IRIS Data Management System
Self Study Report

Years 2000-2001

Summary

In 2000-2001 the DMS Standing Committee (DMSSC) undertook a self-study of the DMS to provide an overview of the current operation, to provide a summary of DMS initiatives and directions for growth as defined by the DMSSC and the DMS, and to define the role played by the DMS Standing Committee. This report is a collection of information provided by the DMS program manager and staff, the DMSSC, and the DMS subcontractors liberally sprinkled with current DMSSC consensus viewpoints on various issues. The purpose of this document is to provide DMSSC members and other IRIS interested parties with a snapshot of DMS activities and DMSSC responsibilities. As such the report should be viewed as a living document that will be updated by succeeding DMSSC chairs and the DMS program manager to keep its information accurate.

The role of the DMSSC is necessarily circumscribed by its biannual meeting schedule, and by its attempts to understand a rather far-flung operation involving personnel at eleven locations working for at least nine different organizations. This report it meant to provide a snapshot of current operations as well as guide for the DMSSC to provide meaningful and practical oversight for the IRIS DMS in the future.

A significant fraction of the DMS budget (~84%) is devoted to maintaining the core operations at the Data Management Center in Seattle and the two Data Collection Centers in Albuquerque and San Diego. The two DCCs are funded by subawards, the DMC is funded directly as a part of IRIS. These three operations appear to be reasonably streamlined. The remaining 16% of the DMS budget is disbursed through seven other subawards. The DMSSC feels that it is imperative that the appropriateness and performance of work performed by the other subawardees be monitored closely, as much of the funding for new developments must come from this 16% of the budget. The DMSSC is instituting an annual review of DMS subawards for appropriateness and effectiveness.

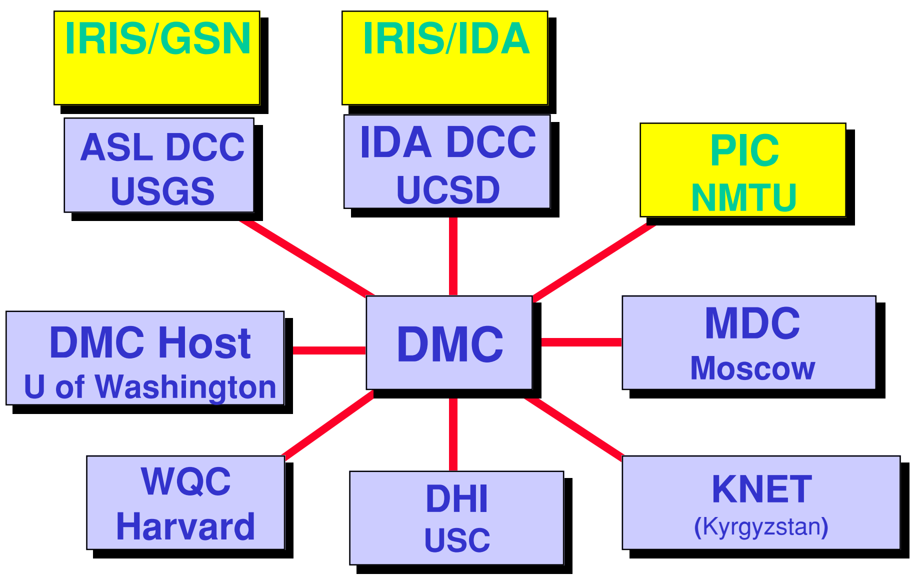
Overall the DMS operates remarkably smoothly, with data flowing from 37 permanent networks and a variety of PASSCAL experiments. It appears that data transfer from PASSCAL investigators to the DMC has improved remarkably in the past few years. PASSCAL data are now being archived at a rate greater than the GSN. Nonetheless, further improvement in PASSCAL and DMC communications, and particularly a bulletproof data synchronization system, are needed.

Table of Contents

I. DMS Structure and Functions (Alan Levander)

The IRIS DMS consists of a Data Management Center hosted by the University of Washington in Seattle, two Data Collection Centers operated by the USGS at the Albuquerque Seismic Laboratory and the IDA program at the University of California at San Diego, and ancillary services provided by a series of subcontracts to the University of Washington, Harvard University, the University of California at San Diego, the University of South Carolina, and the Synapse Science Center, Ltd., at the Institute for Problems of Mechanics (Moscow Data Center) (Figure 1).

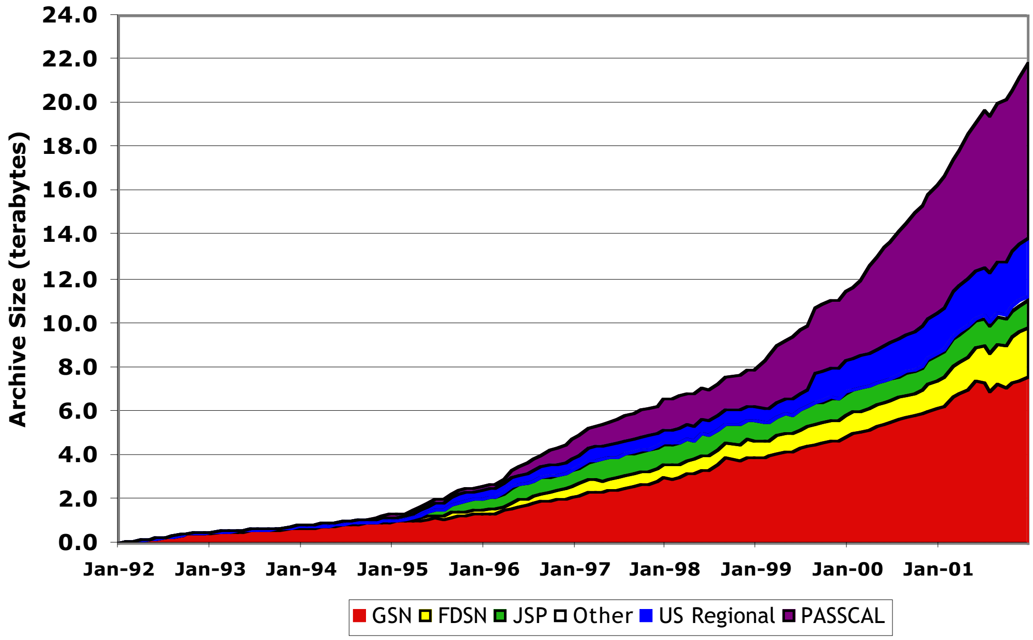
Figure 1: IRIS DMS nodes (blue) and IRIS data sources (yellow)



