FIPS PUB 42-1

FEDERAL INFORMATION
PROCESSING STANDARDS PUBLICATION
1977 MAY 15

U.S. DEPARTMENT OF COMMERCE / National Bureau of Standards

Guidelines

FOR BENCHMARKING ADP SYSTEMS IN THE COMPETITIVE PROCUREMENT ENVIRONMENT

CATEGORY: ADP OPERATIONS
SUBCATEGORY: BENCHMARKING
FOR COMPUTER SELECTION

U.S. DEPARTMENT OF COMMERCE • Juanita M. Kreps, *Secretary*
Dr. Betsy Ancker-Johnson, *Assistant Secretary* for Science and Technology
NATIONAL BUREAU OF STANDARDS • Ernest Ambler, *Acting Director*

Foreword

The Federal Information Processing Standards Publication Series of the National Bureau of Standards (NBS) is the official publication series relating to Federal standards and guidelines adopted and promulgated under the provisions of Public Law 89–306 (Brooks Act) and under Part 6 of Title 15, Code of Federal Regulations. Under P.L. 89-306, the Secretary of Commerce has important responsibilities for improving the utilization and effectiveness of computer systems in the Federal Government. In order to carry out the Secretary's responsibilities, the NBS, through its Institute for Computer Sciences and Technology, provides leadership, technical guidance, and coordination of Government efforts in the development of technical guidelines and standards in these areas.

Workload definition and benchmarking for the competitive procurement of Federal computer systems has proved to be a very costly process for both the computer user and the computer vendor. For some types of systems, such as those supporting large numbers of interactive terminals, current benchmark procedures have technical shortcomings in addition to their high costs.

In December 1972, the Commission on Government Procurement specifically addressed the high cost of benchmarking as a significant Federal cost problem and recommended the development of standard benchmark programs by the National Bureau of Standards. This conclusion was affirmed by the Proposed Executive Branch Position in March 1974 on implementing this particular recommendation. NBS has been designated the lead agency by the Office of Management and Budget for ascertaining and reporting progress of Federal activities responsive to this recommendation. Additionally, the Institute for Computer Sciences and Technology at NBS has initiated an on-going program to identify, define, and reduce both technical and cost/performance problems of benchmarking for comparative evaluation in procurements of computers. The National Bureau of Standards is pleased to make these guidelines for benchmarking available for use by Federal agencies in the computer selection process.

RUTH M. DAVIS, *Director*
Institute for Computer Sciences
and Technology

Abstract

This publication provides general guidelines to best practice for use by Federal agencies in benchmark mix demonstrations for validating hardware and software performance in context with processing an expected actual workload. The publication provides an overview and general discussion of the benchmarking process; guidelines for reducing the problems in benchmarking at the management level and at the technical staff level including a discussion of how these problems can be resolved or minimized; and procedural benchmarking guidelines, a discussion of the four phases of benchmarking, workload analysis, construction and validation of the benchmark, procedural documentation and preparation of the benchmark for the vendors, conducting benchmark tests. The document is written so that the various hierarchical levels in an organization's structure can be directed toward applicable sections of these guidelines.

Key Words: Benchmark mix demonstration; benchmarking; computer selection; Federal Information Processing Standard; workload representation.

Nat. Bur. Stand. (U.S.), Fed. Info. Process. Stand. Publ. (FIPS PUB) 42–1, 27 pages (1977)
CODEN: FIPPAT

For sale by the National Technical Information Service, U.S. Department of Commerce,
Springfield, Virginia 22161

Federal Information
Processing Standards Publication 42-1

1977 May 15

ANNOUNCING THE

GUIDELINES FOR BENCHMARKING ADP SYSTEMS IN THE COMPETITIVE PROCUREMENT ENVIRONMENT

Federal Information Processing Standards Publications are issued by the National Bureau of Standards pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973), and Part 6 of Title 15 CFR (Code of Federal Regulations).

Name of Guideline. Guidelines for Benchmarking ADP Systems in the Competitive Procurement Environment.

Category of Guideline. Benchmarking for Computer Selection.

Explanation. These guidelines provide basic definitions and recommended practices to assist Federal agencies in organizing their benchmarking efforts. Guidance is presented in the form of four Chapters. Chapter I, Introduction, places benchmarking in its proper perspective and identifies its relative position within the procurement process. Chapter II, Overview of the Benchmarking Process, provides an overview of the complete benchmarking process. Chapter III, Guidelines for Reducing the Problems in Benchmarking, provides guidelines for reducing major problems which have been encountered in past benchmarks. It is included with the expectation that it can be used as a checklist. Chapter IV, Procedural Benchmarking Guidelines, provides more explicit procedural guidelines for steps in the benchmarking process.

This guideline is directed to all levels of an organization's management and technical staff. Chapter I, II, and III.A are directed towards top management. In addition to the above, mid-level management should be aware of the contents of Chapter III.B. Project leaders and technical staff who will prepare the benchmark should find the entire document useful. This multi-dimensional format also becomes useful as a check list to ensure that benchmarks are devoid of the problems listed in Chapter III.

Approving Authority. Department of Commerce, National Bureau of Standards (Institute for Computer Sciences and Technology).

Maintenance Agency. Department of Commerce, National Bureau of Standards (Institute for Computer Sciences and Technology).

Cross Index. NBS Special Publication 405, *Benchmarking and Workload Definition: A Selected Bibliography with Abstracts.* (Available from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Order by SD Catalog No. C13.10:405). This document supersedes FIPS PUB 42, *Guidelines for Benchmarking ADP Systems in the Competitive Procurement Environment.*

Applicability. These guidelines are intended as a basic reference document of recommended practices for general use throughout the Federal Government in planning, organizing, and conducting benchmark mix demonstrations for competitive computer system procurements.

Qualifications. These guidelines represent recommended good practices for benchmarking in the competitive procurement environment based upon the collective judgment of a task group composed of members from the Federal Government, computer vendor industry, and other organizations. The philosophy and emphasis throughout is directed toward achieving a measured benchmark mix demonstration which is representative of the user's predicted actual workload requirements at minimum cost to the computer user and competing computer vendors. This goal is predicated on reasonable good practices. These guidelines *do not* attempt to define the domain of representativeness or reasonableness. These are *user* determinations and should be so established upon individual circumstances and requirements. Similarly, the guidelines acknowledge but *do not* address other portions of the procurement process such as functional demonstrations, contractual safeguards, procurement regulations and policy, Federal ADP management policy, validation of Federal standards or other ADP procurement considerations and user requirements. Thus, in order to be consistent with overall Federal policy, the user should seek current guidance from applicable Office of Management and Budget and General Services Administration policy and procurement directives.

In light of the above, the user should keep three basic principles in mind in reading and using these guidelines. First is that since all aspects of procurement are not herein treated, the user should develop a procurement plan that covers all needs. This should include functional demonstrations if appropriate, all needed documentation, and such contractual provisions as are necessary to protect the Federal interest. The user should also ascertain current Federal ADP management, procurement, and standards and guidelines policy and conduct the procurement accordingly. The user is reminded that all standards and technical guidelines of the Federal Information Processing Standards program may not be reflected in Federal Procurement Regulations or Federal Property Management Regulations and that the user should thus self-determine user requirements accordingly and ascertain vendor capability to satisfy these user requirements. Second is that guidelines are general descriptions of good practices for the normal situation. They do not cover nor are they applicable in all situations. The third and last principle is that these guidelines stress reasonableness in all practices and procedures. Reasonableness, in general, is a user determination. The user is solely responsible for determining his organization's requirements, for constructing a benchmark mix demonstration reflecting these requirements, and for ensuring that all decisions made during the entire process maintain the integrity of a representative benchmark mix demonstration. Any question of procedure or technique should be evaluated in this context and ultimate decision should protect the Government's interest.

Guidelines are not procedural steps that can be followed as a "recipe" with successful results. Instead, they are a discussion of good practices associated with areas of concern. In this sense, guidelines are useful as a checklist and, to some degree, identify areas where special competence, expertise, or particular attention is indicated.

These guidelines will need to be expanded and/or modified as further knowledge is gained of the techniques involved. Comments, critiques, and technical contributions directed to this end are invited. These should be addressed to the Associate Director for ADP Standards, Institute for Computer Sciences and Technology, National Bureau of Standards, Washington, D.C. 20234.

Where to Obtain Copies of the Guideline.

a. Copies of this publication are for sale by the National Technical Information Service, U.S. Department of Commerce, Springfield, Virginia 22161. When ordering, refer to Federal Information Processing Standards Publication 42–1 (NBS–FIPS–PUB–42–1), and title. When microfiche is desired, this should be specified. Payment may be made by check, money order, or deposit account.

Federal Information
Processing Standards Publication 42–1

1977 May 15

SPECIFICATIONS FOR

GUIDELINES FOR BENCHMARKING ADP SYSTEMS IN THE COMPETITIVE PROCUREMENT ENVIRONMENT

Contents

GUIDELINES FOR BENCHMARKING ADP SYSTEMS IN THE COMPETITIVE PROCUREMENT ENVIRONMENT

I. INTRODUCTION

A. Background

In 1973, the Secretary of Commerce approved the formation of the National Bureau of Standards sponsored FIPS Task Group 13, entitled Workload Definition and Benchmarking, to serve as an interagency forum and central information exchange on benchmark programs, data, methodology, and problems. The principal focus of Task Group 13 is on procedures and techniques to increase the technical validity and reduce the cost and time of benchmarking as practiced in the selection of computer systems by the Federal Government.

Task Group 13 developed FIPS PUB 42, Guidelines for Benchmarking ADP Systems in the Competitive Procurement Environment, which was published December 15, 1975. FIPS PUB 42 was an interim guideline issued for the purpose of establishing an initial baseline while these more extensive guidelines were being developed. FIPS PUB 42 is incorporated in this guideline.

B. Organization of FIPS PUB 42-1

- Chapter I places benchmarking in its proper perspective and identifies its relative position within the procurement process.
- Chapter II provides an overview of the complete benchmarking process.
- Chapter III provides guidelines for reducing major problems which have been encountered in past benchmarks. It is included with the expectation that it can be used as a checklist.
- Chapter IV provides more explicit procedural guidelines for steps in the benchmarking process.

This guideline is directed to all levels of an organization's management and technical staff. Chapters I, II, and III.A are directed towards top management. In addition to the above, mid-level management should be aware of the contents of Chapter III.B. Project leaders and technical staff who will prepare the benchmark should find the entire document useful. This multi-dimensional format also becomes useful as a check list to ensure that benchmarks are devoid of the problems listed in Chapter III.

C. Guidelines in Perspective

These guidelines are directed toward Federal ADP management and staff, referred to as "users" throughout this document, who are responsible for computer system procurements. The objective of this document is to achieve high-quality benchmarks and benchmark mix demonstrations at minimum cost to the user and computer vendor.

