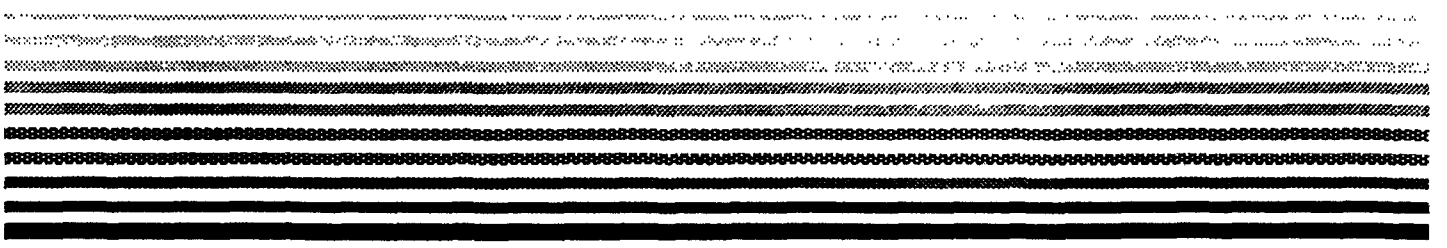
U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF INFORMATION RESOURCES MANAGEMENT
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

AUTOMATED LABORATORY STANDARDS: EVALUATION OF THE USE OF AUTOMATED FINANCIAL SYSTEM PROCEDURES

CONTRACT 68-W9-0037, DELIVERY ORDER 035
JUNE 1990



Automated Laboratory Standards:

Evaluation of the Use of Automated Financial System Procedures

Prepared for:

Office of Information Resources Management
U.S. Environmental Protection Agency
Research Triangle Park, North Carolina 27711

June 25, 1990

Prepared by:

BOOZ•ALLEN & HAMILTON Inc.
4330 East-West Highway
Bethesda, Maryland 20814
(301) 951-2200

Contract No. 68-W9-0037

Computer Sciences Corporation
79 T.W. Alexander Dr.
Research Triangle Park, North Carolina 27709
(919) 541-9287

Contract No. 68-01-7365

Acknowledgments

This report was the combined efforts of Computer Sciences Corporation, Booz•Allen & Hamilton Inc., EPA staff, and outside experts. Richard Trilling of CSC researched and prepared the draft for public review. Jennifer Abrams, Marguerite Jones, and Marcia Balestri of Booz•Allen evaluated the comments and completed this final report. Numerous EPA staff and outside experts provided substantial critical reviews and valuable technical comments. Richard Johnson of the Scientific Systems Staff of EPA's Office of Information Resources Management directed the contractors' work and managed the review process.

Table of Contents

Executive Summary

The U.S. Environmental Protection Agency (EPA) has initiated a program to ensure the integrity of computer-resident data in laboratories performing analyses in support of EPA programs. The possession of sound technical data provides a fundamental resource for EPA mission to protect public health and the environment.

This report describes the findings of a review of standards used in existing automated systems in the financial industry. EPA has chosen to study financial system standards because the financial industry has had many years of experience with automated systems, and reliability and validity of financial data is critical to the success of financial institutions. In addition, auditors have developed a broad system of controls designed to ensure the integrity of financial data compiled in automated systems.

The financial industry's experience in preserving data integrity in automated financial systems can well be applied to the automated laboratory environment. The main sources of risk to data integrity are present in both automated financial and automated laboratory systems. These include the following:

- Adding incorrect data to the data base
- Having multiple applications affecting a single data base
- Intentional or inadvertent acts
- Failure of the data base management system to function as specified.

Because automated financial systems have been developing for a number of years, much research has been conducted on controls to help protect a system

against these risks. Controls that can aid in minimizing these four significant risks include:

- Verifying input data
- Securing data by restricting access
- Permitting write access to only one application/user at a time
- Instituting a program for detecting/reporting/correcting system problems.

Finally, the need to back up the data base regularly and maintain an accurate and complete audit trail, in addition to employing these data integrity controls, has been demonstrated in the financial environment and will be equally important in the automated laboratory environment.

Background

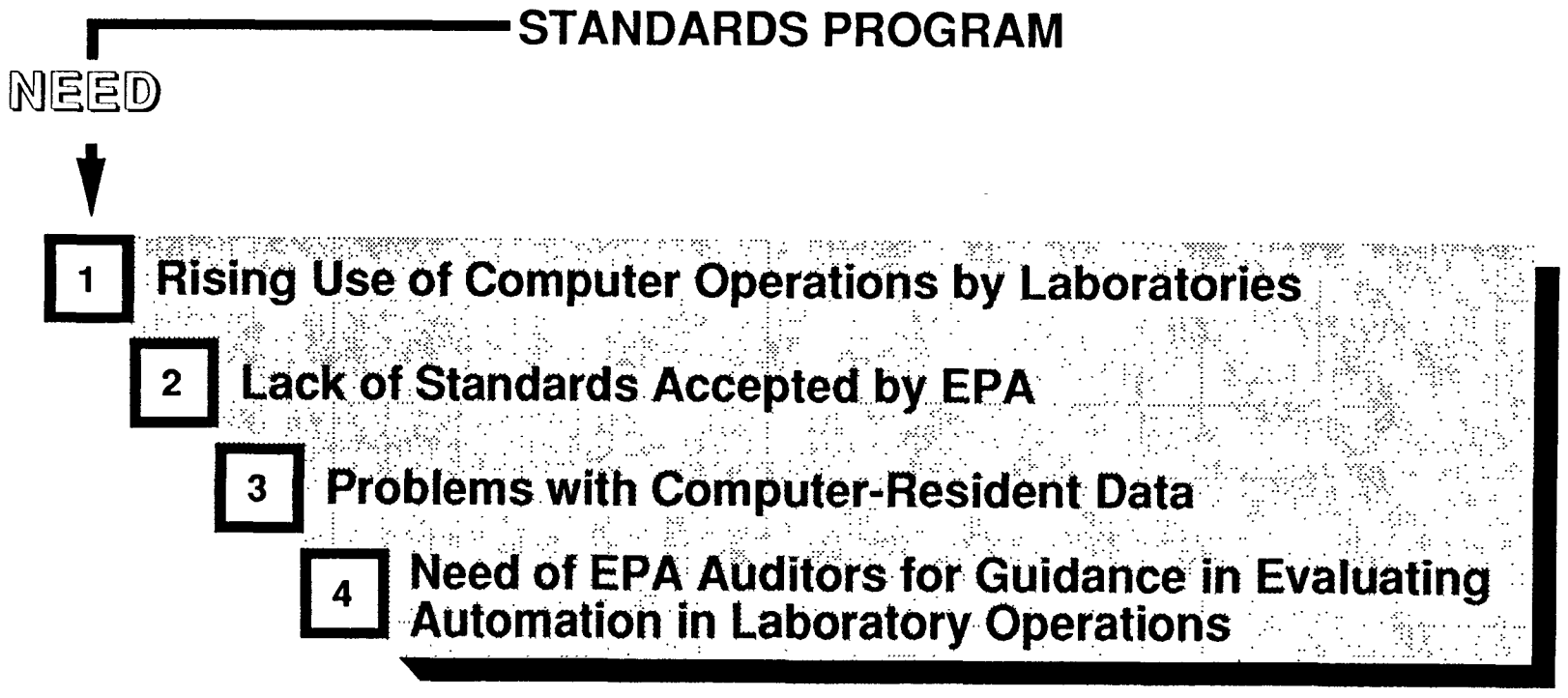
The U.S. Environmental Protection Agency (EPA) has initiated a program to ensure the integrity of computer-resident data in laboratories performing analyses in support of EPA programs by developing standards for automated laboratory processes. The possession of sound technical data provides a fundamental resource for EPA's mission to protect the public health and environment, regardless of the activities of the specific environmental programs. The activities of these environmental programs are diverse, and include basic research at EPA's environmental research centers, environmental sample analyses at EPA's regional laboratories and contractors' laboratories, and product registration relying on analytical data submitted by the private sector.