Data from the IRIS GSN is transferred to the DCC's via Internet and tape where it is quality controlled, time-corrected, reformatted, and passed along the DMC largely by frame relay circuit (TCP/IP on a private telecomm line). In addition, data from the PASSCAL program is transferred from the Socorro PIC or from PASSCAL PIs to the DMC by tape or Internet. [A frame relay circuit will be installed between the Socorro PIC and the DMC when traffic requires it]. Data from other global and regional networks (FDSN, CDSN, etc) is transferred to the DMC by tape and Internet. Data from the USNSN has been transferred by tape and will be transferred by frame relay circuit.

At present the IRIS DMS archive consists of approximately 22 terabytes (dual sorted by time order and by station order), and is accumulating at a rate of ~5.5 terabytes/year, as shown in Figure 2. [The DMC actually manages ~55 terabytes, as additional copies of the data are held on the StorageTek, and a backup copy at UNAVCO).

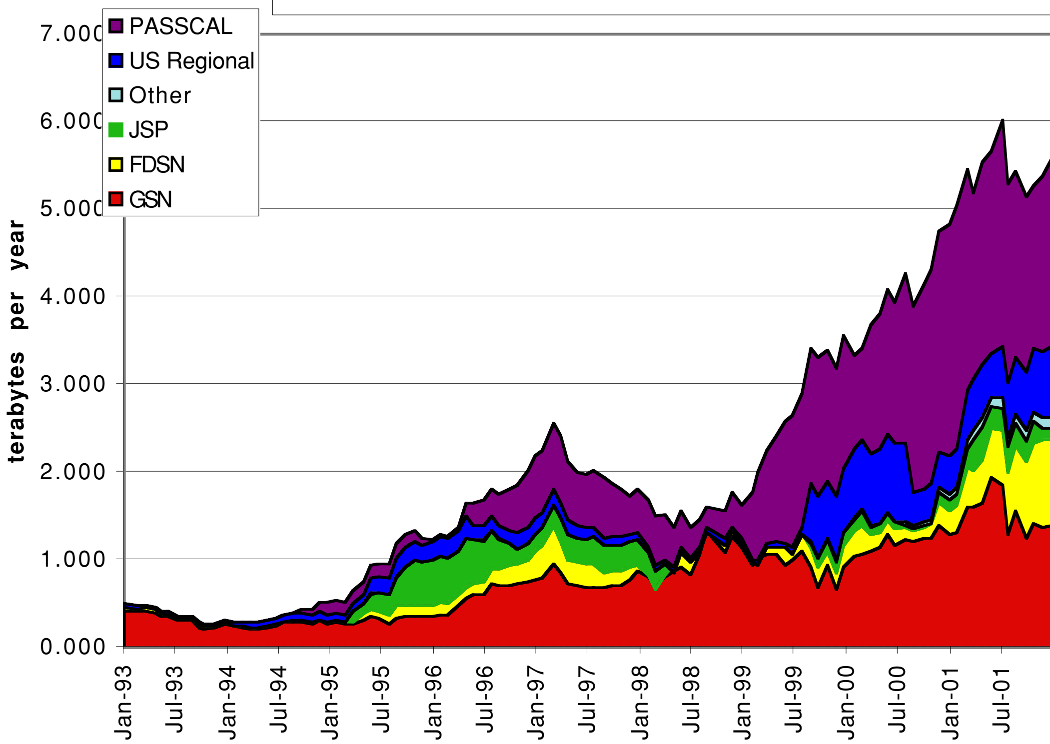
Figure 2. The Archive at the IRIS DMC
as of December 31, 2001



The rate of archiving at the DMC is presently about 5.5 terabytes/year (14 terabytes when all the redundant copies are included) and we believe the capacity might well be an order of magnitude greater than this with the existing hardware and software configuration. The systems are easily scaled and capacity and capability can be added as needed assuming financial resources are brought to bear on the problem.

PASSCAL data are presently arriving at the DMC with the highest inflow rate. At the end of 2001 we were receiving data from PASSCAL at 2.1 terabytes/year, from the GSN at 1.4 terabytes/year, from the FDSN at 1.0 terabytes/year, from US regional networks at a rate of 0.8 terabytes/year and about 0.1 terabytes from other sources (See Figure 3).

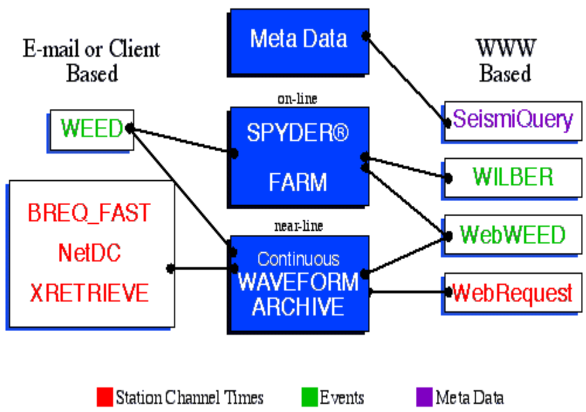
Figure 3. The Rate of Archiving Data at the DMC



During 2001, DMS subcontracts include several quality control centers, one at the University of Washington and one at Harvard University, as well as a Data Center in Moscow, a software development program at the University of South Carolina, and network support for two central Asian broadband networks. The IRIS DMS also provides a low level of support for a Data Center in Beijing. Work statements that summarize these activities are given at the end of this report.

Data Request Tools: The ultimate goal of the DMS is to provide the scientific community with seismograms. To this end the DMS has developed a broad suite of data request tools that are summarized in Figure 4. Development of data request tools has largely been the responsibility of the DMC with assistance and guidance of the research seismologists at the University of Washington.

