creation of 4th cycle
17			an index to record material such as correspondence; legal hearings and decisions; patents, trademarks; and record copy of publications	disposition by individual agency authorization * <i>ch.</i>
18	Federal loan and grant program master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file (initial data includes excerpts from forms placed in case files)	cumulative data of funds made available through Federally supported loan and grant programs	scratch after creation of 4th cycle
19			non cumulative periodic file of status of Federal loan and grant activity	disposition by individual agency authorization * <i>W.</i>
20	"Housekeeping systems: master data file	magnetic media containing data for such "Housekeeping systems" as fiscal accountability, supply management, payroll administration	not required for General Accounting Office site audit	scratch in accordance with standards applicable to the equivalent "hard copy" records
21			required for General Accounting Office site audit	scratch in accordance with General Accounting Office requirements

	File Designation	Consisting of	Which are	Then
22	Economic statistics Master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative data such as status of banks and insurance institutions; production, consumption and monetary status of industry and agriculture; value of foreign commerce and other economic indicators construction of houses and buildings; motor, rail and air travel; communications including broadcasting, telephone and telegraph	scratch after creation of 4th cycle <i>freeze?</i>
23			noncumulative; used to prepare reports covering a limited period of time	disposition by individual agency authorization *
24			noncumulative recurring periodic surveys including wholesale and consumer price indexes; annual industry; housing vacancy and other economic indicators	disposition not authorized
25			noncumulative economic census taken during five year intervals	disposition not authorized

* Standard Form 115

	File Designation	Consisting of	Which are	Then
26	Social statistics Master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative social and demographic data such as ? births, deaths, and marriages; income taxes paid; social security accounts; employment information; law enforcement, crime and civil disturbance, and other social indicators	scratch after creation of 4th cycle
27			noncumulative; used to prepare reports covering a limited period of time	disposition by individual agency authorization *
28			noncumulative recurring periodic surveys including current population statistics; annual industry; housing vacancy; voter participation; statistics of income sample.	disposition not authorized
29			noncumulative demographic censuses	disposition not authorized

* Standard Form 115

	File Designation	Consisting of	Which are	Then
30	Natural Resources Master file	continuously updated magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative data on characteristics, use and ownership of natural resources such as land titles, water, mineral and timber resources	scratch after creation of 4th cycle
31			noncumulative; used to prepare reports covering a limited period of time	² disposition by individual agency authorization *
32	Longitudinal studies Master data file	magnetic tape containing data recorded over time from one or more sources	a series of observations relating to individual units (persons, places, things)	disposition not authorized

* Standard Form 115

	File Designation	Consisting of	Which are	Then
33	Scientific data files	magnetic media source data recordings received from experimental sensor instruments for scientific measurements such as outer space orbiting spacecraft, oceanographic and geophysical phenomena and medical research (including analog tape)	converted to raw data digital magnetic tape media	scratch after meaningful data has been analyzed
34			not converted or converted only in part to raw data digital magnetic tape media	scratch after determination has been made that the data will not be converted to raw data digital magnetic tape media
35		magnetic media containing data created either from analog magnetic tape or recorded directly on magnetic digital tape for scientific measurements of astronomic, outer space, oceanographic phenomena; air and water quality, and medical research measurements	held in national data centers	disposition not authorized
36			not duplicated in national data centers	disposition by individual agency authorization *
37			duplicated in national data centers	scratch after determination is made that data is not required outside of the data centers
38		not calibrated or validated	scratch after subsequent magnetic tapes containing the accepted data have been created and proven to be satisfactory	

* Standard Form 115

	File Designation	Consisting of	Which are	Then
39	Summary data file	magnetic tape containing aggregates of individual observations from valid transaction or master data file which are disaggregates of published data	substantially unpublished such as tapes containing data that are disclosure free	disposition by individual agency authorization *
40	Publication tape	magnetic tape containing source output data extracted from the system (without destroying the source tapes)	reproduced and disseminated as a publication or used for reproducing a printed publication	record copy not authorized for disposal
41	Print tape		used for producing required printouts of tabulations, ledgers, tables, registers and reports	scratch after output has been released and approved
42	Reformatted data file	magnetic tape containing essentially duplicate data from the master data file but which is created for use with other computer hardware systems	created for the specific purpose of information interchange	disposition as provided for master data tape
43			of specific application for agency computer hardware systems	scratch when determination is made that such format is unnecessary

* Standard Form 115

	File Designation	Consisting of	Which are	Then
44	Security back up file	magnetic tape which is identical in format to master tape retained as security in case master tape is damaged or inadvertently erased	updated	scratch after creation of 4th cycle
45			a one time study or survey	scratch or retain in accordance with standards for scratching of corresponding master file
46	Other agency files	magnetic tape created by other agencies	not altered substantially by the receiving agency	scratch after determination is made that retention of the data is no longer necessary

PART III

NOTES:

Items 3 and 4. This type of data is differentiated from simple debugging test data in that the data set is used to exercise all possible data system options within the complete set of programs. System debugging test data means data used to debug individual programs or groups of programs prior to final acceptance testing. It must be retained until the related program is discontinued.

Acceptance test data may also be a contractually defined specification or item in software systems being procured and it or a listing of it may have to be kept with the contract file. For details in this case, see General Records Schedule 3, item 4.

In other cases, particularly in systems where accounting for funds is involved, the files may be required to be kept until a particular version of a system has been audited and approved by the General Accounting Office. Retention periods in this case will be in accordance with the specific functional file in one of the other General Records Schedules. This means that specific acceptance test data sets might have to be kept for the life of the particular version of a software system or until all records produced under that system have been disposed of.

Item 7. Just as the acceptance test data may be required to be kept beyond its useful life for auditing purposes, programs which processed that data may also be kept for audit purposes beyond the operational life of the particular system. Disk packs are relatively expensive for long term storage and there is usually a back-up copy of the system on magnetic tape. In these cases, the tape copy of the program together with all relevant documentation may be used in lieu of the disk pack version. Either source or object versions of the system may be used for this purpose.