EPA recognizes that the implementation of an automated laboratory standards program will require each laboratory to allocate resources of dollars and time for the program's execution. Experience has shown that in developing and using a proper standards program, a net savings may be achieved, as acquisition, recording, and archiving of data will be improved with a net reduction in test duplication.

Within EPA, the Office of Information Resources Management (OIRM) has assumed the objective of establishing an automated laboratory standards program. The need for this program is evidenced by several factors. Exhibit 1 illustrates these factors, which include the rising use of computerized operations by laboratories, the lack of uniform standards developed or accepted by EPA, evidence of problems associated with computer-resident data, and the evolving needs of EPA auditors and inspectors for guidance in evaluating automated laboratory operations.

Laboratories collecting data for EPA's programs have taken advantage of increasing technology to streamline the analytical processes. Initially, automated instrumentation entered the laboratories to increase productivity and enhance the accuracy of reported results. Computers maintaining data

EXHIBIT 1
Need for EPA's Automated Laboratory Standards Program

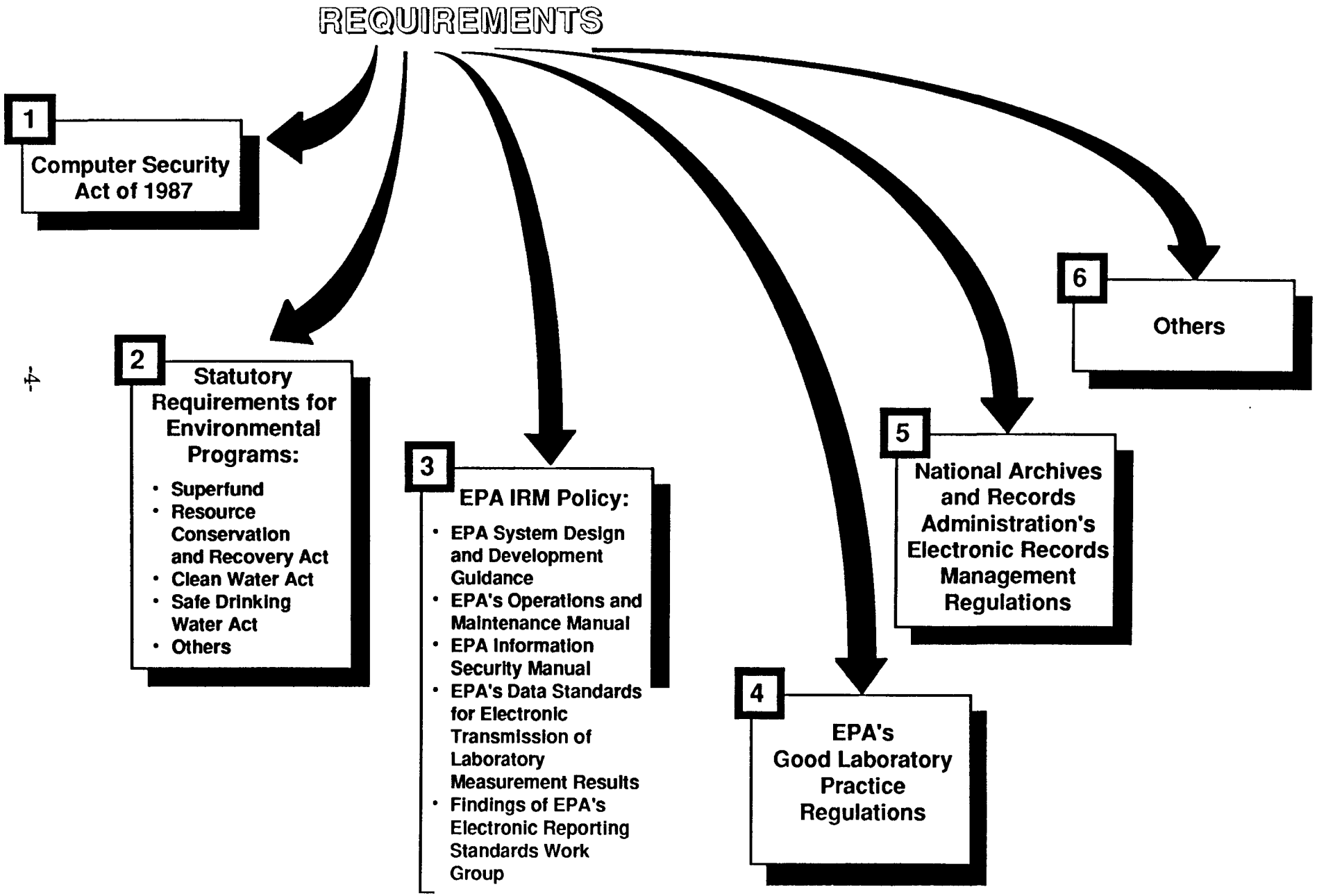


bases of results were then used for data management and sample tracking. These computer systems were integrated into more sophisticated laboratory information management systems (LIMS). Methods for data reporting include electronic mail, electronic bulletin boards, and direct links between central processing units. Each of these advances necessitates thorough quality control procedures for data generation, storage, and retrieval to ensure the integrity of computer-resident data.