Figure 4: Data request tools



II. DMS Priorities and Resource Allocation

In keeping with the committee charge (see below), the DMSSC strongly feels that the primary DMS mission is accomplished by the Seattle DMC and the Albuquerque IRIS/USGS and San Diego IRIS/IDA DCCs. As PASSCAL data are now accumulating at a faster rate than GSN data, a DCC type function for the PASSCAL program is being worked out between the DMS and PASSCAL. **The operation of the DMC, the two DCC's, and the resources necessary for effective transfer of data from PASSCAL to the DMC take precedence in budgetary considerations.**

The additional large items in the budget are subawards, generally to Universities, to provide services that make the DMC and the two DCC's more effective. It is the view of the DMSSC that these subawards should be viewed as temporary arrangements subject to periodic review, as outlined below, and can be terminated as program emphasis and technological change make the current services unnecessary, or as tasks are completed.

In developing the DMS budget for the COCOM, the primary budgetary items are always the effective operation of the DMC, the two DCCs, and the effective transfer of PASSCAL data to the DMC.

III. DMSSC Oversight Role of DMS Operations/Charge to the IRIS Data Management System Standing Committee

The DMS Standing Committee operates in an oversight capacity to ensure that the IRIS Data Management System (DMS) is effective in archiving and making available data from the GSN and PASSCAL programs, as well as other data. The DMS standing committee charge is to:

1. Develop and maintain policies that ensure that data flow into the DMC in a timely fashion from the Data Collection Centers (in the case of GSN data) and from individual

investigators (in the case of PASSCAL data). These duties will be closely coordinated with the GSN and PASSCAL Standing Committees.

2. Ensure completeness of the data archive. Develop policies for the archival of non- IRIS data, particularly FDSN data needed to ensure the global recording of teleseismic events by broadband seismometers.
3. Develop, set and maintain data quality standards for GSN, PASSCAL, as well as other data stored by the DMC.
4. Ensure that users have easy and rapid access to the data archive.
5. Advise in the development of software tools for the display, management, and processing of seismic data by users.
6. Provide oversight for the operation of the IRIS DMC, the IRIS /IDA DCC, the IRIS/USGS DCC, and other DMS components as needed.
7. Develop new initiatives to enhance the effectiveness of the DMS.
8. Advise the Program Manager, the IRIS EXCOM, and the IRIS President on program planning and the yearly budgets.

Related tasks of the DMS Standing Committee are

1. To educate members of the IRIS community about the operation of IRIS as a consortium through active participation in the DMSSC.
2. To ensure a high level of IRIS community input to the decision making process within the DMS.
3. To prioritize development efforts and allocation of resources within the DMS
4. To review the work of the various components and subawardees of the DMS and provide feedback on their appropriateness.
5. To maintain a strategic plan for the activities of the DMS.

Through consultation with the president of IRIS, the executive committee, the IRIS program managers, and the IRIS community, the DMSSC will identify and prioritize current and future data management related needs of the IRIS community. The DMSSC will work with the DMS program manager and the scientific community to assess the effectiveness and efficiency of ongoing DMS operations and to define future DMS hardware and software requirements and development goals needed to meet current and anticipated data management needs. The DMSSC should also provide an assessment of the DMS on regular basis using input from the IRIS community.

In particular, to aid the committee in fulfilling its responsibilities we recommend that:

1. In consultation with the IRIS president, executive committee, DMS program manager, and IRIS community, the DMSSC will write a brief report at least once every two years summarizing the DMS strategic plan and ongoing work to address specific goals. The DMSSC will discuss and re-evaluate the strategic plan annually.
2. In order to evaluate the effectiveness and efficiency of DMS operations and sub-awards the DMS program manager and sub-award managers should provide an annual brief written overview of ongoing and future operations at the DMSSC meetings including:
    a. A list of major tasks with a summary of its importance to the DMS and the IRIS community.

b. A cost breakdown for each task in terms of manpower, capital equipment, and overhead.
c. The DMSSC will provide a brief written evaluation for each component/subawardee including, if appropriate, suggestions for changes or improvements. These written evaluations will be available to the components/subawardees.

3. At each DMSSC meeting, the DMS program manager or appropriate staff will present a brief summary on new initiatives. The DMSSC will provide a brief written evaluation of each initiative including an assessment of the significance to the IRIS community, the probability for success, and the relative cost benefit in the context of other on-going or proposed projects.

4. The DMSSC will review significant expenditures on a regular basis, identify significant variations from anticipated spending, and re-evaluate current/proposed budget in the context of prior years funding and expenditures.

5. DMSSC budgeting recommendations be based on preservation of primary activities, expansion of services, and evaluation of effectiveness of existing contracts in relation to the needs of new developmental efforts. Subcontract awards should be based on both past performance and future needs.

Committee makeup: It has been customary to draw committee members from the GSN and PASSCAL active and passive seismology groups roughly equally to ensure adequate community input to the decision making process. It also has been customary for the DMSSC to have at least one member from the USGS and one from the Department of Energy National Laboratories to ensure proper liaison with other large seismic data collection and analysis groups in the United States. We have recently added a member representing Marine Science seismology groups who are developing new instrumentation whose data will be archived by the DMS.

IV. DMS Initiatives

The DMSSC recognizes that the bulk of the IRIS budget is used for operations and maintenance of the core IRIS programs, leaving a small fraction of the IRIS budget available for new initiatives. As such all expenditures not related to core activities need to be thoroughly scrutinized, and new initiatives need to be properly researched and vetted before funding.

a). Review of current DMS subawards : As the DMS currently has a number of projects that are not directly part of the core functions provided by the DMC and the two DCC's, we provide a summary of the current committee's view of the ongoing projects.