Part IV ADP Punched Card and Paper Tape

	File Designation	Consisting of	Which are	Then
1	ADP program card files	punched cards containing common language source program data (source deck)	processed with a processor or utility program to produce a machine coded object program	dispose of individual cards when replacement by new ones. destroy program deck after program has been removed from system See Note Part III Item 1
2		machine punched cards containing coded machine language instructions arranged in proper sequence (object deck)	read into computer memory before running a program to cause the computer to perform data processing functions	destroy after successful completion of a program revision or after related program has been removed from system See Note Part III Item 1
3		prepunched utility or processor program card decks furnished by computer manufacturers (systems subroutines or supplemental programs written by agency programmers)	provided to assist in computer operation "housekeeping" functions	destroy after receipt and successful use of new cards from the manufacturer or programmer, or 1 year after discontinuance of program or system
4		job stream (job stack, job control) card decks	used to activate program processing modules performing a data processing job	destroy individual cards or sets of cards when replaced by new cards and when necessary changes (if any) have been made to appropriate data processing manual

	File Designation	Consisting of	Which are	Then
5	ADP program control cards	punched cards containing data for program control generated by the producer and/or user (See Note)	pertinent to a specific run or cycle	destroy individual cards or sets of cards when replaced by new cards and when necessary changes (if any) have been made to appropriate data processing manual
6			for repetitive use and are updated by ADP and/or user	dispose of individual cards after replacement by new cards destroy control deck 1 year after program has been removed from system, or after system has been discontinued
7	ADP source data cards (or paper tape as applicable)	punched cards or paper tape containing data abstracted from source documents and used for conversion to magnetic tape or processing on (EAM) electric accounting machine equipment created after January 1, 1970	retained by ADP operational elements as backup to magnetic tape or disk	destroy when related magnetic file has been proved satisfactory and has grandfather back up
8			EAM output listings and reports	dispose after 180 days if used in processing without being converted to magnetic tape
9			on tape	destroy after verification of data on related magnetic tape
10		punched cards that have source data entered directly on them; with film inserts; with written information	source documents	dispose of in accordance with instructions applicable to non-card (hard copy) files documenting same transaction or case

PART IV. ADP Punched Card and Paper Tape

NOTES:

Items 5 and 6. These items refer to parameter cards associated with the execution of various options of operational programs. These ~~things~~ include date cards, periodic (monthly or quarterly) options executed only occasionally and queries to information retrieval systems. They do not include card decks for generalized interpreter systems used with computer simulation software packages such as SIMSCRIPT, GPSS, DYNAMO, and similar systems. These decks have the status of program source decks. Similarly, all except report generation decks in file management systems are considered to be source program decks and should be retained or destroyed in accordance with the criteria of Items 5 through 7 of Part III.

PART V. PROCEDURAL ANALYSIS OF DATA PROCESSING SYSTEMS -- GUIDELINES FOR APPRAISING FILES AND DATA SETS FOR PERMANENT RETENTION

X Introduction

This section is a guide to ADP systems analysts, records officers and archivists, for determining the nature of data files (also called data sets) generated by computers. Factors which influence the selection of specific data files for permanent retention in machine readable form (currently, ~~predominantly~~ *chiefly* on magnetic tape) are indicated and explained here.

In examining a variety of ^{*data*} documentation files for different ADP systems, substantial differences were found in the usage of technical terms between agencies, and in some cases, within agencies. These differences are being resolved by several vocabulary standardization groups, among them Federal Information Processing (FIP) task Group 5 and its successors and the American National Standards Institute (ANSI) X.3.5 Committee on Vocabulary. However, the definitions in the vocabulary have not been standardized to the extent that flow chart symbols have been in ANSI Standard X3.12-1968, Flowchart Symbols and Their Usage in Information Processing.

Accordingly, ~~better guidance for appraising data files can usually be achieved by studying the high-level system flow charts in addition to the narrative description found in the system documentation files.~~ *be more easily understood*
The system files are enumerated and described in Part II of the schedule. ~~This section has been written based on the fact that virtually all Automatic Data Processing systems are composed of a small number of basic procedure types connected in sequences that can be called modules.~~
The ~~text in the~~ following sections is organized around a small number of prototypical flow charts into which almost all existing ADP systems can be resolved. A small glossary is appended to this section to cover most of the commonly used terms and synonyms found in both the charts and narrative descriptions.

1.2 The Elements of Data Processing Systems

Data processing systems are composed of four basic elements. These are hardware, software, people-ware and data files. The hardware consists of the central processing unit and all of its peripheral devices and ~~recording media.~~ The software consists of the machine instructions which direct the hardware to perform the processing. People-ware is listed in Parts I and II and consists of specifications, hard-copy documentation and user manuals for all personnel involved in running a system. The data files themselves are listed and described in Part III. Appraisal criteria for them will constitute the bulk of sections 3 and 4, below.

Section 1 - the section etc.
The following sections are organized in the order of the basic flow charts for ADP systems.

A.1 Hardware

For records retention purposes, the main problem hardware presents is ~~the one of~~ compatibility and obsolescence. Computer hardware and recording media are still undergoing relatively rapid evolution and this presents a problem in attempting to find equipment which can successfully read some older machine readable files. Files to be permanently retained may have to be periodically recopied onto newer media or totally converted in format and most other physical characteristics. Since costs for this type of work are ~~still~~ declining, this situation presents no undue burden to the holder of this data. In general, the property value and conversion costs of machine readable records are less than one-tenth of one percent of the data collection and editing costs of the information recorded on it. ~~Upon consultation,~~ the Office of the National Archives, ~~of the National Archives and Records Service,~~ will ~~furnish the proper~~ *recommend* procedures and techniques needed for the physical preservation of the recording ~~medium~~ beyond its record content.

A.2 Software

Sections 5.

This is divided into two main types. The first is systems software. This is furnished by the computer manufacturer and is concerned primarily with efficient and accurate operation of the central computer and all the peripheral devices attached to it. In general, this type of software is not related to any specific record or file within an ADP system beyond specialized control characters of data elements mixed in with the record content of a file. It would, therefore, ~~almost never have any permanent value~~ *except* in its own right. Subclasses of system software which have little permanent value are ~~called~~ utility, sorts, merges and compiler software.

Application software consists of computer programs which are oriented towards the processing of data on specific files. It is ~~invariably~~ either written by employees of the installation or procured to work on specific data sets. The retention value of application software is ~~extremely variable.~~ *variable,*

In some cases, software must be kept regardless of its form for auditing purposes. If the software has been written in machine language, it is generally not executable on another computer and, therefore, ~~cannot be~~ *of temporary* permanently retained. There is a capability on some computers to emulate programs written for other computers, but this is generally a temporary expedient, useful for only a few years in particular installations.