Currently, EPA has no Agency-wide guidelines for laboratory information integrity that laboratories collecting and evaluating computer-resident data must follow. The requirements that must be considered in developing automated laboratory standards come from a variety of sources, as Exhibit 2 illustrates, including the requirements of the Computer Security Act of 1987 (P.L. 100-235, January 8, 1988) and various EPA program-specific data collection requirements under Superfund, the Resource Conservation and Recovery Act, the Clean Water Act, and the Safe Drinking Water Act, among others. Additionally, OIRM has developed electronic transmission standards and is developing a strategy for electronic recordkeeping and electronic reporting standards that will affect all Agency activities. The development of uniform principles for automated data in EPA laboratories, regardless of program, will take into account the common elements of all these data collection activities, and provide a minimum standard that each laboratory should achieve.

There is increasing evidence of problems associated with the collection and use of computer-resident laboratory data supporting various EPA programs. To illustrate, as of November 1989, EPA's Office of the Inspector General was investigating between 10 and 12 laboratories in Superfund's Contract Laboratory Program (CLP) for a variety of allegations, including "time traveling" and instrument calibration violations. In "time traveling," sample testing dates are manipulated, by either adjusting the internal clock of the instrumentation performing the analyses or manipulating the resultant computer-resident data. (Hazardous waste samples must be assayed within a prescribed time period or the results may be compromised.) Additionally, calibration standard results have allegedly been electronically manipulated

EXHIBIT 2
Considerations in Developing Automated Laboratory Standards



and other calibration results substituted when the actual results did not meet the range specifications of the CLP procedure being followed.

Because the introduction of automation is relatively new and still evolving, no definitive guidelines for EPA auditors and inspectors have been developed. Inspectors must be alert to the steps in those procedures used by the laboratories generating and using computer-resident data where the greatest risk exists. These critical process points indicate the magnitude of control that should be placed on each step of the process. If adequate controls are not present, the remainder of the process cannot correct a deviation, and the entire process will provide no reliable conclusions. Automation introduces many new variables into a system, each with its own set of critical process points. Inspectors must verify that laboratory management has recognized the various risks and has instituted an appropriate risk management program.

As part of the EPA's program to ensure the integrity of computer-resident data, EPA reviewed the policies and procedures in place in automated financial systems. The purpose of the review was twofold. First, EPA hoped to learn possible risks to data integrity and controls to protect against them. Second, it was hoped that some of the standards in place in automated financial systems could be applied to the automated laboratory environment. As a result of its research in automated financial systems, EPA identified many procedures that could be applied to automated laboratories.

Other areas of evaluation in developing the standards program include a review of current technology, a survey of current automated laboratory practices, and an analysis of the applicability of EPA's Good Laboratory Practice regulations to automated laboratories. The findings of each of these evaluations are provided in separate reports.

The purpose of this paper is to report on the findings of EPA's research into the financial systems and associated procedures. It is intended to provide guidance for people developing standards for the automated laboratory setting, by indicating sources of risks and procedures to control risks.

The findings reported in this paper are based on both library research and interviews with laboratory and financial systems experts. Because little has been written on the preservation of data integrity in automated financial systems specifically, interviews with local bank systems operations managers were used to learn more about practical application of the theory.

This paper is organized as follows. First, a side-by-side introduction and comparison of a generic laboratory system and a generic automated banking system will be presented to illustrate the similarities between the two systems and thus the applicability of procedures used in the financial arena to the laboratory environment. Then, the paper will present a discussion of risks to data integrity and controls to counter the risks. Next, the paper will discuss the importance of backup, audit trail, and auditing functions. Finally, a brief introduction to the questions of the legal validity of computer-resident data will be presented.

Findings

System Introduction and Comparison

An example of a generic laboratory system is presented in Exhibit 3. In practice, the degree of automation varies widely across laboratories; however, at a base level, most systems have the general components and interfaces depicted in the exhibit. In summary, when a sample is received at a laboratory, sample identification data is entered into the system (usually by manual keying, but the use of a bar code is becoming increasingly common). Analyses are then performed on the sample; these will be done completely without human interaction in the most automated of systems, or entirely "by hand" in the most elementary of systems. Sample analysis results are then validated (again automatically or manually, depending on system capabilities) and either posted to the data base or reprocessed. The data base is then accessed to produce desired reports, such as sample analysis reports.

Exhibit 4 illustrates a generic automated savings account system. As with automated laboratory systems, the degree of automation will vary across banks, although not nearly to the degree it will in laboratory systems because basic automated financial systems have been developed and put in place for a number of years. The system depicted in Exhibit 4 details the manual deposit and withdrawal system; several other applications, such as automatic deposit, automatic mortgage payment withdrawal, and wire transfer, may affect the account balance data base as well, and are therefore noted in the exhibit.

The deposit or withdrawal process begins as account change data are keyed to the system either by a teller or by the account owner at the Automated Teller Machine (ATM). The account change data are verified by daily teller and ATM reconciliations. The account change data are then posted to the main data base, where changes to the appropriate account balances are made. The data base is then accessed to produce desired reports, such as monthly savings account statements.

