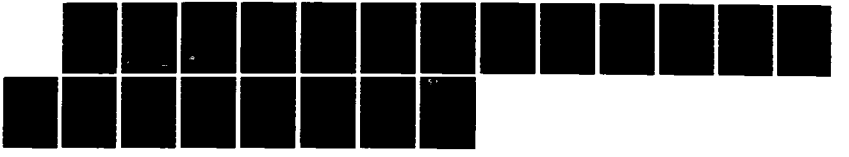
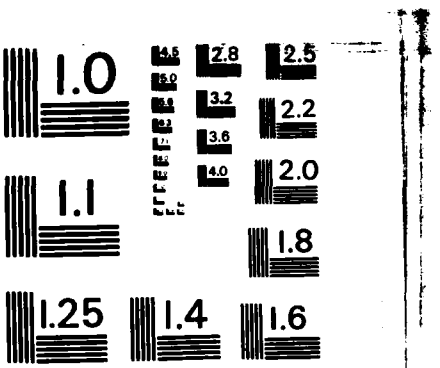
AD-A139 863 R & D SUPPORT FOR THE CENTER FOR SEISMIC STUDIES(U) 1/1
SCIENCE APPLICATIONS INC MCLEAN VA C F ROMNEY ET AL.
15 JAN 84 SAI-84/1077 MDA903-84-C-0020

UNCLASSIFIED F/G 8/11 NL





(12)

R & D SUPPORT FOR THE CENTER FOR SEISMIC STUDIES

QUARTERLY TECHNICAL REPORT

3 OCTOBER 1983 - 30 DECEMBER 1983

SCIENCE APPLICATIONS, INC.

(12)

R & D SUPPORT FOR THE CENTER FOR SEISMIC STUDIES

QUARTERLY TECHNICAL REPORT

3 OCTOBER 1983 - 30 DECEMBER 1983

SCIENCE APPLICATIONS, INC.

Post Office Box 1303, 1710 Goodridge Drive, McLean, Virginia 22102, (703) 821-4300

R & D SUPPORT FOR THE CENTER FOR SEISMIC STUDIES

QUARTERLY TECHNICAL REPORT FOR THE PERIOD 3 OCTOBER 1983 THROUGH 30 DECEMBER 1983

Sponsored by:
Defense Advanced Research Projects Agency
Under ARPA Order No. 4882
Monitored by:
Defense Supply Service - Washington
Under Contract No. MDA903-84-C-0020

Science Applications, Inc.
1710 Goodridge Drive
McLean, Virginia 22102

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A139862	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* R&D Support for the Center for Seismic Studies		5. TYPE OF REPORT & PERIOD COVERED Quarterly Technical
		6. PERFORMING ORG. REPORT NUMBER SAI-84/1077
7. AUTHOR(s) Carl F. Romney Louis Huszar Gerald A. Frazier		8. CONTRACT OR GRANT NUMBER(s) MDA903-84-C-0020
9. PERFORMING ORGANIZATION NAME AND ADDRESS Science Applications Inc. 1710 Goodridge Drive McLean, VA 22102		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, VA 22209		12. REPORT DATE January 15, 1984 13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* Defense Supply Service - Washington Room 1D245, The Pentagon Washington, DC 20310		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

Unlimited

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Seismic
Data Center

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

An analysis of the state of development of the physical facilities and software at the Center for Seismic Studies has identified areas of uncompleted work. Actions were initiated to complete this development. Planning began for an experiment to test the ability of the center to function as a "National Data Center" and an "International Data Center" as defined by the Group of Seismic Experts, U.N. Committee on Disarmament.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

TABLE OF CONTENTS

	Page
INTRODUCTION	1
ACCOMPLISHMENTS	2
APPENDIX 1	1-1
APPENDIX 2	2-2

INTRODUCTION

The Center for Seismic Studies has been under development for several years under a project directed by the Defense Advanced Research Projects Agency (DARPA). Development of the physical and software facilities of the Center, and the research to be accomplished at the Center using these facilities, are part of DARPA's program to develop techniques that can be used to improve national capabilities to verify compliance with potential nuclear test ban treaties.