An exception is application software written in one of the standardized machine independent programming languages. COBOL, FORTRAN, and PL-1 are the three most widely used such languages. In most cases, application software written in these languages may be considered for retention along with the files. However, only a small portion of the total software written for an application need really be retained permanently. For example, a file that has been closed off and covers a specific period of time will not be updated. Therefore, the update software is unlikely to be ever required again and can be safely destroyed.

If the file is a large complex data base designed to service many inquiries, retention of this portion of the software may be warranted. However, much good query software is available to handle the problems of file inquiry. Therefore, retention of this software is less important than retention of user documentation described in section 2.3, below.

The final class of software which may have permanent value is that used in computer simulation work. There are several software systems which have been used in policy formulation and evaluation work for high level management in agencies. The three best known such software systems are SIMSCRIPT, DYNAMO and GPSS (General Purpose Systems Simulator. Like COBOL and FORTRAN, these systems are available for most computers on the market. ~~It is also likely that they will continue to be available for the foreseeable future. What is important to save in such applications are the source programs decks. The policy alternatives and much of the information on a project is contained in these decks and they often constitute records of intrinsic value for historical purposes. Economic and financial projection models and war game software are typical examples.~~

2.3 People-ware

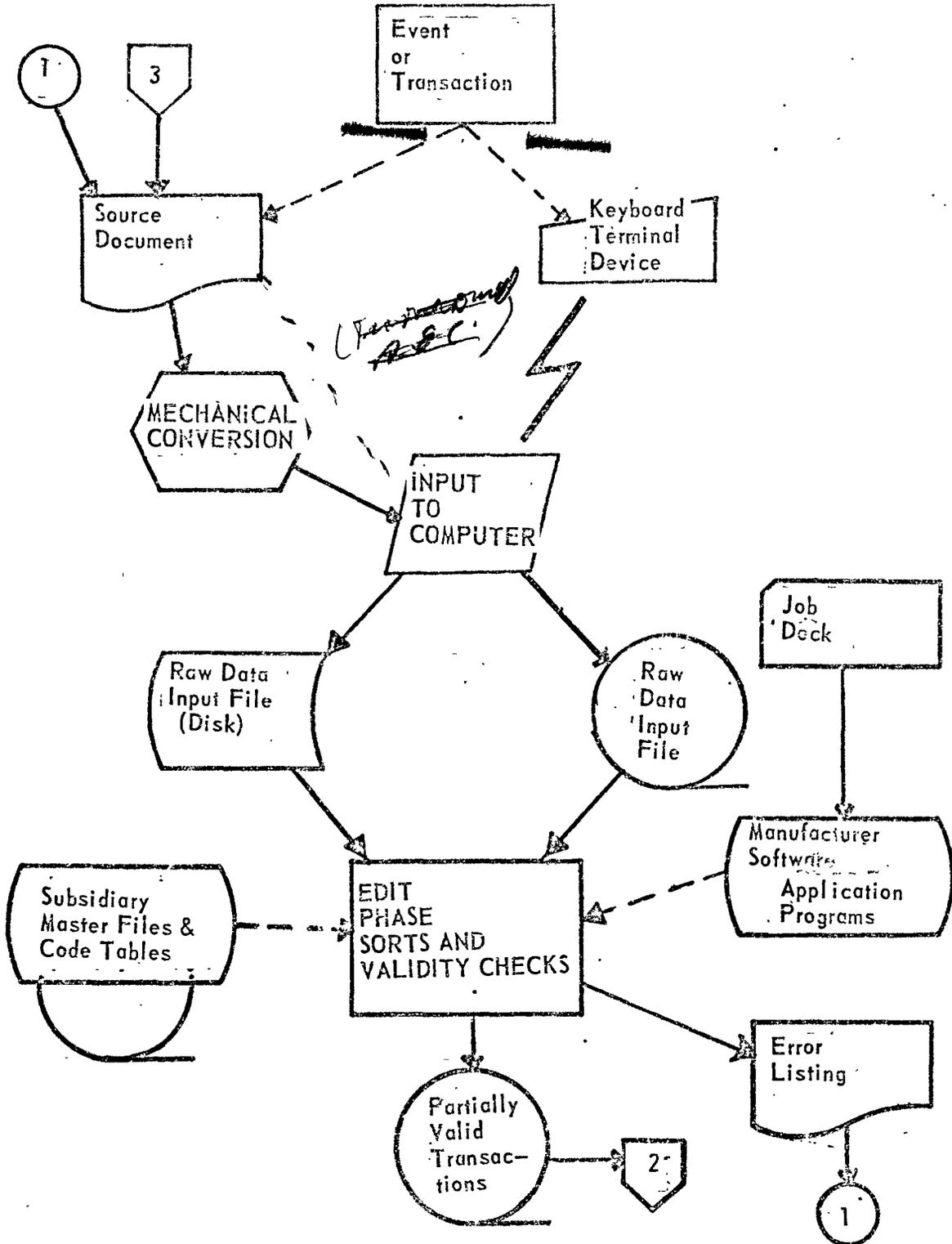
A wide variety of hard-copy documentation is produced in data processing systems. People-ware is that documentation required by the personnel involved in the design, development, operation and maintenance of ADP systems. The files are listed in parts I and II of this schedule. Of interest in this section are primarily those files required for the direct serving of files declared permanent.

The basic concept to grasp in data processing is that the record constitutes a representation of an event and not the event itself. As such, the representation or record may have been recorded by a sensor (as in scientific measurement) or may have been transcribed and encoded from some other document or document group as in all transaction reporting. In either case, a researcher needs to know what kinds of transformations occurred between the actual event and its representation on magnetic tape. This knowledge is in the documentation described in part II of this schedule.