EXHIBIT 3
Generic View of an Automated Laboratory System

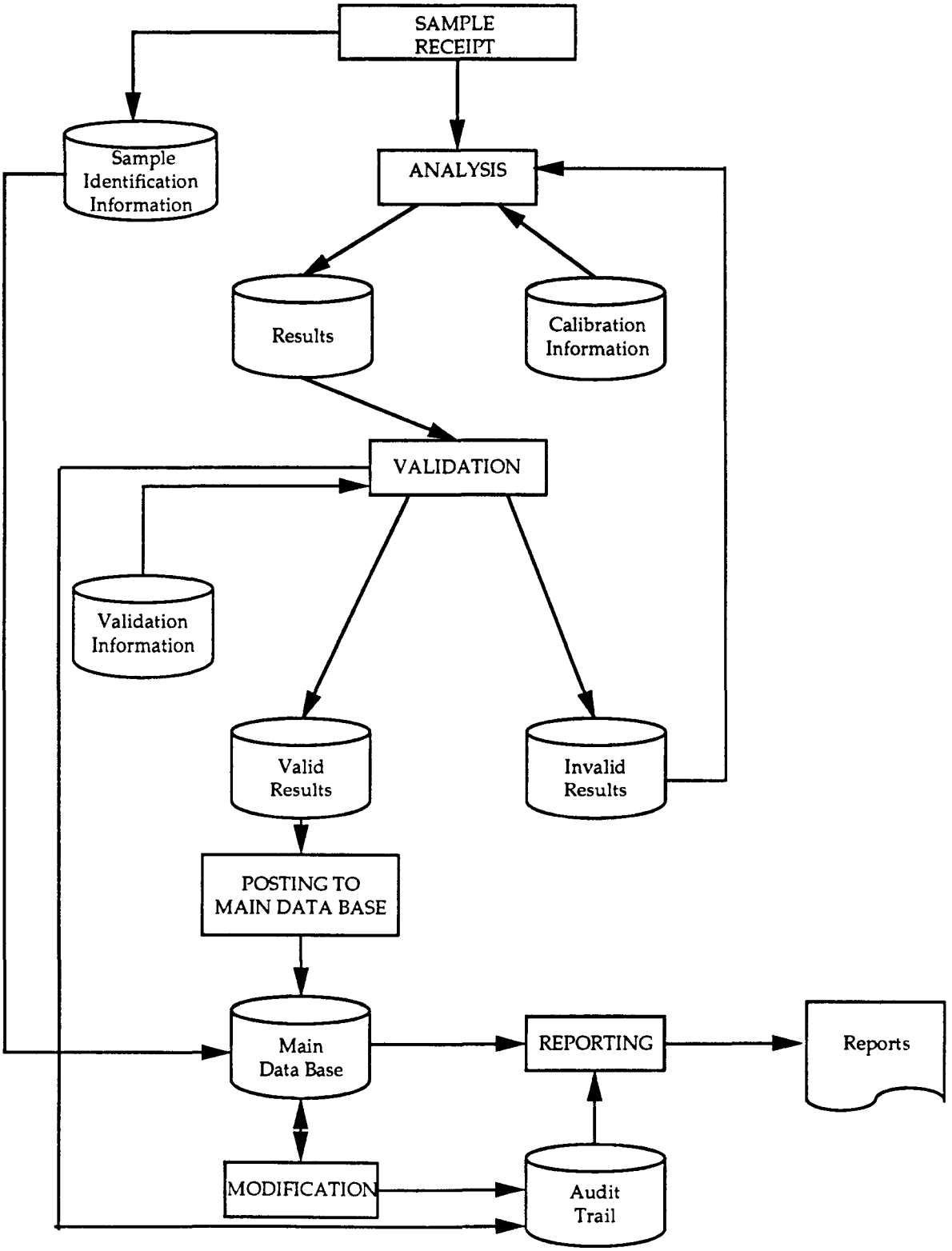
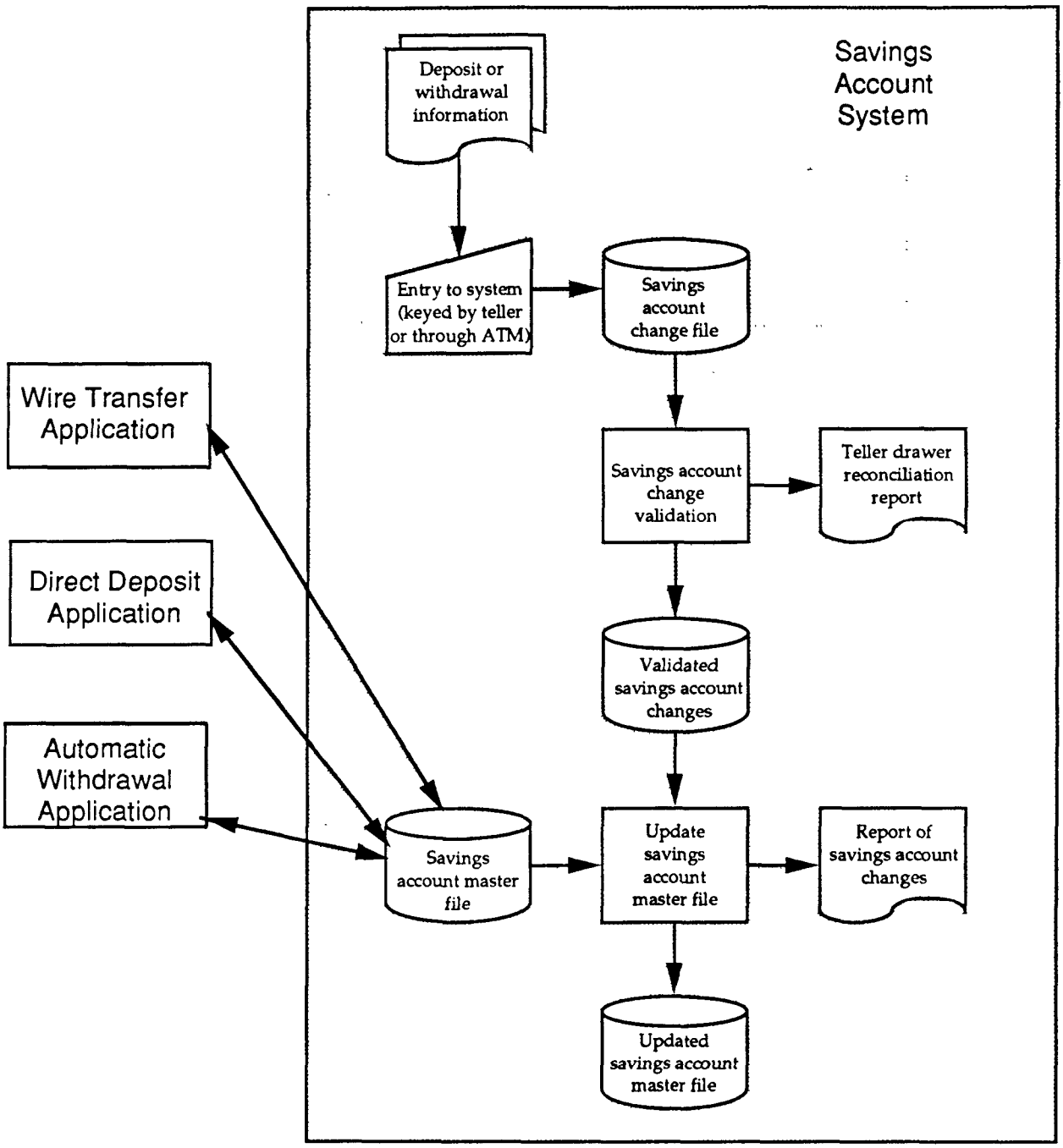