The user should keep two basic principles in mind in reading and using these guidelines. One is that guidelines are general descriptions of good practices for the normal situation. They do not cover nor are they applicable in all situations. The second principle is that these guidelines stress reasonableness in all practices and procedures. Reasonableness, in general, is a user determination. The user is solely responsible for determining his organization's requirements, for constructing a benchmark mix demonstration reflecting these requirements, and for ensuring that all decisions made during the entire process maintain the integrity of a representative benchmark mix demonstration. Any question of procedure or technique should be evaluated in this context and ultimate decisions should protect the Government's interest.

Guidelines are not procedural steps that can be followed as a "recipe" with successful results. Instead, they are a discussion of good practices associated with areas of concern. In this sense, guidelines are useful as a checklist and, to some degree, identify areas where special competence, expertise, or particular attention is indicated.

D. Benchmarking in Perspective

Before considering "Guidelines for Benchmarking," it is first necessary to realize that "benchmarking" is a term that has been used to describe a number of different functions. For these guidelines the term "benchmarking" is used to convey the same meaning as the more explicit term "benchmark mix demonstration." A "benchmark mix demonstration," sometimes referred to as a Live Test Demonstration (LTD), consists of a user-witnessed running of a group (mix) of programs representative of the user's predicted workload on a vendor's proposed computer system in order to validate system performance. Another type of demonstration that is frequently called "benchmarking," more properly should be referred to as

either a capability demonstration or a functional demonstration. The latter type of demonstration is intended to show only system or functional capabilities in some specific areas without regard to total system performance.

NOTE: Validation of the system's performance is meaningful only if the programs selected are representative of the work to be processed and are combined into representative mixes which reflect the user's workload and are consistent with the solicitation document requirements.

Since benchmarking is a very expensive undertaking for the vendors and the Government, a general guideline should be considered before addressing guidelines specific to benchmarking:

> A benchmark demonstration should include only requirements which contribute information needed for the selection process.

Specifically, the vendors should not be asked to demonstrate system capabilities which: (1) can be validly ascertained in other ways; (2) have not had any evaluation criteria assigned by the user agency; or (3) only demonstrate the vendor's ability to handle some worst case program(s) or situation(s) which are not representative of or critical to the user's requirements.

E. The Procurement Process

The competitive procurement of a new computer system is a lengthy and time consuming process. Its objective is to provide the most cost-effective computer system which will meet the present and future requirements of a user. The initial step in the process is the determination of a need to procure a computer system which is substantiated by the appropriate internal justification. This justification is followed by an agency approval cycle before proceeding further.

Once this approval is accomplished, the agency must then follow the applicable procurement regulations, Federal policy circulars, etc., prior to release of a solicitation document. Figure 1 depicts a management overview of this process. These guidelines for benchmarking do not include an in-depth discussion of the competitive procurement process. However, it is important to illustrate (fig. 1) how the benchmark fits into its proper context within the entire Federal Government procurement process.

Figure 1 is generalized and applies to all Federal Government computer system procurements which come under the Brooks Act (PL 89–306). Variations and delays may occur during the process due to a variety of factors ranging from incomplete justification to Congressional involvement. In any case, it is important to review procurement regulations with your ADP contracting office upon identification of the need to procure in order to establish timely planning, scheduling and other information pertinent to the procurement process. The requirements depicted in the chart are current as of May 1976.

II. OVERVIEW OF THE BENCHMARKING PROCESS

This chapter provides a summary description of the major phases of the benchmarking process. Five phases are discussed: workload definition and analysis; construction, validation and documentation of the benchmark; procedural documentation and preparation for vendors; vendor construction of the required demonstrations; and conducting benchmark tests. More detailed guidelines for conduct of the five phases are provided in Chapter IV.

A. Workload Definition & Analysis

Objective

The initial phase of work leading to a benchmark mix demonstration is the detailed definition and analysis of the workload to be performed by the new system. A number of complexities may be expected during this analysis, including a workload which changes in volume and composition over time, and is at the same time characterized by repetitive and recursive peaks and valleys. The objective is to define these workload characteristics, and to determine the trade-off options between levels of performance and related cost. This information enables agency managers to decide what level of performance to provide within overall agency cost constraints. This, in turn, allows the completion of benchmark developments.

System Life

Before workload analysis can be completed, the planned system life must be decided. This is the same period used for costing of the new system, and in many past instances has ranged from five to ten years. Future requirements should be analyzed and workload projected over this period of time.

Functional Workload

Workload should be quantified in terms of agency functions and objectives, user performance objectives, and work volumes.

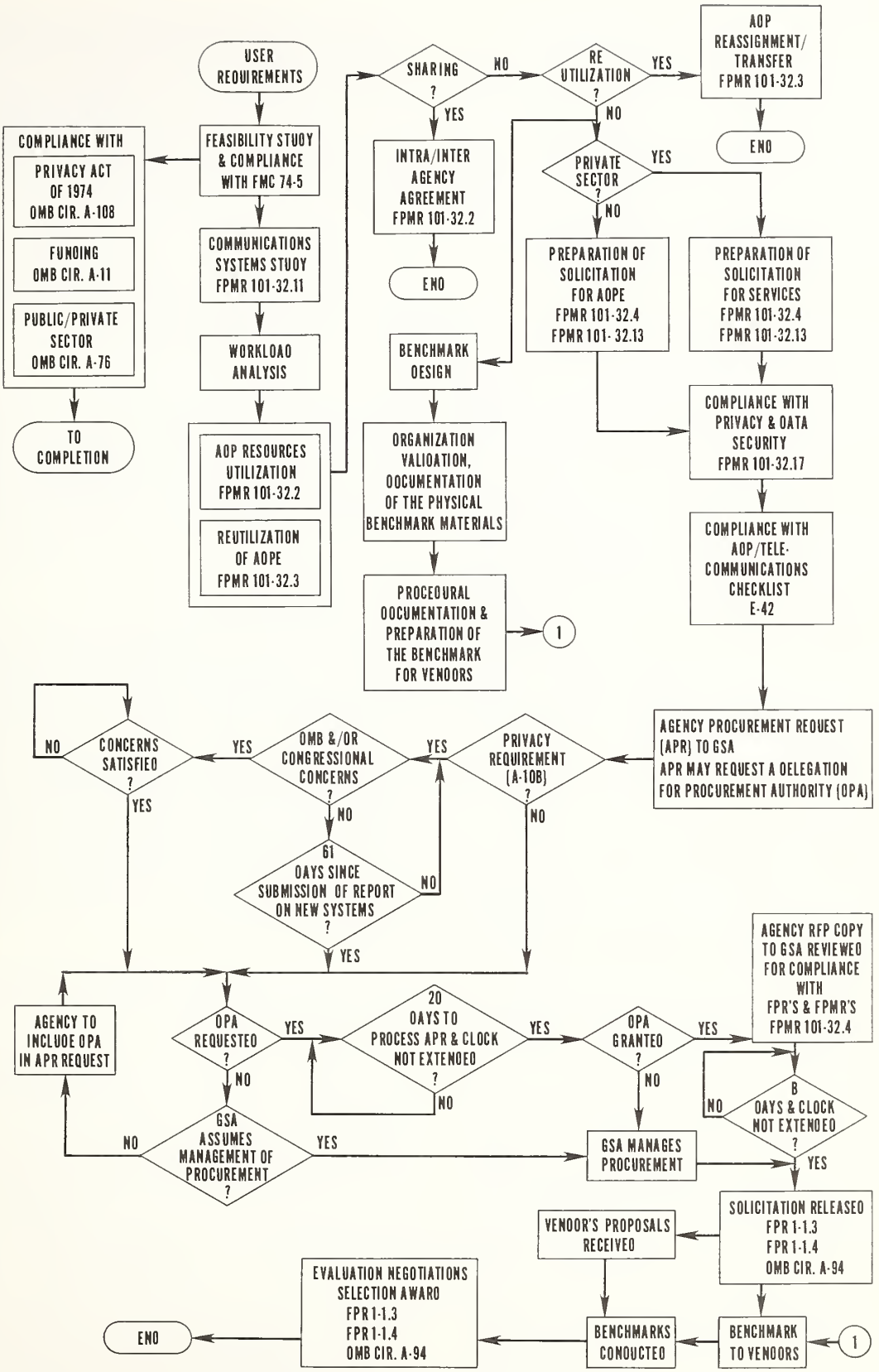

FIGURE 1. *Federal ADPE Procurement Policy*

B. Construction, Validation, and Documentation of the Benchmark

Purpose and Context

The second phase in preparing for the benchmark mix demonstration is to construct the set of programs, transactions, data, and documents which together will represent the workload established for the new system. This phase can begin, but cannot be completed prior to agency determination of requirements.

Complexity

The task of accurately representing a complex workload over a period of several years and developing the representation in a relatively brief time span necessitates a well-disciplined approach. The steps include selection or construction of a set of representative programs, combining them in the representative mix(es), producing corresponding transactions and data in appropriate volumes, and determining the minimum benchmark equipment configuration (primarily peripheral equipment). All of the material must be carefully validated, cross-checked and thoroughly documented. This phase is most likely to be an iterative process, as analysts identify deficiencies in their initial work and are able to adjust and tune the benchmark. Care should be taken to avoid the mistake of selecting programs only on the basis of the fact that they are easy to prepare for the benchmark. Representativeness should be the chief selection criterion.

Functional Tests

If functional tests are necessary in addition to the benchmark mix demonstration, they must be specified and additional test material may need to be constructed.

C. Procedural Documentation and Preparation of the Benchmark for Vendors

Procedural Documentation

The benchmark material must contain full documentation for the benchmark mix demonstration; programs and required data files should be provided. Additionally, the benchmark material must be accompanied by a procedural document detailing for the vendor how the benchmark will be run. It should specify the maximum permissible run time for each benchmark mix, and otherwise relate the benchmark programs to the system life cycle. It should include an overview of what comprises the benchmark, including planned functional demonstrations. It should also treat such subjects as the sequence in which the benchmark mix programs will be run, the minimum acceptable subset of the proposed equivalent configuration and how the benchmark outputs will be validated. The vendor should not be prohibited from exercising the proposed system in a manner which shows its best performance as long as the representativeness and integrity of the workload are maintained.

What will be demonstrated should be explicitly defined. Any latitude allowed for making changes in the benchmark programs, the data files, or in systems operation should be specified to the vendors. All required output data, such as source listings, accounting log data, console logs, etc., should be specified.

Ensure Validity

Verify that each vendor's copy of the benchmark programs and data is accurate. When a second set of data or other modifications is to be used during the benchmark validation runs, both sets of data and modifications must be tested and validated by the user, prior to the release of the benchmark materials.

Be Informative

A policy and mechanism should be established for rapid exchange between the user and the vendor of information such as benchmark changes, questions, configuration substitutions, etc.