For example, most housekeeping systems usually encode events using elaborate code tables rather than narrative fields on the record. A payroll system may have dozens of deduction code possibilities as well as an equal number of pay plans. Typical codes would represent bond deductions, local tax rates for states and municipalities, bond and charity deductions, overtime and premium shift differential rates, etc. In scientific work, usually instrument voltages are read to represent physical phenomena actually occurring. ~~rather than other occurrences~~

Each time a transaction is encoded or ⁱⁿ instrument reading is made, there is a possibility of an error or distortion taking place in the process. The error that may occur may be simple random occurrences such as digit transposition by keypunch operators or transcribers, or systematic because of some bias in the recording instrument or observer. In general, the scientist

GENERALIZED INPUT UPDATE (I)



Check in my book

transaction items are ~~listed~~ on a printer along with the error indicators for immediate correction on a batch basis. ~~In other systems,~~ *the file* erroneous records ~~are added~~ to indicate the presence of certain errors, but ~~are not deleted~~ from the transaction file. Instead, ~~they are kept~~ *them* in the transaction file for still further checks in the update program itself. This gives a consolidated error listing ~~at one time~~ for a given batch of transactions. The most common additional tests performed on the data would be ~~tests~~ *tests* against the master file key itself. Examples are transactions which attempt to delete a nonexistent record or insert a duplicate record into a file; others may be quantities which ~~are checked~~ for "reasonableness".

3.1.3 Update Master File or Data Base Phase

The output of the edit phase is input into the updating run or subsystem. If the file contains only one application such as accounts payable or receivable, it is normally called a master file. If it contains data from a series of applications or is primarily summary data from a variety of sources, it is usually called a data base.

Individual transaction and status reports are not always excluded from being called data bases. This is particularly true when individual status reports are the direct concern of an organization's top management. Examples are status reports for a series of research and development projects, construction jobs, and loan files.

There are several types of master files. In some systems, only currently active records are found on the file. Purged master file records are periodically transferred to dormant account or history files. Personnel and payroll files are typical of this class; the purging generally occurring at year end. Many other files are cumulative and continue to grow in size indefinitely depending upon the application. Just as is the case of paper records, the frequency of reference is the major design criterion when deciding upon the length of time ~~detail transactions~~ *the* reside on the master file.

HP
data

In the case of periodically updated files where ~~the~~ transactions are deleted, the purged records are often merged into historical files. These are often quite valuable for long-term retention and historical files. However, ~~these transaction files often lack much~~ *these* required detail found on the master file record. In such cases, both the merged or periodic transaction file plus the master file for the given as-of date ~~should~~ *should* be retained for historical and research purposes.) ?

Master files are seldom updated for a given period in a single update pass. Some errors cannot be detected until the actual posting attempt is made. ~~The result is that~~ a series of 'interim' master files ~~is created~~. The only valid file would then be the one from which the periodic reports were run. Usually, processing deadlines determine which version of such a file is 'valid'. Interim master files are usually retained for short periods as back up tapes for the final master file itself. This retention system

output plan

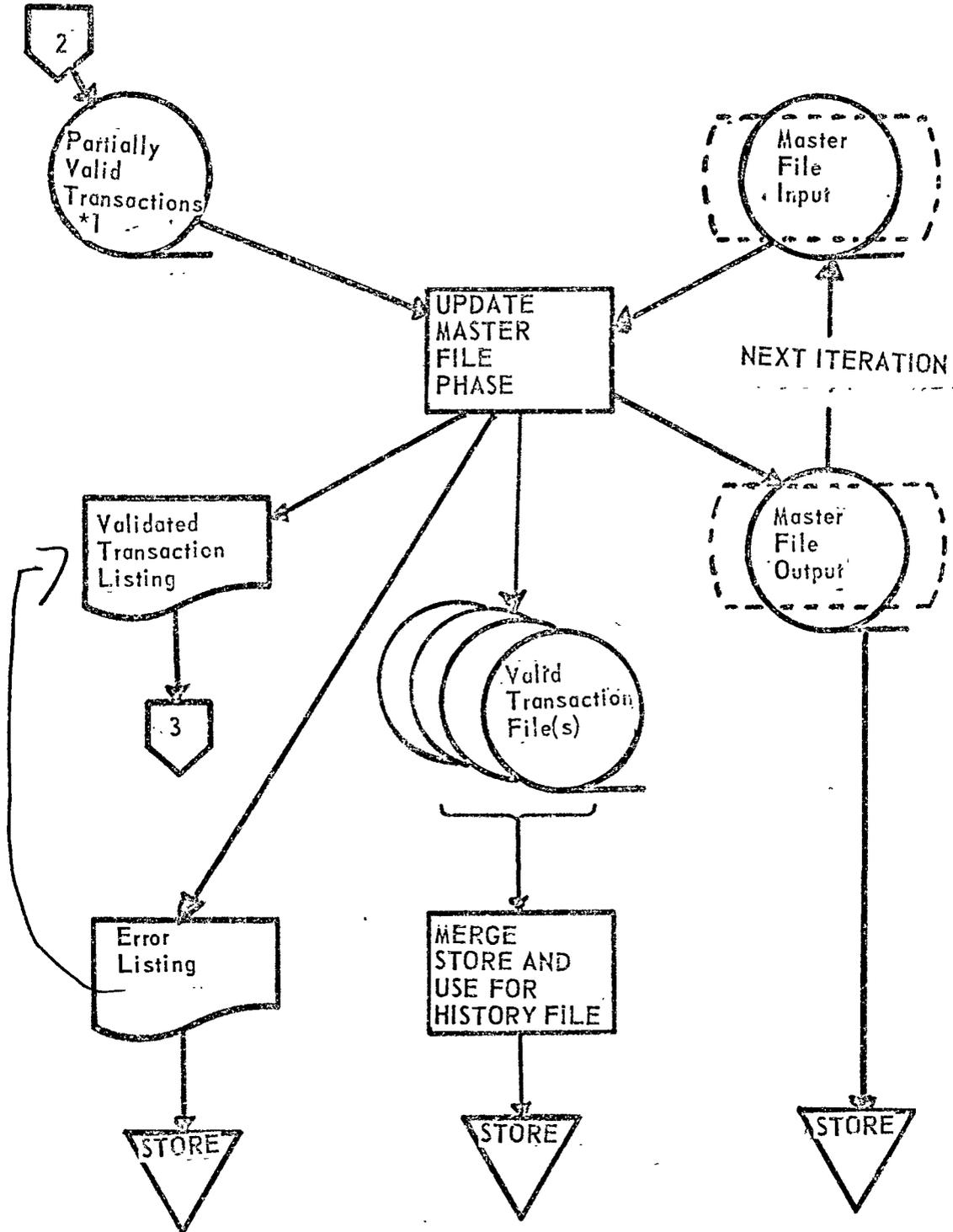
for --

TITLE(S)

GENERALIZED INPUT UPDATE (II)

TYPE OF CHART

CODE NO.