Non-IRIS Regional Networks

During discussions of IRIS budget, members of DMSSC frequently ask questions about the role of non-IRIS regional networks receiving IRIS support, i.e. networks in foreign countries started with IRIS or NSF (or some other) support, for which IRIS has assumed some financial responsibility. For example, what should the policy be regarding the support of regional networks KZNET and KNET. KZNET is a 5-station broadband network in Kazhakistan (plus 3

IRIS GSN stations), started by Lamont, and currently receiving modest support from the IRIS DMS for data transfer. KNET is a 10-station broadband network in Kirghizistan (plus 1 IRIS GSN station), started as a part of the now defunct Joint Seismic Program, and in the past has received modest support from the IRIS DMS for data transfer.

One of the goals put forward for the DMS in the IRIS 2000 proposal is to tangibly encourage development of regional networks within the US and abroad by providing PDCC and NetDC software, hardware, and, if necessary, direct financial assistance for data transfer to the DMC. NetDC is presently installed at two US and three European data centers and three Asian data centers and is functioning well. The viewpoint of the DMSSC is that **any financial assistance for a regional network should be of limited duration, i.e. 5 years or less, to enable the startup of a particular network and move it to a self-sustaining mode, except in extremely unusual cases.**

This policy is intended to cover the two networks that are currently receiving DMS support for data transfer operation, KNET and KZNET (Support for KZNET is now being phased out).

University of California at San Diego: KNET: General impressions from the community of both seismologists and geodynamicists working in Asia, are that the KNET array is invaluable. At the Spring 2000 AGU meeting in Washington DC, in a special session that focused on the CTBT and related seismological issues, Terry Wallace gave a talk emphasizing the value of KNET in his research. The proximity to Lopnor, Pakistan and India makes KNET extremely valuable for enhancing resolution for discrimination between explosions and earthquakes and for estimating yield of explosions. The inclusion of KNET data enabled a local detection capacity in a 2500 km radius around the KNET area of M=2.0-2.5, which is considerably better than what is obtainable from just the use of other data, such as GSN. Wallace emphasized that the inclusion and availability of data from networks such as KNET will really open up possibilities in future for significantly lowering yield detection capability. The discrimination between explosions and earthquakes will in the future rely on the open data assets, such as provided by KNET.

KNET is viewed as a bargain network, in that it provides 11 stations of real time data for the cost of only one GSN station. The KNET facility is very reliable at 98% data recovery and station uptime. It has also been used as a testbed network for realtime data delivery from a remote site. The data from KNET are used by 10's of seismologists and geodynamicists to study the regional tectonics and structure of an active mountain belt, as well as occupying an important monitoring site for the CTBT.

In summary, continued support for KNET is important and valuable. It appears to be a network that is in very healthy operation and is fulfilling all that it is supposed to do in terms of data flow and data availability. In light of the opinion amongst some that arrays such as KNET are the future of global seismology, support should continue for KNET since it represents a great example of what is possible in the future. Currently KNET will not ask for support from the IRIS DMS starting in July 2002, as funding of the network will be picked up by other sources.

Should KNET be considered an IRIS/GSN network? Given the widely held view that KNET is a valuable seismograph network operating in an important part of the world, should it be considered to be an IRIS-affiliated network? Should IRIS assume a long-term responsibility for DCC functions for KNET, the de facto situation now?

University of Washington: SPYDER® system: There is strong support for maintaining the SPYDER® system until such time as real-time telemetry is in such widespread use as to make it unnecessary. Currently the DMC is in the final stages of moving the SPYDER® system support and operation totally within the IRIS DMC. The University of Washington continues to provide advice and guidance as to how the system should perform.

Phase Picking at ASL, IRIS/IDA, and University of Washington: As a routine part of QC operations a large number of P-wave picks have been made by the two DCC's on GSN data. This has been formalized so that the DCC's are now picking P-wave travel times and reporting them to the NEIC and the ISC. The University of Washington has been examining various aspects of picking GSN data and is embarking on a program to pick PASSCAL BB network data as well as data from all other networks whose data reach the DMC.

Harvard University: Waveform Quality Control: It appears that Harvard's efforts could be redirected along more profitable lines, as current waveform quality products appear to be used rarely by the DCCs. The DMSSC is unclear as to whether the Harvard/DMS contract in part supports the Harvard CMT catalogue, and wonders if the result of reducing the DMS award to Harvard would negatively impact the CMT solution effort. The WQC subcontract will be put up for competitive bidding in the near future.

University of South Carolina Data Handling Infrastructure Development: This has remained one of the more controversial initiatives of the DMS. Originally conceived of as the FISSURES development to provide user data management and analysis tools written in a platform independent object-oriented language, the initiative has been scaled back to provide very generic data exchange and management programs. The development has been heavily leveraged by a grant from the state of South Carolina to USC to develop a state-wide seismograph system for high school education.

Moscow Data Center: If appears to us that this subcontract is largely in place to keep a data link open to Russia. As such it is an important subcontract, but its details appear largely irrelevant. The software efforts at the MDC are now totally directed towards the development of a client-application that will run within the Data Handling Infrastructure as described above.

Future DMS Projects: Two data related problems recently brought to the attention of the DMSSC are:

1) The need for more detailed, and easier to access station histories for GSN stations, and
2) The need for better installation procedures and stricter orientation standards for GSN and PASSCAL sensors, and stricter calibration standards for GSN and PASSCAL instruments.

The former appears to be an extremely important issue that is now under consideration by a subcommittee of the DMSSC with DMC and DCC staff members. The latter issue has been forwarded to the COCOM and GSN committees for consideration.

b). IRIS 2001-2005 DMS Initiatives:

The following was written as background for the IRIS 2000 proposal and represents the DMS program manager's perspective on development efforts underway and needed in the next 5 years.