During an early phase of the work, a system for receiving, preprocessing, storing and retrieving seismic data was designed and subjected to proof-of-principle testing. The primary technical challenges were associated with the coherent management of large volumes of data received in numerous different formats, arriving by several different means ranging from mailed tapes to instantaneous electrical transmission, and including both digital waveform and alphanumeric data.

The following phase included establishing a facility and constructing a prototype system to test the design concepts, and developing selected data bases to support seismic research. A test of the ability of the prototype data center to handle high-data-rate information was provided by receiving on-line digital waveform data from five North American seismic stations operated under auspices of the Department of Energy (the so-called Regional Seismic Test Network-RSTN). Similarly, arrangements were made to receive alphanumeric data from the National Earthquake Information Service (NEIS), from selected stations operated by Canada and the United Kingdom, and from other sources. This phase also included the development of software for preprocessing the incoming data and for performing elementary scientific analyses of the data. Most of this work

has progressed to the point that all essential capabilities can be demonstrated, but there are engineering or software problems remaining throughout the system.

The goals of the current phase are to complete the development, test and evaluation of the prototype data center, and to establish the center as a focal point for research and support to the DARPA nuclear test verification program. This entails establishing a management structure to guide the Center's activities, a small resident research staff, and a developmental staff. Principal developmental tasks include completion of the hardware and software systems associated with routine data acquisition, evaluation and documentation of the performance of these systems in an essentially automatic mode, and the development of user-friendly software for seismic analysis. Principal research tasks include application of the Center's data and computing resources to the development and testing of seismic analysis methods, designing special research data bases, and conducting studies and evaluations of U.S. seismological capabilities.

This report describes technical activities directed toward these goals during the first quarter of FY 1984.

ACCOMPLISHMENTS

Work during the first quarter concentrated on the establishment of priorities for further developmental activities and the initiation of work for advancing Center objectives. A systematic effort was undertaken to evaluate the status of the various capabilities and resources, namely:

- computers (three VAX 11/780's and two PDP 11/44's) and a range of peripheral devices (mainly, printers, storage devices, graphics devices, and remote terminals);

- computer software, including operating systems, on-line data processing programs, data base management software, graphical display and demo packages, applications programs, and research tools;
- data acquisition systems, notably, systems for receiving telemetered digital data;
- archived data bases; and
- documentation of functions, procedures, and makeup of the various system components.

This systems evaluation resulted in the identification of several important problem areas where further developmental work was needed. The performance of the data acquisition system was found to be unreliable, and the documentation for the troublesome components (the communications interface system) was found to be inadequate for timely remedial action. Computer software was not developed, tested and documented to the level that was expected or needed for the Center's participation in an upcoming test of an international exchange of seismic data and determination of source parameters. Large parts of the data on hand had not been archived in the Center's standard retrievable data bases. Operations of the computer-based systems were generally found to be difficult to master, although, once mastered, these systems provide capabilities and ready access to data unrivaled in the seismic research community.

In short, considerable effort was found to be needed to make the various systems function reliably and to provide documentation and user-friendly interfaces to facilitate their use by researchers unfamiliar with the particular software operating at the Center.

Appendix 1 summarizes the status of developments at the Center as of December 1983 with respect to DARPA objectives for FY 1983.

While the primary activity during the first quarter was devoted to the evaluation of existing systems, several important advancements were made, including specific measures that were implemented to remedy deficiencies noted above. The most significant accomplishments are presented below.

a. Steps were taken to improve the utilization of computer resources at the Center. A decision was made to upgrade the operating system to UNIX version 4.2, which contains important enhancements such as a needed facility for managing allocations of storage space on the computer disks. However, a companion decision was to make no further upgrades beyond version 4.2 for the forseeable future, because several deficiencies found at the Center were attributable to the past practice of implementing successive versions of the operating system as they were issued, which detracted from completion of development work already underway.

