4. TECHNICAL PLANS

4.1. GSN Technical Implementation

4.1.1. Introduction

The scientific objectives of the proposed new global seismographic network and the technical plan are described in detail in Appendix 1A: The Science Plan for a new Global Seismographic Network. The principal features of the proposed network are:

- Standardized modular equipment
- Approximately 100 globally distributed stations
- Broadband frequency response (from D.C. to 10 Hz)
- High dynamic range: 140 db at 0.04 Hz
- Real-time digital data telemetry whenever possible.

What is described in the following sections is a specific plan to achieve the objectives of that proposal. The ten-year project schedule illustrated in the "Milestones" chart in Figure 4.1 consists of three phases:

1. Design Specification
2. Production and Deployment
3. Network Operation

Phase One culminates in a contract for the production of seismographic station systems. At this time, the seismic system will have been specified and some of the modules will have been tested. Real-time data transmission from prototype stations to the USGS data collection center in Albuquerque will have been demonstrated. Questions of data structure will have been resolved by interaction with the IRIS Data Management Committee.

Phase Two begins in year two with site preparation and continues through year eight with the deployment of the 100 station network. The last two years of the ten-year program, Phase Three, consists of routine network operations.

The principal activities are divided into four main categories:

1. Design, Test, and Evaluation
2. Network Deployment
3. Communications System Deployment
4. Data Collection

The Budget (Figure 6.3) projects the entire program through a ten-year period.

While this proposal focuses on the establishment of a state-of-the-art network, it seems both desirable and possible to improve significantly the quality of currently available data by an early deployment at the existing stations of tested modules, such as seismographs or data loggers, which satisfy the design goals. The additional expense incurred would be minimal. At the same time, the early availability of better data is critically important to the seismological community and, in addition, would provide the GSN Project with the much needed extensive field test.

4.1.2. Research and Development

An ongoing research and development program is a key element of this plan. The objective is simply to avoid planned obsolescence in instrumentation, data handling, and analysis techniques. In this regard, the IRIS program differs significantly from previous seismographic network programs. The goal is to complete the proposed ten-year program with as modern a system then in place as the science requires. Only through constant development in such areas as seismic instrumentation, data telemetry, data base management, and analysis techniques can the program keep abreast of technological and scientific advances. In addition, an equipment

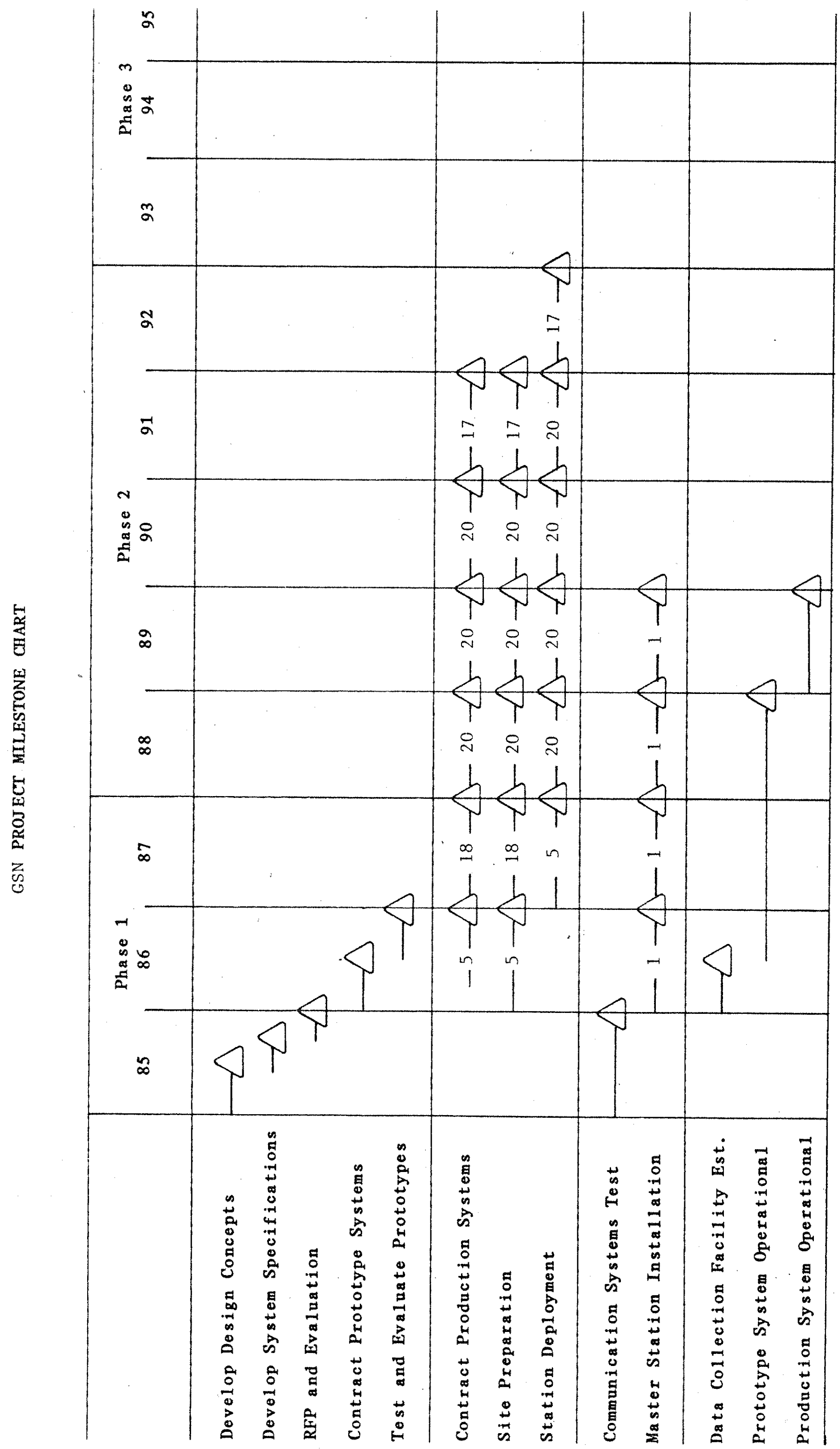

Figure 4.1. GNS Project Milestones

reserve account is planned to provide funds to keep the network up to date.

In the early years of the program, approximately 15% of expenditures are budgeted for research and development, leveling off to approximately 10% for the remaining years. These funds are distinguished from programmatic research funds in that the former are meant to support a wide range of unsolicited and innovative basic research in key areas whereas the latter address specific and directed needs of the GSN project.

4.1.3. Design, Test, and Evaluation

The overall goal of Phase One of the project is to design and produce a system that meets the specifications needed to achieve the scientific objectives. At the same time, the system must meet the practical goals of standardization and modularity that are necessary for cost-effective network operations. In this section, the equipment and physical plants of the remote seismic stations will be considered exclusive of the communications system and the data collection system.

4.1.3.1. Develop Design Concepts

This is the initial stage of the process to translate the scientific needs outlined in Appendix 1A into equipment for the new global network. The result will be a short document that raises the pertinent questions and issues to be addressed in the design study below; a draft of such a document is included as Appendix 1C, "Design Considerations for a New Global Seismograph Network". It is presently planned for this concept to be developed by various committees and distributed widely for comment and review.

4.1.3.2. Conduct Design Study and Develop System Specifications

Guided by the results of the study in the item above, a design study would research the technical issues, reviewing the availability in the commercial market of system modules.

Design study sub-tasks include:

- Begin with Design Concepts Document
- Review candidate designs for system modules
- Research off-the-shelf versus custom system hardware modules
- Address maintainability, reliability, exportability, and operability of candidate systems
- Review technical requirements and cost trade-off
- Discuss interfaces of seismic system with other elements of the proposed network
- Conclude with recommended system specifications

This design study will most likely be contracted to an outside firm.

4.1.3.3. Prepare RFP

After review of the design study recommendations, an RFP could be issued for proposals to build the system. Depending upon the system complexity and the amount of custom modules, the RFP could call for the prototype system to be tested and evaluated prior to production of multiple units. If the system specifications look fairly straightforward, however, the procurement stage could begin immediately. This is a fairly critical decision and can only be made after careful review of the design study and a thorough analysis of the risks involved in immediate system production. For the purposes of this budget discussion, it is assumed that there will be a separate prototype procurement phase.

4.1.3.4. Proposal Evaluation

A panel of scientific and technical experts will be convened to review the results of the proposal solicitation. Most likely, the responses will deviate in some details from the design specifications and/or present new ideas that merit consideration. Typically, this situation leads to a revision of the RFP technical section and a second round of solicitations. With a properly conducted RFP and review process, the next stage of procurement should be straightforward.

4.1.3.5. Prototype Test and Evaluation

Some of the modules of the seismic system are fairly conventional, multi-use items such as the tape recorder or microprocessor. Production units for these items can be readily specified without extensive experimentation. Other items, however, are rather special-purpose or state-of-the-art. In this category are:

1. The seismometer package
2. The analog-to-digital encorder
3. The satellite communications system

Programs for test and evaluation of candidate modules are scheduled to begin in Year One of the project, and funds are included in the budget for this purpose. In fact, work is already in progress on several items:

- seismometer systems by the USGS Albuquerque Seismological Laboratory
- analog-to-digital encoders by Sandia National Laboratories
- satellite communications by Stanford University, UCSD, USGS, and Sandia National Laboratories

These activities can be conducted in parallel with the Design Study with the goal of specifying the entire system by early in the Second Year. The seismometers, after being tested, can be used to upgrade selected existing stations.

The contract award will result in the delivery of one or more prototype (or pre-production) systems to IRIS. These units will undergo extensive tests and evaluation under a range of operating conditions. Most likely, these tests will reveal the necessity of some modifications of the design specifications to which the units were built before full production begins.

4.1.3.6. Contract Production Systems

At this stage, the system specifications are finalized and the production of the fully-tested units begins. The funding and deployment schedules must be coordinated with the production facilities to insure adequate supply of systems to meet schedules.

4.1.3.7. Prepare Full Documentation

While the production systems are being manufactured, full documentation of all aspects of the system must be prepared. This is a critical item that must precede deployment and training of field-installation, operational, and maintenance personnel.

4.1.4. Network Deployment

Deployment of production systems is not scheduled to begin until Year Three of the project. Prior to this, however, a number of prototype modules will have been deployed as part of the test and evaluation program.

Network Deployment consists of a four stage process:

- Negotiation Stage: prospective sites are identified by general geographic location; local scientists and agencies are contacted; final details negotiated. This stage should start at the earliest possible time. An initial list of 20 to 30 stations will be prepared at the beginning of this project. It is anticipated that many of these will be stations of existing seismic networks where operating agreements are in effect.

- Site Survey: once the location has been determined for a new site, detailed surveys are needed to determine an exact location. Relevant factors for site selection include seismic noise conditions, local power supply, accessibility, local assistance, etc. This stage should be conducted a least a year in advance of deployment.
- Site Preparation: when the location is determined, site preparation can begin. This stage will include some construction of the stations, physical plant, and training of local personnel. This stage must be completed before station deployment.
- Station Deployment: when the site preparation is completed and personnel trained, a team will install the seismic station equipment, calibrate it, and insure its correct operation.