Robust Real Time Data Delivery
At the present time the DMS supports delivery of seismic data in real time or near real time. The first efforts at near real time data delivery began with the UofW GOPHER system, which is now SPYDER®. The IDA group at UCSD developed the NRTS system, that emphasizes the return of seismic data over TCP/IP circuits to the IDA hub, or other hubs such as OME, in near real time. Much of the NRTS system relies upon telephone lines using PPP, but is generic in its approach to using TCP/IP. The ASL group developed the LISS system shortly after the NRTS system, but this system offered links to IRIS/USGS GSN stations in real time. The IRIS DMC began working with the US National Data Center at AFTAC in 1998. This system uses the CD-1 format developed for monitoring the CTBT to transfer data in real time from a variety of AFTAC operated stations and arrays. The ORB technology developed first by the University of Colorado with JSP and IRIS DMS financial support and later by BRTT of Boulder is used to transfer data from the PASSCAL broadband array

Clearly there are a multitude of systems presently being used within IRIS, which from a users perspective are rather diverse. Most of the systems have limitations affecting general use.

A DMS priority is to launch a project to provide access to as many real time data sources as conceivable. Features this system will include are:

- Multiple connection methods such as CD-1, CD-2, LISS, NRTS, autoDRM will be hidden from the end users.
- Data will be buffered on high-speed disk for a few months providing fast Internet access.
- Data will ultimately be archived as GSN and PASSCAL data are today
- A uniform connection method will be developed to the on-line data buffers
- The system will support a playback mechanism whereby an earthquake can be played back over the Internet well after it occurs.
- Support of near real time request mechanisms like autoDRM
- A new method of delivering continuous data via the Internet in which the distribution of the software needed at the remote end is transparent, system independent, and based on JAVA.
- The system would allow both continuous connections as well as segmented data transfers.
- Requests for continuous connections will allow connection times at any time in the buffer.
- SPYDER® will generate event volumes for all data sources

The DMC currently has the Buffer of Uniform Data (BUD) system running and offers data from many networks and over 400 stations through it. Developments continue in this area.

Generation of Multiple Network FARM Products: In 1998 nearly half of all data shipments were of GSN data in FARM volumes. The IRIS DMC has data from 38 permanent networks representing 316 network years. The DMC also has 75 PASSCAL Network Years. There are about 250 FARM events per year and each FARM event has 7 files associated with it. This translates to a multiple network total of about 100,000 SEED files which with supporting files totals nearly 700,000 files.

This approach is straightforward and the only enhancements that are needed are

- A way to identify when the time series need updating in a FARM volume either due to newly arriving data or when data have been replaced
- A method to determine when the dataless SEED volume needs to be updated

This project is now nearly complete and data in the New FARM and the New SPYDER® systems provide data and information as specified. The WILBER II Web based interface provides easy access to these data.

Reformatting of Georeferenced Data into Standard Formats: The DMS has developed efficient mechanisms for handling time series data. A great deal of geophysical data comes in the form of raster data, vector data and point data. USArray is an example of a project in which the need to handle non-time series data is necessary.

The most common way to handle georeferenced data is to use a Geographic Information System. The DMS should enhance its expertise to include capabilities in GIS technologies. Specifically the DMS can hire staff or issue a subaward to a university GIS program in order to gain GIS capabilities.

Specific suggested projects are to

- Systematically acquire other geophysical or geological data sets and translate them into ARC/INFO or other appropriate GIS formats.
- Act as a source of knowledge for users wishing to access other types of information, both those held within the DMS and those held by other organizations, a cross institutional archivist
- Keep abreast of developments in 4D GIS systems

Parameterization of Waveforms: The DMC should develop a variety of data mining algorithms that characterize waveforms. In this way, classes of waveforms can be characterized and stored in a DBMS without having to extract waveforms to determine their characteristics. These characteristics can be open-ended but might include such things as

- Peak amplitude per hour,
- Average RMS amplitude per hour,
- Dominant frequency per hour

and other parameters that might be indicative of a signal being present in the data stream.

Phase Picking and Integration with NEIC and ISC efforts: The two DCCs and the UofW have already begun making picks of various subsets of data. We need to formalize and coordinate activities between IRIS and these groups. We should leverage the strength of these facilities in hypocenter determinations and our strength in managing large volumes of waveform data in order to create a better integrated system. Given the likelihood of having the ANSS and USArray on-line in the next five years the phase-picking projects must be closely coordinated with the USGS and NEIC. As the DMS Program Manager is a member of the ANSS Technical Intergration Committee (TIC which is defining the technology including instruments, network architecture, output products, and data distribution for the ANSS, this should be a natural development.

Revise the Data Flow within the IRIS DMS to take advantage of global communication networks: We should begin putting the infrastructure in place to take advantage of changes in global communications that are likely to occur in the next 5 years. Specifically we can expand the current high speed communication network between the DMS nodes. As the quality of waveform data steadily improves, new methods of data flow from stations into the DMS will be adopted. We will strive to have real time data flow into the IRIS archive. Quality control will still be applied to those data by the various DCCs using resources at the archive. As high speed networks improve in reliability and speed, where the waveforms physically reside becomes less relavant. The DCCs can logon to the DMC computers, and access the waveforms or alternatively just access waveform files remotely, since the real time data will be on disk at the DMC.

Software Acquisition or Development for Real Time Data: Software will need to be developed so end users of data can easily handle these data. Many of the concepts of the Virtual Seismic Network software system proposed in the unsuccessful Earth On-Line proposal submitted to the NSF KDI proposal must be implemented. This system will still be used to select the stations to be included in the VSN, handle the data flow within a relational DBMS, display the data, and be expandable by an end user to meet their specific needs.

Platform Independent Software Development: All DMS software should move toward platform independent approaches. The DMS is now rewriting much of its software in JAVA.

Development of an OSN Data Collection Center: Several different initiatives are under way to extend the deployment of seismometers into the oceans. Initiatives such as OBSIP, DEOS and Neptune are real or very possible sources of seismometer data from the oceans. The data flow from some of these initiatives is significant (> 1 Tbyte/year each). The infrastructure at the DMC can be extended to handle the data flow through incremental funding. However the quality control of these data and management of the associated metadata will require the development of an Ocean Seismic Data Collection Center. Given the structure of the Marine Sciences community the OSN DCC functions will likely be distributed among a number of institutions. The Ocean sciences community has agreed to incorporate these DCC functions within their facilities.