*1 or work tape

*3 May be cumulative or noncumulative file.

*2 A series of these files may be merged into a continuous history file.

*4 Until final version is approved, all prior versions are interim master files.

is called the grandfather system. In appraising master files for permanent retention, it is preferable that the official files for an as-of date be ~~submitted~~. While this is usually possible, there are many cycle billing systems in which the master file is never completely purged of detail transactions or in a final state. In such cases, it may be preferable to retain extracts of the master file made for reporting purposes and audit trails rather than the basic file itself. *retain* *quite complete*

C 8/29/43

3.2 Report Generation Phase

This section describes the files, processing and software used in producing output products from ADP systems. The chart labelled as Report Generation Phase shows the usual processing sequence in such modules from the machine-readable ~~only~~ record to the final printed report or listing. Since mass random access storage is being increasingly used, the ~~flow~~ chart shows tape and disk files used interchangeably, although in practice, one or the other medium will predominate.

As ~~was~~ indicated previously, it is not necessary to retain output ~~module~~ oriented software unless its usefulness and transferability to future computing systems is assured. This ~~type of~~ evaluation should be made on a case-by-case basis. *use*

3.2.1 Report Data Extract and Format Phase

If the printed report or tabulation is in the identical sequence that the master file is in, the data selection, tabulation and printing phase may occur in one program. This ~~situation is true in~~ billing, payroll and most housekeeping systems. However, it is often desired to print the listing in ~~some~~ sequence different from the way the master file is kept. In these cases, a sorting pass must be used to sequence the selected records. The result will be a series of intermediate work files between the master file and the printed output. The flow chart labelled Report Generation Phase shows both tapes and disk files in the processing sequence. *characteristic of*

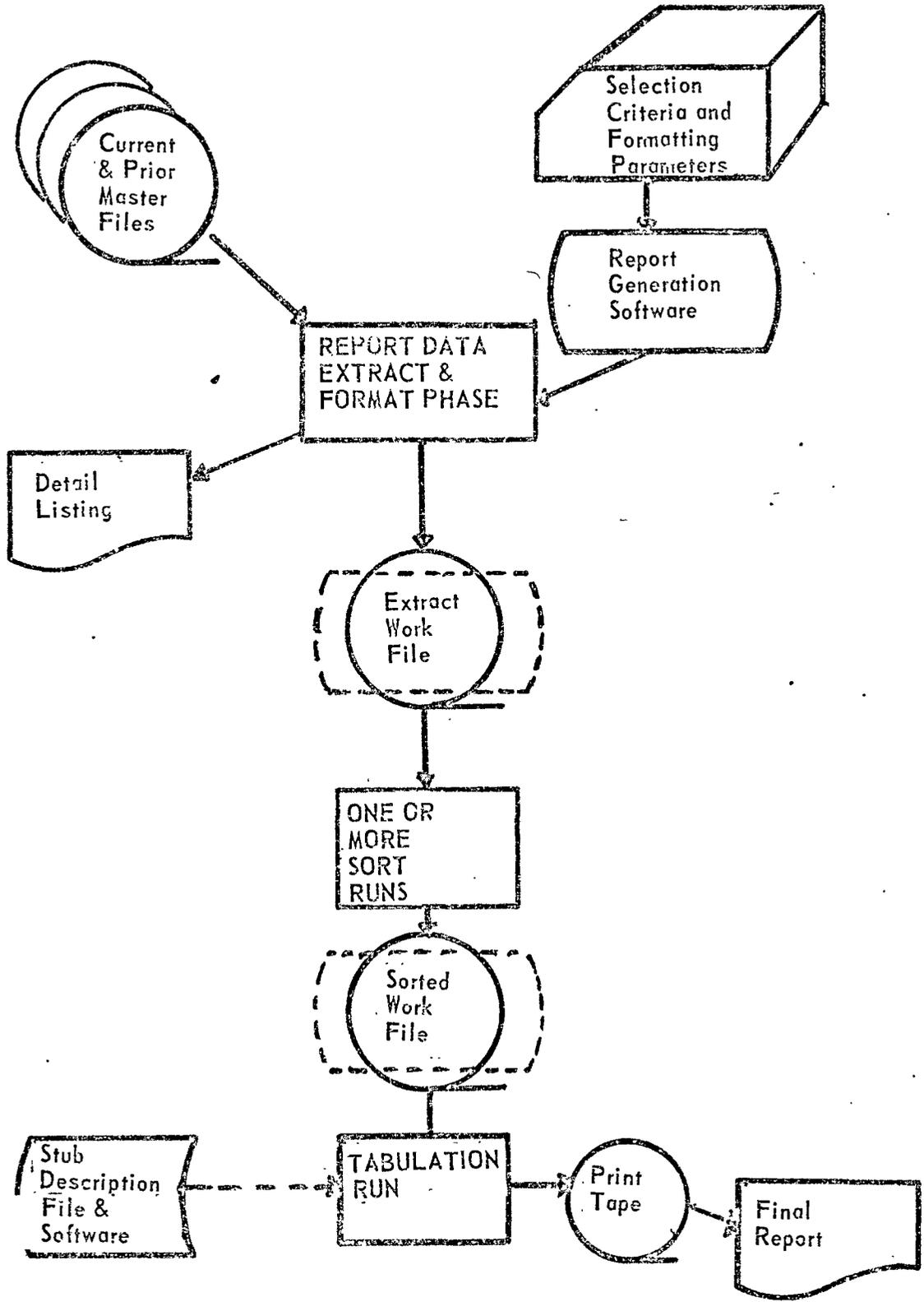
create

The newer large-scale computer and operating systems seldom produce work tapes except for the largest of multi-reel files. The intermediate files reside on disk as transient files within the 'Job Stream'. To an outsider watching the computer, only the input file, the job control deck, and the final printed output are visible. In such cases, there is seldom any need to retain any intermediate files. This is because they can always be recreated from the master file itself.