Stage one will commence at the start date of this project. There are already a number of possible candidate sites that could easily and quickly accept deployment of GSN systems. These include some of the existing sites of the GDSN, IDA, and other digital networks.

In Year two, site surveys will begin for production system deployment Three, 5 stations of the GSN can be deployed. The schedule for the following years calls for the deployment of 18 stations in Year Four, 20 stations in Years Five through Eight with the completion of this Phase by the end of Year Eight.

4.1.5. Communications System

The most likely candidate system for the real-time telemetry of the GSN data on a global scale uses the recently introduced commercial application of spread-spectrum technology to satellite communications. Currently funded by the National Science Foundation is a pilot study being conducted by Stanford University and the University of California San Diego for IRIS. This study will include a field experiment in the US, using a domestic satellite to transmit compressed data from a single remote station to a recording site. To validate the technique for the entire network communications, however, other critical tests need to be conducted. See Section 4 of Appendix 1A for details.

4.1.6. Data Collection System

In Year one, as part of the design study contract, specifications will be developed for the Data Collection System. The specific functions of this system depend upon how much of the data arrive in near real-time and how much arrive much later. In the limit of all data being broadcast in real-time, this system need only serve the quality control and network control functions. However, in the more likely scenario, where a significant fraction of data arrive much later, the system will also need to serve the function of data merging. In this case, the real-time data is stored at the Data Collection Facility until the arrival of all other data so that a merged complete data set can be produced.

During Year two, hardware for the implementation of this system will begin to be purchased. It is presently envisioned that this system will be built around off-the-shelf computer modules and that only software development will be needed. There may be some special purpose modules associated with the real-time aspects of the network telemetry, but even these should not require significant hardware development.

In subsequent years of the project, this system will grow as the network is deployed and as the data gradually reach its full flow.

4.1.7. GSN Management Structure

The management chart of the GSN program is shown in Figure 4.2. Solid lines imply organizational dependence and subordination, broken lines indicate flow of information and cooperation.

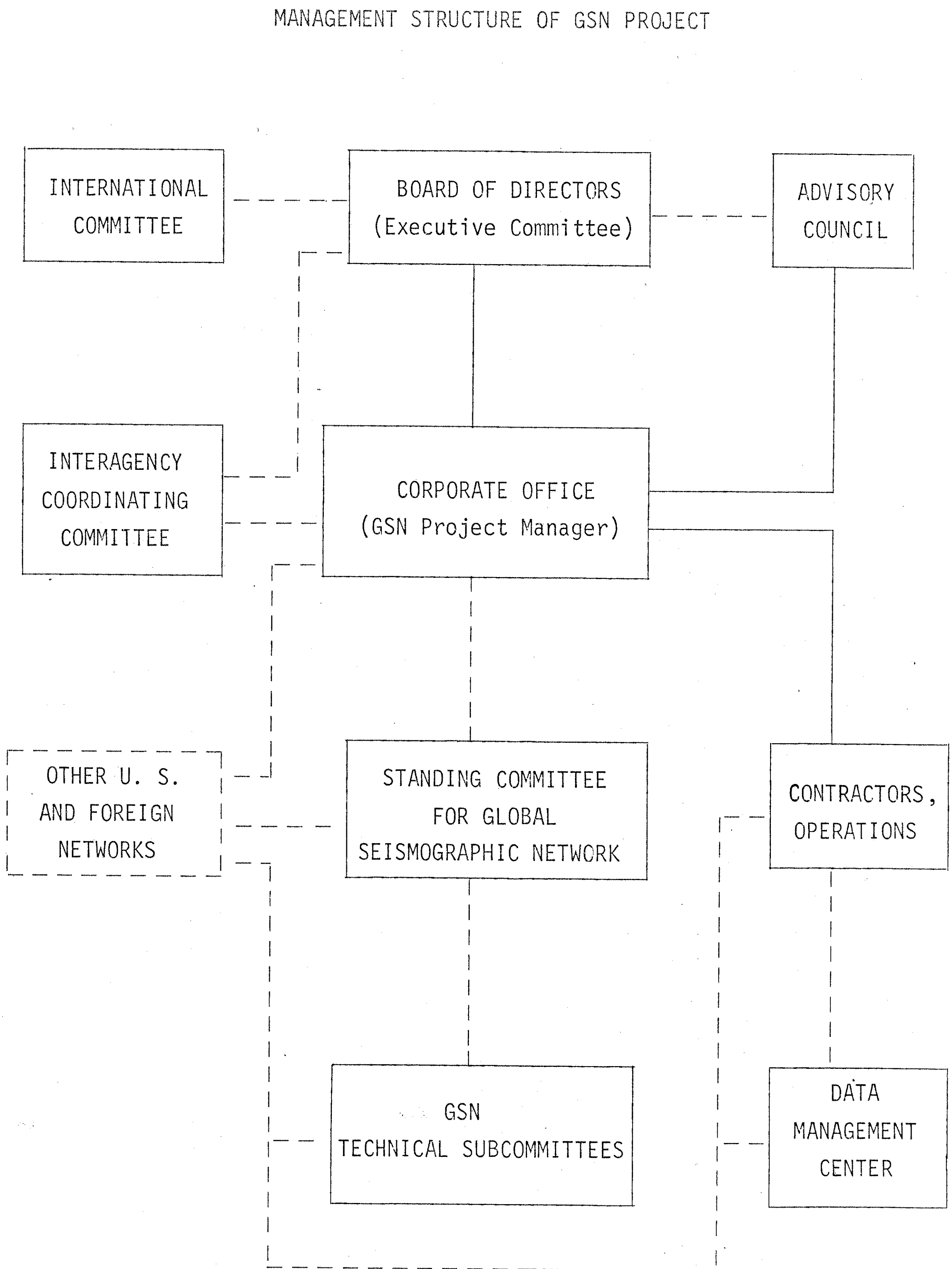

Figure 4.2. GSN Project Management Structure

IRIS Board of Directors and, when it is not meeting, the *Executive Committee* have the responsibility for achievement of the goals for which IRIS had been established. They have authority to establish programs and approve proposed projects and their budgets. They direct the President, who as the chief executive officer heads the *Corporate Office*, to execute the programs within the established budget.

IRIS Corporate Office provides the contracting, legal, accounting and administrative support of the programs proposed by the *SCGSN* and approved by the *Board of Directors*. *Corporate Offices* provide also the logistical support for the activities of the SCGSN and its subcommittees, such as organization of meetings, travel arrangements, secretarial and publishing services. Administration of the GSN Program will be conducted within the *Corporate Office* by the *GSN Program Manager*. His, or her role is to interface the activities of *SCGSN* and its subcommittees with the *Corporate Offices*, to monitor performance of the *Contractors*, to cooperate with *Other U.S. and Foreign Networks* that contribute data. Program Manager may be assigned small professional and administrative staff for assistance in these multiple and diverse duties. For example, during the Year 2 and 3 of the program such staff might consist of:

- Engineers (2). Engineering coordination and oversight of IRIS contractors. Engineering support for test and evaluation of contractor-delivered systems;
- Technicians (2). Technical service;
- Programmer; advice and coordination on software and hardware issues.

Contractors. It is intended that most of the major items in the budget will be contracted to outside vendors. It is not planned for IRIS to employ directly a large staff of scientists, engineers and technicians. Various aspects of the operation, including the design studies, construction of instruments, deployment, field operation and maintenance, operation of the Data Collection Center will be sub-contracted. Such contracts, following appropriate procedures, may be issued to universities, both U.S. and foreign, private corporations or government agencies, although in the latter case some other form of agreement may be appropriate. It is expected that regardless of the nature of a contracting agreement, the funds awarded to IRIS by the NSF for execution of this program would be administered by IRIS.

The final product of the GSN program is a set of data, such as a network-day-tape, transferred to the *Data Management Center* (see Sections 3.3 and 4.3) for archival and distribution to the users.

Other U.S. and Foreign Networks. There are roughly 60 digitally recording stations operated by the IDA, GDSN, RSTN and GEOSCOPE projects. Other countries or groups of countries have announced plans to establish a number of high quality broadband stations. These networks represent an invaluable resource for IRIS. Representatives of the USGS serve now on *SCGSN* and all its subcommittees. An invitation to join IRIS in its planning activities has been issued to the (West) European Working Group for Global Seismology. and the Seismology Division within the Canadian Government. Improvements and integration of these networks is the principal objective of one of the subcommittees of *SCGSN*.

Depending on the specific agreements and timeliness of delivery, early data from contributing networks or individual stations could be included in the "network-day-tape" (Operations), while data arriving late would only be archived by the Data Management Center.

Figure 4.2 contains three additional boxes; these are not directly connected with the GSN program, but their activities have an impact on its operation. The *IRIS International Committee* will include representatives of international scientific organizations (such as IASPEI), executive level officers of foreign equivalents of the NSF program in addition to the representatives of IRIS and U.S. agencies. Its role will be to advise the Board of Directors in the International aspects of IRIS operating and the GSN project will undoubtedly be an important part of the activities of this Committee. The *IRIS Advisory Council* is expected to provide an independent assessment of all IRIS activities, but if it were to decide to establish subcommittees overseeing the individual programs, the link between such a subcommittee and *SCGSN* could be much more direct. The *Interagency Coordinating Committee* will advise the *Board of Directors* and cooperate with the *Corporate Office* on the issues that involve activities of branches of the U.S. Government other than NSF.

Standing Committee on Global Seismographic Network (SCGSN) is appointed by the *Board of Directors* to formulate the proposal for the design, construction, deployment and operation of a permanent network of globally distributed seismographic stations. During execution of the program, the Committee will monitor the progress with the aid of *Program Manager* and, if necessary, it will recommend adjustments and revisions in the implementation plan.

Technical Subcommittees are created by the *SCGSN* to advise it on particular issues. At this time there are four such subcommittees:

1. Subcommittee on Instrumentation is charged with development of the design goals, recommendation of technical specifications, evaluation of tests of the system modules as well as the entire modules, recommendation of specific equipment for network deployment.
2. Subcommittee on Data Collection Center is charged with development of the design goals, presentation of recommendations on the hardware and software configuration of the Center, which is responsible for merging and quality control of the data from remote stations.
3. Subcommittee on Satellite Telemetry is charged with presentation of recommendations on the communication system to be used for real-time telemetry. At this time, the Subcommittee is also monitoring the progress of communication tests carried out with the support of NSF by Stanford and UCSD.
4. Subcommittee on Preservation, Integration and Improvement of Digital Networks is responsible for recommendations with respect to transition from the current digitally recording stations to the planned level of GSN technology. Its charge is to assure continuity of the digital data base, development of recommendations how to improve, as soon as possible, the quality and quantity of the data and how to merge the data produced by the existing networks into a common data set available to the users. As the new network develops, the role of this Subcommittee could become that of monitoring the network operations.

