John W. Lainhart IV
Inspector General

Office of Inspector General
U.S. House of Representatives
Washington, DC 20515–9990

MEMORANDUM

TO: Jeff Trandahl
Acting Chief Administrative Officer

FROM: John W. Lainhart IV
Inspector General

DATE: May 8, 1997

SUBJECT: Audit Report - Direct Access Storage Device Management Can Be Improved (Report No. 97-CAO-10)

This is our final report on the efficiency and effectiveness of management controls surrounding House Information Resources' (HIR) Direct Access Storage Device (DASD) environment. The objectives of this audit were to (1) evaluate the current management procedures that have been established to control DASD resources, (2) determine the adequacy and effectiveness of the implementation and administration of the Data Facility Product (DFP), and (3) determine to what degree management has implemented security controls to ensure enforcement of DASD policy.

In this report, we identified fundamental management concepts and corresponding controls that are not only essential in the DASD mainframe environment, but transcend to all automated platforms. We found that HIR did not have a dataset naming convention standard to support effective data management practices. These practices include security controls that require the identification of the critical elements of a dataset such as its purpose, application, origin, and type. We also found that HIR management is not taking advantage of the DFP system by implementing and enforcing compliance with DFP's Storage Management Subsystem classes. Finally, we determined that HIR lacked an effective data retention policy addressing user, legal and regulatory retention requirements. Instead, data retention policy was developed based exclusively on operational concerns of storage management. We have addressed our concerns in the following audit report and made specific recommendations for corrective action.

In response to our December 5, 1996 draft report, your office concurred with our findings and recommendations. The March 17, 1997 management response is incorporated in this final report and included in its entirety as an appendix. The corrective actions taken and planned by your office are appropriate and, when fully implemented, should adequately respond to the recommendations. Further, the milestone dates provided for implementing corrective actions appear reasonable.

We appreciate the courtesy and cooperation extended to us by your staff. If you have any questions or require additional information regarding this report, please call me or Robert B. Frey III at (202) 226-1250.

cc. Speaker of the House
Majority Leader of the House
Minority Leader of the House
Chairman, Committee on House Oversight
Ranking Minority Member, Committee on House Oversight
Members, Committee on House Oversight

DIRECT ACCESS STORAGE DEVICE MANAGEMENT CAN BE IMPROVED

Report No. 97-CAO-10
March 8, 1997

RESULTS IN BRIEF

CONCLUSIONS

Although this audit focused on the Direct Access Storage Device (DASD) environment, the fundamental management concepts and corresponding controls that are discussed within this report are not only essential to the mainframe environment, but transcend all automated platforms as well.

The existing House Information Resources (HIR) dataset naming convention standard (Standard) does not support effective data management practices in today's computer environment. This Standard, which defines the existing HIR dataset naming conventions, was created for the primary purpose of identifying the owner of the dataset and not to establish effective site naming conventions. As a matter of policy, HIR adopted the philosophy of delegating the management and control of application datasets to the supporting application staff. As a result, HIR management is unable to implement effective data management and security controls that require the identification of the critical elements of a dataset such as its purpose, application, origin, and type.

HIR management has not taken advantage of the capability of the Data Facility Product (DFP) by fully implementing and enforcing compliance with DFP's Storage Management Subsystem (SMS) classes. This was due to the lack of an HIR storage management policy addressing the development and enforcement of site standards. An effective storage management policy requires the continuous study of the computer environment to identify and implement solutions for resource bottlenecks, as well as to foresee and satisfy future requirements. The development and enforcement of an effective storage management policy would have assisted HIR management in achieving full utilization of DFP resulting in increased performance and capacity.

HIR lacks an effective data retention policy to ensure that datasets are not deleted or migrated prematurely, or retained beyond their legal or accounting requirements. As noted, HIR has adopted the philosophy of delegating the management and control of application datasets to the supporting application staff. As a direct consequence, HIR management has not exercised a proactive approach concerning a dataset retention policy and believes its current management practices are sufficient. This negates the benefits of a more structured policy, which we believe would increase the effectiveness of storage management; identify dataset backup and recovery criteria; and institute effective dataset storage placement. A well established dataset retention policy accompanied by oversight controls would prevent datasets from being deleted or migrated prematurely contributing to the inefficient use of computer resources.

RECOMMENDATIONS

Key recommendations to the Chief Administrative Officer include establishing (1) enterprise-wide dataset naming convention standards and procedures; (2) a formal storage management policy and an accompanying storage administration function; and (3) a data retention policy.

MANAGEMENT RESPONSE

In his March 17, 1997 response, the Acting Chief Administrative Officer (CAO) formally concurred with the findings and recommendations in this report, and indicated that corrective actions have been initiated and are planned to address the report's recommendations. Key actions to be undertaken by the CAO include defining enterprise-wide dataset naming convention standards for new, MVS-based projects that will include, where appropriate and with standards that are unique to the platform addressed, the dataset qualifiers identified; constructing a plan to bring all new datasets into compliance with the new naming convention standards; and working closely with HIR Security to develop standards and procedures that will require that ACF2 dataset rules be established to enforce compliance with the new naming convention standards. The CAO also agreed to develop a formal storage management policy to better define how DASD resources are managed; define a data retention policy that ensures all MVS-based datasets have a defined retention period that meets the needs of the owners; address retention questions in the formal policies; revisit the issue of SMS management classes after the reassessment of the MVS-based environment is conducted; continue to run the DASD dataset aging report that HIR developed; conduct a re-evaluation of its usefulness; and make adjustments as necessary. Finally, the CAO agreed that non-SMS managed datasets will be more carefully monitored and reports will be created to assist staff in these efforts.

OFFICE OF INSPECTOR GENERAL COMMENTS

The CAO's planned actions are responsive to the issues we identified and, when fully implemented, should fully satisfy the intent of the recommendations.

TABLE OF CONTENTS

[This Page Intentionally Left Blank]

I. INTRODUCTION

Background

The House Information Resources (HIR) Enterprise Computing Group (ECG) provides both operational support and systems software support for HIR and external time-sharing customers including installing, maintaining, and configuring operating system software. Some of the products they support include the automated direct access storage device (DASD) management, MVS operating system, job entry system, automated report distribution, automated direct job scheduling, security software management, and robotic tape management hardware and software. Also, the staff is responsible for monitoring performance and analyzing various performance metrics in the areas of central processing unit capacity planning, DASD performance, benchmarking, online response time tracking/tuning, application design review, application installation and maintenance, and software configuration used for performance analysis. The annual budget for ECG management including all hardware, software, personnel and operating incidentals is approximately $6.0 million.

The Central Systems section within ECG management is responsible for installing, maintaining, and configuring the MVS operating system which includes the Data Facility Product (DFP) on the central processor (IBM 9021-720). MVS/DFP is the central component of both the system managed and non-system managed storage environments. MVS/DFP simplifies the management and use of DASD storage resources by providing a device-independent means of requesting services by electronic data (dataset). The component within DFP known as Storage Management Subsystem (SMS) responds to these requests by managing performance, availability, space, and allocation services according to hardware/software capabilities. SMS allows the ability to automatically delete obsolete datasets, remove wasted space, and migrate infrequently used datasets.