EXHIBIT 4
Generic View of a Savings Account System



Automated laboratory and financial systems are similar in many ways. Both systems build a data base of information on which important decisions will be based, so the integrity of the data is critical to the usefulness of both systems. Both systems are dependent on human input. Savings account systems receive data input from tellers and ATMs; most laboratory systems require human input of either sample identification information or test results. In addition, both systems require a well-developed system support plan, which includes staffing requirements and procedures, and defined performance control criteria. Finally, system planning for both systems must include procedures for establishing and maintaining an audit trail.

The systems differ in ways that may at first appear to be relatively minor but ultimately have important implications for the application of financial system data integrity standards and procedures to the automated laboratory environment. First, and most importantly, an audit trail is easier to establish with an automated financial system than with an automated laboratory because paper backup of transactions (such as deposit and withdrawal slips) are generated in the course of initiating a change to the financial data base in the first place. With automated laboratory systems, unless a well defined process is adhered to (such as a manual logging process that tracks sample identification and test results data), it may be harder to establish an audit trail. This is primarily for two reasons: first, the sample is thrown away (or naturally degraded), and second, records of the results of analyses may be inadequately documented.

The second difference between the systems is that financial systems are more subject to fraud and embezzlement than laboratory systems; therefore, some of the procedures used to ensure security of the data base in financial systems may not be required in automated laboratory systems. However, Pincus (1989) recommends that several financial auditing techniques be applied to scientific data to detect fraud.

The final significant difference between the systems is that a financial system will have a much higher volume of transactions than will a laboratory system.

Risks to Data Integrity

After reviewing available information, we have identified four primary risks to data integrity that would be present in both automated laboratory and automated financial systems. The following, summarized from Perry (1983), presents a description and defines the implications of each type of risk.

1. **Incorrect data are added to the data base.** Data added to the data base can be incorrect due to data entry error and/or lack of data verification and validation. Data entry errors are easy to make, especially when data are manually keyed in. And unless data are verified or validated before being posted to the main data base, incorrect data can be keyed in correctly, and still be incorrect when added to the data base. The implications of adding incorrect data to the data base are obvious. At best, with a good audit trail, corrections can be made to the data base, and the integrity of the data preserved (however, not without considerable extra effort). At worst, the incorrect data may never be discovered, and the data base could provide bad information and ultimately lead to incorrect decisions.

2. **The data base is interfered with and data integrity is damaged.** Through intentional or inadvertent acts, the integrity of the data in the data base may be damaged. Fraud and embezzlement in financial systems are examples of intentional acts; as discussed on page three of this paper, "time traveling" and instrument calibration allegations are examples of incidents of potential laboratory fraud currently being investigated by EPA's Office of the Inspector General. It is also easy to imagine accidental interferences with the data base that would compromise data integrity. Again, the implications of these risks are that they would affect data integrity and that their occurrence could easily go unnoticed.

3. **Data base is affected by multiple applications, and the data resident in data base may not reflect the effect of all applications at all times.** The use of multiple applications on a single data base is common in the

banking industry and is introduced in Exhibit 4. The risk to integrity is that one application will change the original data, and another application will access and make decisions based on data that do not reflect changes made via the first application. For example, consider the number of applications that can affect a savings account balance (see Exhibit 4). Suppose an automatic withdrawal application (for an automatic mortgage payment) is initiated to effect a withdrawal from an account. However, at the same time, the account owner, using an ATM, attempts to withdraw all of the money from the account. If there were not procedures in place to ensure that the account balance data reflect the effect of all the applications at all times, clearly, more money could be withdrawn from the account than is actually available. While the use of multiple applications on the same data base may be less common in laboratories, it could certainly be the case that the type of analysis run would depend on the results of a previous analysis, and, if the data base doesn't reflect those results, the wrong experiment could be run.

4. **Failure of the data base management system to function as specified.** The data base management system includes all of the algorithms that are required to update, sort, reproduce, and maintain the data in the data base. If the system does not function as it is designed to or if there are flaws in the data base design so that it does not achieve the same results as a manual system would have (for example, if withdrawal information is posted to the wrong account), the resulting data will not be useful as a source of information because their integrity will not be assured.

Controls That Can Help Protect Against Loss of Integrity

The following, also summarized from Perry (1983), presents a list of possible controls that can be used to help protect against the loss of data integrity for each of the types of risk identified above. Where relevant, actual procedures used by local banks in the administration of savings account data bases are included. The controls included are intended not to be prescriptive but to provide suggestions. The system configuration, potential risks, and

data characteristics must be considered when designing controls for a specific system.

1. **Incorrect data are added to the data base.** Several control procedures exist that could be used to help ensure that the data added to the data base are correct:

   - Re-key the data
   - Produce system control report listing updates to be posted to the data base
   - Include in the system design a validation step using computer-resident validation data and/or verification routines and procedures.

   Banking systems ensure that the data are correct by applying a version of the third suggested control: requiring tellers to "balance their drawers" at the end of every shift, before the data are posted to the main data base. To do so, the teller must reconcile the deposit and withdrawal slips with the data entered on the terminal. The same reconciliation process is conducted for ATM activity.