4.2. PASSCAL Technical Plan

4.2.1. Introduction

The technical implementation of PASSCAL is centered around its plan to develop, procure, and operate a facility based on 1000 portable digital seismic recording systems. The backbone of the facility is the instrumentation. For the past 8 months, following the special meetings on instrumentation in Salt Lake City (April 1983)† and at Los Altos (November, 1983) and the PASSCAL organizational meeting in Madison (January, 1984), the Committee on Instrumentation and its panels have been meeting to develop a plan which will lead to the new instrumentation being defined, designed, and produced, for acquisition by PASSCAL. The Committee has found that existing designs, even those which use the most current technology, have been conceived to meet specific needs of a particular type of experiment. The needed instruments do not exist, and available instruments are few, unmatched, and obsolete. The demand that the PASSCAL instrumentation actually have the flexibility to serve for the wide variety of studies which exploit the frequency band .02 – 50 Hz has led the Committee to propose the development of a class of compatible bus-oriented data acquisition devices which go beyond those now available, but which make use of technology which is either off the shelf or imminent. A detailed exposition of the conclusions of the Instrumentation Committee to date, and its plan for the development, testing, and deployment of the new instruments are given in section **4.2.2.**

The mechanics of maintaining a large network in working condition, and of insuring that deployment and retrieval of instruments is properly handled require that PASSCAL maintain facilities, vehicles, and support staff. The complex series of steps by which data are retrieved from the temporary memory or recording medium of the field instruments and converted into a fully documented, edited collection of event data sets in a data archive requires coordination not only with the support operations, but with the PI's responsible for the data. In section **4.2.3** we present an analysis of the closely related questions of field support and data retrieval. A plan is presented for the establishment of support facilities, including a central support facility, support vehicles, field computers, and preprocessing services at at the Data Management Center of IRIS. We also present a management plan for large multi-institution cooperative field experiments, in which PASSCAL provides the support for participation by PI's in the data acquisition.

The IRIS Data Management Center (DMC) is the subject of a separate Proposal and Plan (sections 3.3 and 4.3 of this proposal, and Appendix 3A). Its purpose is to provide a digital data archive for IRIS and related geophysical data to which all investigators can have access. The large quantities of digital data which may be expected from the PASSCAL array, operating at high data rate modes with all 1000 instruments, define the very high performance which must be planned into this program from the start. The particular needs of investigators for access to data sets of substantial size and for browsing or searching of the data bases equally serve to define the performance which is required. In section **4.2.4** we specify a model for these demands on the DMC. This includes a model for the quantities and types of data to be expected from the PASSCAL system, and a model for the demands which seismologists will make on the DMC in the course of their analysis of PASSCAL experiment data.

In sections **4.2.5—6** we present digest of the calendar for the implementation of the technical plan, and a brief discussion of both the ten year budget projection (see Figure 6.4) and the budget request for FY85— 87. In these budgets, the annual PASSCAL cost rises from $4.5 million in FY85 to $13 million in FY89 and remains essentially constant after that. The FY85 budget level is based on the assumption that PASSCAL begins critical operations promptly and aims to complete them in a timely fashion: namely, instrument development and prototyping, and the first interim experiment. While the budgets rise for the first five years as instrument acquisition proceeds, there is little drop to an "operations only" mode, for the annual costs of maintaining the full instrument complement just about equals the annual cost of procuring 200 new instruments.

4.2.1.1. Management

PASSCAL will operate as a distinct scientific program under IRIS. Its management structure is described in Chapter 10 of the PASSCAL Program Plan (Appendix 2A), with Figure 10-1). Two major lines of responsibility flow downward from IRIS:

(1) From the IRIS Board of Directors to the Standing Committee for PASSCAL and to the PASSCAL Subcommittees. This structure will have responsiblity for setting policy, for planning PASSCAL activities, and for preparing and approving plans and budgets. The Standing Committee will assume overall responsiblity for the Program.

(2) From the President and Executive Director of IRIS downward to the Chief Scientist of PASSCAL; thence to the operational groups. People in this line are either IRIS employees or contractors. The Chief Scientist will be responsible to the Standing Committee for satisfactorily executing policy, and for insuring satisfactory coordination between the operational groups and the relevant Subcommittees. The PASSCAL Chief Scientist's office will include staff to handle contracting, purchasing, and other purely administrative tasks, which will report directly to the Executive Director of IRIS.

The *review of proposals* and *selection of projects* will be necessary as a final stage in the program planning activities of the Science Planning and Coordination Committee. It is planned that an independent reviewing mechanism be established with the cooperation of the Earth Science Division Director at NSF, to prevent problems with conflict of interest.

4.2.2. Instrumentation

An expanded discussion of the instrumentation plan appears in the Program Plan (Appendix 2A) as Chapter 7, and budgetary and implementation matters are more fully developed in Chapters 13 and 14.

At the heart of the program to meet the scientific objectives of PASSCAL is a new generation of highly versatile matched portable broad-band digital seismographs. Advances in recent microprocessor and microelectronics technology now make feasible a modular digital instrument of unprecedented versatility, and seismologists worldwide are moving rapidly to capitalize on these technological advances. With existing technology it is now possible to construct a rugged, fieldworthy portable seismograph with the following capabilities:

- high dynamic range digital recording (120 dB)
- highly accurate phase-lock time system (10 μsec, relative; 100 μsec, absolute
- variable recording bandwidths, user specifiable, within a very broad overall frequency band (0.01– 200 Hz)
- multi-channel recording, with channel capacity expandable in modules.
- triggered event recording based on local, regional, or teleseismic wave onsets, or turn-commands based on preset or radio codes
- high speed memory to permit buffering and preprocessing of data
- continuous recording of decimated data
- automatic self-calibration of the instrument and tape storage of complete system parameters including sensor type
- self-diagnosis, including identification of problems at the module level
- high density recording media, with 160 Mbyte cartridge magnetic tape available now, 300 Mbyte magnetic tape cartridges forecast for the near future
- microelectronically controlled broad-band portable seismometer
- duplex data transfer between user and seismograph

The functional needs for field seismology and the general design principles have been discussed at some length in the 1984 National Academy Report, in the Proceedings of the 1983 Utah workshop on Guidelines for Instrumentation, and in the Proceedings of the Workshop on Instrumentation held in Los Altos, California, November 29– December 2, 1983. The systems we propose to bring into being are in accord with the strong recommendations of these two studies. Table 4.1 gives the requirements for a versatile, portable system which can record both controlled and natural source signals.

We propose that such a seismograph system be designed, built, and procured as the core of the instrumentation for large, portable lithospheric arrays. To meet these requirements for versatility the seismograph system is being designed around a communication bus structure (Figure 4.3). Various functions (analog-to-digital conversion, filtering, triggering, recording, arithmetic operations, etc.) are performed by modules which have their own microprocessor "intelligence". The modular structure allows the the system configuration to be varied by adding or deleting modules without reprogramming the central microprocessor. Moreover, modular design allows the instrument configuration to be tailored to specific experiments; it is probable, for example, that seismograph units will be available in more than one package. In none of the proposed instrument configurations, however, is the seismograph unit expected to exceed about 30 lbs (exclusive of seismometers) and about 0.5 cu. ft. in volume.

A major consideration in the decision to employ a bus architecture is that it permits the system to "evolve" as new modules incorporating new features or improvements replace old modules without significant modification of the instrument or its software. By defining a comprehensive bus communication protocol, we can assure that general and special purpose modules, compatible with the instruments, can be developed industry-wide and marketed in the competitive marketplace. It is noteworthy that many in the seismograph industry regard the

Requirement	Artificial Source: Vibroseis	Artificial Source: Explosion	Earthquakes: Short-Period Local and Teleseisms	Earthquakes: Broad-Band Body and Surface Waves
Time accuracy of any sample	250 us	1 ms	1 ms	>1 ms
Minimum number of channels (maximum 12)	1	3	3	3
Bandwidth, Hz (exclusive of transducer)	5-200	2-200	500	0.01-20 Hz maximum; 0.1-20 Hz portable
Typical sample rate	4 ms/channel	10 ms/channel	10 ms/channel	40 ms/channel
Event length	20-60 s	1-5 min	10 s - 5 min	30 min to several hours
Dynamic range (exclusive of transducer)	120-140 dB	120-140 dB	120-140 dB	120-140 dB maximum
Minimum resolution (exclusive of transducer)	12 bits	12 bits	12 bits	12 bits
Service interval	10 days (maximum)	10 days (maximum)	10 days (minimum)	10 days (minimum)
Typical number of records per service interval	1,000 sweeps or stacks	100	1,000/day worst case; 20/day typical	10/day or fewer
Special requirements	Radio command to initiate stacking	Programmed turn-on	"Smart" trigger	"Smart" trigger Digital filtering Special seismometers
Master communication	Radio turn-on and sync for T_0	Radio turn-on	--	--
Position	1 m	1 m	10 m	--
Minimum total capacity	10 mbytes	--	25 mbytes	--

Table 4.1

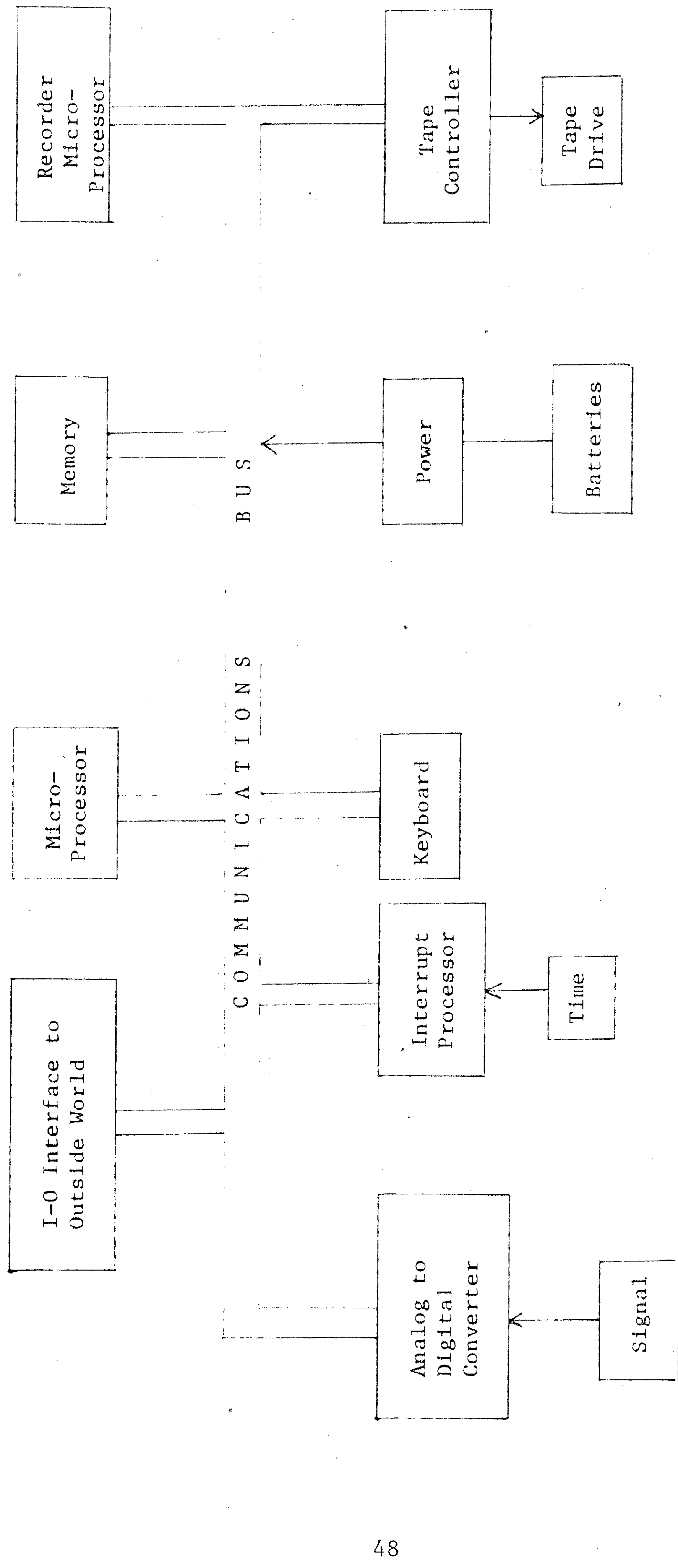

Figure 4.3. Architecture of a datalogger designed around a communications bus.