The DASD medium in the HIR mainframe environment stores information and is readily accessed and updated. All of the current disk storage was acquired by the House before May 1994 with most of it having been installed during the 1991 through 1993 time period. The HIR DASD environment contains 16 disk storage devices, consisting of a total of 522.8 Gigabytes (Billion Bytes - GB) of capacity. HIR utilizes an 80/20 traditional DASD configuration for which they assign approximately 20 percent of the allocated space to executive systems and 80 percent to users. As of August 15, 1996 there were 411GB allotted to the user area of which 248GB were being used leaving 163GB of available user storage capacity.

Objectives, Scope, and Methodology

This audit assessed the efficiency and effectiveness of the management controls surrounding the HIR DASD environment. We reviewed current policies, procedures, and practices used to perform and measure capacity and storage management. The objectives of this audit were to:

- evaluate the current management procedures that have been established to economically and

effectively control DASD resources,

- determine the adequacy and effectiveness of the implementation and administration of DFP, and

- determine to what degree management has implemented security controls to ensure enforcement of established DASD policy.

We conducted our review in accordance with *Government Auditing Standards* issued by the Comptroller General of the United States. The audit work included such tests and auditing procedures as were considered necessary under the circumstances. We conducted our field work during the period July 1996 through September 1996. Our review included an evaluation of disk storage utilization and DASD management and acquisition planning. We also reviewed HIR's implementation and management of DFP, the primary automated DASD management system utilized by HIR, and looked at other systems, methods or practices utilized by HIR to support and maintain DASD.

In conducting this review, we performed the following specific tasks:

- Reviewed applicable government-wide internal control criteria that addresses controls in computer-based systems. Reviewed industry related documentation and reference materials that address DASD storage devices and related issues, specifically focusing on control and utilization.

- Evaluated the adequacy and effectiveness of the implementation and administration of MVS/DFP.

- Interviewed HIR staff: ECG management, application programmers, and Computer Security management involved in DASD performance, space utilization, security, and capacity planning.

- Assessed the presence of sufficient and adequately trained personnel assigned to the management and administration of DASD resources.

- Evaluated the effectiveness of:

    - the current management procedures that have been established to control DASD resources.

    - the existing dataset retention policy governing migration, retention and deletion of datasets.

    - the existing standards and procedures for application online purge criteria.

- Assessed the efficiency of space allocation and use of DASD resident datasets.

- Determined the effectiveness of security controls and dataset naming conventions that have been established to ensure the enforcement of DASD policy.

- Evaluated the accuracy and timeliness of generated management reports relating to performance, availability, and space utilization of the disk storage environment.

- Reviewed user management reports for accuracy and usefulness in evaluating internal charges to identify cost-reduction opportunities.

Internal Controls

During this review, we evaluated internal controls over the management of DASD resources. The internal control weaknesses we identified are described in the "Findings and Recommendations" section of this report.

Prior Audit Coverage

No prior audits have been conducted regarding the efficiency and effectiveness of management controls relating to the HIR DASD environment.

[This Page Intentionally Left Blank]

II. FINDINGS AND RECOMMENDATIONS

Finding A: Dataset Naming Convention Standard Is Inadequate

The existing HIR dataset naming convention standard (Standard) does not support effective data management practices in today's computer environment. This Standard, which defines the existing HIR dataset naming conventions, was created for the primary purpose of identifying the owner of the dataset and not to establish effective site naming conventions. As a matter of policy, HIR has adopted the philosophy of delegating the management and control of application datasets to the supporting application staff. As a result, HIR management is unable to implement effective data management and security controls that require the identification of the critical elements of a dataset such as its purpose, application, origin, and type.

Discussion

The established philosophy, as professed by Enterprise Computing Group (ECG) management, is the foundation for which the current HIR dataset naming standard has been constructed. This standard only requires the identification of the owner of the dataset through the Resource Identification Code (RIC) as the first high level qualifier[1]. During exit conference discussions, ECG management advised that from their perspective they only consider the programmer to be the critical element in the dataset name, not elements such as purpose, application, or type. They further emphasized that it is the programmers who are responsible for the management of their respective application system, and not HIR management.

This approach runs contrary to the fundamental principles embedded in the System Development Life Cycle **(**SDLC) methodology, of which the CAO and HIR management have established as policy and are in the process of implementing. Further, as reflected in GAO's 1983 release of "Standards for Internal Controls in the Federal Government" (now known as Appendix II to GAO's Title 2-Accounting), it is management who should be responsible for the establishment and dissemination of written policies and procedures which govern the design, development, and modification of Computerized Information Systems (CIS); acquisition of information resources; operation of CIS; and implementation of internal controls over CIS activities, including confidentiality, integrity, availability requirements, and security, privacy, and freedom of information requirements.

As the fundamental building block of all CIS controls, the establishment of an effective dataset naming standard will allow management to define a set of procedures and detailed guidelines that can be followed when creating and naming a new dataset. This will not only allow management the ability to identify various system and application files, but also supply the critical information necessary to implement the management controls outlined within the June 1996 "U.S. House of Representatives Management Policy for System Development Life Cycle" document. In order to ensure the effectiveness of these controls, the identification of a dataset

[1] To ensure their uniqueness, a dataset name can contain up to 44 characters including periods. Each level of a name, called a qualifier, must conform with the naming convention standard and must be separated from other names by a period.

sponsor, application, origin, type and use is required. This information can be most efficiently communicated through the name of the dataset.

The more structured and standardized the name of a dataset, the more knowledge that can be gained. In a sense, the names of datasets can be equated to individual account numbers within a general ledger of accounts. The scheme used for numbering accounts can bring order or chaos to financial management. The same is true for data. A Standard should organize dataset names from an enterprise-wide perspective. This allows not only the immediate visual recognition of the purpose of the datasets, but also promotes the effectiveness of the supporting executive systems[2]. Industry best practices point to a Standard that includes a set of specific statements embodying control requirements suitable for achieving management's goals to promote and support a more effective and secure data management environment. Standard naming conventions for datasets vary depending on the category of the dataset. Such categories are depicted in Figure 1 below. A Standard is the principle building block on which all fundamental management controls depend, such as security, data storage administration, change management, and disaster recovery.

SYSTEM	Executive software (acquired) files, system catalogs, and general, shared libraries.
APPLICATION	Application system data files and libraries.
GROUP	General purpose files for a Department, Office, Member, or Project.
USER	Files for specific USERID.
TEMPORARY	Temporary files created on job execution.

Figure 1 Categories of a Dataset

Security

As management's primary preventative and detective tool, the ACF2[3] security system places significant reliance on the dataset name in order to ascertain the validity of an access request to a dataset. When dataset names clearly identify the sponsor, system, subsystem, type, and use of the data, they enable the creation of fewer, more effective ACF2 access rules[4]. Hypothetically, if an environment contained a population of 100 datasets, this could translate into 100 separate access rules. Whereas, these same datasets renamed to comply with a new naming convention, identified by category, could be reduced to ten access rules. Typically, such rules give more consistent and comprehensive protection. Fewer rules also reduce the workload for the system and the administrative burden on the security staff and application system owners.