2. **The data base is interfered with and data integrity is damaged.** The following procedures help prevent against this sort of risk:

   - Restrict access to the system; allow different people different degrees of access; employ password control over access
   - Bond employees
   - Enable a security officer to oversee all activity and report suspicious activity
   - Train personnel on how to use the system (to avoid unintentional acts).

Bank systems typically function using all four of these controls, because financial data is especially susceptible to fraud and embezzlement. Although it is less likely that a laboratory system would require bonded employees or a security officer function, the other controls listed above may be beneficial.

3. **Data base is affected by multiple applications, and data resident in data base may not reflect the effect of all applications at all times.** Several control procedures exist that could be used to help ensure against the risks associated with multiple applications to a single data base:

   - Development of data base control standards for entry, validation, use, and deletion of data

   - Concurrent data control whereby if one user is changing a data element, another user is temporarily denied write access

   - Data ownership, where one individual is responsible for each data element

   - Development of a data element reconciliation utility, including a library of locations for redundant data elements.

   Bank systems have well-developed data base control standards for different applications that interface with and affect a data base. For example, in the automatic withdrawal scenario described earlier, banks have daily ATM and teller "shutdown" periods during which other applications (such as automatic mortgage payment withdrawals) are performed. As a result, the data are never being affected by more than one application at the same time. Additionally, procedures for ensuring that a wire transfer is reflected in the data base are well defined, and only staff with a higher level of responsibility are able to execute transfers.

4. **Failure of the data base management system to function as specified.** The primary control for protecting against the loss of integrity resulting from this risk is the development of a tool for formal reporting of data base problems. The automated financial systems we considered had routine procedures and forms to report problems with the data base, such as reporting to the data base administrator any unusual or problematic occurrence in posting data to the main account data base.

The Importance of Backing Up the Data Base and Maintaining an Audit Trail

In addition to using data base integrity controls, the importance of backing up the data base and maintaining an audit trail cannot be overemphasized. Despite even the most careful use of controls, there is always the possibility that something could happen to disrupt data availability, which may necessitate that data be recreated. In addition, data resident in a data base do not suffice as legal evidence; the court looks to hard-copy data documenting the complete chain of custody for legal evidence. Regular maintenance of backup files, and strict adherence to procedures designed to maintain an audit trail is fundamental to ensuring that a data base can be relied upon for correct information now and in the future.

Two aspects of providing backup for a data base are important. First, the backup should be made as often as is required to maintain the usefulness of the backup data base. In theory, the frequency of backing up the data base depends on the cost of losing the data. In the banking industry, backups are made at least daily; in automated laboratories, the frequency with which backups should be made depends on the degree of activity in the laboratory, the sensitivity and difficulty of the activity being tracked, and the size and type of hardware. Second, the backup copy of the data base should be maintained on a different storage device from the original file; in other words, the backup should not be made on the same disk (or tape), but on a different disk (tape). It is also advisable to store the medium with the backup in a different location to protect against risks that might affect both the original and the backup, such as fire or intentional acts.

In designing procedures to maintain an audit trail, it is important to consider hard-copy data to document the complete chain of custody for every data element added to the data base. In the banking industry, that hard-copy chain is documented by deposit and withdrawal slips, teller reconciliation reports, monthly statements, and the like. In the laboratory environment, procedures must include maintaining backup of all sample identification information, analysis results, validation results, calibration results, and other ancillary information to recreate the complete chain of custody of the laboratory results. From an evidentiary standpoint, a hard-copy chain of custody surpasses a computer-resident trail.

The Importance of the Auditing Function

Although the emphasis in auditing is on controls, the typical steps in the process of auditing resemble the steps in traditional automated data processing (ADP) systems analysis, including EPA's own system methodology, as the following demonstrates:

1. The role of the auditor in controlling ADP activities to ensure data integrity:

    - Understand the purpose of the system and its requirements — the need for the system;
    - Design, implement, and test an effective system of control; and
    - Test the system for compliance — that is, determine how well the outputs of the system meet expectations.

2. The role of the systems analyst in controlling ADP activities to ensure data integrity:

    - Understand the purpose of the system and specify its requirements;
    - Design, implement, and test a system to satisfy those requirements; and

- Evaluate the ability of the system to satisfy those requirements.

The EPA System Design and Development Guidance (OIRM, 1989) describes and documents the steps in traditional systems analysis and design. EPA's methodology for systems development specifies the following stages in the software life cycle:

1) Mission Needs Analysis
2) Preliminary Design and Options Analysis
3) System Design
4) System Development
5) System Implementation
6) System Improvement Plan
7) Software Improvement Increment
8) Software Obsolescence and Disposal

The ADP auditor must "safeguard software, trace computer transactions, review the systems development cycle, and monitor adherence to administrative policy" (Gallegos and Bieber, 1986, p. 2).

> To make sound managerial decisions, organizations need properly authorized, complete, accurate, and reliable data. Achieving the data integrity objective requires that systems have adequate controls over how data are entered, communicated, processed, stored, and reported. (GAO, 1986, p. 11).

Auditing imposes controls to maximize one's assurance of data integrity. The concepts of risk and control provide a useful perspective for reviewing laboratory automatic data processing.

Legal Validity

Ultimately, data integrity is important so that one may have confidence in the conclusions that are drawn from the data. In general, automated data base systems must maintain audit trails and provide a complete chain of custody for data from which conclusions with legal implications will be

drawn (see Glover *et al.*, 1982). At present, it is typically the case that the chain of custody must be documented in court using hard-copy data.

In automated financial systems, it is currently the case that banking institutions are required by Federal statute to generate paper trails of all financial transactions. In other words, no matter how carefully financial systems implement procedures to control risks, they must provide written documentation of all activity. The Electronic Fund Transfer Act, which regulates ATMs and other aspects of the "cashless society," states:

> For each electronic fund transfer initiated by a consumer from an electronic terminal, the financial institution holding such consumer's account shall, directly or indirectly, at the time the transfer is initiated, make available to the consumer written documentation of such transfer [EFTA, 1979: Paragraph 906, Section (a)].

For a listing of the Federal Reserve Board's regulations in conformance with the Electronic Fund Transfer Act, see the Board of Directors of the Federal Reserve (1979). For further comment on the Act, see Schroeder (1983).