Any Government required special purpose equipment should be reviewed with the vendors. Any vendor proposed equipment substitutions planned for use during the benchmark mix demonstration must be approved by the Government. The Government should inform the vendor of the planned procedures concerning acceptability, validation and certification of the system.

D. Vendor Construction of the Required Demonstration(s)

During this phase of the benchmarking process, vendor questions and comments inevitably surface. It is critical that rapid dissemination of Government responses be made to all participating vendors. It is also during this time that modifications suggested by the vendors should be resolved. Questions of special purpose equipment and/or equipment substitutions should be reviewed and certified as acceptable to the Government or rejected, as the case may be. The previously established policy and mechanism for accomplishing such review must be available for rapid implementation. All questions of acceptability, validation, certification, etc., should be resolved prior to conduct of the benchmark mix demonstration.

E. Conducting the Benchmark Tests

Review Procedures

A meeting should be held with the vendors to confirm what will be required at the benchmark mix demonstration, and to review the test and observation procedures to be followed.

Regulate Teams

The size of the benchmark teams should be kept to a minimum, and vendors should be requested to keep the demonstration area free of all but essential personnel. Prior to the benchmark, the Government and the vendor should each have one individual designated as their point of contact for all communications regarding the demonstration. These individuals should possess expertise in all phases of the procurement including solicitation document requirements, proposal contents and benchmark requirements.

Benchmark Demonstration Management Plan

This documentation details the procedures and organization for conducting the benchmark. A primary objective of documenting the planned process is to ensure a smooth running benchmark demonstration and to minimize misunderstanding between the vendor and the user. The plan should detail the responsibilities of the user benchmark team members, method of performance measurements, validation procedures, and output to be gathered for each task. The plan should include forms necessary for recording measurements and validations.

Validate System

A certified description of the configuration(s) benchmarked should be obtained and any substitution(s) of equipment or software for the proposed system should be noted. Physical inspection and software validation checks at the time of the benchmark are necessary to supplement the certification.

Prior to the benchmark mix demonstration, some users may require each of the vendors to run, in sequence, all of the programs in the benchmark mix in order to validate their performance and to ascertain the resources required by the benchmark programs on each of the proposed systems.

Run the Benchmark Mix

Vendors should be permitted to generate and load large data bases or perform other time consuming activities prior to the benchmark mix demonstration. The benchmark programs should then be run in the designated mixes, with performance measurements (e.g., timings) made as defined in the benchmark package documentation. All external performance timings must be measured and recorded by a user representative. In some cases a vendor representative will also make and record such timings. Any discrepancies should be resolved immediately, before continuation to another phase of the benchmark mix demonstration.

Collect Materials

Appropriate benchmark materials such as output listings, console logs, accounting log data, and secondary storage listings should be collected and identified after each run. This material will assist in validating the results of the benchmark mix demonstration.

Communicate Results

Prior to departure from the vendor's site, an exit debriefing should be held with participation by both the user agency and vendor benchmark teams.

III. GUIDELINES FOR REDUCING THE PROBLEMS IN BENCHMARKING

This chapter comprises two sections. Management Highlights, Section A, identifies potential problem areas and provides guidance for their avoidance. Top and mid-level management can use Section A as a check list against which a benchmark package can be evaluated. It also provides a Table of Contents for Section B. Section B is a more detailed discussion of each individual problem including recommendations for avoiding or minimizing the problem.

Problems encountered in benchmarking can cause major delays in the procurement process. They also contribute heavily to the costs incurred by the Government and the vendors. For these reasons it is important to identify potential problems in an effort to avoid as many as possible. However, it is also important to recognize that since problems do occur, every effort must be made to resolve them in a prompt, fair and practical manner. The problems which are encountered are attributable to actions of both the Government and vendors. In fact, there is no practical way to address benchmarking problems from one side only because neither the Government nor the vendor operates in a vacuum. Much of the expertise required to achieve a "good" benchmark is the skill exhibited in reacting and responding to problems as they arise.

A. Management Highlights

Benchmarking Philosophy

1. Require the vendor to physically demonstrate only the peripheral and terminal equipment needed to process the actual benchmark programs and data.

2. Require only necessary logs and listings.

3. Allow adequate time for vendor conversion of benchmark programs.

4. Avoid lengthy, equipment-only, reliability test runs unless they reflect unusual user requirements.

5. Preplan user requests of the vendor. Settle questions of reasonableness prior to arrival of the benchmark team.

6. Design the length of individual runs and the length of time for the benchmark to be representative of the user's workload. The total time for each run of the benchmark mix demonstration should be approximately two hours or less.

7. Request functional demonstrations only when such tests demonstrate features which are not an integral part of the benchmark mix.

    a. Be clear and concise in your statement of requirements for functional demonstrations.

    b. Be specific and reasonable on hardware configuration requests for functional demonstrations.

    c. Specify clearly whether each functional demonstration is mandatory or desirable.

    d. Limit your requests for demonstrations to those which are actually required and which you plan to witness.

Analysis, Design, Construction, and Documentation of Benchmark Package

1. To the extent possible, avoid mandatory requirements for hardware not manufactured by a vendor being benchmarked.

2. Avoid use of vendor specific hardware/software features.

3. Code the benchmark programs in compliance with Federal Information Processing Standards (FIPS) for languages.

4. Do not use programs and data bases tailored to a specific vendor's system features.

5. Use standard character sets as defined in applicable FIPS publications for distribution of program code and input/output data.

6. The degree of complexity of benchmark programs should be representative of the projected workload.

7. Realistic consideration should be given to the workload planned for the future. A realistic workload is one that reflects the projected requirements of the agency during the required system life.

8. Test all programs in the benchmark mix with the data to be furnished to the vendors (including any program modifications and alternate data) to be used at the benchmark demonstration.

9. Adequately consider precision requirements. Use floating point data in ways that yield predictable and definable results.

10. Clearly define all timing constraints associated with the benchmark mix demonstration. Do not state a series of time constraints on various interrelated pieces of the benchmark in such as way as to permit various interpretations.

11. Be consistent in conventions for naming of programs and associated data files.

12. Clearly define requirements for the pretimed, timed, and post-timed portions of the benchmark demonstration.

13. Rely on benchmark performance rather than specific statements of desired hardware characteristics.

Benchmark Package

1. Provide complete program documentation including source code listings, compilation listings, job control information, and all output generated.

2. Provide complete documentation for all files, including intermediate files, and program/file cross references.

3. Utilize system block and flow diagrams to indicate system flow, including program order dependencies.

4. Provide estimates of computer system resource requirements for all programs.

5. Carefully define system conditions at the start of the benchmark timed runs.

6. Specify the use of multiple copies of inputs for multiple executions of the same benchmark program.

7. Provide clear instructions for vendor preparation of programs and data required to process the benchmark.

8. Include a glossary of terms to reduce probability for misunderstanding.

9. Minimize use of punched cards. When cards are necessary, utilize a mechanism for assuring their proper sequence.

10. Carefully control the environment in which cards and tapes are stored and handled.

11. Ensure accuracy of files through comparison with copies of the original file.

Planning, Conducting, and Managing the Benchmark Demonstration

1. Establish a user benchmark coordinator who is accessible to the coordinator for each vendor for providing answers to technical questions, providing replacement of missing material, and coordinating the dissemination of all other information pertinent to the benchmark demonstrations.

2. Develop an overall schedule of on-site vendor visits. Once a schedule is established, maintain its integrity to the extent possible.

3. Organize the benchmark team and dry-run the benchmark prior to arrival at the vendor's location.

4. Determine and adhere to a scheduled agenda for each on-site benchmark demonstration.

5. Develop and document expedient methods for making changes to data files at the benchmark test demonstration.

6. Plan and state procedures for validating the hardware configuration and the specific systems software to be used in the benchmark mix demonstration.

Evaluation of the Benchmark Results

1. Ensure benchmark team understanding of the difference in terminology and meaning of the output results from the vendor's resource utilization logs.

2. Clearly define criteria for evaluating the results of the benchmark demonstration.

3. Indicate the benchmark results.

B. Problems

This section details each of the highlighted parts in Section III.A above.

Benchmarking Philosophy

User requests of the vendor should be made only if the resultant actions can be objectively evaluated. The cost of vendor personnel and equipment is high and ultimately is recovered by the vendor in higher equipment costs. Therefore, every effort should be made by the Government to minimize requests for demonstrations or services in order to lessen the resource requirements of the vendors. Likewise the vendor should make every attempt to be responsive to the requests made by the Government.

1. Require the vendor to physically demonstrate only the peripheral and terminal equipment needed to process the actual benchmark programs and data.

   Extra equipment which does not assist in the evaluation process or substantially add to the representativeness of the benchmark should not be required. A Remote Terminal Emulator (RTE) can be effective for demonstrating large numbers of terminals. If a RTE is used, require no more than one live terminal of each type specified.

   NOTE: At the time these guidelines were written the use of a Remote Terminal Emulator (RTE) in competitive procurements was under study by a joint GSA/NBS study group. The agency considering the use of an RTE should consult with the local contracting office to obtain the latest information on the use of RTE's.

2. Require only necessary logs and listings.

   Only logs or listings which have been established as necessary for evaluation or validation prior to the running of the benchmark should be required. Care should be taken in analyzing any accounting system log data to ensure that the information evaluated is the same for all vendors, i.e., differences in definitions of the data elements such as CPU time should be considered.

3. Allow adequate time for vendor conversion of benchmark programs.

   The amount of time permitted for the vendor to convert the benchmark programs should be proportional to the complexity and number of the benchmark programs.

Sufficient time should be allowed for special preparation if such as data communications or data management systems interfaces are required of the vendor.

4. Avoid lengthy, equipment-only, reliability test runs unless they reflect unusual user requirements.

   Test runs prove only the reliability of a specific piece of equipment at a given point in time and cannot be used to predict how such equipment might act in a particular user environment.

5. Preplan user requests of the vendor. Settle questions of reasonableness prior to arrival of the benchmark team.

   Do not request documentation or demonstrations which were not specifically delineated in the benchmark instructions. Exceptions to this should be made only in unusual circumstances. If it is determined that a functional demonstration, other than those preplanned, is required, the vendor must be allowed time to prepare.

6. Design the length of individual runs and the length of time for the benchmark to be representative of the user's workload. The total time for each run of the benchmark mix demonstration should be approximately two hours or less.

   Select programs for the benchmark mix that are representative of the user agency's projected workload. Avoid programs that are very short in duration unless they are representative. In such cases multiple copies of these programs can be run to ensure that the benchmark mix demonstration represents a valid workload. Conversely, unless the user agency has an actual or projected requirement, inordinate run lengths for processing can create problems which may not exist in the actual installation. The elapsed time for the longest benchmark mix demonstration run should be approximately two hours or less.