During the exit conference, ECG management acknowledged that fewer than 10 percent of the existing datasets have an associated ACF2 access rule. ECG management stated that because the

[2] Executive systems control, configure and/or maintain the environment in which the application software reside, e.g., ACF2, MVS/DFP.

[3] Access Control Facility 2 (ACF2) provides enhanced system access and data security for the HIR mainframe environment. ACF2 prevents illegal attempts to access system resources or data and logs those attempts.

[4] ACF2 Access Rule - the sharing of data based upon a set of data access rules. There are four types of access rules: Read; Write; Allocate; and Execute.

ACF2 security system's approach is "access denied unless granted", the security access controls are considered adequate. ECG management further noted that it is the application project leader or the supporting application programmer who is responsible for establishing further access controls. This rationalization stems from ECG management's philosophy that advocates a more passive position regarding the management and control of application datasets. This is further reflected in the current configuration of ACF2 rules which translates into the granting of unlimited access to all project members who are identified with a corresponding RIC. This ACF2 system default has been allowed to become the defacto security control to restrict access.

Best practices dictates that security access controls should be designed as a layer of barriers, and reliance should not be placed on any one barrier. Access controls for a dataset should not default to only one global access rule, but rather should address the activity (i.e., read, update or delete), type of dataset (production versus test), and identity of the user. These levels of access control points can translate into a corresponding ACF2 rule. The absence of a definitive set of access rules defeats the ability to implement critical management controls such as separation of duties, data integrity, security detection and monitoring, and change management.

Conversely, the implementation of a viable dataset naming convention will provide the application system owners (i.e., Finance, Office of Procurement and Purchasing, Human Resources, etc.) and HIR management the ability to implement effective ACF2 rules which will then represent a solid foundation for other management controls to be enforced.

Data storage administration

The HIR mainframe environment has installed a number of key executive systems to support and automate DASD management which collectively are referred to as DFP. DFP optimizes the efficient use of limited DASD resources by organizing the storage of datasets based on their type and utilization. These products examine the name of the dataset to distinguish between the different types and uses of data in performing their management functions. They can increase system throughput and reduce the need to purchase additional, expensive DASD. In order for HIR to obtain the greatest benefits from these products, the comprehensiveness of the Standard must be improved.

Change management

As part of the SDLC process, an information system will continue to evolve over its lifetime as user and site requirements change. As the term "change" implies, this process will transform the existing production program(s) to reflect those requirements. To ensure the accuracy of this change, exhaustive testing is performed before the modified version is allowed to be migrated over to the production environment. In order to maintain the integrity of the production environment, it is critical to be able to differentiate between the program that is in the testing phase versus the current production version.

To accurately identify and control access to a "test" dataset versus a "production" dataset, the ACF2 security software requires a unique identifier (qualifier) contained within the dataset name

to allow rapid and precise authentication. Since HIR has not effectively established and strongly enforced a Standard, it is possible for actual production software to become contaminated with inaccurate and theoretically unauthorized changes. Without this pivotal control at the very outset of the change control process, all other control procedures become moot.

Disaster recovery

In recent years HIR users have grown to depend on the continued, uninterrupted processing of information by their data processing department. This dependence has also increased the responsibility of HIR management to maintain a viable disaster recovery plan that ensures the continued availability of its computer resources. Any interruptions of this service can cause untold havoc and may result in the loss of a critical system and potentially have a negative impact on public confidence. One of the key components of a disaster recovery plan is the identity or naming convention assigned to the critical systems classified as essential to the survival of the organization. An installation's minimum requirements in this crisis necessitates the portability of datasets, as well as the smooth transition of system re-initialization. Without a well entrenched Standard, the successful recovery from a major disaster may be in jeopardy.

In conclusion, each of the issues reviewed requires well established and strictly enforced dataset naming conventions to ensure their effective implementation. We recognize that identifying and converting all existing datasets to the new Standard at one time can be cost prohibitive and it is clearly not our intention to make such a recommendation. Other institutions, faced with this same task, have taken a phased-in approach, whereby compliance is required for all new datasets and a segmented conversion plan is developed to incorporate all existing datasets using a logical system by system conversion process.

Failure to adopt a site Standard will adversely impact the implementation of management controls such as security and DASD storage management. Equally important, the inability to adequately implement these management controls will prevent HIR from fully benefiting from their investment in the existing executive systems such as ACF2 and DFP.

Recommendations

We recommend that the Chief Administrative Officer:

1. Establish enterprise-wide dataset naming convention standards which require, at a minimum: uniquely identified datasets; the identification of the owner of each dataset; and data management and security controls that distinctly identify the category, system, subsystem, environment, function, type, and content of each dataset in the system.

2. Require compliance with the dataset naming convention standards for all newly created datasets.

3. Establish and commence execution of a plan, including interim target dates, to systematically convert all dataset names in a phased approach, to comply with the new naming convention

standards. Consideration should also be given to starting this exercise with the scheduled maintenance process that is already in place.

4. Establish standards and procedures that require each ACF2 dataset rule to comply with naming convention standards.

Management Response

The Acting CAO concurred with the recommendations in this finding. In his March 17, 1997 response, the Acting CAO indicated ECG will (1) define enterprise-wide dataset naming convention standards for new projects that will include, where appropriate and with standards that are unique to the platform addressed, the dataset qualifiers identified in this recommendation, (2) include only those MVS-based applications that are not planned to be discontinued or migrated, (3) construct a plan for the inclusion of all new datasets to be in compliance with the new naming convention standards, and (4) by working closely with HIR Security, develop standards and procedures that will require that ACF2 dataset rules be established to enforce compliance with the new naming convention standards. The Acting CAO will submit the new dataset naming standards to the Committee on House Oversight (CHO) by December 31, 1997, and will complete the corrective actions identified in approximately 90 days from the date of CHO approval.

Office of Inspector General Comments

The planned actions are responsive to the issues we identified and, when fully implemented, should satisfy the intent of our recommendations.

Finding B: Absence Of A Storage Management Policy Limits The Effectiveness Of The Data Facility Product

HIR management has not taken advantage of the capability of DFP by fully implementing and enforcing compliance with DFP's Storage Management Subsystem (SMS) classes. This was due to the lack of an HIR storage management policy addressing the development and enforcement of site standards. An effective storage management policy requires the continuous study of the computer environment to identify and implement solutions for resource bottlenecks, as well as to foresee and satisfy future requirements. The development and enforcement of an effective storage management policy would have assisted HIR management in achieving full utilization of DFP resulting in increased performance and capacity.

Discussion

When used effectively, SMS, through the use of data, management and storage classes, can automatically delete obsolete datasets, remove wasted space, and migrate infrequently used datasets. It can also balance space utilization across volumes in storage pools. SMS allows application programmers to exploit the performance features of advanced hardware and place datasets on devices most likely to meet their performance requirements. It also allows application programmers to specify availability, and space utilization requirements for different types of data, as well as help automate storage management tasks to provide effective service levels.