concept of modular bus design as it is to be incorporated in the new seismograph system to be of such general versatility that it may serve as an industry standard for a wide class of data-loggers and control instrumentation.

It would be extremely attractive if satellite telemetry could be incorporated into the system design, to permit direct downloading of array data to a data collection point. Extensive discussions with industry experts have brought out that no satellite telemetry technology now available meets industry needs for high data rate, acceptable costs, and low weight and power. All telemetry systems now in use by industry use line-of-sight radio, electrical cable, or optical cable. These do not fill PASSCAL requirements, since the spacing of instruments and the overall size of the arrays deployed will frequently be too great to make use of these methods. Single side band is not an acceptable technology due to the large antenna required and the unreliability of the propagation paths to the ionosphere. For data storage and retrieval, therefore, we make use of the rapidly developing magnetic cartridge technology. Systems now being sent to China by the U.S.G.S. use a 67 Mbyte 3M cartridge. Another vendor is currently developing a 160 Mbyte cartridge. Amoco, which developed the Seismic Group Recorders (SGR), uses digital tape cartridges which are collected from the field daily and played back in special vehicles. Most PASSCAL experiments would not require retrieval of cartridges more often than once a week.

The present design decision, therefore, is to adopt the magnetic cartridge technology for at least the first five year cycle in the development of this kind of instrumentation. While frequent visits to the instruments to retrieve data are not flashy high-technology, the SGR's provide an example where this approach works well. This decision will become firm at the time the first round of prototypes is bid in mid-1985. The modular nature of the instruments would permit adding telemetry systems to them at a later time when the capabilities and costs become appropriate.

Implementation plan for instrumentation. The complex process of coordinating the definition, specification, design, prototyping, and procuring of a new generation of instruments is being managed by an Instrumentation Design Team. Members of this team, in addition to these three academics, consist of engineers from universities and private industry†. The involvement of private industry representatives in these discussions from the start has been shown to be essential; without this, no assurance exists that the instrument specifications are of any interest at all to potential bidders.

The instrument definition process is proceeding well, and should be completed by early 1985. Funding of this process has been provided through an NSF grant to the Carnegie Institution of Washington ("Phase zero") for PASSCAL planning. Additional funding for FY85 is needed to complete the bus design, and is discussed below, as well as in Chapter 14 of the Program Plan (Appendix 2A). In meetings of subgroups held over the past several months, comprehensive reports have been produced on each aspect of the system. General (but not unanimous) agreement was reached that the task of the design team is to set both engineering specifications of the modules and complete hardware and software specifications for a communications bus.

A meeting with oil industry representatives was held 19–21 September 1984 in Dallas, organized by Ken Larner of Western Geophysical (who is also liaison representative of the Society of Exploration Geophysicists). A full presentation and review will be made of major instrumentation systems in use by the oil industry. The meeting will include discussions with manufacturers of exploration instruments to determine which parts of their technology are appropriate for PASSCAL. Discussions will also be held to determine the level of interest among manufacturers of exploration instrumentation in participating in the PASSCAL effort.

As of September 1, 1984, the Instrumentation Committee, with the unanimous consent of the Standing Committee, plans to go ahead with specifications for:

† A list of members may be found in the full PASSCAL Plan, Appendix 2A.

- engineering characteristics of the modules
- hardware and software definition for a communications bus

Following the Dallas meeting with industry representatives, the Standing Committee will appoint an outside panel of three non-interested experts to critique the instrumentation plans and provide feedback to us before the December AGU meeting.

Present plans call for specifications, including the hardware and software specifications of the communications bus, to be released April, 1985. At that time, RFPs will be issued to business for bidding to construct prototypes. Companies will be asked to manufacture a prototype of the basic instrument in two packages, a superlight package for backpacking applications and a "standard" package for operation in rugged but not unusually hostile environments.

A special evaluation panel will select three companies to construct prototypes or parts of prototypes for at least 30 instruments which can be tested in the field. We anticipate prototype delivery in late fall or early winter of 1985, and estimate the price will be about $30K/unit. Depending upon the results of testing and the extent of required modifications during an anticipated 6 month evaluation period, the instruments could be in production by late spring, 1986, in time for limited use (with the prototypes already in existence) in some of the interim experiments being planned.

To meet this ambitious schedule, we are requesting additional funds starting Feb. 1, 1985, for design of the communications bus. The bus design is fundamental to all module designs, and in the view of the instrumentation committee must be public to ensure that modules can be produced by different manufacturers. Bus specification is, however, by far the largest of tasks remaining to be completed by the design team. The funds requested ($150K) are in addition to ones already provided for in the Phase zero proposal*. Based on work now complete† Phase zero underestimated by at least $150K the resources necessary to complete the full design work on the bus. We are also requesting funds to design and prototype an Omega receiver with phase-lock time accuracy of 10 μsec.

The instrument design team considers 30 prototypes a necessary number of instruments to allow for testing under a variety of realistic field conditions while sufficient to allow useful scientific data to be obtained in actual small experiments. Members of the design team will oversee the laboratory and field testing of the instruments. We anticipate that laboratory testing will be done primarily by university and private engineers. Field testing will be carried out as part of the PASSCAL program of interim lithospheric experiments, in conjunction with planned deployments of presently available instruments (section 3.2 of this proposal).

The design team will have a number of continuing tasks that extend beyond the period of testing and modification of the prototypes. Technological advances, particularly in microprocessor technology, are occurring so rapidly that there will be opportunities for substantial upgrades in the performance and versatility of the instruments. Even now, it is possible to foresee that engineering investigations will continue into such areas as:

- evaluation of new or redesigned modules
- improvements in dynamic range
- lower power modules
- satellite telemetry
- alternatives to magnetic tape storage
- internal stacking of data

* Phase zero is a pre-IRIS grant to Carnegie Institution of Washington to fund design work on instrumentation and data management. See chapter 14 for a summary of this project.

† Technical Reports on Bus System, A to D, Recording, Arithmetic Unit, Filters and Triggers; PASSCAL subcommittee on Instrumentation, July, 1984. (S. Sacks, Dept. of Terrestrial Magnetism, Washington, D.C.)

Among the PASSCAL program needs crucial to the final development of the multicapability seismograph is the ability within its basic module to uniformly trigger on the set of known seismic-event types, including the onsets of P waves and S waves at local, regional and teleseismic ranges. While a number of algorithms are published/and/or available a relative evaluation is needed and consideration must given to their implementation within a physically small, low power portable digital seismograph.

Sensor Development:

The versatile dataloggers now under development and described in the previous section are being designed to accommodate the widest possible range of portable and semi-portable sensors currently available or under consideration for development. The datalogger will be capable of handling frequencies from 0.01 to 200 Hz, over 120 dB of dynamic range. In actual use, however, the datalogger is inherently limited by the capabilities of the sensors. The limitations take at least three forms:

- lack of portability
- spurious response characteristics
- narrow frequency and/or dynamic range

Portable 1 Hz seismometer: An important instrumentation need of the PASSCAL program is the acquisition (and/or development, if necessary) of a light weight, rugged, portable 1 Hz seismometer with acceptable response characteristics. It gains us little to design a highly compact light weight datalogger package if the accompanying seismometers can not meet the same standard. The popular 1 Hz HS-10 geophones, for example, weigh nearly 50 lbs per 3 component package with carrying case. Other lighter weight 1 Hz sensors currently available may prove satisfactory, but at least some have been suspected of poor response characteristics. In fact, portability and response fidelity are commonly rather exclusive, one being achieved at the expense of the other.

A PASSCAL panel on sensor development consisting of Steim (Harvard), Sacks (Carnegie), Wielandt (Zurich), and O. D. Starkey (Teledyne/ Geotech) has begun a program of sensor evaluation, using a set of shake tables at DTM and at Teledyne/Geotech. To be investigated is the entire suite of portable vertical and horizontal sensors in current use in wide angle reflection-refraction, surface wave and normal mode investigations of continental lithosphere as well as strong motion transducers. Of special concern is linearity as a function of amplitude, the possible presence of parasitic resonances, and cross-coupling response as a function of amplitude and frequency. The effects of temperature and external magnetic fields, and some quantitative estimates of durability will be sought. The results of these investigations will be published after consultation with the manufacturers. Continuing negotiations will be carried on with manufacturers to incorporate desirable design modifications in existing instruments. We anticipate that these negotiations will lead to improved transducers including those having lighter weight and significantly better response characteristics.

Broad-band portable seismometer: One important long-range objective of the PASSCAL program is to develop a broad-band long period **portable** seismometer. By "broad-band" we mean an instrument that ideally will give a usable response from 100 sec (perhaps 20 sec in the initial stages) periods to at least 5 Hz. Such a sensor would be the basic tool for recording regional and teleseismic earthquakes with PASSCAL arrays. It is needed for a wide range of studies that cannot be done with existing portable sensors, such as surface wave studies, studies of attenuation, and many kinds of earthquake source studies. While the development of a broad band sensor is not as urgent a priority as development of the datalogger, it is likely to require a longer time period and depend more critically on innovative design.