7. Request functional demonstrations only when such tests demonstrate features which are not an integral part of the benchmark mix.

   Functional demonstrations conducted during the course of a benchmark provide the vendor the opportunity to demonstrate hardware, software or system features which are required to meet the user's operational requirements, but are not demonstrated in the timed portion of the benchmark demonstration. Requiring a functional demonstration of a feature exercised during the timed portion of the benchmark demonstration serves no useful purpose.

   Functional demonstrations may also be requested in those instances where problems have occurred during the benchmark demonstration which indicate that the vendor may not be in compliance with requirements of the RFP and indications are that the questionable issues can be readily resolved through a functional demonstration.

   a. Be clear and concise in your statement of requirements for functional demonstrations.

      Improperly stated or unclear statements can lead to a wide variety of interpretations by different vendors. In some cases, a large allocation of resources could be required which are really unnecessary to meet the Federal user's requirements.

   b. Be specific and reasonable on hardware configuration requests for functional demonstrations.

      Limiting the vendor to the proposed configuration could preclude an effective demonstration.

   c. Clearly specify whether each functional demonstration is mandatory or desirable.

      A desirable feature demonstration which is incorrectly identified as a mandatory requirement may reduce vendor competition or increase the cost of the equipment.

   d. Limit your requests for demonstrations to those which are actually required and which you plan to witness.

      Avoid requiring vendors to prepare to meet a lengthy list of requirements which will be narrowed down after arrival at the vendor's site.

Analysis, Design, Construction, and Documentation of the Benchmark Package

A benchmark package is a precision system with a particularly important function. Its engineering must follow the strictest standards for software development. Unlike some production systems, the benchmark will not have a long breaking-in period in which to eliminate errors. The magnitude of the decision to be based upon results of the benchmark programs

makes it essential that it be accurate in its imposition of workload and degree of complexity. The fact that it must be simultaneously implemented by multiple vendors, through written instructions (rather than in-house with hand-holding by the programmers), necessitates the clearest of documentation and instructions. Inadequate or unclear documentation may extend the selection process.

1. To the extent possible, avoid mandatory requirements for hardware not manufactured by a vendor being benchmarked.

   Equipment not manufactured by the vendor being benchmarked causes additional costs to the vendor and user.

2. Avoid use of vendor specific hardware/software features.

   Functional needs can be specified without resorting to make and model numbers of equipment. Similarly, avoid specifying functions in such a way that only one vendor will meet the requirements as this inhibits competition.

3. Code the benchmark programs in compliance with Federal Information Processing Standards (FIPS) for languages.

   Commonly used higher level computer languages should be used whenever possible where they adequately represent the user's workload. Deviations which are legitimate requirements of the user's organization should be justifiable. Avoid use of vendor dependent compiler features or extensions.

4. Do not use programs and data bases tailored to a specific vendor's system features.

   A data base or program which is tailored to the architecture or features of a specific vendor restricts competition. Specify the data base requirements in terms of functions required to accomplish the work.

5. Use standard character sets as defined in applicable FIPS publications for distribution of program code and input/output data.

   Nonstandard character sets are potentially costly and time consuming.

6. The degree of complexity of benchmark programs should be representative of the projected workload.

   If worst case or best case programs must be included in the benchmark, they should be proportionate to their occurrence in the projected workload. Excessive reliance on worst case programs requires the vendor to propose equipment capabilities in excess of what the user requires. Overly simplistic programs, unless an adequate number of copies are run, may provide the user with insufficient system capacity.

7. Realistic consideration should be given to workload planned for the future. A realistic workload is one that reflects the projected requirements of the agency during the required system life.

   The benchmark programs, data and transaction volumes should reflect the workload projected for the computer at the time of installation and during the computer life cycle. The analysis of current and future workload requirements is an important step to developing the benchmark. Unrealistic workload projections for the future can cause over or under specification of hardware. Over specification of hardware will result in unnecessary expenditures. Under specification of hardware may result in upgrades or additional procurements to meet the true requirement.

8. Test all programs in the benchmark mix with the data to be furnished to the vendors (including any program modifications and alternate data) to be used at the benchmark demonstration.

   Programs which have not been tested with the actual data that is to be used during the benchmark often cause unpredictable results or results which invalidate the running of the benchmark. This lack of sufficient testing may cause a prolongation or repetition of the benchmark mix demonstration.

9. Adequately consider precision requirements. Use floating point data in ways that yield predictable and definable results.

   Precision requirements for floating point numbers must be used with caution due to variations in word size and compiler implementations. Results of floating point operation might be different on various machines. Take care to ensure that the answers desired during the benchmark are what are actually needed by the user, and the answers used for the test comparison are indeed correct. The precision of results required in the benchmark must not exceed the specifications in the solicitation document. Vendors may be held nonresponsive for failure to meet the degree of precision

232-489 O - 77 - 2

specified, but not for exceeding the precision specified.

10. Clearly define all timing constraints associated with the benchmark mix demonstration. Do not state a series of time constraints on various interrelated pieces of the benchmark in such a way as to permit various interpretations.

    The precise timing requirements for the benchmark demonstration should be stated to allow the vendor to propose a cost effective configuration. The precise methodology for obtaining timings, the initiation of the timing, termination of the timing, and the processing that is required during this time period should be explicitly stated and clarified to ensure no misinterpretation by vendor and user personnel during the benchmark. If credits are to be given for reduction in run times, these should be clearly stated in the solicitation document. Consideration should be given to other times which may be required to represent the user's workload. For example, times specified for input data from interactive terminals should consider think times of appropriate and realistic human and terminal performance. Statistical techniques should be considered when defining timing requirements for terminal workloads.

11. Be consistent in the naming conventions of programs and associated data files.

    A convention for naming programs and their related files should be used throughout the benchmark package. This will assist the vendor in relating specific files to specific programs and assist the user in evaluating the results of the benchmark demonstration.

12. Clearly define requirements for the pre-timed, timed, and post-timed portions of the benchmark demonstration.

    A detailed agenda for the benchmark demonstration is necessary to provide the vendor and the evaluation team with the requirements for the pre-timed, timed, and post-timed portions of the benchmark demonstration. An incomplete agenda can cause confusion and make evaluation of the benchmark results difficult.

13. Rely on benchmark performance rather than specific statements of desired hardware characteristics.

    Minimum hardware characteristics such as tape and disk transfer rates should not be specified. If specific hardware characteristics are required, they should be validated by the benchmark mix demonstration. The requirements dictated by the benchmarks should be consistent with those specified in the solicitation document.

Benchmark Package

The benchmark package consists of the procedural documentation, test programs and data files. Failure to provide a complete, tested benchmark package causes delays and errors. The benchmark documentation must include all the information necessary for the vendor to implement the programs on his machine and must be checked and rechecked to eliminate omissions and errors. The documentation should also be examined by a third party. An excellent test of the benchmark package can be conducted by sending the programs, data, and documentation to another Government computer site and asking them to examine the package, and if possible run the benchmarks. Failure to provide a complete and accurate benchmark package is one of the biggest causes of delays in the procurement process.

1. Provide complete program documentation including source code listings, compilation listings, job control information, and all output generated.

   Source code listings should include comments and/or be accompanied by external descriptive documentation including flowcharts. System parameters normally specified through operating system control statements must be provided in English, including essential device assignments. The actual control statements used for testing the benchmark may also be provided if they will be informative to the vendors. A listing of the output including console and terminal messages should be included.

2. Provide complete documentation for all files, including generated intermediate files, and program/file cross references.

   File documentation includes file structure, format, data element definition, file labels, recording mode, density, etc. Provide sample record listings including the first and last record of each file. Indicate the number of records in each file and provide hash totals in order to ensure integrity of the file. Indicate all the input, intermediate, and output files associated with each program.

3. Utilize system block and flow diagrams to indicate system flow, including program order dependencies.

The interrelationships between the various benchmark tasks and programs should be described with the use of flowcharts. Specify required run sequences for programs which interact.

4. Provide estimates of computer system resource requirements for all programs.

   Identify the base computer and indicate the average run time, memory, and other system resources required for each program on that configuration.

5. Carefully define system conditions at the start of the benchmark timed runs.

   There are many possible starting conditions for the beginning of a benchmark mix demonstration. Failure to specify in what state the computer should be readied can affect the processing time and provide unfair advantage to a vendor. In the benchmark package, the user should specify whether or not programs can be loaded, tapes mounted and readied, cards stacked in the readers, the state of the operating system as well as other initial conditions for the beginning of the benchmark run.

6. Specify the use of multiple copies of inputs for multiple executions of the same benchmark program.

   The availability of multiple copies of data can directly affect processing time. It is the responsibility of the user to determine the acceptability of multiple copies and explicitly state this requirement in his benchmarking instructions.

7. Provide clear instructions for vendor preparation of programs and data required to process the benchmark.

   The vendor should be explicitly told what he is permitted to modify within the benchmark programs. Providing permission from the user agency is obtained, it is not unreasonable to allow the vendor to determine such things as blocking factors to allow for his machine architecture and processing efficiencies. Optimization of program code should be prohibited except that which is routinely performed within the vendor supported compiler. Any changes not explicitly authorized by the benchmark package must be approved by the user prior to the benchmark demonstrations.

8. Include a glossary of terms to reduce probability for misunderstanding.

9. Minimize use of punched cards. When cards are necessary, utilize a mechanism for assuring their proper sequence.

10. Carefully control the environment in which cards and tapes are stored and handled.

11. Ensure accuracy of files through comparisons of copies with the original file.

Planning, Conducting, and Managing the Benchmark Demonstration

The planning, conducting, and managing of a benchmark is a severe test of organizational ability. The user is specifying a set of tasks to be done by a number of geographically and managerially separate vendor organizations. Inadequate planning, unspecific procedures and unclear requirements for these tasks become much more difficult to resolve than for an "in-house" project. Concurrent resolution of these problems with several vendors causes expensive delays.

1. Establish a user benchmark coordinator who is accessible to the coordinator for each vendor for providing answers to technical questions, providing replacement of missing material, and coordinating the dissemination of all other information pertinent to the benchmark demonstrations.

   These individuals are responsible for answering all questions and providing solutions to all problems associated with the benchmark demonstration.

2. Develop an overall schedule of on-site vendor visits. Once a schedule is established, maintain its integrity, to the extent possible.

3. Organize the benchmark team and dry-run the benchmark prior to arrival at the vendor's location.

   The benchmark team should be organized and trained prior to the first benchmark demonstration at a vendor's site. Each of the members of the team should be skilled at his job, understand what is required during the test, and understand his responsibilities. If possible, the entire team should dry-run the entire benchmark at another installation for all phases of the benchmark demonstration prior to the first actual benchmark demonstration.