Most users are not experts in choosing the most efficient device for their application, yet they are actively involved in matching data space requirements to storage requirements and in moving data from one device to another. Application programmers manage storage only for their own needs, so they do not always know how their actions affect system performance, availability, or space utilization. In contrast, the storage administrator knows that files should be deleted, yet has no authority to delete them; knows that there is a large amount of wasted space, yet cannot reclaim it; knows that the system needs tuning, yet cannot control the position of specific datasets; or knows that policies and guidelines are being ignored yet can do very little about it. The operating system simply processes data as fast as it can, without making control decisions. This is because the system is often constrained by what the application programmer specifies for a particular job.

With a properly implemented, fully functioning SMS, the storage administrator can match the logical needs of the application programmers data to the physical characteristics of storage devices without requiring them to know or understand the HIR data center hardware configuration. When properly utilized, SMS by system default is allowed to automatically calculate the optimum block size. Empowering SMS to calculate the blocking factor also means that the block size will be recalculated if the dataset is moved to a different device. Although HIR has implemented the SMS data class block size attributes, compliance by application programmers is strictly on a voluntary basis. If redefined as a site standard, this would ensure that those qualifying datasets would be in compliance with the attribute, taking full advantage of SMS.

However, HIR management has not actively taken full advantage of the SMS data class capabilities. A data class is a named collection of dataset and space attributes that are assigned to a dataset when it is created. Dataset attributes contained within the JCL (Job Command Language), can be managed effectively with these data classes. Such attributes are depicted below.

BLKSIZE: System-determined block size will be used when the dataset is allocated. This ensures optimum block sizes, improves performance and is device independent.	AVEREC: In conjunction with the SPACE parameter, this parameter tells SMS to allocate space, thereby removing overallocation concerns and ensuring device independence.
UNIT/VOL=SER: This parameter is no longer needed, volume selection is automatically controlled by SMS.	DATACLAS: This parameter is used to identify the established data class name containing specific DCB* and space attributes, such as AVEREC, SPACE* and DSORG*.

Figure 2 Data Class Attributes

Through discussions with application programmer staff, we noted that the awareness of the availability and usefulness of these SMS attributes was not apparent. As a result, management of HIR DASD resources has become more labor intensive than need be, defeating the efficiencies offered by SMS. Compounding this deficiency, HIR management has not created any monitoring or administration mechanisms (i.e., trend analysis, capacity planning) to determine the degree to which efficiencies of the SMS system are being employed.

HIR's storage management procedures need to be changed in order to achieve a more efficient storage management environment. Based on programmer-specified needs, the storage administrator needs to create corresponding SMS classes that define space, availability, performance services and data definition attributes. Application programmers could then store and retrieve data without needing to be aware of device characteristics, or the type of storage media they were utilizing. As the SMS classes are defined, datasets can then be linked to the appropriate class and ACF2 dataset security rule to ensure compliance. The SMS subsystem, in concert with ACF2, would then allow the storage administrator to automate and optimize storage management.

By developing a storage management policy and defining site standards, HIR will be able to enforce compliance with SMS data classes. This will allow HIR to take full advantage of DFP resulting in a more efficient storage management environment today and assist HIR in effectively adapting to environment changes in the future.

Recommendations

We recommend that the Chief Administrative Officer:

* DCB - Dataset Control Block is a control category that contains subparameters used for the creation of non-temporary datasets; SPACE - requests quantity and type of space to be allocated; DSORG - specifies the organization type of space to be allocated.

1. Establish a formal storage management policy that economically and effectively addresses DASD resources, to include, but is not limited to:

    a) management reports relating to performance, availability, and space utilization of the DASD environment.

    b) procedures to determine the appropriate mix of DASD technology to best meet processing requirements.

    c) the capacity planning process that projects future DASD needs, but requires effective utilization of DASD resources as a precursor to future acquisition.

2. Develop oversight procedures to ensure site compliance with the data retention standards as recommended in Finding C, to include:

    a) determining that a retention period has been assigned for all production datasets.

    b) ensuring that SMS management classes have been created to ensure compliance with established retention periods.

    c) generation of user retention expiration reports, as well as the corresponding distribution and reporting procedures.

3. Establish a storage administration function whose duties and responsibilities will be to oversee the development, implementation and enforcement of the HIR storage management policy; the administration and control of DFP; and the:

    a) development of a methodology for conducting trend analysis of data storage utilization which will present management with a reliable predictor, to include performance, availability, and space utilization of the DASD environment.

    b) establishment of and compliance with application on-line purge procedures.

    c) establishment and maintenance of effective and efficient SMS data classes.

    d) establishment and maintenance of effective and efficient SMS management classes, as discussed in Finding C, Recommendation 2.

    e) implementation of procedures to monitor to what extent HIR datasets have been converted to DFP control as discussed in Finding C, Recommendation 3.

    f) implementation of procedures for the systematic monitoring and migration of datasets to tape as discussed in Finding C, Recommendation 5.

    g) establishment and implementation of a storage management awareness program.

4. Ensure the storage administration function identified in Recommendation 3 above will oversee the implementation of security controls to ensure enforcement of established DASD policy that includes the integration of:

    a) dataset naming conventions.

    b) ACF2 dataset rules.

    c) SMS management classes.

    d) SMS data classes.

    e) SMS storage classes.

Management Response

The Acting CAO concurred with the recommendations in this finding and indicated ECG will define a formal storage management policy to better define how DASD resources are managed and submit this new policy to the CHO by December 31, 1997. Implementation of this policy as well as the corrective actions identified, will be completed approximately 30 days from the date of CHO approval. In addition, ECG will develop oversight procedures to ensure site compliance with the data retention standards as recommended in Finding C, and will (a) advertise in the "Computer Center Handbook" the availability of additional retention periods beyond the one-year default--formal retention policies will also be defined, (b) work with the users to insure that SMS management classes provide the level of retention they need, and (c) develop and generate hard-copy user retention reports and the corresponding distribution and reporting procedures for users of MVS-based applications, and continue the three types of paperless notifications currently in place. These corrective actions will be implemented 60 days after CHO approval of the DASD Storage Management Policy.

ECG will also develop a methodology that encompasses the duties and responsibilities of a storage administration function identified in this recommendation. This methodology will be implemented and in place six months after CHO approval of the DASD Storage Management Policy. Further, ECG will ensure that the storage administration methodology as previously discussed, includes oversight procedures that, after the new dataset naming conventions are established, will ensure the linkage of the ACF2 dataset rules to the new datasets. For distributed systems, ECG will also conduct a cost benefit analysis to determine the impact of linking distributed system datasets (system file names) to a security system. The results and accompanying recommendation of this study will be completed and forwarded to the CHO within 90 days after the DASD Storage Management Policy has been approved. Work will be completed approximately 90 days after Committee approval of the study recommendation. Finally, ECG will address the security controls surrounding the SMS classes on the MVS platform.

Acknowledging that the current direction of the House is toward full migration from the mainframe to the client-server platform, ECG believes it may not be cost beneficial to invest a high priority level of effort into linking the ACF2 security system with the SMS classes. However, once the first step has been completed, they will conduct a reassessment of the environment, security controls, and migration issue to determine the most effective course of action. At the conclusion of this assessment, HIR/CAO will present the results to the Inspector General and request his comments and opinion. Both Enterprise Computing and HIR Security will work together to ensure that whatever policy is proposed will be enforced.

Office of Inspector General Comments

The planned actions are responsive to the issues we identified and, when fully implemented, should satisfy the intent of our recommendations.