There is one case where extract files are useful and should be retained by the originator. This is where the file is a carefully developed sample or subsample which is considerably shorter than a very large (over ten reels of tape) file. These sample files, along with appropriate weighting factors and stripped of identifying information disclosing individual persons or establishments, are very useful for release to public researchers. *I would use*

Another item needed
I can get it, breakdown file extract
(74)
has to be provided

TITLE(S) REPORT GENERATION PHASE	TYPE OF CHART	CODE NO.
---	---------------	----------



The next question is to determine which of several work files is best suitable for retention. In general, this depends upon the degree of encoding of stub descriptors required to interpret the file. Heavily encoded files with little or no narrative descriptions are suitable provided the stub description files and tables required for human reading and interpretation are of reasonable length. In cases where the code is an actual Federal Information Processing standard, the length of the code table is unimportant. This would be the case for state and county codes of the United States; a table of over 3,000 entries. For individual agencies or one-time studies, tables of less than 200 entries are reasonable to leave in encoded form as they can be reinserted by simple computer programs. Larger code tables should be inserted into the records themselves. Although there is no hard and fast rule in this regard.

3.2.2 One or More Sorting Runs

Extract files are often in the wrong sequence for the final report or tabulation required. In fact, the same file may be sorted into as many as ten different sequences for different types of analysis and tabulation. The criteria for retention of sorted work tapes are the same as for extract tapes described above. The output of this phase is a sorted work tape of file ready for tabulation, summarization, or editing.

This type of file on tape is often a useful research file, particularly if there has been some editing and interpretation work performed on it. In general, administrative systems which run the same files on a periodic basis, have relatively few processing steps between the first extract run and the final output pass. This is different from one-time reports, as is described in section 4, below.

A common variation in many administrative systems with relatively short files and no discs is to generate multiple copies of the selected records for a series of reports. Usually, the first sorting key field is the report number itself. By having a series of files on one reel, with the key fields rearranged to generate different sorting sequences from the same computer run, large savings in machine time ensue. This is due to the elimination of multiple setup operations in the machine room. In general, these files are seldom of permanent value.

3.2.3 Tabulation Run

The inputs to the tabulation run are the sorted work files and usually a stub description file. The stub description file is invariably present in those cases with a very large list of codes which must be displayed in plain text. If it is on magnetic tape, one or more sorting runs are typically required of the extract tape in order to apply the descriptions. If on disk, most of the table lookups can be performed during the tabulation run. Although these lookup programs are much slower than a straight tabulation and list program, they are usually more efficient on an overall basis when several sorts are eliminated.

The final output of a tabulation run is the printed report itself. ~~In most early computers, a print tape was generated and later run off on a smaller computer or peripheral device. In newer systems, a tape is not even produced unless specifically requested.~~ The print file is stored on a disk as a transient file and is listed whenever the high speed printer is free. Systems which can do this work in a multiprogram mode. That is, they execute several computer programs simultaneously. The primary justification ~~for this is to achieve~~ much higher computer efficiency by minimizing idle time of peripheral devices.

Quite often, ~~it is~~ these print tapes or files ~~which~~ are very useful in further studies after the report is printed or published. Usually, they are disclosure-free, having been modified in subsequent operations (~~this is~~ described below in section 4). Much of this type of data, especially from the Census Bureau, is now being retained and sold. In most administrative systems, the usual retention period for such tapes when they are still being produced is about five work days after the printing is complete. Retention criteria for print tapes are simple: They should be less than one percent of the length of the parent file or have had considerable manual and computer processing in their preparation.

4. One-Time Surveys and Report Generation Systems

The sequence of operations in one-time surveys, censuses, and tabulations is shown in the following two charts. When the flow process charts are compared to the typical continuous file maintenance system, the similarities are evident. The basic differences between continuously running systems and one-time jobs is the much higher amount of manual editing and encoding required. Unless the job is a very large effort, with many thousands of observations, the forms used allow somewhat more variability in field entries than accounting type documents. Since little feedback occurs between respondents to these surveys in case of error, much more manual editing and encoding is required to clean up a file prior to use in tabulations.

As the second sheet shows, there is a file buildup process which occurs with no maintenance occurring to individual records after they have entered the file. Where a multiplicity of systems and sources feed the file, the individual records are usually of variable length for the sake of minimizing storage requirements.