A review was conducted of individuals and organizations having access to computers at the Center. A new authorization list was developed, which limits access to those having research requirements validated through DARPA and to individuals providing direct support to Center functions.

The installed version of the Ingres data base management system was compared with RTI Ingres for performing typical Ingres tasks. Benchmark tests, reported in Appendix 2, show clear superiority of RTI Ingres over the installed version; RTI Ingres executes a common task involving the selective retrieval of data at more than 50 times the speed of the installed Ingres. This improved efficiency is fundamental for accessing data at the Center, and therefore, action was initiated to acquire and install the RTI version of Ingres without interfering with data accessibility.

b. Planning began in December for an experiment to be conducted during the third week of January for the purpose of assessing the Center's capabilities and testing procedures for a subsequent multi-national test of a systematic exchange of seismic data, and using it for determination of source parameters.

One major element of the January experiment was to determine the rate at which seismic data reports are received over the Global Telecommunication System of the World Meteorologic Organization (WMO/GTS) and the earliest times at which valid locations can be obtained for the seismic sources of the reported signals. A second major element of the planned experiment was to detect and analyze seismic signals from telemetered data received at the Center from the five stations that comprise the Regional Seismic Test Network (RSTN), and to transmit the measured signal parameters over the WMO/GTS network. The third element of this experiment was to compare two competing programs for automatically determining source events from the reported seismic signals.

Procedures for the January experiment were designed to emulate those anticipated for the Center's participation in the subsequent large-scale test, namely, to emulate the operations of a National and an International Data Center as described by the Group of Scientific Experts (Third Report to the Committee on Disarmament of the ad hoc Group of Scientific Experts to Consider International Co-operative Measures to Detect and Identify Seismic Events).

c. Action was taken to repair and upgrade the Center's capabilities for analyzing waveform data using automatic and semiautomatic (analyst-assisted) methods. The primary objective was to provide computer-assisted capabilities for measuring signal parameters as needed in the upcoming multi-national test

of a systematic exchange of seismic data and determination of source parameters. These enhanced capabilities should also prove useful as tools for conducting research at the Center.

An installed program for detecting signals in continuous digital data was found to be malfunctioning and was fixed. Additional work is needed to extend the capability for selectively distinguishing local, regional and teleseismic signals and for automatically extracting additional signal parameters.

The seismic analyst station at the Center was designed to function using an outmoded version of the operating system and, consequently, was only partially operational. Action was initiated to upgrade the system and make it operational under UNIX version 4.2.

Preliminary work had previously been done to develop a remote seismic terminal for routine measurement of signal parameters using a SUN micro-computer with interactive graphics. Our analysis of the system showed that the terminal was inherently capable of measuring complete signal parameters in a semiautomatic manner, and could also provide data base access and computational capabilities useful for research. Action was initiated to complete the original development for routine measurement of signal parameters and to provide access to the Center's data bases. Additional work will be initiated at a future date to develop research programs on the remote seismic terminal.

d. Additional accomplishments included:

- Rewriting and updating programs to convert data to the standard data base formats;
- assisting in the installation of a special data base composed of close-in recordings of explosions;

- making programs more friendly to non-expert users;
- installing the SNAP/D network evaluation program; and
- assisting various visiting scientists.

APPENDIX 1.

CENTER STATUS WITH RESPECT TO MAJOR OBJECTIVES FY 1983

1. DATA BASE DESIGN

Objective: Complete database design

Status: Completed with adequate flexibility to accommodate current needs including nonstandard database constructs.

Comments: No changes to the database structure (Version 2.6) are anticipated; however, upgrades in the implementation software (RTI INGRES and 4.2 UNIX) will necessitate database upgrades.
The 'arrival' relations are not adequate for providing Level 1 parameters for international data exchange; a special construct will be needed.

Version 2.6 does not restrict the order of records in the assumes particular data sequences.

Needs: Additional user-oriented documentation is needed to describe the database structure, including conceptual design, implementation status, and operational guidelines which identify addition documentation and potential pitfalls.