Finding C: A Viable Data Retention Policy Needs To Be Developed

HIR Datasets are migrated and deleted based on operational concerns of storage management, not based on user, legal or regulatory retention requirements. Best practices observed by most Federal agencies and private industry prescribe an effective data retention policy that ensures datasets are not maintained beyond their legal requirement or usefulness. As was discussed in Finding A, HIR has adopted the philosophy of delegating the management and control of application datasets to the supporting application staff. As a direct consequence, HIR management has not exercised a proactive approach concerning dataset retention policy and believes its current management practices are sufficient. This negates the benefits of a more structured policy, which we believe would increase the effectiveness of storage management; identify dataset backup and recovery criteria; and institute effective dataset storage placement. A well established dataset retention policy accompanied by oversight controls would prevent datasets from being deleted or migrated prematurely contributing to the inefficient use of computer resources.

Discussion

Data retention for the classification of datasets is based on function and purpose. As part of the development process of an application system, end-users should determine and document the legal and functional life cycle of all generated datasets. These attributes can be identified through management classes defined within DFP which specify the length of time a dataset may remain on a medium (i.e., DASD) until migration, backup, retention, or deletion. Overall, data retention policy takes several unique factors into consideration when building management classes. These factors include:

- how frequently the datasets are accessed,
- legal or organizational requirements for retention,
- the cost to move the datasets to tape or other storage media and the cost of restoring data back to DASD,
- computer system performance considerations, and
- loss of productivity due to the time required to access data sets that are not immediately available on DASD.

HIR currently migrates datasets to tape using the Hierarchical Storage Manager (HSM), a subsystem of DFP, which manages DASD space availability through a set of storage device hierarchies or management levels. Although HIR has set default migration periods based on dataset size or Customer Information Control System[5] region using HSM, HIR has not established a data retention policy which would have included management classes to assist in the migration of DASD to tape. It should be noted that current default migration periods (the longest being 40 days) have been created strictly to address operational concerns**.** Once these thresholds are met, datasets are subsequently migrated to tape and tagged for deletion after a

[5] Customer Information Control System (CICS) is a general purpose mainframe based data communication system that can support a network of many hundred of terminals.

"default" retention period of one year of inactivity. Operational policies like these are based primarily on the concerns of DASD resources versus the performance and retention needs of an application dataset. Rather than having a "default" retention period, HIR should take a proactive stance and implement a policy that addresses management classes that are available to the supporting application staff who can more closely deal with the needs of the end-user.

As was discussed with ECG management during the exit conference, they believe that most supporting application staff are not experts in choosing efficient device placement and performance considerations and that the supporting application staff, in self-interest, will choose high performance DASD devices and/or the "Never Delete" file management option. If this concept was factual, then a "default data retention policy" would suffice. However, we believe that this argument disavows the expertise of the application programmer's ability to choose appropriate performance and resource parameters. These are the same programmers that HIR management depends on for the efficient and reliable execution of the various application production systems. We believe that by taking a proactive approach, ECG management will realize significant benefits from a joint review and selection of production dataset criteria comprised within a Formal Dataset Retention Policy.

The absence of a formal data retention policy may cause critical datasets to be deleted or migrated prematurely. These datasets may include significant transaction files required for backup and recovery requirements as well as for audit trails. Often, pertinent data files such as administrative and financial records need to be retained to adhere to Federal and State statutes, and the loss of this significant information could negatively impact management. Further, performance and cost should also be considered in determining on which storage medium (DASD/tape) a dataset should reside. For example, a monthly transaction dataset with a high access rate would most likely be stored on DASD, whereas a quarterly transaction file that is retained for auditing purposes would more than likely be stored on tape. In contrast, datasets retained beyond their legal or accounting requirements can contribute to the inefficient use of computer resources. Increased storage costs such as additional hardware, floor space, power, air conditioning, and personnel may be incurred.

Ultimately, failure to properly manage DASD may lead to the continuing maintenance of inactive or outdated files and contribute to the premature procurement of additional unneeded storage. Proper management over disk storage should ensure that datasets are monitored and that data has been migrated according to an established data retention policy. Low activity data should be migrated either automatically or by DFPHSM commands. To demonstrate the significance of this point, we reviewed all (20,935) cataloged datasets (not including VSAM

datasets[6]) residing on DASD as of July 16, 1996. The objective of this test was to determine the effectiveness of the default migration policy of 40 days that requires the migration of all inactive

[6] A key-sequenced dataset or file with an index containing extensive dataset and volume information that Virtual Storage Access Method (VSAM) requires to locate datasets or files, allocate and deallocate storage space, verify the authorization of a program or operator to gain access to a dataset or file, and accumulate usage statistics for datasets or files.

datasets to tape. As depicted below in Figure 2, we discovered that a total of 2,850 datasets (14 percent) ***had not been accessed*** for over 40 days and of these, 937 (5 percent) ***had not been accessed*** for over two years.

	Status as of July 16, 1996		*Status as of August 15 1996*	
	Number of Inactive Datasets	**Percent of Datasets Inactive**	**Number of Inactive Datasets**	**Percent of Datasets Inactive**
GT 40 days	1,913	9%	977	5.5%
GT 2 years	937	5%	135	.76%
Total Datasets Inactive	2,850	14%	1112	6.26%

Figure 3 Status of Inactive Datasets

After we had brought these results to the attention of HIR, we were told that most of these datasets remained on DASD as a result of an interface failure that occurred in May 1994 between the Control-D[7] and SMS systems. The absence of DASD monitoring controls exacerbated this problem and was a determining factor that allowed these files to remain long after they should have been deleted. We conducted a follow-up review of DASD datasets as of August 15, 1996, and noted that HIR had subsequently migrated and/or deleted 802 of the 937 datasets.

Figure 3 below depicts that in addition to the failure that occurred approximately two years ago, there were still a number of older datasets that had not been accessed for some time, some of these datasets were as old as 11 years.

Time Elapsed Since Datasets Were Last Accessed					
Time	***2 Years***	***3-4 Years***	***5-7 Years***	***8-9 Years***	***10-11 Years***
Datasets	867	58	7	3	2

Figure 4

Proper management controls over disk storage would ensure that datasets stored on disks (whether they are SMS-managed or non-SMS managed) are reviewed periodically to maintain that those that have not been accessed for a certain time period should be considered inactive and thus migrated.

The current default migration and deletion procedures, which are based on resource management requirements, have been allowed to become the defacto data retention policy. During our discussions with ECG management, they stated that "HIR has effective and efficient SMS management classes, but they currently are not advertised". We believe that if management classes are not published in any formal HIR standard documentation or made available for technical reference to HIR staff, then for all practical purposes they do not exist. Also, these unpublished management classes are not available to the supporting application staff who are therefore left uninformed. DFP has been implemented by HIR in order to enhance the effectiveness of data management practices relating to DASD resources. Without a well

[7] A report distribution system that manages the printing of user reports. Reports may be decollated, moved from spool to compressed datasets, and archived or restored as requested by users.

documented policy supporting fundamental DASD management practices such as data retention, HIR cannot take full advantage of the benefits that could be achieved by DFP.