The development effort is being lead by Steim and Sacks, both of whom have extensive experience in the hardware development of broad-band seismometers. Steim has recently built a state-of-the-art broad-band system of non-portable design and is committed to developing a similar portable sensor. He will devote about six months, beginning in the fall of 1984, working with Sacks at DTM. Wielandt is expected to collaborate for at least two weeks during the fall or early winter of 1984 with Steim, who has already spent several weeks collaborating with Wielandt on permanent broad-band seismometers. Wielandt has indicated, however, that he is prepared to commit as much as 6 months or more of his time to developing a portable broad-band sensor, provided that effort is directed toward production of a fully operational field unit on a commercial scale (i.e., is not another laboratory instrument). There will also be extensive collaboration with some of the instrumentation firms, notably Teledyne/Geotech (O.D. Starkey), who are interesting in developing a commercial version of a broad-band seismometer.

We have no estimate in this early stage of the program of a timetable for development of a broad-band sensor, although we do expect that a 20 sec instrument can be developed relatively rapidly, while the 100 sec instrument will require considerable developmental effort. For the long term, the broad-band sensors will be used more sparingly than standard seismometers. For experiments where spatial aliasing is a concern, long-period instrument responses mean that station spacings may be proportionately greater. In the long term budget we are requesting funds for 400 three component instrument sets (1200 seismometers), which we judge adequate for most anticipated applications.

Positioning: The Global Positioning System (GPS), when fully operational in the late 1980's, will make highly accurate seismometer positioning practical for large-scale PASSCAL experiments, where thousands of array elements may be deployed. The dataloggers are being designed with an input port, through which geographic coordinates and elevation can be downloaded from portable GPS (or other satellite) receivers and stored directly onto the permanent recording medium of the datalogger. The GPS represents a significant improvement over any previous positioning system, and its potential for revolutionizing geodetic surveying is well known. For PASSCAL, the advantages of GPS include portability, simplicity, and speed. With a single receiver, GPS gives instantaneous absolute positioning accuracy to better than 10m (latitude, longitude, elevation) using the classified military (P) code. Instantaneous relative positioning to several meters is also possible using the unclassified civilian GPS code, provided that two instruments are used, and that absolute positions are computed later.

4.2.3. PASSCAL Facilities: Operation and Maintenance

An expanded discussion of support and maintenance facilities appears in the Program Plan (Appendix 2A) as Chapter 8, and budgetary and implementation matters are more fully developed in Chapters 13 and 14.

Effective operation and maintenance of a pool of 1000 instruments requires that essential support services and facilities be provided. We propose to establish as an operational division of PASSCAL known as *PASSCAL Facilities*, whose day-to-day operations will be supervised by the Chief Engineer[1] The PASSCAL Facilities will have responsibilities in the areas of maintenance, calibration, deployment, training, communications, data retrieval and preprocessing, transportation, and land work. Justification for centralized facilities in PASSCAL. Scope of facilities and services supported by PASSCAL.

The PASSCAL array can be seen to be analogous in many ways to a large oceanographic research vessel. There are many different institutional research people who need to use their time effectively, leaving the operation and maintenance of the ship and its basic facilities to a full-time staff. Centralization of support is easily justifiable on the ground that very few institutions could afford to provide the number of personnel required, and that the maintenance of many separate support staffs would be far more costly than the provision of a single central facility. The strength of the new system lies in having standard instruments with a low failure rate forming each array deployment. The requirement for **standardization** is therefore a very strong constraint which leads to the central facility. Finally, a number of operational chores

1. See Appendix 2A, Figure 10-1.

which arise during work with a large array can be done well by university personnel only in rare cases... the provision of optimum communications, the services of an experienced landman, and management of dynamite shooting, for example.

On the other hand, it is widely agreed that participation by the PI and associates must be an intrinsic part of the process of experimenting with the PASSCAL instruments. It turns out that when one models the number of personnel required in the field, that this figure varies widely, depending on whether a large-scale cooperative experiment is being conducted or not. Consequently, it is planned to utilize PASSCAL staff at a level required to maintain and operate the instruments at a low-level baseline of activity. This would include essential engineering, programming, and maintenance people, as well as a basic number of per-instrument staff for field operations and maintenance. At this baseline level, it would be possible to supply a single maintenance/training technician to a PI who has borrowed 100 instruments for a small-scale experiment.

The basic operational principles of the PASSCAL Facilities are therefore:

(1) Equipment will be centrally procured and managed, for all cases where standardized performance and centralized management is required. This would probably encompass nearly all the needed facilities except for vehicles owned by participating institutions and special gear incorporated into an experiment by a PI.

(2) In general, personnel costs will be minimized by:

- use of contracts to provide short-term services, like drilling and shooting, and extra field personnel.
- use of contractors to provide continuing services, such as operating the maintenance facility.
- use of PASSCAL employees only for supervision, contract monitoring, and liaison with PI's.

(3) For large scale cooperative experiments, the funding and logistical issues are more complex. We exploit the fact that such activities will come up about once a year for two or three months only.

- *Funding* for the data acquisition itself will come through PASSCAL... as a part of its ongoing effort. Such support will come as an annual activity, in which the extraordinary costs of the annual large scale experiment are met for the two months of the experiment.
- *Field hands*, for deployment and retrieval work will be provided by the participating PI's. PASSCAL support would cover their travel costs. Wages of extra labor would also be covered.
- *Permitting, drilling, shooting, communications*, and other specialized services would be provided through the Special Services branch of PASSCAL, with close collaboration between the PI and the Logistics Officer. The use of contractors or of PASSCAL employees would follow the guidelines set above for keeping personnel levels (and costs) down.

The philosophy for operation of the Facilities, therefore, leads to an economically optimal approach which requires that PI's be closely involved in the conduct of the field work. This requirement follows from grounds of simple economy, in addition to the basic requirement that the PI exercise responsible oversight in all phases of his research.

Functions and organization of the PASSCAL Facilities

Consider (Figure 4.4), a hypothetical deployment of a 1000-element array.

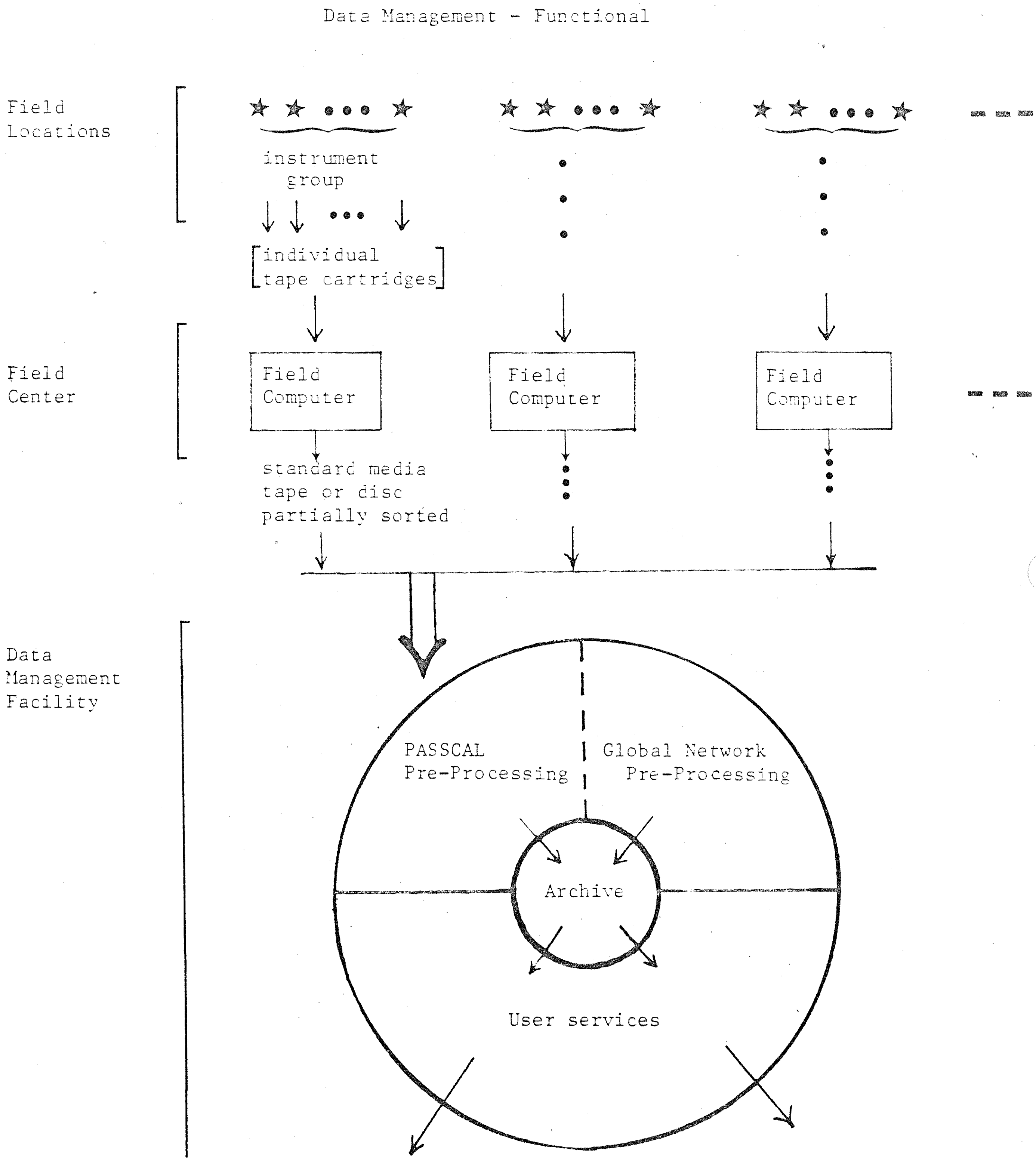

Figure 4.4. PASSCAL Data Management — Functional

(1) The instruments themselves are dispersed geographically over an area of linear dimension as great as 1000 km. Subsets of 100 or so may be situated in another area altogether. An appreciable number will be in maintenance status elsewhere, at the Central Maintenance Facility.

(2) The support crews, their vehicles, and computers must be headquartered in field centers in towns near the deployed array.

(3) The Data Management Center, in which the edited event data are to be archived, is in yet another distant central location.

There is, therefore, a flow of hardware outward from the central instrument facility, to the field centers, and to the field sites. A similar reverse flow of digital data inward from the field sites to the field centers and on to the Data Management Center.

It will be required, therefore, that the per-instrument support operating out of the field centers be vehicle-based, with adequate support personnel. Following the discussion of the different divisions of the support Facilities, we develop quantitative estimates of the magnitude of the support which will be required per-100 instruments as a basis for the budget plan.

We adopt the following four-fold division of the PASSCAL Facilities (as described below).

(1) The Central Maintenance Facility

(2) The Field Deployment and Maintenance Group

(3) The Data Support Group

(4) The Special Services Group

It is generally agreed that a contractor should be engaged to run the facilities and provide staffing for the maintenance and deployment functions. The key supervisory people... the Chief Engineer, the programming staff, and the Logistics Officer, at minimum, should probably be PASSCAL (IRIS) employees.

Central Maintenance Facility (CMF)

In order to maintain high standards of instrument performance and low down-time, the CMF will be required to perform the following functions.

(1) Maintain inventory of instruments, and records on deployment and maintenance histories. Administer distribution of instruments and training of users.