Summary and Conclusions

The financial industry's experience with automated financial systems standards is well applied to the automated laboratory environment. The financial industry demonstrates that reasonable levels of assurance of that integrity can be achieved through careful use of controls and backups. Integrity of computer-resident data in automated laboratory systems depends simply on whether systems are designed with appropriate controls. Although specific controls applied will differ from between the two applications, in general, the risks the data bases are subject to and controls that can be considered to counter the risks are similar. Examples of these are shown below:

RISKS	CONTROLS
Incorrect data are added to the data base.	Re-key the data Produce system control report Include a validation step
The data base is interfered with and data integrity is damaged	Restrict access Employ a security officer Train personnel
Data base is affected by multiple applications	Develop standard procedures Use concurrent data control Create data ownership Develop data reconciliation utility
Failure of data base management system to function as specified	Develop routine procedures and forms to report problems

Integrity of computer-resident data is important so that one may draw valid conclusions from such data — including conclusions that will hold up in court. Even ADP systems in the financial and banking industry must maintain paper trails of transactions should those transactions be challenged in court. Despite rapid and complex developments in automated data processing, the need and expectation still exist that only hard-copy paper trails

constitute generally accepted evidence in support of conclusions drawn from computer-resident data.

Systems concerned with financial matters have developed standards to achieve reasonable levels of assurance of data integrity. Thus, it seems fair to conclude that systems concerned with issues of public health and the environment — e.g., automated laboratory data processing — would require similar levels of assurance of data integrity and the controls required to achieve such levels.

GLOSSARY

Application controls - one of the two sets or types of controls recognized by the auditing discipline. They are specific for each application and include items such as data entry verification procedures (for instance, re-keying all input); data base recovery and roll back procedures that permit the data base administrator to recreate any desired state of the data base; audit trails that not only assist the data base administrator in recreating any desired state of the data base, but also provide documentary evidence of a chain of custody for data; and use of automated reconciliation transactions that verify the final data base results against the results as reconstructed through the audit trail.

Application software - a program developed, adapted, or tailored to the specific user requirements for the purpose of data collection, data manipulation, data output, or data archiving [Drug Information Association].

Audit trail - records of transactions that collectively provide documentary evidence of processing, used to trace from original transactions forward to related records and reports or backwards from records and reports to source transactions. This series of records documents the origination and flow of transactions processed through a system [Datapro]. Also, a chronological record of system activities that is sufficient to enable the reconstruction, reviewing, and examination of the sequence of environments and activities surrounding or leading to an operation, a procedure, or an event in a transaction from its inception to final results [NCSC-TG-004].

Auditing - (1) the process of establishing that prescribed procedures and protocols have been followed; (2) a technique applied during or at the end of a process to assess the acceptability of the product. [Drug Information Association]; (3) a function used by management to assess the adequacy of control [Perry]. That is, auditing is the set of processes that evaluate how well controls ensure data integrity. As a financial example, auditing would include those activities that review whether deposits have been attributed to the proper accounts; for example, providing an individual with a hard-copy record of the transaction at the time of deposit and sending the individual a monthly statement that lists all transactions.

Automated laboratory data processing - calculation, manipulation, and reporting of analytical results using computer-resident data, in either a LIMS or a personal computer.

Availability - see "data availability."

Back-up - provisions made for the recovery of data files or software, for restart of processing, or for use of alternative computer equipment after a system failure or disaster [Drug Information Association].

Change control - ongoing evaluation of system operations and changes during the production use of a system, to determine when and if repetition of a validation process or a specific portion of it is necessary. This includes both the ongoing, documented evaluation, plus any validation testing necessary to maintain a product in a validated state [Drug Information Association].

Checksum - an error-checking method used in data communications in which groups of digits are summed, usually without regard for overflow, and that sum checked against a previously computed sum to verify that no data digits have been changed [Drug Information Association].

Cipher - a method of transforming a text in order to conceal its meaning.

Confidentiality - see "data confidentiality."

Control - "that which prevents, detects, corrects, or reduces a risk" [Perry, p.45], and thus reasonably ensures that data are complete, accurate, and reliable. For instance, any system that verifies the sample number against sample identifier information would be a control against inadvertently assigning results to the wrong sample.

Computer system - a group of hardware components assembled to perform in conjunction with a set of software programs that are collectively designed to perform a specific function or group of functions [Drug Information Association].

Data - a representation of facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, or processing by human or automatic means [ISO, as reported by Drug Information Association].

Data availability - the state when data are in the place needed by the user, at the time the user needs them, and in the form needed by the user [NCSC-TG-004-88]' the state where information or services that must be accessible on a timely basis to meet mission requirements or to avoid other types of losses [OMB]. Data stored electronically require a system to be available in order to have access to the data. Data availability can be impacted by several factors, including system "down time," data encryption, password protection, and system function access restriction.

Data Base Management System (DBMS) - software that allows one or many persons to create a data base, modify data in the data base, or use data in the data base (e.g., reports).

Data base - a collection of data having a structured format.

Data confidentiality - the ability to protect the privacy of data; protecting data from unauthorized disclosure [OMB].

Data element (field) - contains a value with a fixed size and data type (see below). A list of data elements defines a data base.

Data integrity - ensuring the prevention of information corruption [modified from EPA Information Security Manual]; ensuring the prevention of unauthorized modification [modified from OMB]; ensuring that data are complete, consistent, and without errors.

Data record - consists of a list of values possessing fixed sizes and data types for each data element in a particular data base.

Data types - alphanumeric (letters, digits, and special characters), numeric (digits only), boolean (true or false), and specialized data types such as date.

Electronic data integrity - data integrity protected by a computer system; automated data integrity refers to the goal of complete and incorruptible computer-resident data.

Encryption - the translation of one character string into another by means of a cipher, translation table, or algorithm, in order to render the information contained therein meaningless to anyone who does not possess the decoding mechanism [Datapro].

Error - accidental mistake caused by human action or computer failure.

Fraud - deliberate human action to cause an inaccuracy.