The most effective way to implement this process, after a data retention policy is established, would be to identify and then link the appropriate SMS management class within the corresponding ACF2 rule. HIR management could then develop a conversion plan to schedule dataset ACF2 rule updates by application system.

In addition, HIR management would also need to develop monitoring controls to ensure regular reviews of the DASD environment. The results of these reviews would provide a basis to evaluate and revise if necessary, the use and benefit of established SMS controls such as dataset migration and deletion as well as space allocation and use. In our view, HIR management would benefit from implementing oversight controls over datasets regardless of whether they are SMS-managed or non-SMS managed.

Recommendations

We recommend that the Chief Administrative Officer:

1. Establish a data retention policy that ensures all datasets have a defined retention period as determined by the owner.

2. Establish HIR-defined SMS management classes that define dataset availability, space and retention attributes.

3. Use SMS management classes that incorporate the factors discussed in this finding to implement the retention period recommended in number 1 above.

4. Establish standards and procedures that require each ACF2 dataset rule to be linked to a corresponding SMS management class.

5. Establish monitoring controls for DASD dataset activity. These controls should include, but not be limited to:

    a. DASD dataset aging report - this report will identify all datasets that are inactive beyond the site default migration period.

    b. Non-SMS managed dataset report- this report will identify those datasets which are under manual control and should correspond to the authorized exceptions.

Management Response

The Acting CAO concurred with the recommendations in this finding. In his March 17, 1997 response, the Acting CAO indicated ECG will establish a data retention policy that ensures all MVS-based datasets have a defined retention period that meets the needs of the owner and will

work with users to insure that they have the retention period they need for their data. The Acting CAO will submit the new data retention policy to the CHO by December 31, 1997, and implementation of this policy will be completed approximately 90 days from the date of CHO approval. ECG will also address retention questions in the formal policies that will be implemented 60 days after the DASD Storage Management Policy is approved. ECG management will revisit this issue after the reassessment of the MVS-based environment is conducted as indicated in the response to Recommendation 4, Finding B. Finally, ECG will continue to run the DASD dataset aging report that HIR developed on November 30, 1996, and conduct a re-evaluation of its usefulness and make adjustments as necessary. The non-SMS managed datasets will be more carefully monitored, and reports will be created by July 31, 1997 to assist staff in these efforts.

Office of Inspector General Comments

The planned actions are responsive to the issues we identified and, when fully implemented, should satisfy the intent of our recommendations.

[This Page Intentionally Left Blank]

III. OTHER MATTERS

Redundant Array Of Inexpensive Disks Procurement (RAID) Issue

During our audit, it came to our attention that HIR submitted a July 8, 1996 draft Request for Proposal (RFP) to the Committee on House Oversight, asking approval to procure RAID storage devices for both the mainframe computer and the client/server configuration which will replace the mainframe when the migration to client/server is completed. The RAID medium for the mainframe has an estimated cost of $1.3 million and consists of a mainframe-attached subsystem with a base configuration of 540GB while the RAID medium for the client/server configuration (costs estimates not provided in the RFP) will provide that platform with 120GB of storage. The mainframe version and client/server version of RAID are unique to their respective platforms, i.e., they are not interchangeable. In this report, we indicated that as of August 15, 1996, there were 411GB of DASD mainframe storage (out of 522.8GB) allotted to the user area, of which only 248GB (60 percent) were being used. The remaining, unused storage is approximately 163GB.

Considering the fact that HIR does not manage or monitor DASD utilization (i.e., it might have more if it were managed properly) and the possibility that additional storage may become free as a result of implementing our recommendations, it appears that HIR has more than adequate capacity to service its storage needs in the short-term. Furthermore, in light of the fact that the RAID medium is platform dependent and considering the House's plans to adopt a network-centric configuration, it is our view that a more prudent storage investment approach would favor a procurement that supports a client/server environment and which is an integral part of the overall client/server proposal.

During the exit conference, ECG management advised that the July 8th draft RFP is now obsolete and has been withdrawn. ECG management further advised that they have received quotes of $1/MB versus $1.85/MB quoted in the now defunct draft RFP, and that they will be proposing the procurement of approximately 320-360GB of RAID to replace the high-performance DASD currently installed at considerable cost savings. They further advised that they will conduct a cost benefit analysis, the results of which will be the major factor in determining what, if anything, will be proposed. We totally agree that this would be the most effective approach to ensure adequate support if justified.

Regarding RAID for the client/server platform, ECG management advised that the prevailing direction is to provide storage for a single platform, and to provide storage from the platform vendor. As an example, the Compaq e-mail servers do not have the ability to share unified storage across multiple servers with other systems This appears to make the argument that RAID for the client/server platform, requiring the sharing of large storage areas with dissimilar products (i.e., messaging and full-motion video), are still in the infancy stage and must await further development. Given these facts, we agree with HIR that RAID for client/server must be more extensively evaluated before procurement can begin with any confidence.

[This Page Intentionally Left Blank]

EXHIBIT

DEFINITIONS OF TECHNICAL TERMS

Terminology	Definition
Access Control Facility 2 (ACF2)	A general security system that can be added to various IBM operating systems. This product provides control over access to all resources under control of the operating system.
Address	A character or group of characters identifying a register, a particular part of storage, or some other data source or destination.
Address Space	The complete range of addresses available to a programmer.
Dataset	The major unit of data storage and retrieval, consisting of a collection of data in one of several prescribed arrangements and described by control information to which the system has access.
Data Facility Product (DFP)	DFP manages performance, availability, space, and allocation services according to hardware/software capabilities.
Direct Access Storage Device (DASD)	DASD refers to magnetic storage devices on which data is stored by magnetic recording on the flat surfaces of one or more disks that rotate in use. Such devices provide high speed, online access to data where rapid access is critical to the effectiveness of a computer application.
Hierarchical Storage Management (HSM)	A subsystem of System Managed Storage. It is responsible for migration of inactive datasets, backups of datasets, and freeing unused space in datasets on SMS packs.
Input/Output (I/O)	Pertaining to the software or hardware process that may be involved in a reading operation and, at a different time, in a writing operation.
Multiple Virtual Storage (MVS)	An IBM virtual storage operating system that gives each user an environment that is defined as an address space. It provides for the isolation and protection of one user from another in a multiprogramming system supporting many users concurrently.
Operating System	An integrated collection of service routines for supervising the sequencing and processing of programs by a computer. The operating system controls the allocation of resources, scheduling of programs, input/output, and the processing of data.
Storage Management Subsystem (SMS)	A component of MVS/DFP that is used to automate and centralize the management of storage by providing the storage administrator with control over data class, storage class, management class, storage group, and ACS routine definitions.

Office of the
Chief Administrative Officer
U.S. House of Representatives
Washington, DC 20515

MEMORANDUM

To: Robert B. Frey III
Deputy Inspector General

From: Jeff Trandahl JT
Acting Chief Administrative Officer

Subject: Draft Audit Report - Direct Access Storage Device Management Can Be Improved

Date: March 17, 1997

Thank you for the opportunity to comment on this draft audit report. We have carefully reviewed the draft audit report and its recommendations and are in general agreement. Specific comments on each recommendation follow.