V. PASSCAL Data and the DMS (Monica Kohler and Alan Levander)

Unlike data from the GSN, PASSCAL data are supplied directly by the principal investigators conducting field experiments. As a result, PASSCAL data tend to be heterogeneous, have different levels of quality control applied, and are submitted in a variety of formats depending on the experiment type. The bulk of PASSCAL data are submitted in SEED format and are collected during earthquake monitoring experiments of one sort or another. Each such PASSCAL experiment adds another new network to the DMS archive, and although generally short-lived, each network has its own unique set of archival problems. In the past four years the PASSCAL and the DMS have jointly greatly improved the archival process so that, as noted above, PASSCAL data are now being archived at a rate greater than the GSN.

A serious current problem appears to consist of imperfect synchronization between data acquired in the field, and that archived at the DMC, i.e, investigators have a hard time knowing if their data are completely archived at the DMC. The DMC and PASSCAL are developing software for data synchronization between the PASSCAL field computers and the DMC archive to address this problem. PASSCAL has now hired a staff member

Ideally, the processing, quality control, and permanent archiving of PASSCAL data begins in the field during the experiment. For passive source experiments PI's are supplied with field computers on which the PASSCAL database software suite *pdbtools* that is interfaced with the PostgreSQL database management system has been installed. The continuous raw waveform data (e.g., Reftek format) are downloaded during the experiment and processed using the database software. Basic processing includes making time corrections to the trace headers and is carried out as new raw data are continuously downloaded replacing previous batches. The primary function of *pdbtools* is to convert the raw data into data-only miniSEED and data-less miniSEED volumes for archival at the DMC. Low-level quality control is occasionally carried out on a subset of the data at the PASSCAL IC before they are sent to the DMC but most traces are not checked before being sent. The miniSEED data can be converted to a variety of formats including SAC trace files that are categorized by event directories using the DMC software *rdseed*. Requests for waveform data can be made via the DMC WWW page using the data request software *weed* or *SeismiQuery*. These tools automatically generate files in a format suitable for automatic data retrieval at the DMC, and the requested data subset can be sent to a user via tape, ftp, or email. These datasets are labeled "Temporary Network" data sets by the DMC.

Due to a variety of firmware problems, the miniSEED conversion by *pdbtools* and thus the final DMC archiving is not as streamlined as it could be. As a result data processing rarely takes place entirely from the field. Most often it is handled by PI lab staff and PASSCAL staff after the experiment is over, sometimes resulting in DMC archiving delays ranging from months to years. The delays are thought to be the source of somewhat uneven data flow from PASSCAL to the DMC in spite of the fact that almost all seismometers are recording data during experiments almost all year round. Part of the solution to the data flow problem could be found by increasing the level of database management assistance provided to PI's and their lab staff, and part of the solution requires rewriting the database software to make it easier to use. Neither of these solutions is necessarily a DMS task and may require more efficient communication with PASSCAL staff. Alternately the database problem could be jointly addressed by PASSCAL and DMS software engineers, requiring a greater level of staff coordination between the two programs than currently exists.

A large number of PASSCAL experiments are active source in which the final format is something other than miniSEED (usually SEG-Y). PASSCAL software ref2segy is used to convert data from the RefTeks to segy format. These non-miniSEED data that are sent to the DMC are labeled “Assembled” datasets and are often archived directly from PIs’ data storage tapes. Although it has been a long-term goal of the PASSCAL IC and DMC to archive all data from active-source experiments, not all of those data sets are completely sent to the DMC. Datasets from combined land-marine experiments and high resolution experiments are notable problems.

A number of experiments involve marine reflection and OBS acquisition offshore and PASSCAL acquisition onshore. The marine reflection data are not subject to the IRIS archival policy; new OBS data acquired with the OBSIP instruments will be archived at the DMS. P.I.s need to be encouraged to archive all MCS data with the DMS to ensure a complete experiment archive at one location.

A second archival problem at present is the lack of archival of many of the dataets acquired with the IRIS Geometrics multichannel seismographs. Except for data acquired as part of class exercises, Geometrics data are subject to the same archival policy as all other PASSCAL data. In response to a letter from the DMS SC and PASSCAL SC chairs data from the Geometrics units are now beginning to be archived with the DMS.

Lastly the data reduction software for the new Texan RefTeks is awkward and in need of rewriting. At present data reduction is an extraordinarily clumsy procedure requiring unnecessary duplication of the datasets, resulting in a waste of P.I. field time and disk space.

VI. IRIS DMC Audit: Tim Ahern, Program Manager

This is a summary of the infrastructure in place at the IRIS DMC in Seattle and is provided as input for the self-study that is being conducted by the DMSSC.

Physical Location:
The IRIS DMC occupies leased space in a building owned by the Wells Fargo Bank. At the beginning of 2001, this office space consisted of 4,190 square feet for which the DMC pays $4,300 per month. This space will be increased in 2002 to roughly 5,390 square feet and the rent will increase to $5,928 per month. This is an extremely attractive rate. The new space will have offices for up to 20 staff and visitors, most with individual offices. It has a conference room that can comfortably seat 20 people and has been used for meetings as large as 30 people. It will also have one large work/project area. There is a small computer room that houses the main DMC servers and terminus for the Frame Relay data communication circuits. We anticipate building a new computer room in the newly acquired space. There are two storage rooms and one small kitchen. Additionally the DMC occupies about 350 square feet of space in the University of Washington's Computer and Communications Division that is located in the building immediately east of the IRIS DMC. This space is secure, fire protected and has diesel backup UPS power. DMC connectivity to the Internet is provided through the C&C system. The DMC pays the University of Washington $2,856 per month for occupying this space as well as for the Internet connection as well as many other computing services. The DMC houses three major subsystems within the space at the University of Washington, the main 360 terabyte mass storage system, 1.0 terabyte of front end RAID, and a 3.5 terabyte DLT based tape library used to create the offsite tape copies.

One emerging hardware priority is to provide hardening for 100% uptime for real time data reception/distribution by providing redundant RAID disk, and separate processing systems, essentially a tandem computer system.