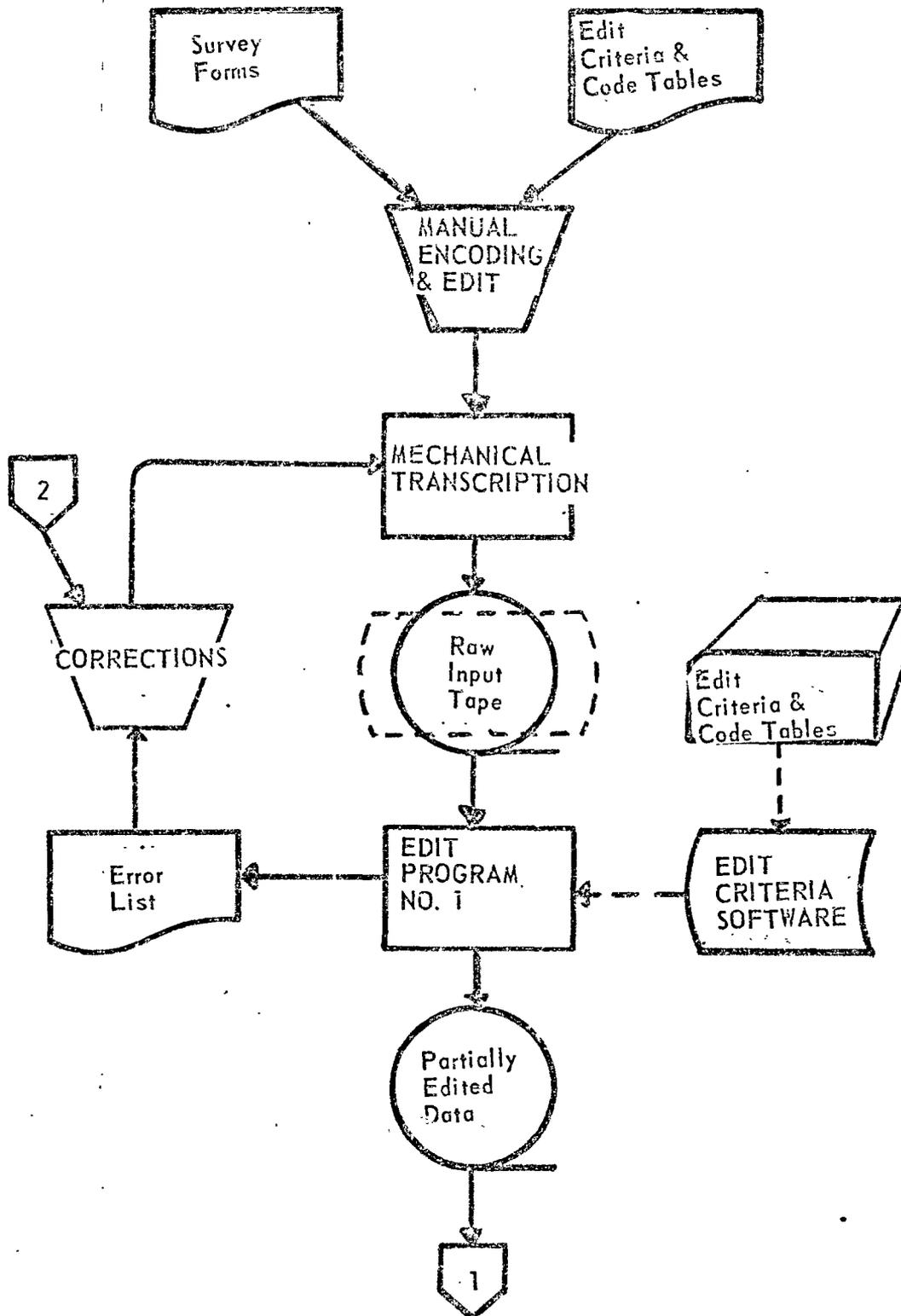
Documentation for such systems generally contains complex record formats but few of the elaborate codes usually found in administrative systems. When these files are retained, it is very important to identify the original sources of information, the instructions to respondents for filling out the forms together with sample forms, and the directions given to the response form editors for proper interpretation and secondary usage of these files. Most such files are described in part II of this schedule.

TITLE(S)

TYPICAL ONE-TIME PROCESSING SEQUENCE (I)

TYPE OF CHART

CODE NO.



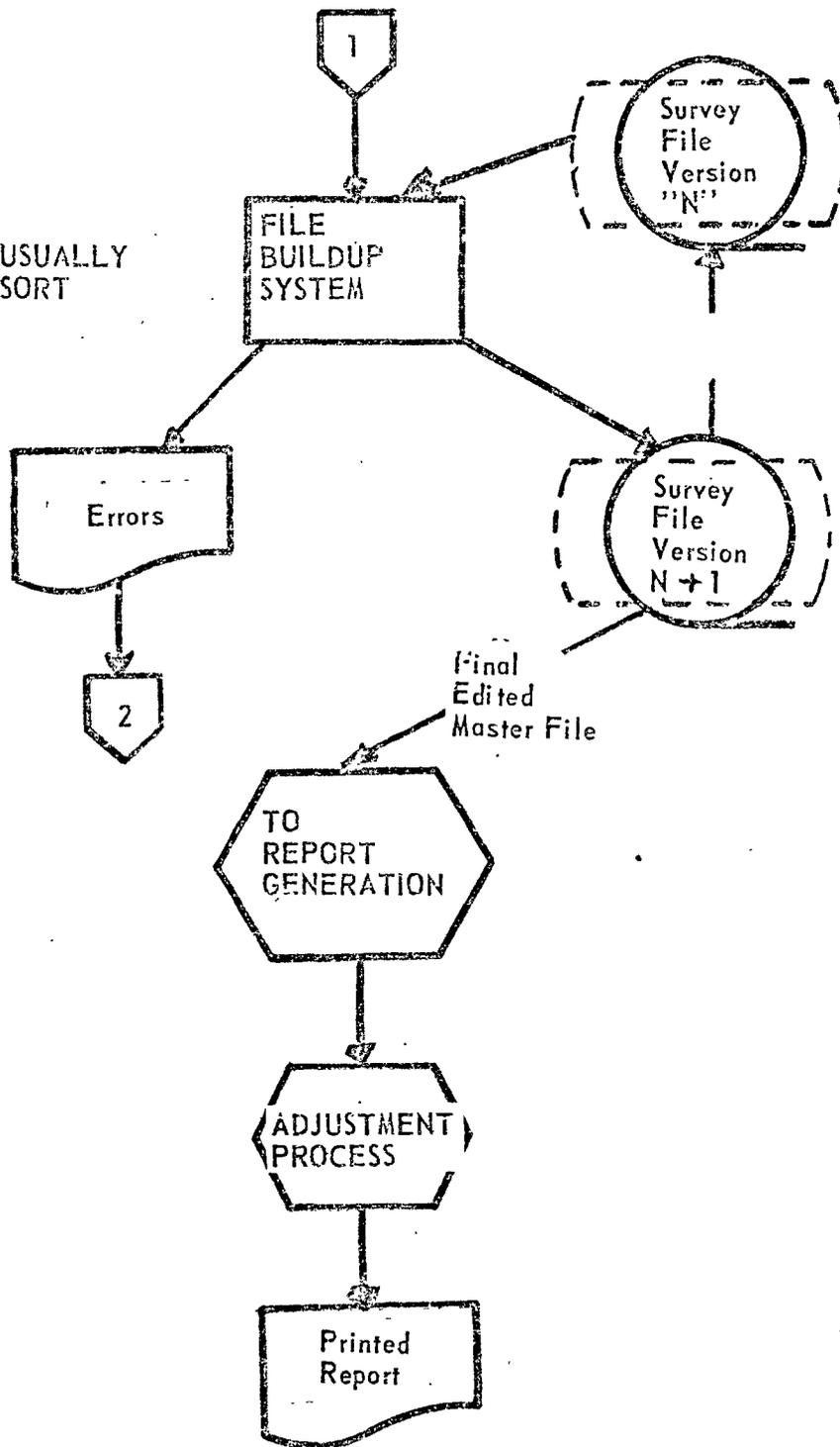
TITLE(S)

OF CHART

CODE NO.