2. DATA BASE ARCHIVING

Objective: Achieve "automatic"/operational capability for routine data logging/archiving into database:
WMO, UK, YKA
RSTN
SRO (GDSN)

Status: The basic software and operating procedures have been developed for archiving three types of data: events, arrivals and waveforms. Only NEIS events are being routinely archived under Version 2.6. Current procedures are summarized below.

Events - NEIS origins are being routinely archived (monthly) under Version 2.6. These origins are later updated using corresponding ISC locations provided by the USGS. With some effort these updates could be generated directly from ISC data.

Arrivals - NEIS, WMO, UK, YKA, CSN alphanumeric data are being routinely archived on HUGO under Version 1.0. These data are then being reformated and stored in a disorganized Version 2.6 which includes duplicate data. Programs are currently being tested for (1) removing duplicates and (2) renumbering the data sequences.

Waveform - RSTN data (without detection) are not being routinely archived because of CIS/Network problems, which are currently being addressed. SCARS tapes are being generated as backup.

Waveform - USGS Network Day tapes denoted GDSN (SRO, ASRO, DWWSN, RSTN) are being archived (~monthly) into a pre-2.6 Version of the database. Existing software makes this incompatibility largely transparent to the user.

3. EXISTING DATA BASES

Objective: Install research databases:
Yield
Explosions
Discrimination
IDCE 1980 Continuous (1-15 October and 1-7 November)
IDCE 1982 Synthetic
Earthquakes (assumed to be waveforms from GDSN)
NEIS, ISC

Status: The majority of these databases exist and are being maintained at the Center on either HUGO or SEISMO. Few, if any, of the databases have undergone final editing. Future improvements include such things as - updates to Version 2.6, updates to accommodate RTI INGRES, ISC updates, corrections to recorder characteristics, and general removal of data glitches as they are discovered. Fragmented information on the characteristics and status of the most significant databases is as follows:

IDCE 1980 Continuous Data - 'IDCE' on HUGO - Contains the U.S. data for the 1-15 Oct Experiment. Non-U.S. data for the 1-15 Oct Experiment and the 1-7 Nov data are not identifiable databases at the CSS.

IDCE 1982 Synthetic - Unidentifiable at the CSS

Discrimination - Unofficial version on JANUS, which does not have INGRES, therefore not usable within the context of the database system.

YIELD - 'YIELD' on HUGO - The identifiers within this database have inconsistencies.

Explosions - 'EXPLOSION' on HUGO - Contains essentially complete data (unclassified) on both chemical and nuclear explosions over the time period January 1964 through 1982, apparently including information contained in 'YIELD' (Undesirable to maintain both databases). Data prior to 1964 are included for underground explosions but not for underwater or atmospheric explosions. The database is probably reliable and should be updated as more recent information becomes available.

GDSN - 'GDSN76'...'GDSN82' and 'master83' on HUGO - pre-2.6 Version with tailored access software.

NEIS(arrivals); ISC - 'pre82'and 'a3' on HUGO - ISC arrivals are not archived.

NEIS(events); ISC - 'events' on SEISMO

DEMO - 'DEMO' on HUGO

RSTN - Not currently being archived (apart from SCARS tapes and GDSN).

Needs: Additional effort should be devoted to the assessment of database status followed by a plan with priorities and responsibilities for updating, correcting, and documenting databases.

4. INFORMATION EXTRACTION AND DATABASE CONVERSION

Objective: Complete automated algorithms conversion; implement "routine" processing/database installation for: DP, Post-DP, AA, LOC(TG,LL)

Status: The basic software exists for performing the various operations involved in detection (DP), arrival identification (Post-DP), automatic association (AA), and location (LOC); however, the combination of errors in the individual software modules and incompatibilities in the formating/network protocol prevent routine processing of continuous waveform data.