(2) Schedule and conduct routine preventive maintenance. Do diagnostic instrument testing before instruments are returned to the field.

(3) Repair, maintain, and calibrate instruments returned from field duty.

(4) Maintain and administer field support crews, including vehicles, computers, etc.

(5) Administer acceptance testing when new instruments or components are delivered. Advise business office on enforcement of quality control in the purchasing and contracting process.

The resource needs of the CMF will grow with PASSCAL, as more and more instruments are brought on line. Since instrument specifications are being developed at this time, the Program Engineer should come aboard as soon as is practical, preferably around Feb. 1, 1986, as detailed in Chapter 14. The Program Engineer is needed to interact with the manufacturer(s) selected for instrument construction throughout the course of production, and with an augmented staff, to perform instrument acceptance tests and to take delivery. Space for all the PASSCAL Facilities will be focused at the CMF, since the Field Group and the Data Support Group will require some central administrative site.

Field Deployment and Maintenance Group (FDMG)

The backbone of the FDMG is a number of support vehicles which are equipped to handle all tasks connected with the deployment and visitation of instruments in the field. They will also be needed to handle field data retrieval, validation, and initial data preprocessing. Staffing of these facilities has to be subject to our policy given above, which minimizes the number of full time staff in accordance with a usage profile which fluctuates substantially over the year.

A structure for the FDMG facilities which minimizes the number of vehicles is as follows:

(1) Instruments will be transported to a field area *en bloc* in Instrument Vans, which are trailers outfitted for storage and cross-country transport. We assume 200 instruments per van. The use of vans minimizes the number of vehicles which must be owned for deployment to actual field sites. Vans to support 100 instruments may also be included as a more effective way of distributing smaller complements of instruments.

(2) Deployment will be done from rugged, field-worthy vehicles outfitted to support about 25 instruments. A single deployment vehicle can support up to 100 instruments, if used in conjunction with the vans.

(3) Test vehicles will contain electronic support equipment, including test gear, possible SSB or satellite communications, and a field computer for initial data playback and monitoring. The number of instruments which can be tied to a single field computer is still a matter of much analysis; we set here a figure of one field computer for every 50 instruments.

With this organization, the FDMG will include 10 deployment vehicles, 5 instrument vans, and 20 test vehicles, for the full 1000 instrument system.

For a large-scale field exercise, staffing the FDMG will require personnel from both the PASSCAL Central Maintenance Facility and from PI groups. For smaller studies being handled by a single PI, most of the personnel would come from the PI's staff.

The Data Support Group

Once data cartridges are retrieved from the field stations, support services related to the playback and preprocessing of the data and to preparing the data for archiving and release will be provided by the Data Support Group. Although the line between this function and the hardware services of the FDMG must be somewhat arbitrary, and may be lost in practice, a separate discussion here serves to highlight the importance of Data Support services. Some of this discussion also overlaps that in the next section, which deals with Data Management for PASSCAL.

Data Support services are needed to convert unedited data recorded on the field instrument media (cartridges) into edited event data resident in the IRIS Data Management Center.

The Data Support functions are:

(1) Physical retrieval of cartridges.

(2) Initial playback of cartridges. Validation and editing of descriptive parameter data. Validation of instrument performance. Creation of initial edited trace data in standard format on standard exchange media (e.g. SEGY tape).

(3) Initial monitoring by the PI of the progress of the experiment.

(4) Final editing for event data sets. PI supervised editing of field data and creation of edited event data sets or profile data sets, suitable for inclusion in the Data Management Center archive.

The field computers. Each test vehicle in the field area will have a computer with facilities for playback, plotting, sorting, and output of data onto standard media. It is expected that the PI will be responsible for spelling out the procedures which need to be done on the data by the field computers. Either the technician in charge of the test vehicle, the PI, or any suitably trained field assistants will carry out the playback procedures, under the supervision of the PI.

Two PASSCAL senior programmers will be needed to supervise the field computers as a standard system. Responsibilities would include the production or purchase of software needed for the baseline needs of quality control, initial sort, and instrument evaluation; also the maintenance of software and software documentation, and the training of participating scientists in the use of the software and hardware.

The baseline facilities and personnel needed for Data Support are estimated at:

(1) One field computer per 50 instruments.

(2) One test vehicle, with driver-technician, per 50 instruments.

(3) One field computer at the CMF, for software debugging and prototyping, for backup service to the field systems, and for general use by the programming staff.

(4) A programming staff at the CMF, consisting of a senior programmer, and at least one other programmer-analyst. Two is an absolute minimum here, for a large experiment involving the entire network would need at least one software support person in the field and one in headquarters.

We emphasize that the initial definition of the field computers, the definition and writing of the baseline software, and whole process of getting these facilities started will require additional technical effort, probably through extramural projects and contract-written software. Details of prototype operations already in the planning stages are given in Chapter 14 of Appendix 2A.

Support services at the IRIS Data Management Center. While the need for field computers to facilitate data flow and serve the PI in monitoring the experiment is overwhelming, the analysis (in section 4.2.4) of the requirements for preprocessing of PASSCAL array data also show clearly that a substantial fraction of this work will still need to be done at a facility with much more storage and speed than the field computers. We have, therefore, specified that the data stream from the PASSCAL array will require substantial centralized data preprocessing services (particularly sorting), and have recommended that this be done at the IRIS centralized Data Management Center. In section 4.2.4 we have summarized the nature and quantity of these needs. By agreement with the IRIS Executive Committee, all these services will be covered under the program of the Standing Committee for Data Management.

The Special Services Group

It is regarded as likely that PASSCAL will need to provide additional special services on the ground that the expertise required is highly specialized and/or intrinsically centralized. Under the Logistics officer, this group would supply such services as (1) communications, (2) permitting and blasting, (3) time and position services, (4) other special needs.

4.2.4. Implementation Calendar

A detailed discussion of the calendar for startup of the instrumentation, facilities, and data management efforts is given in Chapter 14 of the PASSCAL Program Plan (Appendix 2A).

The overall form of the ten year calendar has three phases:

(1) FY85 through FY87: Program startup, instrument design, prototype testing, initial procurement, support facilities definition and startup. Interim scientific experiments using available instruments, with testing of prototype PASSCAL instruments.

(2) Mid- FY87 through FY91: Operation of the PASSCAL instrumentation and support facilities through a period of growth.... at the rate of 200 instruments per year.

(3) FY91 and later: Instrumentation at full complement of 1000. Full operation of system. Time to begin second generation technical enhancements to system and plans to iterate system configuration for more efficiency, lower cost, and higher performance.

Figure 4.5 shows a summary timeline for the operations of PASSCAL.

PASSCAL - Proposed Ten Year Schedule for Implementation.

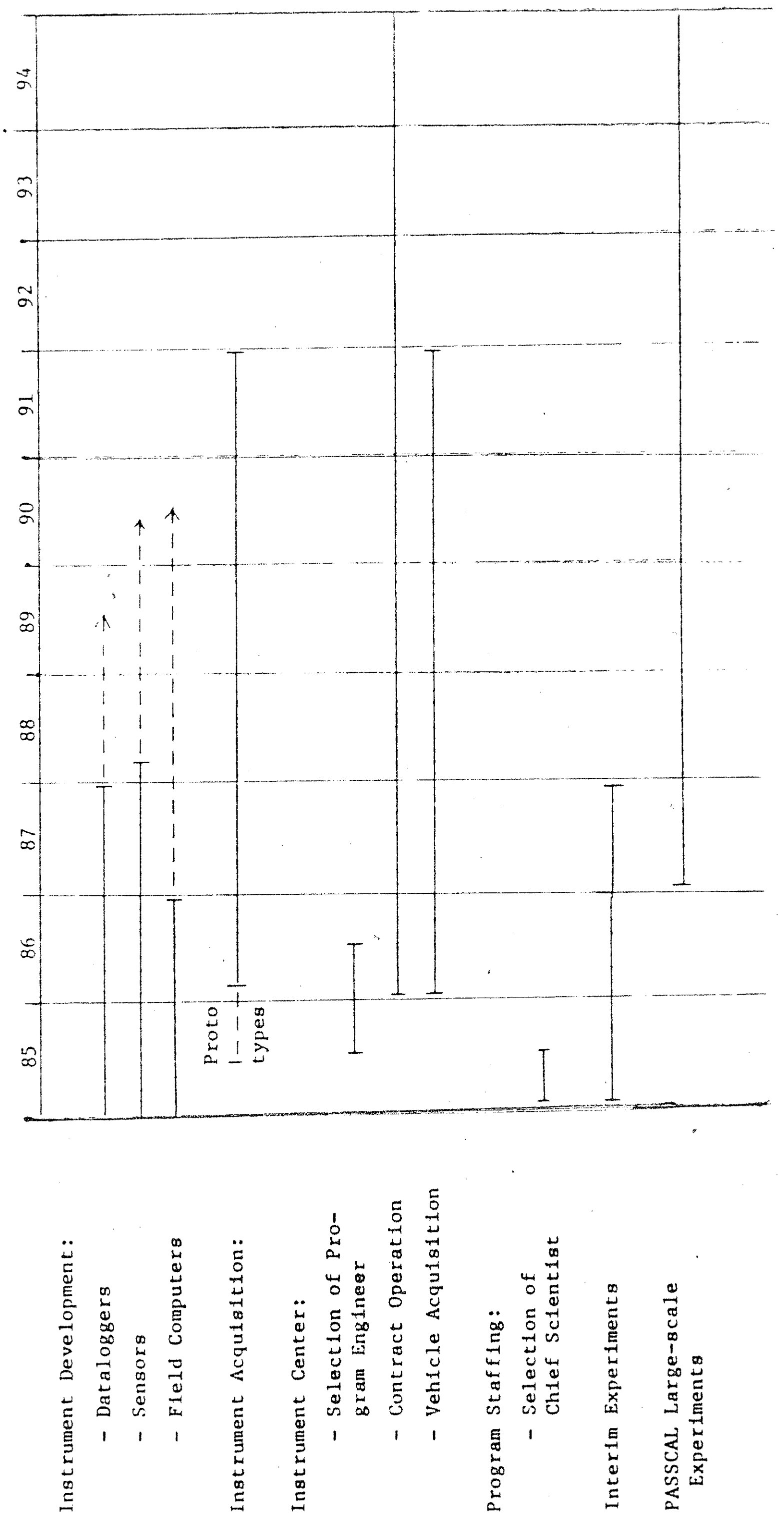

Figure 4.5. PASSCAL Proposed Ten-Year Schedule for Implementation

The key dates on the schedule are:

Datalogger definition	by 6/85
Datalogger prototype delivery	by 2/86
Begin datalogger acquisition	8/86
Datalogger acquisition period	through 1991
Sensor evaluation period	9/84 through 2/86
Sensor acquisition period	2/86 through 1991
Advanced sensor development	by 1987
Central Maintenance Facility	
specifications	by 1/86
startup	6/86
Chief Scientist	begin 2/85
Chief Engineer	begin 9/85
Programmers-Logistics Officer	begin 9/85
Field computer evaluation	9/85
Initial field computer acquisition	2/86
Interim experiments	9/85 through 9/87
First experiment with new production instrumentation:	1/87

The time line for the buildup of PASSCAL is a fairly aggressive one, which reflects a high degree of enthusiasm on the part of many seismologists across the country for this program, and their commensurate willingness to contribute their efforts. If adequate funding is made available, as described by the budget plan, we have little doubt that the above schedule is reasonable.