General controls - one of the two sets or types of controls recognized by the auditing discipline. These operate across all applications. These would include developing and staffing a quality assurance program that works independently of other staff; developing and enforcing documentation standards; developing standards for data transfer and manipulation, such as prohibiting the same individual from both performing and approving sample testing; training individuals to perform data transfers; and developing hardware controls, such as writing different backup cycles to different disk packs and developing and enforcing labelling conventions for all cabling.

Integrity - see "data integrity."

Journaling - recording all significant access or file activity events in their entirety. Using a journal plus earlier copies of a file, it would be possible to reconstruct the file at any point and identify the ways it has changed over a specified period of time [Datapro].

Laboratory Information Management System (LIMS) - automation of laboratory processes under a single unified system. Data collection, data analysis, and data reporting are a few examples of laboratory processes that can be automated.

Password - a unique word or string of characters used to authenticate an identity. A program, computer operator,or user may be required to submit a password to meet security requirements before gaining access to data. The password is confidential, as opposed to the user identification [Datapro].

Quality assurance - (1) a process for building quality into a system; (2) the process of ensuring that the automated data system meets the user requirements for the system and maintains data integrity; (3) a planned and systematic pattern of all actions necessary to provide adequate confidence that the item or product conforms to established technical requirements [ANSI/IEEE Std 730-1981, as reported by Drug Information Association].

Raw data - ". . . any laboratory worksheets, records, memoranda, notes, or exact copies thereof, that are the result of original observations and activities of a study and are necessary for the reconstruction and evaluation of that study. . . "Raw data" may include photographs, microfilm or microfiche copies, computer printouts, magnetic media, . . . and recorded data from automated instruments." [40 CFR 792.3] Raw data are the **first** or **primary** recordings of observations or results. Transcribed data (e.g., manually keyed computer-resident data taken from data sheets or notebooks) are **not** raw data.

Risk - "the probable result of the occurrence of an adverse event..." [Perry, p.45]. An "adverse event" could be either accidental (error) or deliberate (fraud). An example of an adverse event would be the inaccurate assignment of an accessionary number to a test sample. Risk, then, would be the likelihood that the results of an analysis would be attributed to the wrong sample.

Risk analysis - a means of measuring and assessing the relative vulnerabilities and threats to a collection of sensitive data and the people, systems, and installations involved in storing and processing those data. Its purpose is to determine how security measures can be effectively applied to minimize potential loss. Risk analyses may vary from an informal, quantitative review of a microcomputer installation to a formal, fully quantified review of a major computer center [EPA IRM Policy Manual].

Security - the protection of computer hardware and software from accidental or malicious access, use, modification, destruction, or disclosure. Security also pertains to personnel, data, communications, and the physical protection of computer installations [Drug Information Association].

System - (1) a collection of people, machines, and methods organized to accomplish a set of specific functions; (2) an integrated whole that is composed of diverse, interacting, specialized structures and subfunctions; (3) a group of subsystems united by some interaction or interdependence, performing many duties but functioning as a single unit [ANSI N45.2.10, 1973, as reported by Drug Information Association].

System Development Life Cycle (SDLC) - a series of distinct phases through which development projects progress. An approach to computer system development that begins with an evaluation of the user needs and identification of the user requirements and continues through system design, module design, programming and testing, system integration and testing, validation, and operation and maintenance, ending only when use of the system is discontinued [modified from Drug Information Association].

Transaction log - also **Keystroke, capture, report, and replay** - the technique of recording and storing keystrokes as entered by the user for subsequent replay to enable the original sequence to be reproduced exactly [Drug Information Association].

Valid - having legal strength or force, executed with proper formalities, incapable of being rightfully overthrown or set aside [Black's Law Dictionary].

Validity - legal sufficiency, in contradistinction to mere regularity (being steady or uniform in course, practice, or occurrence) [Black's Law Dictionary].

References

Black, Henry C. (1968), *Black's Law Dictionary*, Revised Fourth Edition (West Publishing Co., St. Paul, Minnesota).

Board of Governors of the Federal Reserve System (1979), "Electronic Fund Transfers," Regulation E (12 CFR Part 205), Effective March 30, 1979 (as amended effective May 10, 1980).

Datapro Research (1989), *Datapro Reports on Information Security* (McGraw-Hill, Inc., Delran, New Jersey).

Dice, Barry, Operations Manager, Sovran Financial Corp., Telephone Interview, April 25, 1990 (Hyattsville, Maryland).

Drug Information Association (1988), *Computerized Data Systems for Nonclinical Safety Assessment: Current Concepts and Quality Assurance* (Drug Information Association, Maple Glen, Pennsylvania).

Electronic Fund Transfer Act (1979), 15 USC sec. 1693 et. seq.

Gallegos, Frederick, and Doug Bieber, (1986), "What Every Auditor Should Know about Computer Information Systems," available as Accession Number 130454 from the General Accounting Office (GAO) and reprinted from p. 1-11 in *EDP Auditing* (Auerbach Publishers, Inc., 1986).

Glover, Donald E., Robert G. Hall, Arthur W. Coston, and Richard J. Trilling (1982), "Validation of Data Obtained During Exposure of Human Volunteers to Air Pollutants," *Computers and Biomedical Research* 15(3):240-249.

National Bureau of Standards (1976), *Glossary for Computer Systems Security* (U.S. Department of Commerce, FIPS PUB 39).

National Computer Security Center (1988) Glossary of Computer Security (U.S. Department of Defense, NCSC-TG-004-88, Version 1).

Office of Information Resources Management (1989). *EPA System Design and Development Guidance*, Vols. A, B, and C (U.S. Environmental Protection Agency, Washington, D.C.)

Perry, William E. (1983), *Ensuring Data Base Integrity* (John Wiley and Sons, New York).

Pinkus, Karen V. (1989), Financial Auditing and Fraud Detection: Implications for Scientific Data Audit. *Accountability in Research* 1:53-70.

Schroeder, Frederick J. (1983), "Developments in Consumer Electronic Fund Transfers," *Federal Reserve Bulletin* 69(6):395-403.

U.S. General Accounting Office (1986), *Evaluating the Acquisition and Operation of Information Systems* (General Accounting Office, Washington, D.C.).

U.S. General Accounting Office (1987), Bibliography of GAO Documents, *ADP, IRM, & Telecommunications 1986* (General Accounting Office, Washington, D.C.).