Overview

HIR agrees that formal policies are needed to control DASD usage on all enterprise-wide computing platforms. However, since the DASD audit specifically addressed the MVS platform, and the software tools and facilities listed (ACF2, SMS) are MVS-based only, Enterprise Computing Group is limiting the response to the MVS platform as well. It is also agreed that the documented procedures in the "Computer Center Handbook" are not formal policies, and have not been distributed as formal policies would be.

In FY97, Enterprise Computing will be looking for opportunities in obtaining training in enterprise storage practices and concepts, to assist in establishing workable policies. It is expected that enterprise storage products will be evaluated for their usefulness at the House as they become available in the coming year. Funds have been budgeted for the acquisition of an enterprise storage product in FY97.

Audit Findings and Recommendations:

Finding A: "*Dataset Naming Convention Standard is Inadequate* "

Recommendations (Page 8)

1. "Establish enterprise-wide dataset naming convention standards which require, at a minimum, uniquely identified datasets; the identification of the owner of each dataset; and

data management and security controls that distinctly identify the category, system, subsystem, environment, function, type, and content of each dataset in the system.'

2. "Require compliance with the dataset naming convention standards for all newly created datasets."

3. "Establish and commence execution of a plan, including interim target dates, to systematically convert all dataset names in a phased approach, to comply with the new naming convention standards. Consideration should also be given to starting this exercise with the scheduled maintenance process that is already in place."

4. "Establish standards and procedures that require each ACF2 dataset rule to comply with naming convention standards.

Response: (1-4.) **Concur** HIR will define enterprise-wide data set naming convention standards for new projects that will include, where appropriate and with standards that are unique to the platform addressed, the fields identified in the recommendations. It is expected that MVS-based legacy applications that are planned to be discontinued or replaced will not be converted. HIR will construct a plan for the inclusion of new data set names to comply with the new naming convention standards. Enterprise Computing, working closely with HIR Security, will develop standards and procedures that will require that ACF2 dataset rules be established to enforce compliance with the new naming convention standards on the MVS platform. A recommendation for a new standard will be submitted to the CHO by 12/31/97, and the corrective action that responds to these recommendations will be completed approximately 90 days after the Committee has approved the recommendations.

Finding B: "*Absence of a Storage Management Policy Limits the Effectiveness of the Data Facility Product*"

Recommendations (Page 11)

1. "Establish a formal storage management policy that economically and effectively addresses DASD resources, to include, but is not limited to:

   a) management reports relating to performance, availability, and space utilization of the DASD environment,

   b) procedures to determine the appropriate mix of DASD technology to best meet processing requirements,

   c) the capacity planning process that projects future DASD needs, but requires effective utilization of DASD resources as a precursor to future acquisition."

Response: (I.) **Concur.** Although HIR defines storage management procedures for users of MVS-based systems in the "Computer Center Handbook," a formal policy will be established to better define how DASD resources are managed. Implementation of the formal policy and other corrective action that responds to this recommendation will be completed approximately 30 days after Committee approval of the recommended actions.

a. HIR systems staff is producing a series of jobs to provide additional information on the DASD environment, to avoid some of the problems that were highlighted in the audit.

b. Procedures will be developed (including SDLC, requirements, product evaluation, and technical evaluation) to try to best meet processing requirements. SMS (System Managed Storage) is currently used to place data where it is most appropriate for its performance requirements.

c. With better reports, HIR systems staff will be able to better project future DASD needs. It is expected that as legacy applications begin to be removed from production that additional space should not be needed.

2. "Develop oversight procedures to ensure site compliance to the data retention standards as recommended in Finding C, to include:

a) determining that a retention period has been assigned for all production datasets,

b) ensuring that SMS management classes have been created to ensure corresponding distribution and reporting procedures,

c) generation of user retention expiration reports, as well as the corresponding distribution and reporting procedures."

Response: (2.) **Concur.** HIR will develop oversight procedures to ensure site compliance to the data retention standards as recommended in Finding C, and implement other corrective action that responds to this recommendation sixty days after Committee approval of the DASD Storage Management policy.

a) The next version of the "Computer Center Handbook" following Committee approval of the policy will advertise to users the availability of additional retention periods beyond the one-year default, formal retention policies will also be defined. Retention periods beyond one year are available, but have not been advertised, and have been handled on an ad-hoc basis.

b) HIR systems staff will work with the users to insure that SMS management classes provide the level of retention they need.

c) HIR provides three types of paperless notification now, including voice mail messages to project leaders, a notify message through TSO (Time Sharing Option), and an SPF panel that gives users access to a file that includes all data sets about to be deleted. Enterprise Computing will also develop and generate hard-copy user retention reports, as well as the corresponding distribution and reporting procedures for users of MVS-based applications. All methods of user notification will be fully documented in the "Computer Center Handbook."

3. "Establish a storage administration function whose duties and responsibilities will be to oversee the development, implementation, and enforcement of the HIR storage management policy; the administration and control of DFP; and the:

a) development of a methodology for conducting trend analysis of data storage

utilization which will present management with a reliable predictor, to include performance, availability, and space utilization of the DASD environment,

b) establishment of and compliance with application on-line purge procedures,

c) establishment and maintenance of effective and efficient SMS data classes,

d) establishment and maintenance of effective and efficient SMS management classes; as discussed in Finding C, Recommendation 2,

e) implementation of procedures to monitor to what extent HIR datasets have been converted to DFP control,

f) implementation of procedures for the systematic monitoring and migration of datasets to tape as discussed in Finding C, Recommendation 5,

g) establishment and implementation of a storage management awareness program.'

Response: (3.) **Concur.** (This responsibility is shared between systems management staff, with different staff members performing different tasks in the total storage administration effort. It is not expected that a new position will be created or that any one existing staff member will be named to this position. DASD management policies will be established with input from management and senior systems staff. Systems management staff will develop additional reporting tools (reports) to be able to take a more active role in the area of storage management.) A methodology, that includes additional reports and procedures will be developed to monitor storage to insure that the inconsistencies identified in the audit on the MVS platform are addressed. This methodology will be implemented and in place by six months after Committee approval of the DASD Storage Management policy.

a. Some of the problems that were brought to light would have been caught sooner if the reports existed, although in many cases very small data sets (1 track) were involved, and because they were not using a significant amount of space, were not noticed.

b. Though on-line purge procedures exist, and are believed to be in compliance with application requirements, HIR acknowledges that they have not been formalized, and they will be incorporated in the DASD policy.

c. HIR systems staff evaluated the establishment of SMS data classes in the migration to SMS, and basically used the low-level identifier to determine data class. However, this issue will be revisited as the recommendations contained in this report are implemented.

d. There are SMS management classes in addition to the defaults. These will be identified in the "Computer Center Handbook," but Enterprise Computing staff will work with users who have requirements outside of the default, and will try to avoid having users waste space by specifying longer retention periods.

e. Data sets are continuing to be migrated to be SMS-managed. Some of the SSD volumes (particularly those owned by Communications Services) remain to be done, but the majority of disk is either SMS-managed, or by design, not under SMS control.

f. Reports will be developed that will be used to monitor whether data sets are being migrated to tape as specified in the published standards.

g. Enterprise Computing offered training on SMS as the migration was taking place, and worked with users as the various projects were effected. This training will be offered again, and will be offered on a regular basis, to help make programming staff aware of the benefits of System Managed Storage.