4.2.5. Comments on the Budget

A full presentation of the ten-year budget plan is given in Chapter 13 of the Program Plan (Appendix 2A), while a detailed discussion of the startup period is given in Chapter 14.

Ten year budget profile:

The budget ramps up steeply for two years to $7.5 million in FY87, then grows more gradually to $13.5 million in FY89, remaining essentially constant thereafter. It has this form, and does not come down, because most of the major costs are proportional to the number of instruments, which characterizes the growth of the program. Among these proportional costs are: instrument maintenance, instrument replacement, numbers and maintenance cost of supporting equipment, and costs of data acquisition in large experiments.

These per-instrument costs reflect the true cost of operating a large facility, and obtaining the benefits of the program. We estimate these at $13,000 per instrument per year. Within uncertainties, this is equal to the original one-time acquisition cost of the instrument. In slightly different terms, a participating university would have to pay $1,300,000 to obtain full use and support services for 100 instruments for a year, if the costs were paid for "up front". This is somewhat less than the comparable cost for a commercial 96 channel crew, or half of a 200-channel SGR crew, prorated downward to comparable personnel levels.

Startup budgets:

During the period 1 February 1985 to August 1986, before acquisition of production instruments or supporting hardware has begun, major items drive the budget level:

(1) Development costs: consultants, engineering salaries, component costs, travel.

(2) Purchase of prototype dataloggers.

(3) Interim large-scale lithospheric experiment.

At the point where acquisition of production instruments begins, then all per-instrument costs also begin, including engineering and maintenance personnel. This point occurs in mid-FY86, which budget also shows the following per-instrument costs associated with the beginning of the growth of the physical facilities:

(1) Equipment acquisition.

(2) Engineering personnel and maintenance.

4.3. IRIS Data Management Center Implementation Plan

4.3.1. General Approach

The need for large-scale, high performance data management services of a similar sort by both PASSCAL and the Global Seismic Network (GSN) has led the IRIS Board of Directors to establish a separate Standing Committee to implement the centralized, common data management facilities. A division of operational responsibility for handling the PASSCAL data flow is thereby defined.

(1) PASSCAL will implement and operate all components of the data management system which are decentralized, and associated with the initial field data acquisition. This will consist of the seismometer packages and the field playback/computer systems which will normally be located at a support center near the deployed packages.

(2) The Data Management Standing Committee will implement all components of the data management system which are centralized, with particular emphasis given to functions which are common to PASSCAL and the Global Network. We call this the Data Management Center.

Generally, then, the data will flow from the instruments to the field support centers under PASSCAL, and be passed on to the Data Management Center.

The size and diversity of the data sets that will be generated by GSN and PASSCAL, as well as the requirement for wide distribution of these data, present problems of a scale beyond any current models of data handling within the seismological community. Development of an adequate system must be undertaken with a full appreciation of the enormous quantities of data to be accessed at any time and the high bus speeds required for this task. The planned and predictable progress in computer processing and storage technology will obviously result in a substantially enhanced capability and reduced costs/unit operation over the several years that will elapse between the planning and implementation phases of this project. Furthermore, the growing trend toward multiple processors in both super-mini and mainframe computers may allow the data management system to be more simply realized than current architectures will permit.

4.3.2. PASSCAL Needs for Data Management Services

Raw data appearing in sequential form on cartridges from individual seismometer stations require a number of processing steps before the information is fully documented, sorted, and associated with other information in a data archive. The archive itself represents a repository of such documented data sets which exists to provide convenient access to the data by participating scientists. We list services (1) to scientists using the PASSCAL array for data collection and (2) to scientists using data archived in the Data Management Center.

Needs associated with data collection

Scientists participating in PASSCAL data collection will have need for the following:

(1) Intelligent functions in the instrument packages: Selection of channels, sampling rates, and filter settings; preprogrammed time windows; simultaneous multiple triggers; simple event picking; storage of parameter information in trace headers.

(2) Retrieval of data cartridges from instrument locations to field centers.

(3) Rapid response services: verification of integrity of cartridge data; verification of performance of recorders; verification of characteristics of recorded data; plotting and preliminary analysis of a limited subset of the total recorded information. In controlled source experiments, plotting of record sections or shot gathers is essential to verify that the design characteristics of the experiment are optimum, particularly the source characteristics. In natural source experiments, an early look at selected instruments is essential to develop a preliminary catalog of events recorded so that subsequent sorting and association of traces with events can be done smoothly.

(4) Sorting and event association: Data recorded on the cartridges is in effect cataloged by its instrument number and may be unidentified, if due to natural sources. For the data to be usable, the traces from the entire network must be resorted and associated with a catalog of events. Where the events are artificial sources, the number of traces will often be small enough and the complexity of the catalog simple enough that this is a simple process. When many natural events are recorded by an array, or when a major 3-d seismic reflection or tomography profile is involved, both complexity and quantity escalate substantially. Since much of the rationale for the PASSCAL array lies with its use for acquiring large, complex data sets, the need for well-thought-out sorting and event association provides a strong demand upon the overall data management system.

(5) Support for documentation activities by the P. I.: The process of converting the field data to a form useful for analysis and appropriate for the Data Management Center archive may also involve a second, more time consuming stage of producing event catalogs, plots of event gathers, and of thorough editing and preprocessing of the data for parameter inconsistencies, errors, and other grossly unacceptable perturbations of the data. This requires interactive access to large quantities of sorted or partially sorted trace data and availability of relevant software tools for this work.

The **sorting** process requires a few remarks. The rearrangement of large numbers of data traces on the basis of event rather than the basis of instrument number is equivalent to a large matrix transpose, say 1000×500 in dimension, where the elements to be rearranged in the transpose are digital data traces of, typically, 10^4 bytes. If it is demanded that the sort be carried out on a small computer, with quite limited memory and disk, then it is required that the original data cartridges be played back many times. Such a method is unacceptably demanding of time, wear on the cartridges, and on the operating personnel. If it is demanded that the data be fully sorted in one pass, to minimize these three penalties, then extremely large scratch disk storage, perhaps 2×10^{11} bytes, or 600 standard drives, would be required. This method poses an unacceptable cost, especially if such data sets occur relatively infrequently. The way of obtaining the sort at acceptable penalty levels for time and facility cost is a two or three-stage sort. In the PASSCAL context the field computers can be utilized during initial playback to provide the first level of sorting. At most data from 100 instruments could be sorted at that stage. The remaining stages of sorting would have to be carried out at a more powerful fixed facility.

Individual P.I.'s who are running experiments with a 100 element deployment and moderate rates of data may well be able to handle all sort functions at the level of the field computer. The experience of the U. S. Geological Survey with their 120 analog instruments confirms this projection.

Requirements for the data archive

What services are required by the scientific community? For scientists intending to retrieve data from the archive of the Data Management Center, the following services are needed:

(1) Catalogs, both printed and on-line, of data in the archive. On-line, the ability to search a catalog structure for basic information about availability and characteristics of data sets.

(2) Ability to preview representative trace displays of a given data set.

(3) Ability to obtain desired data sets or subsets in standard exchange media format in a reasonable time. With present technology, this would be overnight mail delivery of up to four reels of 6250 density 9 track tape, or about 500 megabytes on a one day basis. In the future, much larger quantities should be feasible.

(4) Dial-up access to the Data Management Center.

(5) Services for high resolution, high volume previewing of the archive, in the Data Management Center. It is premature to estimate the level of this kind of support. An obvious minimum would be that which would support the needs of the management of the Center.

(6) Adequate support of the software which provides the above services. This includes help features, manuals, a professional programming staff, and the ability to assess the need for and to develop new software tools.

Needs for hardware and software at PI's institutions:

Ultimately, research on the data sets produced by the PASSCAL array will be done by scientists working at their own institutions. Unless each participating research group has access to a certain level of computer hardware and software for data analysis, much of the research will be frustrated. In Chapter 15, we discuss this issue at greater length.

4.3.2.1. Functional Components of a Data Management System

From the preceding sections, we can define (Figure 4.4), the different components of the total PASSCAL-IRIS data management effort:

1. The instrument packages are to have a limited, but critical set of logical functions. These would be incorporated into the firmware of the different modules. Intelligent algorithms for event detection are now part of the arsenal of network seismology, and are the most general tool for restricting the volume of data at its source. For many experiments, however, multiple detection algorithms may be necessary, acting in an *or* mode; some development work lies ahead. In general, parameter information about each sensor will be loaded into trace headers, and information about the trace data itself might be extracted and also saved in the headers. The application of a P-picking algorithm for use in local earthquake situations and the storage of the picked times on trace headers could greatly simplify the subsequent preprocessing of the large data sets involved.

2. The field computers will be systems equipped for playback of the data cartridges, perhaps with multiple input devices; otherwise they would carry memory, scratch disk, plotting hardware, and a device for the output medium. Such systems can be assembled with available 32-bit microprocessor-based high performance computer workstations and peripherals, and fit in the physical space which would be available on a PASSCAL test vehicle. Identification of specific hardware will depend on what is available at the time bids are sent out, and will depend on qualification for the field environment. Immediate plans for testing existing field computer prototypes are described in Chapter 14 of Appendix 2A. One field computer will go into the field center to work with each 50 instruments (estimated).

A general philosophy for minimizing bottlenecks in data flow involves handling as many chores routinely, and as early in the data preprocessing as possible. The field computers are intended to meet this need and to meet the need for the P.I. to conduct rapid inspection and evaluation of the data. It is adequate to set performance specifications for the field computers as follows: equal to those presently obtainable and qualified for the purpose (or better), at present price levels (or better). Clearly, availability of Gbyte levels of disk capacity and multi-megabyte memory is anticipated as a means of enhancing the ability of the field computers for first stage sort of trace data.

In terms of the needs posed above, the field computers will handle (1) rapid response functions and (2) first stage sort and event association. In perhaps a few instances, where passive deployments of the PASSCAL array produce a steady, but moderate flow of data, it is possible that the mobile field computers can be dispensed with almost entirely, in favor of sending the field tapes directly to the playback facility at the Data Management Center.

3. **The Data Management Center**, to be implemented by the IRIS Standing Committee on Data Management, needs to have facilities and support for completion of data preprocessing, ahead of the archive and data distribution functions (Figure 4.4). This preprocessing facility would be a compatible superset of the field computers, and would permit completion of the sorting, event association, and documentation chores of the PI. The performance specifications for the IRIS-operated facility would need to be based on typical recurring data volumes which are expected from a typical mix of field experiments. Where extraordinarily large, but relatively uncommon data sets become involved, the economics of cost and effort suggest that preprocessing is better handled by an outside contractor.