5. DATABASE ACCESS

5.1 Objective: Complete FORTRAN interfaces into applications programs

Status: Not implemented. Currently all of the applications programs work from "external" file structures ("external" to INGRES). Some time in the past there was thought to implement subroutines which would access data directly from the INGRES data bases but someone deemed that the resulting programs would be slow, thus the whole idea was abandoned. Special programs now exist which extract data from INGRES data bases and write special files to be used by applications programs. The drawback of this process is that one must know the subset of data to be used by an applications program before that program is run.

5.2 Objective: Design/implement user scripts for common database queries

Status: Some of these exist. It is difficult to predict a common set of possible queries. Perhaps the emphasis should be placed on the documentation of the current data base design from a seismological rather than a computerese point of view.

5.3 Objective: Design/develop user tutorials in anticipation of research support in FY 1984

Status: Some standard system related tutorials exist (e.g. UNIX, INGRES) but the orientation is toward the computer professional. Even so, it takes a significant amount of time and effort to locate relevant software subroutine libraries for many of the peripheral devices at the Center (e.g. Megatek, Tektronix, Versatek). An effort was made to document Center specific software (Volume II of the S-cubed document), unfortunately, the method emulates the UNIX Manual page style which turns out to be notoriously unfriendly to users that are unfamiliar with UNIX. A significant amount of effort is needed to turn the Center's orientation toward seismological usefulness.

6. INTERNATIONAL DATA EXCHANGE

Objective: Insure Center capability to run operational international datacenter/treaty participant

Status: To date, special procedures have been tailored to accomplish this task with significant analyst intervention.

6.1 Task: Data acquisition

Status: The Communications Interface System (CIS, 11/44) is being used to handle WMO, UK, etc. parameter data. The CIS software is undergoing evaluation in order to identify the problems. The problem may be with the CIS software or the local network software or possibly both. The acquisition of parameter data is being accomplished on a routine basis but the acquisition of waveform data has some serious problems.

6.2 Task: Parameter extraction/MSMT

Status: Parameter extraction is taking place with certain qualifications. A FORTRAN program was obtained from NEIS and modified for the CIS, some of the routines were rewritten in C language in order to speed up the process. Some of the parameters are not being extracted from the incoming parameter data.

6.3 Task: Data exchange

Status: Data is received at the Center, none is sent out.

6.4 Task: Automatic Association (Level 1)

Status: A systematic evaluation of the AA is in order, because the various versions in use at the Center give conflicting results.

6.5 Task: Location (Level 1)

Status: There is no operational location program at the Center apart from that contained in AA.

6.6 Task: Bulletin preparation/distribution

Status: Software exists for generating bulletins, however, standards have not yet been established. Bulletins are not produced on a routine basis. The distribution mechanism exists.

6.7 Task: Incorporation of waveforms in "routine" processing

Status: Because of the problems associated with the CIS-to-HUGO transfer of waveform data, it is not being used on a routine basis.

6.8 Task: Documented procedures

Status: Documentation is lacking for much of the processing software. A task oriented procedure has been implemented for reducing this deficiency.

7. RESEARCH SUPPORT

Objective: Develop plan for supporting researchers

Status: Considerable expertise or perseverance is currently needed by researchers to utilize the Center facilities.

7.1 Objective: Provide data/applications support
Status: This is done on an ad hoc basis.

7.2 Objective: Integrate third VAX; other equipment
Status: Completed.

7.3 Objective: Initiate classified processing
Status: This is routinely performed.

7.4 Objective: Support remote researchers

Status: Generally remote users fall into two categories:
(1) logging in from remote sites on a continual basis,
(2) coming to the center to perform specific tasks for a specific duration. Users in the first category are fairly knowledgeable about the Center's operating system and its resources. Their requests relate to specific subsets of data from the Center's data bases. The second category of users need more specialized attention because they are new to UNIX or they have tapes which they need to put on the CSS computers or they have special communication requirements. All these functions are satisfied by the Center's staff each according to his/her expertise.

APPENDIX 2.