Computing Hardware Infrastructure:
Primary Servers
The DMC operates the following servers

UW Computer Room

- **JBOD - Enterprise 3000**
  Function: This system controls the front end RAID for the Mass Storage System. It allows us to buffer data for a few months while it is repaired, and made complete. It then moves from here to the tape based mass storage system in multiple sort orders. When data for a request come from this RAID system, this file server performs the time slicing of the data.
  - Executing Software - Veritas Volume Manager and Veritas File System and Time-slice
  - Disk - 1650 gigabyte of disk that is being phased out as well as 1 terabyte of newly acquired Hitachi RAID.
  - Processors 2 x336 MHz
  - Memory 512 Mb

- **SILO - Enterprise 4000**

Function: This system controls the StorageTek Powderhorn. 9940 and Timberwolf/DLT7000 tape based mass storage systems.
Runs SAM-FS (or ASM as StorageTek calls it), the software that controls the 360 terabyte mass storage system and the 3.5 terabyte DLT library.
When data for a request come from this RAID system, this file server performs the time slicing of the data.

- Executing Software - SAM-FS and time-slice
- Disk - 414 gigabyte
- Processors 4 x250 MHz
- Memory 1.5 gigabyte

Mass Storage Systems

- **Tape Based**
  - Powderhorn Tape Library
    - A 6000 slot tape robot
      - 350 to 450 tape exchanges per hour
    - 9 T9940 StorageTek drives
    - ASM (SAM-FS) control software
  - 9714 Timberwolf Tape Library
    - A 100 slot tape robot
    - 2 DLT 7000 drives
    - ASM (SAM-FS) control software
  - ACSLS on a Sparc 5. The software system that controls both robots.
- **Disk Based RAID**
  - On SILO
    - Hitachi RAID system
      - 1 terabyte active
      - 7.5 terabyte expansion capacity
    - StorageTek (Clarion)
      - 1 terabyte
      - being phased out due to difficulty to repair
    - COMPAQ Storage Works
      - 500 gigabytes
      - used as scratch work space

DMC Office Space

- SOB - Enterprise 4000
  Function: An Enterprise 4000 that serves as both the Oracle Database server as well as the machine that builds all of the SEED volumes leaving the DMC.
  - Executing Software - Oracle and DMC SEED generation software
  - Disk - 438 gigabytes
  - Processors 6 x 336 MHz
  - Memory - 1.5 gigabyte

- ARCHIE - Enterprise 4500
  Function: Main data archiving system. Almost all of the actual waveforms entering the DMC eventually come through ARCHIE.
  - Executing Software - All DMC produced and maintained archiving code

- Disk 72 gigabyte
- Processors 2 - 400 MHz
- Memory 1.3 gigabytes

- DMC - Enterprise 4000
  Function: User Interface Computer. This executes most front-end processes that users see at the DMC. It is the WWW server, and is tightly integrated with SOB so that SeismiQuery can return information from Oracle. It is also the FTP server. It holds all of the FARM data.
  - Executing Software - Veritas File System and Volume Manager. Apache WWW Server software. Generic Mapping Tool. Lots of DMC developed and maintained software.
  - Disk 1 terabyte FARM and SPYDER
    162 Gigabyte Work space and ftp space
  - Processors 6 x 250 MHz
  - Memory 1.5 gigabyte

- NAFS - Network Appliance F720
  Function: This is the Network Attached File Server that holds all the user partitions at the DMC. This allows the DMC to maintain function even if significant servers fail.
  - Disk - 32 Gigabyte

- BUD - Enterprise 4000
  Function: The BUD processor handles all of the real time data entering the DMC. It provides the computational power for all of the data reception as well as all of the data reporting and dat aexporting capabilities.
  Executing Software – IRIS DMC developed software for the BUD system, Antelope software from BRTT, ew2mseed from ISTI, Wiggles Wavefrom display applet, and a variety of other DMC developed softw3are
  - Disk 1 terabyte RAID system from LSI
  - Processors 8 x 250 MHz
  - Memory 2.0 gigabyte

- NAFS - Network Appliance F720
  Function: This is the Network Attached File Server that holds all the user partitions at the DMC. This allows the DMC to maintain function even if significant servers fail.
- Disk - 32 Gigabyte

Real Time Data Systems

Function: The DMC is receiving significant amounts of data in real time. We receive data from other sources electronically, even though it might not be in real time. We have several systems involved in this transmission scheme. Lots of history here. All of these systems are workstation class computers, not servers.

- **IMS** Data **-** We have an Ultra 1 at Patrick AFB running AlphaForward Software from SAIC. We have an Ultra 2 in Seattle receiving this data, this system has 100 gigabytes of disk for buffering. This system is running the Antelope Extreme ORB software.

- **GORE.** This machine is an Ultra 1 that runs the Antelope ORB. It receives data from the PASSCAL Broadband array, ANZA, Nevada Regional Network. It hands data off to a cross-mounted file system on BOB for DMC archiving.

- **BOB** (Big on-line Buffer). This is an Ultra 2 with 2 x 200 MHz processors. It receives data from ASL by ftp, from IDA by ftp and a variety of other data sources such as H2O, Taiwan, Netherlands, etc.

- **SPYDER®.** This is an Ultra 1 that runs the UW developed and maintained SPYDER® and BADGER systems. All of the SPYDER® data at the DMC ultimately comes from this machine.

- **B**UD. See a descripton of this in the server section above.

Visiting Scientist Systems
The IRIS DMS has 4 systems that are available for use by IRIS seismologists. Sometimes these researchers come to the IRIS DMC to use the systems, sometimes they use them over the Internet, and sometimes the systems are sent to the researchers. The 4 systems reside at the DMC when not in use.

System	Computer	Disk Size	Tape	Software
2-D Omega Processing System	Ultra 10	18 Gigabyte disk	DLT 4000	OMEGA 2D
2-D Omega Processing System	Ultra 10	18 Gigabyte disk	DLT 4000	OMEGA 2D
Tape Copy System	Sparc 5		2- 10 Exabyte Stackers 2- DLT 4000 2- 6 4mm Dat Stackers	Tape Robot Control
3D Omega Processing System	Enterprise 3000	100 Gigabyte	DLT 4000	OMEGA 3D MATLAB

Staff and Visitor Workstations
As workstations are replaced from operational use, they often rotate to the desks of DMC staff. At the present time we have the following inventory of SUN, PC and Mac desktop systems that reside in the offices of DMC staff.