4. Determine and adhere to a scheduled agenda for each on-site benchmark demonstration.

The time period within which the actual benchmark demonstration is to be conducted should be determined in advance with the vendor. The work day for the benchmark team should be clearly defined and the maximum duration of the test in terms of hours or days should be stated to allow the vendor to schedule his computer system and personnel for other activities.

5. Develop and document expedient methods for making changes to data files at the benchmark test demonstration.

   After arrival at the benchmark site and prior to the start of the timed portion of the benchmark mix demonstration, it is in the best interests of the government to change or cause to be changed the data or parts of the data used by the benchmark programs so that it differs from the data originally supplied to the vendor for testing and validation of correct program execution purposes. Time consuming methods for making these changes should be avoided. Where data generators have been provided to the vendor, the parameters for the data generator should be changed to provide different data for the benchmark test. All data should have been tested with the benchmark programs prior to the actual benchmark demonstration and should change results without changing processing characteristics and timings.

6. Plan and state procedures for validating the hardware configuration and the specific systems software to be used in the benchmark mix demonstration.

   The user benchmark team should require the vendor to provide a list of all hardware and software present on the specific configuration being benchmarked. This list should be certified by a responsible vendor representative. A physical inspection of the hardware is required. A list of all software used during the benchmark should be obtained as an output from the computer system and certified by a responsible vendor agent to be the exact software specified in his proposal. Inspection and validation may often be conducted in parallel with other benchmarking activities.

Evaluation of the Benchmark Results

1. Ensure benchmark team understanding of the differences in terminology and meaning of the output results from the vendor's resource utilization logs.

   Vendor provided resource utilization log systems are complex and the definitions of the output results are not the same from vendor to vendor. Care must be taken in understanding the terminology and meaning of the output results. The benchmark team can, however, require reasonable listings and summations of timing results.

2. Clearly define criteria for evaluating the results of the benchmark demonstration.

   The criteria to be employed in evaluating the results of the benchmark demonstration should be clearly stated and defined in the benchmarking plan. Inform each vendor of the criteria to be used and how his system will be evaluated. The user has the responsibility to adhere to the plan and to provide a fair and unbiased evaluation.

3. Indicate the benchmark results.

   If at all possible, the benchmark should be designed to permit evaluation of the benchmark results at the vendor's site shortly after the benchmark mix demonstration is completed. If the benchmark results can be evaluated at the vendor's site, indicate to the vendor whether he has passed or failed the benchmark prior to the benchmark team departure. In every case, whether the evaluation of the benchmark results takes place at the vendor's site or at the user's facilities, the benchmark evaluation should be completed as expediently as possible and formal notification made to the vendor by the responsible contracting officer. It is important that this notification be made as soon as possible since the vendor has considerable resources committed waiting for the Government's decision.

IV. PROCEDURAL BENCHMARKING GUIDELINES

Chapter IV provides more detailed guidelines for the five phases of work described in Chapter II. These two chapters follow the same general outline. The reader is urged to read Chapter II, which provides an overview of the benchmark process, before continuing with this chapter.

The initial two phases, discussed in Sections A and B below, involve a substantial amount of research and development. For convenience, the steps in these phases are described as if they occurred serially. In practice, they are likely to be parallel efforts and iteration of some steps

is usually required in order to improve and tune the benchmark.

A. Workload Definition and Analysis

This section expands upon the section in Chapter II by the same title. Its purpose is to provide more explicit guidelines for quantification of the workload to be represented by the benchmark demonstration mix.

System Life

Workload definition and the benchmark construction should be consistent with current policy for financial analysis and the life of the system. An initial objective of workload definition and analysis is to prepare data, covering the period in question, which represents the projected workload over time.

Quantifiable Variables

The workload should be quantified in terms of its own characteristics and performance objectives and not in terms of computer hardware. For example, the amount and characteristics of data to be stored should be specified; not the number and capacity of disk drives. The throughput requirements must be specified; not CPU instruction rates. The objective of this approach is to encourage innovation and variety in vendor responses as to how the requirements are met.

The aggregate workload for the system to be procured is likely to consist of too many different ADP functional operations to allow each one to be included individually in the benchmark. If so, then workload quantification will first necessitate the grouping of functions into a manageable number of categories. The functions included in a given category must be sufficiently consistent so that they can be represented by a single set of quantifiers, and eventually by one or more copies of a single benchmark program. Major categories exist within each of these. Compilations, sorts, and other utility functions are legitimate categories if they constitute significant workload. Typical factors which characterize a category are, where applicable:

Mode of performance, i.e., batch, on-line, remote entry;
Structure of program;
Number of source instructions executed per transaction or use;
Volume of I/O activity;
Characteristics of data files;
Priority.

Following identification of categories of workload, the specific variables to be quantified for each one must be determined. Some categories will have fewer variables than others. For example, programs of COBOL compilations will primarily require quantification of the frequency of compiles and the number of statements compiled each time. Following are some variables which may apply to other categories:

Frequency of execution;
Input volume and media;
Response time;
Output volumes and media;
Size of data files and media.

Independent quantification will also be necessary for aggregate data storage needs, if data storage equipment is to be benchmarked.

Sources of Data

The primary source of quantification data will usually be the users of the service. Current system usage statistics should be obtained and used as a baseline and to validate user estimates of current workload. The criticality of workload data necessitates ensuring its accuracy. In instances where replacement equipment is being procured for operational systems, the workload reported by users may be validated by monitoring equipment or software in the current system. Analysts must understand the nature of outputs from these sources thoroughly in order to know how to allow for various overhead factors.

Level of Support

Cyclical workload peaks are likely to occur and short cycles are likely to occur within longer cycles. For example, daily workload may peak at 2 p.m., monthly workload may peak during the final two or three days of the month, and annual workload may peak in July. If the ultimate peak period were used to configure the system, there would be excess capacity throughout the rest of the year. The alternative is to configure for somewhat less than the peak, thus imposing turnaround delays during this period and flattening the peaks. The extent of turnaround delay which is tolerable depends on the criticality of the work, and lends itself to cost-benefit analysis. Agency managers must decide how much excess capacity they will buy to achieve the necessary level of performance.

It is not always clear where peaks occur, especially if different kinds of work peak at different times. That is, when the composition of the workload varies sufficiently during different periods of high volume, the analyst may

have difficulty determining which workload imposes the greater burden on the system. Such questions need to be answered by analytical means wherever possible. The alternative is to define a different mix for each of the workload compositions. The use of multiple mixes, particularly when introducing new programs, should be kept to a minimum.

System Upgrades

The workload may be projected to change sufficiently in composition or characteristics over the life of the system so that upgrades may be appropriate following the initial installation. The growth of the projected workload will indicate the points in time when upgrades may be needed. The procuring agency must be prepared to benchmark each of these workloads, which will require a different input-output volume and possibly a different workload composition. If a single workload composition can be used, a benchmark for any point over the system life can be provided simply by adjusting the allowed running time for the mix, or by varying the input volumes. If the composition must change, a technique must be designed to enable the correct benchmark mixes to be assembled to represent the predicted future workload changes.

B. Construction and Validation of the Benchmark

This section provides guidelines for the construction of a benchmark based upon the workload quantification, and validation that the benchmark represents the workload within tolerable limits.

Selection and Design of Programs

In the interest of simplicity a single program should be selected, if practical, to represent each category of workload. All programs provided by the agency should be written in commonly used higher level languages, in compliance with existing Federal Information Processing Standards (FIPS). If the quantified workload mix includes compilations and utility functions such as sorting, vendor software should be used to perform these functions.

It is common practice to select representative programs from operational application systems for benchmarking. The source code of such programs must be reviewed and any nonstandard code removed.

An alternative to the use of operational programs is to develop, or obtain from existing sources, synthetic programs to represent each of the workload categories. The two program types, operational and synthetic, may be mixed. Synthetic programs may be especially useful for representing functions which are not currently automated, or which will be performed in a substantially different way in the new system.

Synthetic programs must be designed and adjusted to accurately represent all of the applicable workload characteristics, such as those listed in Section A above. They need not perform any other useful function. Program size may be controlled by including a data array of appropriate size. Caution must be taken to ensure realism in how the program is treated by operating systems; for example, all of the parts of each program should be used in order to avoid the possibility that some would not be read into memory.

If synthetic programs are to be used, they must perform the same I/O and instruction mixes as the programs they are to represent if the benchmark mix demonstration is to be representative of the user's workload.

It is imperative that all benchmark programs be individually and thoroughly tested using all sets of benchmark data, to ensure their accuracy. User-written application programs sometimes contain bugs when they go into production. Because of the peculiar circumstances under which benchmark programs are run, which do not readily facilitate programmer assistance, they should be carefully and thoroughly tested. Operational programs which have been updated for removal of nonstandard code or for other reasons must be re-compiled and re-tested.

Workload Mix

A plan must be devised to combine the selected benchmark programs with transactions and data in the mixes required to represent all workloads which are subject to benchmarking. The benchmark mixes should be thoroughly tested.

The longest timed run should be approximately two hours or less for each of the benchmark mixes. It is appropriate to use multiple copies of any or all selected programs to provide the proper number of programs or functions for the time period chosen. A number of other variables also must be properly chosen, including transaction volumes per program or function, data volumes, and the parameters of synthetic programs, if used.

Data

Where data volumes to be delivered to vendors are high, the use of a data generation program is desirable. Where synthetic programs are used, data generation is especially facilitated. All data should be in compliance with Federal Standards for media and interchange codes.

Attention must be given to the distribution of matching keys in transaction records and associated file records, as their relationship normally is one of the most significant determinants of workload. One must control the proportion of transactions which have matching file records and also the number of multiple matches for individual transactions.

Generated transactions and data records must be realistic in terms of all factors which determine the amount of storage required, and processing performance characteristics. These factors include number of records, record lengths, field lengths, data types, and construction of key fields. The actual storage media and indexing techniques, unless a particular indexing technique is required, should be left to the discretion of the vendor.

Data is often furnished to vendors to benchmark the capacity of storage equipment. This requires much larger volumes of data in order to add to the realism of the benchmark mix demonstration. It sometimes happens that the amount of storage equipment needed exceeds what would be reasonable to be required in the benchmark. An acceptable practice in such cases is to furnish a stated percentage of the aggregate data volume required (e.g., 10% to 50%). Caution must be exercised to ensure that the file accessing workload is accurate, and that the storage required by the benchmark can be legitimately extrapolated to ascertain the amount of storage necessary for the aggregate requirement.

Configuration

The minimum configuration for running the benchmark demonstration mix must be determined and specified.

Determining the number of terminals required for benchmarking on-line systems presents a complex problem. While terminals may not be a part of the procurement, they are often necessary for demonstration of the system. If this is the case, the vendor should be allowed maximum flexibility in selecting the demonstration terminals. In all cases, the number of live terminals in the benchmark should be minimized. As a practical matter, few transactions can be entered via terminals during the demonstration.