EVALUATION OF INGRES AT THE CENTER FOR SEISMIC STUDIES

INGRES is a relational data base management tool being used at the CSS for seismic data bases. There are at least two versions: one that comes with the standard UNIX distribution (so-called Berkeley INGRES), and one that has been used at CSS as a beta test-site for Relational Technology, Inc. (RTI). During the latter part of Oct. and early Nov., 1983, CSS personnel exercised INGRES in several different configurations on different data bases and operating systems. The following configurations were used:

1. Berkeley 4.1A UNIX with RTI 2.0 INGRES (RI),
2. Berkeley 4.1C UNIX with Berkeley 7.10 INGRES (B1),
3. Berkeley 4.2 UNIX with Berkeley 7.10 INGRES (B2).

The reason for the last benchmark was that the 4.2 UNIX 'promised' that the input/output (I/O) execution speed would be faster under 4.2 than under earlier versions. We found quite the opposite to be true in most circumstances.

Three different 'typical' tasks were posed to INGRES:

```
T1. modify wftape to cisam on date
T2. retrieve (origin.orid,arrival.arid)
        where origin.orid=assoc.orid and
            assoc.arid=arrival.arid
T3. replace wftape (remark="this is non blank")
```

The first task is similar to one that would be performed by a data base administrator at the CSS, it is ordering a data base by a certain attribute; in this case, date. The second task is a query of the data base searching for certain attributes from three relations, namely: origin, assoc and arrival. This type of query may be invoked by a user of the data base. The third task directs INGRES to replace the value of the 'remark' attribute by the specified string. This type of task may be given to INGRES by a user or data base administrator.

The first data base attepmted to be used was the GDSN76 which consists of nearly 44,000 records. Unfortunately, for this data base, T2 and T3 under the B1 configuration had to be terminated after about 12 hours of wall-clock time. An attempt was made to run all of these benchmarks at night so as to create a favorable environment for the completion of the tasks. A smaller data base (EXPLOSION) was selected for the rest of the benchmarks which contains about 9,300 records. Actually, T1 ran to completion under B1 but the results are no different than the following, so they will not be presented.

The statistics presented by the UNIX 'time' command were recorded. They include the CPU execution time along with the wall-clock time and the percentage of CPU allocated to the task. The manner in which the computer presents the two types of times makes one or the other uninteresting because the wall-clock time weighted by the percentage of CPU allocated is nearly the same as the CPU execution time. The minor variations seem to be due to the fact that over the period of the task's execution the percentage of CPU can be different. The times reported here will be the CPU execution time and the user can project potential wall-clock times by considering that during peak usage periods a task can get as little as 10% of the CPU versus late night rates of about 80%. Other statistics examined concerned the amount of memory used and the number of I/O operations expended during the execution of the tasks. There seems to be no straightforward conclusion of the sort: "This task is taking longer because it is doing more I/O and it is just waiting for the disk to spin." For future sizing

considerations, RI can consume as much as 500MBytes of memory for query operations. The following table summarizes the CPU execution times for the various tasks under the examined configurations.

	Configurations				
	RI	B1	B2	Ratio(B1/RI)	Ratio(B2/RI)
T1	168.1	370.9	355.0	2.2	2.1
	2:48.1	6:10.9	5:55.0		
T2	132.6	7828.0	9004.5	59	68
	2:12.6	2:10:28	2:30:04.5		
T3	1475.2	1030.7	1149.0	0.7	0.8
	24:35.2	17:10.7	19:09.0		

The CPU execution times are given in seconds as well as in hours:minutes:seconds form. The common denominator for the ratio is the CPU time consumed by the RI configuration, thus a ratio of 1 or larger means that RI runs faster by that ratio. It is evident that RI is faster for data base administrator type of function as in T1 and can be somewhat slower as for T3. The most startling fact is that for query type operations (T2) RI is faster by a factor of 60 or better. That means that queries taking a matter of minutes for RI wind up consuming hours under B1 and B2. The data base used here is rather small by most standards, thus making the Berkely INGRES rather prohibitive as a meaningful research tool. Since the paramount purpose of a data base management software is to make use of queries in a linguistically 'natural' manner, waiting for the mean time between computer hardware failures to manifest themselves seems to be more of a problem rather than a solution.