4. Ensure the storage administration function identified in Recommendation 3 above will oversee the implementation of security controls to ensure enforcement of established DASD policy that includes the integration of:

    a) dataset naming conventions,

    b) ACF2 dataset rules,

    c) SMS management classes,

    d) SMS data classes,

    e) SMS storage classes."

Response: (4) **Concur** - Enterprise Computing concurs that integrating ACF2 security to the future data set naming conventions is an issue that should be addressed. Enterprise Computing recognizes that after the new data set naming conventions are established, the first step will be the linkage of the ACF2 dataset rules to the new datasets.

For distributed systems, HIR will conduct a cost benefit analysis using the principles established through the SDLC (systems development life cycle) to determine the impact of this effort after approval of the DASD Storage Management policy. The results and a recommendation will be forwarded to the CHO within 90 days of the approval of the DASD Storage Management policy and work will be completed approximately 90 days after Committee approval of the recommendations based on the SDLC/Cost Benefit Analysis has been obtained.

Once this is accomplished, the next step will be to address the security controls surrounding the SMS classes on the MVS platform. Acknowledging that the current direction of the House is toward full migration from the mainframe to the client-server platform, it may not be cost beneficial to invest a high priority level of effort into linking the ACF2 security system with the SMS classes. However, once the first step has been completed, Enterprise Computing will conduct a reassessment of the environment, security controls and migration issue to determine the most effective course of action. At the conclusion of this assessment, HIR/CAO will formally advise the Office of the Inspector General of the result and request their comments and opinion. Both Enterprise Computing and the HIR Security will work together to ensure that whatever policy is proposed will be enforced.

Finding C "*A Viable Data Retention Policy Needs to be Developed*"

Recommendations (Page 17)

1. "Establish a data retention policy that ensures all datasets have a defined retention period as determined by the owner."

Response: (1.) **Concur**. Enterprise Computing will establish a data retention policy that ensures all MVS-based data sets have a defined retention period that meets the needs of the owner and will notify users through the "Computer Center Handbook" that additional retention periods exist beyond the default. Staff will work with users to insure that users have the retention period they need for their data. A recommendation will be forwarded to the Committee on House Oversight by 12/31/97, and the implementation of this policy will be completed approximately 90 days after Committee approval of the recommendations.

2. "Improve upon HIR-defined SMS management classes that define dataset availability, space, and retention attributes."

Response: (2.) **Concur**. SMS-managed data sets already are routinely backed-up. For availability, some volumes are dual-copied, but if they are required to be dual-copied, they must reside on particular volumes of DASD in the current environment. The dual copy feature is not intended for general use, and in particular is not meant to provide a redundant copy of every application. Space is determined through Storage Class, not the Management Class, and is already managed automatically. Again, retention questions will be addressed in the formal policies that will be implemented by 60 days after Committee approval of the DASD Storage Management policy, with a mention in the updated "Computer Center Handbook" that additional retention periods are available.

3. "Establish standards and procedures that require each ACF2 dataset rule to be linked to a corresponding SMS management class."

Response: (3.) **Concur** - See response to Recommendation B 4 above.

4. "Establish monitoring controls for DASD dataset activity. These controls should include, but not be limited to:

    a) DASD dataset aging report - this report will identify all datasets that are inactive beyond the site default migration period.

    b) Non-SMS managed dataset report - this report will identify those datasets which are under manual control and should correspond to the authorized exceptions.'

Response: (4.) **Concur**.

a. A report that identifies inactive data sets was created on 11/30/96, although this might not be the ultimate tool in tracking inactive data. Enterprise Computing will re-evaluate its usefulness and make adjustments as necessary. The two-page report

run on 12/12/96 indicates that of a total of 69 data sets, two were created in 1985, which are both model DSCB (data set control block) data sets, which SMS does not migrate, and both data sets need to be retained on the system.

b. The non-SMS managed data sets will be more carefully monitored, and reports are being created to assist the staff in these efforts. Once the remaining SSD volumes have been moved to SMS, only those volumes that by rule of thumb are not recommended to be SMS-managed (SYSRES, DLIB, some system volumes, and in HIR's case ADABAS volumes) will need to be monitored. The first of these reports on non-SMS managed datasets will be created by 7/31/97.

III. **Other Matters**

Redundant Array of Inexpensive Disks (**RAID**) Procurement

The acquisition of a RAID storage subsystem would radically change the way in which data is managed. Now, volumes are needed for SPARE and BACKUP. These volumes are reserved so that in case of a failure, they can be brought quickly into use and recovery efforts begun. These volumes (other than those in the string of 3390-9 high-capacity drives) would not be needed in a RAID environment. Today, some volumes are dual copied, either in anticipation of a possible failure, because of the availability requirement, or because rebuilding after a failure would be too time-consuming and would impact too many users. There would be no requirement for dual-copy in an environment that provides redundancy. And in the system being recommended for procurement, data compression is being used. This could also change the manner in which HSM (hierarchical storage manager) will be used to migrate data.

RAID would also change some of the SMS management classes, where performance is determined by "must cache..... may cache," or "never cache" specifications. Under RAID, all of the data is cached, and this specification becomes moot. Additional monitoring will be necessary by the systems staff to insure that the transition from DASD to RAID is efficient and effective.

Effective space usage is a concern, as it should be. However, even the most detailed reports do not accurately picture how DASD is actually being used, or how the volumes are reserved. The report for 12/12/96 shows that only 493GB are visible to the system, and that of that total, only 292GB are allocated. The system does not see the copy of volumes that are being dual-copied, which adds another 16.5GB to the total capacity number (or 509.5GB). The additional 12GB of total capacity that should be available does not show up at all. But more importantly, system volumes in the allocated space show up as being partially full, or in some cases (SPARE volumes), nearly empty. In fact, they are dedicated for system use, and should be considered "full."

If those volumes' capacity were calculated as being full, that would add another 79GB to the allocated total, bringing it to approximately 371 GB. However, disk usage at this time is believed to be fairly low. Since Members have been unable to send mailings to their constituents for 60 days before an election, less disk storage space is being used for NCOA (National Change of Address) jobs than normal. That application uses a substantial amount of disk storage, although data in that application is migrated to tape more quickly than the default period. To actually determine how much space is available to users, the three different types of DASD

(3390-3 high-performance, 3390-9 high capacity, and 3380K volumes) need to be viewed as separate and not equal storage environments.

Finally, Enterprise Computing expects to have some relevant plans and policies in place over the next six months. The audit has pointed out some perceived weaknesses in the area of DASD management. With the scheduled removal or migration of the largest legacy services, it is probably not cost-effective to invest money and personnel time to rename data sets in legacy applications. However, these issues are highly complex, and the work effort needs to be thoroughly evaluated. It may be necessary to contract for an acknowledged DASD authority, both to assess the effectiveness of the current SMS implementation and to suggest what recommendations in the audit could be cost effectively performed, which tasks must be done and the time frame for completion. However, for the actions that must be taken, time and funds will have to be budgeted to complete the necessary tasks.