Any other peripheral equipment required for the benchmark which is different from the solicitation document specifications should also be specified. Remote batch terminals, line printers, and magnetic tape drives are likely candidates for inclusion in the benchmark configuration specifications.

Validation

The complete benchmark demonstration mix(es) must be validated by running it (them) on at least one system, and preferably on two systems. One reason is to validate that the programs, transactions, data, and equipment configuration as specified are correct. A second reason is to gain as much insight as possible into the magnitude of the system likely to be bid in order to avoid surprises. A third reason is the mapping of the workload requirements and performance objectives into the benchmark time frame. There are other validation techniques which should also be performed wherever practical, to confirm the expected results. Among them are analytical methods, e.g., projecting instruction and data accessing rates, and simulation.

Procurements which include systems software not available on the agency's present systems, may handicap the benchmark developers in attempting to validate the benchmark. That is, benchmark programs may be developed in such a way that they depend upon the missing software (to be provided as part of procurement) in order to execute properly. However, the benchmark components supplied by the Government still must be validated by execution, even if in a degraded mode. One technique which has been used successfully has been to use emulators which provide the missing functions. Timed runs for sizing purposes are still possible by excluding or otherwise allowing for inefficiencies caused by the emulation software.

Functional Tests

In addition to the benchmark mix demonstration, programs and data may be required for functional tests. Some functional tests may require only vendor demonstrations, and not agency-supplied materials. Documentation of functional test material should make it clear whether it is for an independent test that is not included in the benchmark mix demonstration timing.

Physical Benchmark Package

The benchmark package includes the physical files containing the programs, data for the benchmark, and their documentation. It should include the following components:

1. Listing of the source code for each benchmark program
2. Compilation listing
3. Execution output
4. Description of data files
5. Listing of data file generator programs
6. Listings of other pre- or post-benchmark programs (those not included in the timed demonstration)

It is preferable to use magnetic tape for delivering program and data files to the vendor. Also, maximum use should be made of putting multiple files on single reels of magnetic tape.

This is especially emphasized for program files to minimize the physical size of the benchmark package. Punched cards may be used if the volume is very small. If cards are utilized, make certain that they include a sequence field or can be sorted on an actual data field. Whenever practical, hard copy listings should also be provided.

Files supplied on magnetic tape should utilize a minimum number of different formats and recording modes for the data. For example, it is preferable that a single label structure, blocking factor, density, and recording mode be used.

Each tape supplied should be carefully labeled externally and cross-referenced to the documentation describing the file contents. Documentation for each file should include: label information, recording density and mode, blocking information, file structure and record format(s) for data elements in each record type. Data should be supplied in ASCII character mode rather than pure binary mode or nonstandard character codes in order to minimize machine dependencies. The Procedural Documentation should indicate (Section C) the mode to be used during the actual test demonstration. For example, even though data is supplied in character form, it may be permissible (or required) that the data be in binary form during the actual demonstration.

Efforts should be made to minimize the volume of test files. There are a number of ways to generate large data files including: use of controlled random number generation programs; and the use of sampling techniques to obtain data elements from small files or sets of tables which are then used to generate the required number of records.

The documentation supplied with the program and data files should indicate the role of each file. For programs, indicate whether the program is part of the benchmark mix demonstration or is a pre- or post-processing program (e.g., a program to generate a test file, program to validate test files, program to collect and summarize system logs, etc.).

Each copy made of a file should be validated against the original to ensure its accuracy. The simplest method of validation is to make each new file copy from the previous copy and to compare the last copy made to the original. In order to assist the vendor in determining the validity of the files received, check sums and hash totals should be provided for each logical file.

C. Procedural Documentation and Preparation of the Benchmark for the Vendors

Documentation for a benchmark demonstration typically involves three components: 1) documentation of the physical files used to distribute the benchmark programs and data, 2) description and definition of tasks and activities making up the benchmark demonstration (Procedural Documentation) and 3) the detailed Benchmark Demonstration Management Plan. The first two components are usually delivered as a single package well in advance of the actual benchmark demonstration. This section describes the second component of these three—the documentation which describes and defines the benchmark demonstration in terms of the various tasks, interaction and interrelation of programs and files, sequence of tasks, resource requirements for the demonstration (including hardware, software, personnel, and time), output requirements, measurements and timings, functional tests, and evaluation criteria. This documentation is referred to as the Procedural Documentation.

The procedural documentation should be delivered to each vendor well before the actual demonstration to provide the vendor adequate time to assemble the required resources and make trial runs of the benchmark demonstration. Early distribution is also important to allow time for modification or clarification of the test.

The benchmark package should be sufficiently comprehensive and clear in order to allow the vendor to prepare for the benchmark mix demonstration. One method of reducing the probability of misunderstanding and "surprises" is for the user to review the output of vendor benchmark programs for verification prior to the actual, on-site benchmark.

The sections of the Procedural Documentation include:

(1) Overview of the Benchmark Demonstration: objectives of the benchmark, the benchmark environment, nature and scope of the test, responsibilities of vendor and Government.

(2) Resource Requirements: Hardware, Software, People, and Time.

(3) System Hardware and Software Configuration: Allowable Modifications and Restrictions.

(4) Benchmark Mix Demonstration Tasks:

   a. Programs, files, and outputs
   b. Terminal activity
   c. Starting conditions
   d. Sequence and repetition of programs and terminal activity

(5) Measurements

(6) Output Requirements

(7) Post-Benchmark Demonstration Tasks

(8) Functional Demonstrations

(9) Evaluation Criteria and Methodology

(10) Glossary of Terms

The following paragraphs detail the sections of the Procedural Documentation.

1. Overview of the Benchmark Demonstration

This overview and summary of the benchmark test should first define the objective of the demonstration: the primary purpose and the significant information to be obtained from the benchmark demonstration. Then it should describe the nature and scope of the test outlining the batch, interactive, real-time processing, and telecommunications activities and any functional demonstrations which will be involved.

This introductory section should also establish the ground rules and regulations for the test, identify and provide information on the Government contact point, define procedures for requesting modifications or clarification, and describe general responsibilities of the Government and the vendor in regard to providing a smooth demonstration. Procedures for coordinating the dates for the actual benchmark demonstration should be established.

2. Resource Requirements: Hardware, Software, People and Time

This section should describe specific resources which will be required to conduct the demonstration. In many benchmark demonstrations, it will not be necessary for the vendor to include the complete complement of hardware during the benchmark that is required by the solicitation document. For example, a subset of the required storage and I/O devices may be sufficient for the benchmark. Allowable deviations from the solicitation document should be indicated, making it clear that the hardware required for the benchmark in no way alters the requirements of the solicitation document. At the same time, some benchmarks may require additional equipment such as terminals or measurement devices to support the test specified in the solicitation document. Vendors should be allowed flexibility with such hardware to the extent that substitution does not unfairly bias the results of the test. For example, terminal requirements should specify functional characteristics only to the extent they affect the test—hard copy and CRT terminals have essentially the same characteristics if communication modes, data rates, transaction characteristics, and timings are the same. For user-supplied hardware, complete interface requirements should be defined. Also the vendor should be given sufficient time to interface user-provided hardware.

Software systems the vendor is expected to provide may include specialized monitoring, measuring, and logging techniques as well as output requirements beyond those explicitly stated by the solicitation document. These, however, should be kept at a minimum.

Personnel requirements include vendor personnel for conducting and managing the benchmark. Any limitations on the number of vendor personnel who may be present in the immediate benchmark area during the demonstration should be defined. Also, the composition of the user benchmark team and their general responsibilities should be defined. (Specific responsibilities of the user team are described in the Benchmark Demonstration Management Plan in Section D.) The period of time during which members, groups of members, or the complete user benchmark team is available should be indicated.

The "time" resource is the general schedule for the benchmark (as opposed to the specific task sequencing and timing) including the number of days or hours permitted the vendor for successful completion of the benchmark and the hours or shifts during which the user benchmark team will be available for consultation and/or the benchmark demonstration.

3. System Hardware and Software Configuration: Allowable Modifications and Restrictions

The solicitation document should specify the level of detail to which the computer system hardware/software configuration is to be described in the vendor's proposal. Any deviation from this description in the system as benchmarked is considered a proposal modification. However, modifications which result in a nonstandard product are usually disallowed. The Procedural Documentation should define specific requirements for documenting vendor hardware/software configuration modifications.

Within the level of specificity required by the user for the proposed hardware/software system, the vendor may optimize configuration and operating system options and parameter selections to take best advantage of his system for the benchmark mix demonstration. However, in order to ensure that the integrity of the benchmark mix demonstration is not altered in context with its relationship to the actual and planned workloads, the user must determine that the level of detail of each vendor's description of the vendor's proposed hardware and software is adequate. This also requires having

a clear distinction between the vendor's description of the specifics of the vendor's proposed hardware/software system and the vendor's description of hardware/software capabilities which may be "available."

In order to protect the integrity of the benchmark mix demonstration, the user should clearly specify software capability requirements and should specify constraints on how the mix is to be run. The latter includes considerations such as dedication of particular resources to specific tasks or activities which may be either required or forbidden, any interrelationships or dependencies between programs in the mix, etc. The user should also specify that all proposed software be online and available during the benchmark mix demonstration.

The user is particularly urged to give detailed consideration to hardware/software proposal requirements and benchmark mix construction in order to ensure that optimization of the vendor's proposed hardware/software configuration for the benchmark mix demonstration will also be beneficial for the actual workload.

4. Benchmark Mix Demonstration Tasks

This section of the documentation should define the workload for the benchmark demonstration. It should describe programs, files, and outputs; should specify parameters such as the number of repetitions of a particular event; should describe the relationship of programs and files, sequencing of programs or events; starting conditions, terminal activity; and should describe the allocation of resources to processes.

The use of graphical presentations such as flowcharts to define task sequencing and interaction is highly recommended. The total benchmark should also be summarized in tables which indicate the maximum running times, resource requirements, input and output volumes, and files used by each task.

4a. Programs, Files and Outputs

This sub-section should describe the various programs and files making up the benchmark demonstration package and provide sample outputs. Program descriptions should include the following:

(1) General nature of the program and types and modes of processing, e.g., file update, matrix inversion, sort; interactive batch, remote job entry.

(2) Allowable modifications and optimization of code.

(3) Files used by each program including requirements for intermediate or scratch files and type of intermediate files, e.g., disk, tape.

(4) Memory requirements and constraints.

(5) Timing limitations. Timing limitations are discussed in Section A.

Where a benchmark demonstration will involve more than a single configuration, for example when the solicitation document allows system augmentations over time, the association of programs, files, and outputs with the workload requirements over time must be fully described.

Information relating to the benchmark test data base includes descriptions and restrictions on such factors as:

(1) Record blocking.

(2) Organization of files on direct access media, e.g., sequential, indexed sequential, direct access.