Long term budget estimates and FY85– FY87 budget figures (chapter 13) for the PASSCAL data management effort reflect the cost of all operations located in the field.... the instrument packages and field computers, but not the central IRIS Data Management Center, which is covered by the program documents for the Data Management Standing Committee (Appendix 3A).

4.3.2.2. Analysis and Projections for Data Flow Through the PASSCAL System

In Chapter 9 of the Program Plan (Appendix 2A), we have developed a series of model experiments designed to illustrate the quantitative attributes of data which will be collected in various types of lithospheric studies. The detailed discussion there leads to a model for the annual data flow into the Data Management Center.

Model for demand on the Data Management Center

We have adopted a model for the usage of the system. It assumes 3 months of deployment of the full array in a large scale cooperative experiment involving controlled and natural sources. For the remaining time, over half the array is in passive mode, recording natural events... most often teleseisms, and the remaining instruments are on loan to smaller groups of P.I.'s in groups of 100. The impact on the field computers is dependent on the particular type of study, but total figures are unimportant, since the field computers are defined by the requirements of size and cost at state of the art technology, and will be utilized as fully as possible. For all but certain experiments with 100 element arrays, however, the field computers will be insufficient to handle all the data preprocessing.

This model leads to the following estimates for the data flow into the Data Management Facility for preprocessing and archiving.

(1) Between 50 and 100 GByte per year from recording of teleseisms.

(2) Data from one large cooperative experiment per year. From 50 to 200 Gbyte per year, depending on the particular experiment.

(3) From six small scale experiments with 100 element arrays. Less than 50 GByte per year.

(4) From an unanticipated swarm of aftershocks or volcanic earthquakes: Rare data sets as large as 100– 500 Gbyte.

The annual data flow into the Data Management Center would then be a fairly steady 150– 350 Gbyte per year, augmented by unpredicted event swarms delivering an additional

100– 500 Gbyte each. The general level could go up in a few years if the use of vibroseis caught on for tomographic and three-dimensional reflection studies.

The management of the large data volumes in an archive we leave to the Standing Committee for Data Management. We are, however, presently operating under the assumption that technologies for storage of multiple GByte data sets (laser optical media, vertical magnetic technology) will permit the archiving of 10^{12} bytes/year without unacceptable demand on the resources available.

Design and policy recommendations

User Data Management Plan for each PASSCAL experiment.

It is recommended by the Standing Committee for PASSCAL that every scientist requesting PASSCAL facilities and services for data acquisition be required to submit a detailed Data Management Plan. This Plan would explain the nature and quantities of data expected, the methods to be used for preprocessing and sorting of data, the nature of the archive to be produced, and the requirements for support from PASSCAL and the Data Management Center, both for facilities and for software support. It is recommended that the Data Management Plan be incorporated when appropriate as a part of a funding proposal to U. S. Government agencies. PASSCAL would retain the power to approve, modify, or disapprove the proposed Plan. PASSCAL would offer to provide technical advice on the capabilities and services available.

Public release of data

The use of facilities funded by the U. S. Government carries an obligation that the data produced be made available to the public. On the other hand, individual PI's require time to prepare usable data for research and in an edited, documented form. The Data Management Plan, therefore, should include a proposed date or schedule for the release of the data to the Data Management Center for archiving and public release. By filing his Plan, the PI agrees to abide with the schedule for data release. Recognizing that the time factors involved will differ from experiment to experiment, PASSCAL will set guidelines regarding data release.

Summary recommendation on the Data Management Center.

The IRIS Data Management Center needs to have the capability to complete the sorting, event association, and editing of most PASSCAL data sets. The expected data flow (150– 350 GByte/yr) from the predictable component of our array utilization model should be used as the basis for planning the baseline support services and facilities.

On Very Large Data Sets

The largest data sets (> 200 Gbyte) must be treated as statistical outliers. The oil exploration industry has several years of experience and the large facilities required to manage such quantities of data. In addition, industry is equipped for complete processing of three-dimensional reflection data sets. It is anticipated, therefore, that PASSCAL will find it appropriate in certain situations to contract with industry for processing services associated with uncommon, but very large data sets.

A caution

The models developed here describe how the arrays would be used today if they were available now. By the time a full 1000 element array is available, (1989, at the earliest), it is likely that evolution of technology and cost of mass storage, of laboratory computers, and satellite telemetry will make the data volumes discussed here seem less problematical. Continued aggressive development by the exploration industry of such technologies and of three-dimensional seismic reflection methods will probably move PASSCAL toward a comparable increase in emphasis of the three-dimensional imaging methods for which the PASSCAL array is most fundamentally suited, and a commensurate increase in the data flow through the system.

4.3.3. Data Management Center Development

As a result of these considerations, our implementation strategy involves a two-stage design process, in which we first will define in detail the functional requirements of an effective data management facility, before undertaking a more detailed design and system specification that will enable these objectives to be met. With the help of experts from both systems design and computer science interacting with seismologists and managers of existing seismological data services, we expect to enter the design specification phase with a set of definitions and functional specifications of sufficient clarity and detail to insure a successful data center operation. We expect to accomplish both stages of the design process by contracting with commercial entities.

Data flow is expected in FY85 and FY86 with interim experiments of both PASSCAL and GSN using existing data distribution facilities. It will be handled partly by the cooperating institutions that generate the data, but data will also flow through the prototype data facility (e.g., DARPA's Center for Seismic Studies) insofar as that is possible. Beginning in FY87 there will be data from some of the new GSN and PASSCAL stations, and the data center will begin to assume responsibility for archiving and dissemination at that time. The initial volume of data will be modest and will be gradually increased to nearly its full design value by FY90. The schedule of design and implementation presented here reflects the projections of data flow described in the GSN and PASSCAL program plans.

The development of hardware and software systems required to bring the Data Management Center up to full operational capability will be acquired in a series of three, staged procurements beginning in 1985 and continuing through 1988 with completion anticipated in late 1988 or early 1989. In the first stage, the preliminary functional system specifications will be developed and the initial contractor selected by a DMC technical committee charged with this responsibility. This stage, during the first half of 1985, draws heavily on the prototype system developed in parallel by the PASSCAL data management activities discussed in above and in Sections 3.2 and 4.2. This initial contract is expected to provide a detailed definition of the DMC requirements by the end of calendar 1985. These requirements will form the basis for a solicitation for a second contract which will, in turn, provide by the end of calendar 1986 a detailed specification document for a fully operational hardware and software system for the Center. A contract for final system design and implementation is expected to be in place by early 1987. Implementation and installation will be carried out in the third stage beginning by mid-1987 with full hardware and software capability expected by the end of 1987 or early 1988. It should be noted that the option of leasing rather than purchasing hardware will be considered for providing the required capabilities for the Center. During subsequent years the development budget primarily reflects software maintenance and development expenses as well as those required to retain "state of the art" hardware capability.

In general terms, our needs will probably be adequately served by a cluster of high-capability minicomputers, or alternatively by several small mainframes with between 10 and 50 gigabytes of direct access storage. With existing technology, from 10 to 50 1-Gbyte Winchester disk drives would be satisfactory. While this configuration can be used to project reasonable budget estimates, substantial improvements in both price and performance will undoubtedly occur before the design document is finalized.

PASSCAL plans to produce some very large data sets, perhaps a fraction of a terrabyte, for three-dimensional imaging problems. Such data sets will require sort and gather processing that will be beyond the capability of the PASSCAL field computers, and thus will either have to be handled by the IRIS Data Management Center or contracted out to commercial services (such as those used in the petroleum industry). A final decision cannot be made until system specifications are available so that a comparison of costs can be made between enhancing the Center's computational resources and performing front-end data reduction offset on a contract basis.

4.3.4. Startup Data Management and Distribution Experiments

Important prototype IRIS data management and distribution experiments will be carried out by taking advantage of existing data center facilities developed by other organizations for other purposes. These experiments include present and anticipated global network data sets — the data sets that will be collected in PASSCAL-related field deployments during FY85–86. These experiments are intended to identify problem areas, to get user input to the Data Management Center design via feedback from using the prototype digital data sets, and to provide existing digital data needed for ongoing seismological research at IRIS member institutions. While distribution of these prototype data sets would not be limited to IRIS member institutions, their needs will have high priority.

Appendix 3A provides a detailed description of the startup data distribution experiments scheduled during the first two years, starting early in 1985. They include:

a. An IASPEI international digital data set consisting of 51 earthquakes recorded worldwide on a variety of instruments and in a variety of formats.

b. Prototype GSN network data sets from approximately 20 stations that will have capabilities approaching those desired for GSN stations.

c. Preparation and distribution of PASSCAL data sets from planned experiments that use presently-available data collection systems. Development of additional data handling requirements for PASSCAL field-processed will be a related objective. Controlled-source experiments (e.g., Long Valley) will provide these prototype data sets.

d. Hard-copy products evaluation that will include various display formats for digital data (e.g., event record-sections, perspective plots) and different media such as paper, microfiche and videodisk.

The data distribution to users during this two-year startup phase will be accomplished by a combination of facilities available at various cooperating institutions or organizations involved in data collection (e.g., LLNL, LBL, and IRIS member institutions) and the use of a prototype data center. We have investigated the possibility of using the facilities available at DARPA's Center for Seismic Studies as recommended in Ref. 7, and it appears that it will be feasible to do so by providing our own dedicated minicomputer, some peripherals, and communications interfaces to users and to other data bases. This IRIS interim system would be networked with the present CSS system including the present GSN-type data base and display capabilities available at CSS. This approach will enable us to develop independently the prototype IRIS data base structure and distribution alternatives, including PASSCAL data sets, while taking advantage of the substantial capabilities available already at CSS; it will also enable us to begin immediately to distribute desired data sets to users for their current research projects. We anticipate that this would be the headquarters of the prototype data management center and the DMC Program Manager. However, the Program Manager will also spend considerable time traveling to coordinate data preparation and certain types of distribution that will be accomplished through other cooperating organizations, and this will serve as a means of evaluating various existing alternative approaches to data handling that might be appropriate for IRIS use.

When the new IRIS Data Management Center is in place, provisions will be made for its orderly startup, such that the prototype data center operation described above will overlap sufficiently to assure no interruption in data distribution to users. The schedule of activities and budget are summarized in the following table.

4.4. Computational Technical Plan

An evaluation of super computing needs for IRIS was based upon estimates by IRIS member institutions in the recent NSF Super Computer initiative for scientific computing and from estimates by individuals and industry for large-scale data processing and data management. These estimates range from 80–100 hrs/yr for large users (approximately one-third of the IRIS members) to none for very small users. Large-scale data processing of three-dimensional data sets by industry suggest about 100 hrs/major experiment. Data archive and management will require 500– 1000 hr/yr. At approximate four years into the IRIS program these estimates suggest a need of: 1) university science computing, 2000 hr/yr; 2) special data processing, 500 hr/yr; and 3) data management, 1000 hr/yr; totaling 3500 hr/yr.