TYPICAL ONE-TIME PROCESSING SEQUENCE (II)

NOTE; THIS USUALLY
 CONTAINS A SORT
 PROGRAM.



5. Adjustments in One-Time Jobs

If a payroll record contains an error, the originating office usually hears about it in short order, particularly when an employee is short changed. In sample surveys, reporting and transcription errors are seldom corrected by the giver of information unless elaborate review procedures have been established calling for a review of the machine-prepared record. This correction and review process almost always occurs in accounting type systems, while cost usually precludes this process in most one-time jobs.

The net result is that such files of recorded observations contain a variety of errors which can lead to nonsense entries in summary tabulations. Male widows might be a theoretically possible category in a tabulation which is difficult to imagine existing in actuality. These errors can arise from a number of causes. Correction of the tabulations can be done at any step in the process between the final survey file and the printed report. The accompanying flow chart shows the points in the procedure where this is usually done.

If the error is thought to be a random event, the illogical counts are generally distributed to all other possible categories and deleted from the tabulation array where they occur. This would lead to a discrepancy between the published table and the final master file. The illogical records would remain in the file uncorrected.

Systematic encoding errors also occur frequently in encoding and processing. In these cases, the summary file may be corrected by moving the entire nonsense count to the correct location in the table. These errors can also be corrected in the final master file using the computer itself.

Another common adjustment operation occurs in cases where disclosure of individual confidential information may occur in the tabulation. In these cases, the specific tabular entry is deliberately deleted and consolidated with enough other entries so that this confidentiality is maintained. If the correction is performed on a summary tape, then this tape and the printed report will be identical. If the correction was made only to the camera-ready hard copy, the file will be accurate, but will not be releasable to the public.

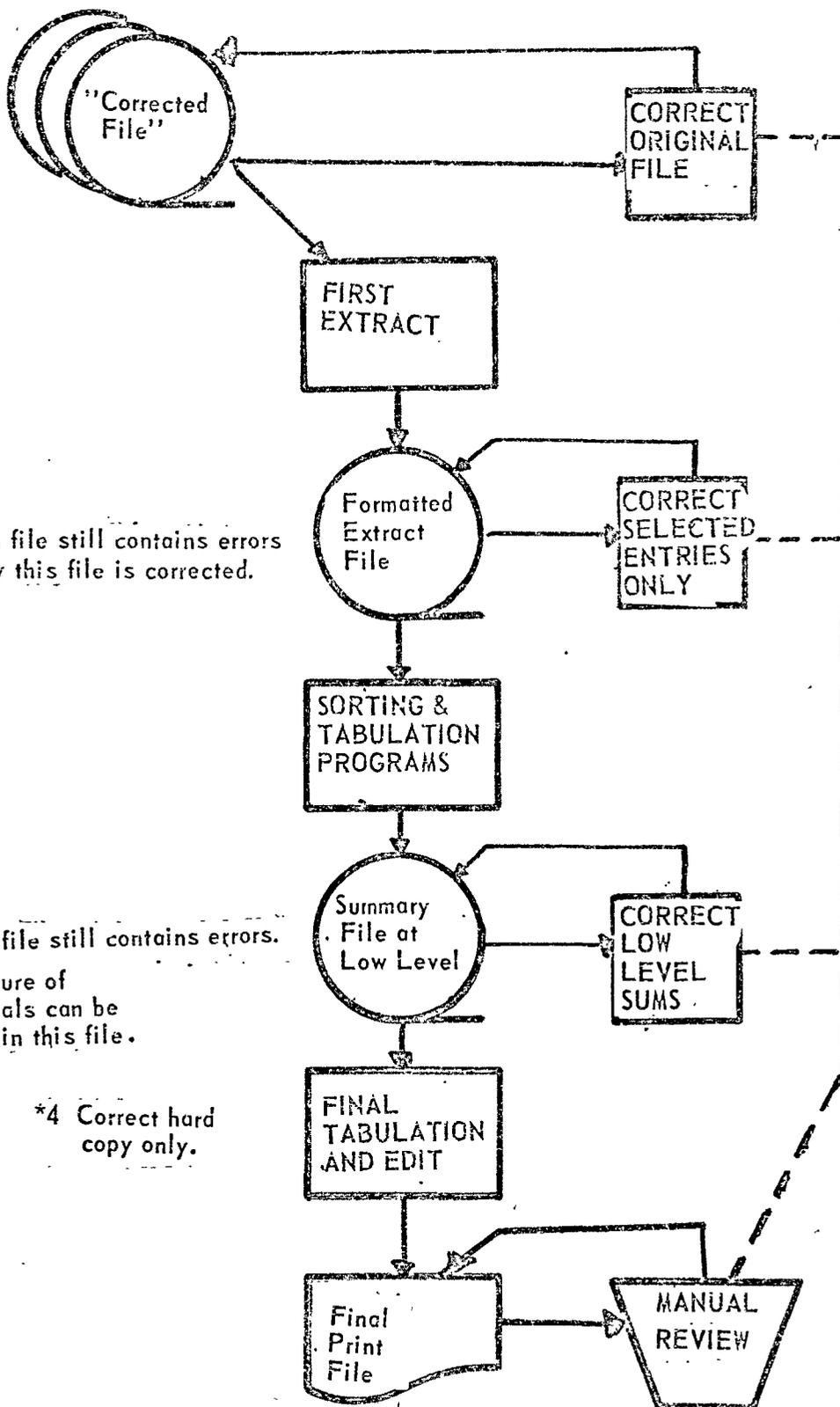
In summary, a variety of additions, changes, and deletions can occur to either individual or groups of entries within a file at any stage of processing. These changes can occur for a variety of reasons and lead to magnetic tape files of different accessibility and validity. It is important to document the errors and adjustments made to such files if they are to be retained permanently. This record of changes constitutes the equivalent of the accountant's audit trail for evaluating the accuracy of a financial file.

TITLE(S)

WHERE OBSERVATIONS AND SUMMARIES
 GET ADJUSTED IN ONE TIME JOBS

TYPE OF CHART

CODE NO.



*1 Origin file still contains errors if only this file is corrected.

*2 Extract file still contains errors.

*3 Disclosure of individuals can be deleted in this file.

*4 Correct hard copy only.