(3) Data representation—for example, can the data representation be selected by vendor or must the data be used as supplied?

(4) Any required recording density.

(5) Restrictions on character sets used—for example, ASCII on output files that will be interchanged in the "real" operation.

(6) Restrictions on rearrangement of data elements within records or more extensive reorganization of files on direct access devices.

(7) Allocation of files to devices—for example, can an index and its file be on separate devices, are multi-file reels or packs required, allowed, or not allowed? Are multiple copies of input files permitted?

As discussed in previous sections of these guidelines, the vendor should be given "reasonable" flexibility to modify programs and files to allow his system to perform efficiently. This includes alteration of file structures and allocation of system resources to programs or processes for optimum performance. It does not, in most cases, include alteration of source code except to the extent that certain coding practices may unfairly bias the performance of a particular vendor's system. This section of the documentation must describe the limits on such

modifications and must specify the procedure for requesting variances from these limits.

The vendor should be provided with a copy of the output from the processing of each benchmark program with its associated files as delivered to the vendor (or to be generated) as the benchmark package. This output should include printed program outputs, console messages, terminal transaction input with its accompanying output, and listings of compile and load tasks where such tasks are part of the benchmark. Many system generated messages will be specific to the source system and each vendor must interpret or translate such messages into information meaningful to the vendor's own system operation in order to validate correct implementation of the benchmark. Nevertheless, such output is valuable to the vendor to confirm the proper functioning of the vendor's system.

4b. Terminal Activity

This section should describe the tasks to be performed from terminals; whether live, emulated or otherwise represented during the benchmark demonstration tests. Documentation for terminal activity should describe each task and include as a minimum:

(1) Description of terminal input and/or output (using the terminology, if possible, of the vendor's system for specific entries such as system commands and editing operations).

(2) Number of repetitions of each series of inputs.

(3) Timing of input messages, e.g., random or fixed interarrival times.

(4) Number of terminals allocated to each activity.

(5) Functional characteristics of terminals.

4c. Starting Conditions

This section should describe the state of the system at the start of the benchmark demonstration, indicating those activities which may be performed prior to the initation of the actual test. The description of the starting state conditions should include:

(1) Allowable premounting of input and output media such as tapes, disk packs, or loading of cards to reader.

(2) The number of terminals and the sequence and schedule for log-on.

(3) The state of the operating system.

Since the objective of a benchmark demonstration is to represent the expected operating environment, it may be advisable to initiate a number of tasks prior to initiation of timing measurements. For example, some terminals may be logged on, and some repetitive background activity may be in progress. Pre-initiation of several tasks should reduce the start-up transients and thereby make the timed portion of the test more realistic.

4d. Sequence and Repetition of Programs and Terminal Activity

This section should describe sequence requirements, if any, for each task, the number of repetitions, and the allocation of processes to specific resources. Where programs are to be executed in a specified order, such as when one program utilizes output from another program, specify the relationships using systems charts, flowcharts, or system block diagrams. When programs are to be executed more than once, specify whether the vendor may select the sequence or if it is prescribed. Where a program is repeated, specify if multiple copies may be in simultaneous execution.

Define the allocation of resources and devices to each task. For example, in terminal and telecommunications environments, specify any required assignment of programs to terminals. It may be necessary to specify sequencing and repetitions for each terminal or other input sources. For example, each terminal or group of terminals may have a particular sequence of program executions and number of external repetitions specified.

Tasks should also include operator action and, in terminal oriented systems, terminal activity. Any restrictions on the timing and/or the order of manual operations should be described. Since terminal activity will usually involve several types of tasks, e.g., editing, transaction entry, program compilation, entry of system commands, the timing of each of the tasks and their interrelationship must be described. The use of flowcharts augmented by timing indications may be a useful method of defining these relationships and sequences.

5. Measurements

The Procedural Documentation should describe the general measurements to be made of the system during the demonstration. The Benchmark Management Plan should detail the method to be used to take these measurements, individuals responsible for recording the measurements, and forms to be used for recording manual measurements.

The procedural documentation should outline measurements to be taken in the following areas: (1) timings including throughput time, terminal response times, and total benchmark time; and (2) resource utilization such as memory requirements and CPU and channel activity levels, if necessary. Such measurements should be described for each task or group of tasks making up the benchmark. Measurements should only be taken if they are used in the evaluation process.

6. Output Requirements

Output to be generated by the benchmark test should be described and the output that will be collected by the Government benchmark team should be indicated. Output may include hard copy printer output as well as files written on magnetic media. Output may be further classified as normal output generated in the execution of the test programs and output which includes measurement information such as system logs or monitor output.

7. Post-Benchmark Mix Demonstration Tasks

Following the timed benchmark mix demonstration it is often necessary to run additional programs or utilities to assist in validating the benchmark. Such tasks may include copying of disk files to magnetic tape for later analysis, computation of check sums and hash totals on updated files, and programs for sampling of file records.

Programs required for post-benchmark mix demonstration tasks should be supplied with the benchmark material including a time estimate for their completion.

8. Functional Tests

While the subject of functional tests is not the major concern of these guidelines, functional tests are often performed as part of the total benchmark demonstration. The functional tests should also be described in the Procedural Documentation including their schedule and time requirements, resource requirements, measurements, output generated, etc.

9. Evaluation Criteria and Methodology

The solicitation document should describe the criteria to be used for evaluation of proposed systems. The Procedural Documentation should be summarized and analyzed consistent with the evaluation criteria stated in the solicitation document. Benchmark tests may generate a considerable amount of timing and resource measurements which may require automated data reduction to arrive at summary figures such as response times, etc. The data reduction procedures or programs should be defined for the vendor so that there would be no ambiguity in how the final measures are to be computed.

10. Glossary of Terms

A glossary should be developed which defines any terms used within the benchmark documentation which may have special, ambiguous, difficult to understand, or user-dependent meanings. The glossary should be included with the benchmark documentation package.

D. Conducting Benchmark Tests

This section provides guidelines for the management of the benchmark demonstration. This phase of the total benchmark activity includes the formation, organization and responsibilities of the benchmark team, preparation for the conduct of the demonstration, the post-demonstration analysis, and validation of the results. The documentation which should be prepared for the demonstration is described in the section Benchmark Management Plan.

Make Up of the Benchmark Team

The benchmark team should be made up of individuals familiar with the requirements of the solicitation document, the structure of the benchmark test, and the benchmark programs. Every effort should be made to keep the size of the benchmark team to a minimum. The actual size of test teams will vary depending on the size and type of system being procured and the complexity of the benchmark test. One individual should be appointed benchmark team leader and held responsible for the conduct of the benchmark test. Individuals familiar with the selected programs should be assigned the task of program validation. Those individuals familiar with the hardware and software requirements should be assigned the task of validating that the system being benchmarked conforms to the system being proposed. The structure of the benchmark team and duties and responsibilities of the members should be delineated in the Benchmark Management Plan.

Trial Benchmarks

The benchmark team should be organized and trained prior to the first live benchmark test at a vendor's site. A valuable training exercise is to perform a trial benchmark in as realistic an environment as possible. Such a trial can serve not only as training for the team but also for

uncovering problems, omissions, and errors in the benchmark package. This trial may indicate the need for modification of the benchmark programs, procedures, or the Benchmark Management Plan. Thus, it should be performed early enough in the procurement process to avoid delay of vendor benchmarking demonstrations. It is advisable to perform this trial benchmark prior to releasing the benchmark package to the vendor. This user conducted trial benchmark will ensure that the package will run on at least one machine and should reduce problems associated with the vendors' conversion of the benchmark package for the vendors' systems.

Basic Ground Rules

The design of the benchmark programs, files, and tasks, the quality of documentation provided to the vendor, and the overall quality control exercised over the benchmark package, will have a major influence on the success of the actual demonstration.

Also of critical importance will be the preparation of the benchmark team and their performance and demeanor during the actual demonstration. The benchmark demonstration is the single most sensitive event in the acquisition cycle. There is no other point when the vendor is more anxious or apprehensive about the possibility of not meeting a mandatory requirement. Because of this, it is important to maintain a good working relationship between the benchmark test team and the vendor personnel.

The recommendations provided to minimize problems which relate to the design, quality control, and documentation of the benchmark package are discussed in previous sections. The following guidelines relate to the preparation of the benchmark team and the on-site demonstration.

- Treat all vendors the same.
- Remain objective at all times, do not help a vendor to pass or to fail.
- Limit the size of the benchmark team to the extent practical.
- Require the vendor to demonstrate a system identical in all aspects to the system as proposed or as officially modified by the vendor. Any exceptions to this should be only those variances specifically allowed by the Procedural Documentation.
- Require the vendor to have a copy of the proposal and the solicitation document available at the benchmark site.
- Require the vendor to provide the benchmark team with a private conference room during the test period.
- Identify focal points of communications during the test period.
- Do not discuss the participation, benchmark performance, or proposals of competing vendors with any other vendor personnel.
- Observe Federal and agency regulations on acceptance of gratuities.

The Benchmark Demonstration Management Plan

The purpose of this plan is to describe the agenda and schedule of the benchmark and to specify the duties and responsibilities of each member of the benchmark team. Sections of the plan relevant to the vendor should be made available to him several weeks in advance of the demonstration. The components of this plan include the following sections:

Test Team Functions and Responsibilities

This section of the plan should include the responsibility assignments of the team members. Specific responsibility functions will usually include: Government spokesperson, demonstration team leader, console timer, other timers, hardware specialist, software specialist, and product validator. The extent to which a single individual has multiple responsibility will depend on the size and complexity of the benchmark. Specific duties and responsibilities of the team members may include:

Government Spokesperson: Presents the official Government position when required and provides liaison between the vendor representatives and the test team.

Test Team Leader: Manages the benchmark demonstration and the benchmark test team, including assignment of duties and functions to team members; serves as the focal point for all recorded information gathered by the team; and is responsible for the satisfactory completion of all benchmark tasks.

Console Timer: Times and records all runs and other events; acquires and identifies console logs; and assists other members in timing peripheral devices when necessary.

Other Timers: Assigned to specific peripheral devices for timing, acquiring and identifying output; and overseeing test data input.

Hardware Specialist: Conducts hardware configuration survey, participates in hardware discussions and obtains hardware certificate from vendor agent.

Software Specialist: Participates in software discussions and obtains software certificate from vendor agent.

Product Validator: Oversees introduction of test data and analyzes output products for acceptability.

In addition to these specific duties, each test team member may be requested to provide a written report of observations; to assist in timing when not involved with other specific tasks; and to assist in organizing and analyzing the output.

Behavior of the Test Team

This section of the plan should describe any restrictions on contact with the vendor personnel; acceptance of gratuities; discussions with vendor personnel; and operation of vendor equipment.

Agenda

This section of the plan should describe the user responsibilities in regard to the vendor's agenda for the benchmark demonstration.

The user should ensure that the vendor's agenda is satisfactory and describes the general required activities during the visit of the benchmark team to the vendor demonstration site. The following sequence of activities is intended as an example; actual activities will depend on the specific type of system being tested and on the benchmark design:

- Introductory Remarks by the Government Spokesperson
- Demonstration Briefing by Test Team Leader
- Vendor Briefing
- System Verification
- Preparation of Test Data
- Benchmark Mix Demonstration
- Functional Demonstrations, if required
- Closing Remarks by the Government Spokesperson

Measurement and Documentation of the Test

This section of the plan describes the specific timing and resource measurements, recording and certification documents, system output, and malfunction recording to be made during the test. It should include:

Timing Measurements

Procedures and definitions for timing various events must be specified in detail. Timings may be obtained in several ways for various events and the procedure should be clearly defined for each event. For example, response times in interactive processing may be measured by a monitor while times for batch execution may be obtained from system logs or by calls to a system clock by executing programs. Clear definition of timing procedure is important and the start and end conditions for each event timed must be carefully specified.

Timing documentation should also describe the number of timings and/or sampling procedures to be used for timings, the number of independent measurements to be made of each timing, the precision of timing measurements, and how the timings will be summarized (e.g., averages, medians, percentiles, ranges, etc.). This section may or may not be distributed to the vendor.

Resource Measurements

Methods for recording and measurements of resource requirements for various tasks and phases of the benchmark should be defined and documented. Such measurements may include memory requirements, number of each type of peripheral devices used, and resource utilization data obtained from software and/or hardware monitors. The role of each of these measurements in the evaluation process must also be stated.

Recording Forms

Where timings and resource measurements are obtained by team members (as opposed to system logs, monitors, program calls, etc.), specially prepared forms should be designed and used. Forms should have space for recording comments to describe malfunctions or other unexpected occurrences. When malfunctions are reported which require vendor corrective action, such action should also be documented.

Forms should be developed for validation and certification of the hardware, software, and test data used in the benchmark demonstration. Recommended steps for validation and certification are described in the section entitled, Conduct of the Benchmark Test.

System Output

The required output from applications, systems, and monitoring programs to be collected as part of the test for each task or

phase of the benchmark should be described. Each team member's responsibilities for obtaining and labeling output should also be clearly stated. Checklists should be included to be initialed by members at the completion of each task. The documentation should define vendor requirements for packaging and mailing benchmark output to the user's facility.

Conduct of the Benchmark Test

The specific tasks to be performed as part of the benchmark will have been previously defined to the vendor in the Procedural Documentation (Section C). The Procedural Documentation describes the programs, files, resource requirements, and sequence of performance for each task. The Benchmark Demonstration Management Plan should provide a detailed schedule for the demonstration, definition of starting conditions for various tasks, modification procedures for the test data, and contingency plans —for example, what happens in the event of a system crash?

Specific topics that should be included are:

Detailed Schedule for the Test

The schedule should expand the agenda to provide detailed timing for the execution of various events. The expected duration for events or upper limits, where appropriate, should be stated.

Test Data Modification

Modifications should be made to the data used for the benchmark test to reduce the effects of any vendor tuning of the system to a specific set of data. These changes should be made after arrival of the benchmark team at the vendor site. Procedures for making these changes should be as simple as possible and should be clearly specified. Methods for altering test data include changing parameters of data generators, controlled randomization of data elements across records, alteration of data elements in selected records, and merge of several files. Such changes should not appreciably affect timing of the benchmark programs.

The Procedural Documentation defines the system prior to the institution of the timed portion of the benchmark mix demonstration. The Benchmark Management Plan should detail the steps necessary to establish and verify the initial condition.

Contingency Plans for Malfunctions

Procedures should be established for malfunctions during the demonstration. Each team member should understand his responsibilities for documenting malfunctions, the allowable corrective action, and the effect of a malfunction on timing and other measurements. Conditions should be defined for determining when the test or particular tasks within the test can be restarted, and determining when the vendor has failed the test.

Validation and Certification

Prior to and following the conduct of the benchmark mix demonstration there are a number of procedures that must be followed in order to ensure that the benchmark mix was processed as intended. These steps are designed to validate the hardware, systems software, test data, and benchmark programs.

Hardware Certificate

A detailed survey of the hardware in use should be conducted by team members under the supervision of the hardware specialist. This inspection should ensure that hardware not included in the vendor's proposal is indeed not in use. Any deviations in the hardware model from that proposed should be noted on the hardware certificate. The hardware certificate should be signed by the vendor's agent and by the hardware specialist.

Software Certificate

A software certificate listing the software packages in use during the demonstration should be prepared by and signed by the vendor's agent. Any variation from the software in the vendor's bid should be noted. Procedures should be established and supervised by the software specialist to verify the software packages in use. Such verification may require central memory dumps or listings, listing of the contents of external storage devices, or specific tests of the software.

Benchmark Program Validation

Procedures should be developed and described to make certain that the benchmark programs have not been modified by the vendor to a greater extent than allowed and documented. This will usually require that programs be available in source form. These same programs may then be combined and the resulting object versions used for the benchmark mix demonstration.

Validity of program logic may be tested by executing the program with test data and comparing the results to known correct output. This will, of course, only enable one to determine that the logic of the programs at the vendor

site is equivalent to the logic of the original benchmark programs on a given set of test data. Provision must also be made for differences in output due to differences in machine precisions.

Procedures should also be described for comparison of the vendor programs with original programs at the source level. This comparison is essential to ensure that any modifications required by the vendor to compile and execute the benchmark have not also resulted in unpermitted optimization to the source code.

Procedures should also be developed to make certain that library functions have not been optimized or modified to reduce run time. Compilation listings, load maps, and dumps may be required to verify that subtle changes do not provide one vendor with an unfair advantage.

Test Data Validation

As described in the section entitled, Test Data Modification, the test data should be modified at the vendor site. The validity of the modified test data should then be determined in one of several ways. The safest but most time consuming and expensive validation involves an element by element comparison with known correct test data. This comparison can be made by obtaining a machine readable copy of the test data. Another method for validation of the test data is to compute check sums and hash totals.

At the conclusion of the benchmark demonstration, updated data files should also be tested to ensure that they have been processed as intended. Again, element by element comparisons, sampling, or computation of checksums and hash totals can be used as validation means. Such validation will also help to ascertain the proper functioning of the entire hardware-software complex.

Benchmark Evaluation

Prior to departing from the vendor demonstration site the benchmark team should make sure that all necessary test results, records, and output have been obtained and are properly labeled.

If possible, the benchmark should be designed to permit evaluation of the results at the vendor site shortly after the benchmark demonstration is completed. If the results can be evaluated at the vendor site, the benchmark team spokesperson should indicate to the vendor whether the benchmark was passed or failed prior to the benchmark team departure. However, where complex data reduction is required to determine the pass/fail question, care should be taken to avoid an ad hoc estimate of a vendor's performance.

In all situations, determination of whether the vendor passed or failed should be made as expediently as possible and communicated to the vendor. The vendor usually has considerable resources tied up in the equipment configured for the benchmark demonstration and needs to know as soon as possible if a rerun will be required.

The benchmark team should prepare an analysis of the output products, the system performance, and resource utilization for inclusion in an objective report of observations and findings. This report should present the team's findings in a form which facilitates evaluation of the vendor's system against the evaluation criteria stated in the RFP. This report may be used to facilitate the preparation of the post-award debriefing.

U. S. GOVERNMENT PRINTING OFFICE : 1977 O - 232-489

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH reports National Bureau of Standards research and development in physics, mathematics, and chemistry. It is published in two sections, available separately:

• **Physics and Chemistry (Section A)**
Papers of interest primarily to scientists working in these fields. This section covers a broad range of physical and chemical research, with major emphasis on standards of physical measurement, fundamental constants, and properties of matter. Issued six times a year. Annual subscription: Domestic, $17.00; Foreign, $21.25.

• **Mathematical Sciences (Section B)**
Studies and compilations designed mainly for the mathematician and theoretical physicist. Topics in mathematical statistics, theory of experiment design, numerical analysis, theoretical physics and chemistry, logical design and programming of computers and computer systems. Short numerical tables. Issued quarterly. Annual subscription: Domestic, $9.00; Foreign, $11.25.

DIMENSIONS/NBS (formerly Technical News Bulletin)—This monthly magazine is published to inform scientists, engineers, businessmen, industry, teachers, students, and consumers of the latest advances in science and technology, with primary emphasis on the work at NBS. The magazine highlights and reviews such issues as energy research, fire protection, building technology, metric conversion, pollution abatement, health and safety, and consumer product performance. In addition, it reports the results of Bureau programs in measurement standards and techniques, properties of matter and materials, engineering standards and services, instrumentation, and automatic data processing.

Annual subscription: Domestic, $12.50; Foreign, $15.65.

NONPERIODICALS

Monographs—Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NBS, NBS annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a world-wide program coordinated by NBS. Program under authority of National Standard Data Act (Public Law 90-396).

NOTE: At present the principal publication outlet for these data is the Journal of Physical and Chemical Reference Data (JPCRD) published quarterly for NBS by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements available from ACS, 1155 Sixteenth St. N.W., Wash. D. C. 20056.

Building Science Series—Disseminates technical information developed at the Bureau on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NBS under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The purpose of the standards is to establish nationally recognized requirements for products, and to provide all concerned interests with a basis for common understanding of the characteristics of the products. NBS administers this program as a supplement to the activities of the private sector standardizing organizations.

Consumer Information Series—Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

Order **above** *NBS publications from: Superintendent of Documents, Government Printing Office, Washington, D.C. 20402.*

Order **following** *NBS publications—NBSIR's and FIPS from the National Technical Information Services, Springfield, Va. 22161.*

Federal Information Processing Standards Publications (FIPS PUBS)—Publications in this series collectively constitute the Federal Information Processing Standards Register. Register serves as the official source of information in the Federal Government regarding standards issued by NBS pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NBS Interagency Reports (NBSIR)—A special series of interim or final reports on work performed by NBS for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Services (Springfield, Va. 22161) in paper copy or microfiche form.

BIBLIOGRAPHIC SUBSCRIPTION SERVICES

The following current-awareness and literature-survey bibliographies are issued periodically by the Bureau:

Cryogenic Data Center Current Awareness Service. A literature survey issued biweekly. Annual subscription: Domestic, $20.00; Foreign, $25.00.

Liquified Natural Gas. A literature survey issued quarterly. Annual subscription: $20.00.

Superconducting Devices and Materials. A literature survey issued quarterly. Annual subscription: $20.00. Send subscription orders and remittances for the preceding bibliographic services to National Bureau of Standards, Cryogenic Data Center (275.02) Boulder, Colorado 80302.

U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE
COM-211

3rd Class Bulk Rate

U.S.MAIL