Number	Type of Desktop Systems
6	Macintosh desktop computers
2	Macintosh Portable Computers
1	COMPAQ desktop computer
1	COMPAQ portable computer
2	SUN Sparc 5
2	SUN Sparc 20
4	SUN Ultra 1
1	SUN Ultra 2

6	SUN Ultra 5
1	SUN Ultra 10

Office Equipment
The DMC has the following major office equipment.

2	Canon Photocopiers
2	color Tektronix Laser Printers
4	Black and White Laser Printers (Apple/Lexmark)
2	Scanners, (1 Epson, 1 HP)
1	FAX machine
1	Computer Projection System
1	ESI Phone system

Communications Infrastructure:
The University of Washington participates in the Amarillo system as well as the Internet 2. As such the input/output rate into the computer room where our primary mass storage system exists is greater than 1 gigabit/second, more than 650 T1 circuits. The DMC is connected to the Internet through a single 100 base-T circuit but can easily be expanded when needed.

The DMC offices are connected to the UW computer space with 12 fiber optic cables. The DMC has an Extreme Networks Summit Switch that supports both gigabit Ethernet, 100 Base-T and 10 Base-T. As such all server-server interactions are at gigabit Ethernet speeds and most workstation connections are at 100 base-T

The DMC has 11 standard telephone lines. 4 are used for normal voice lines, 1 for FAX, 2 for dial-in modem access, 1 for alarm signal dial out, 2 for SPYDER® dial out, 1 for computer paging.

Software Infrastructure:
The major commercial software that we use at the DMC includes

- Oracle, a commercial database management system that is at the heart of most of what we do at the DMC. We are considering installing an additional Oracle system to provide complete DBMS redundancy.
- SamFS, software that controls the large tape mass storage systems
- Veritas
  - File System - allows us to correctly handle terabyte size file systems
  - Volume Manager - allows for software RAID control of some disk subsystems
- MATLAB, used by Corel, to generate QC plots in the FARM and SPYDER® volumes
- Apache WWW server software
- Omega, 2D and 3D, software for researchers to use for processing reflection/refraction data
- Antelope Extreme - Currently being used to receive IMS data from AFTAC.

The heart of the DMC software is now centered upon Oracle. The applications that process information coming into the DMC in SEED format or in various event catalog formats are written and maintained by IRIS Staff at the IRIS DMC. The DMC supports a few hundred different applications and software tools. The most significant utilities are

- Operations
  Archiving
    Sow - archives miniSEED data into Oracle
    Seedpack
    Seedunpack
    Rd - read dataless, puts metadata from SEED headers into database
- Request Processing
    Time_window - time slicing to a data record boundary
    Harvest - pulls metadata form Oracle and produces POD directory
    POD - outputs full SEED volume from output of time-window and harvest
    Verseed - checks validity of output SEED volumes
- User Request Tools
    WWW Tools
        SeismiQuery - accesses metadata from Oracle
        WILBER - extracts SEED data from on-liine SPYDER® and FARM volumes
    Email based tools
        Breqfast -specification by station channel time windows
        NetDC - distributed data center software that seamlessly links multiple data centers
    Client Server Tools
        WEED - requests data from SPYDER®, FARM or the archive based upon complex station, event, or station-event parameters
        XRETRIEVE - specification of station channel time windows
    Other Tools
        CROP - bulletin board extraction of event data from FARM
        SOD- standing order for data based upon station channel characteristics
        Assembled data products are requested via email
    Real Time Data Transfer
        LISS – the BUD system has USGS developed LISS protocol software installed
- Supported Distributed Software
    Rdseed - the primary tool to convert SEED data into analysis formats
    Verseed - verification of the validity of SEED volumes
    Evalresp - evaluation of SEED responses
    RELISH - a TCL/TK tool for displaying instrument responses
    NetDC - a distributed data center application
    PDCC - Portable Data Collection Center, software that helps a data center manage their station histories and
- Limited Distribution
    We have distributed and/or installed the following as well
    - WILBER

- SeismiQuery
- NetDC

Personnel
As of 1/1/2002 the IRIS DMC has a staff of 14 professionals. 13.5 of these are funded form the DMS budget and 0.5 is funded from the E&O program. The following summarizes their functions.

Administration and Coordination

- IRIS DMS Program Manager - management of the entire IRIS DMS, including the DMC, the DCC's and supervision of subaward activities. (Tim Ahern)
- Office Manager - responsible for operation of the IRIS DMC office, organization of local meetings, provides some support for the DMS function as well. (Leanne Beach)
- Systems Administrator, overall administrator for all systems software, hardware and LAN (Rick Braman).
- Information Services Coordinator (Webmaster), primary function is maintenance and development of WWW. Also involved in user help, development of new tools, and some data product generation. (Deborah Barnes)

Software Engineering Group (6 software engineers)

There are six software engineers at the DMC. While all of them are available to solve any problem within our entire, complex, system, each of them does tend to focus their efforts in certain areas. These areas are identified below

- 1 database developer/adminstrator (Sue Schoch)
- 1 external applications engineer, (NetDC, PDCC, WILBER, etc) Rob Casey
- 1 internal utilities, POD, rdseed, verseed, Java rdseed, Java pod, etc. (Chris Laughbon)
- 1 real time data ingestion/distribution developer (Sandy Stromme)
- 1 real time quality control developer (Bruce Weertman)
- 1 E&O specialist working on providing DMC information to the K-16 and general public (Russ Welti)

Operations Group

The operations consists of four people. There is one Director of Operations (Rick Benson) and three Data Control Technicians (Anh Ngo, Stacy Fournier, and Mary Edmund). This group is responsible for the archiving of all data reaching the DMC via electronic or tape. The operations group also services all requests for data from the research community. Members of the group are all proficient in at least shell programming and one member of the operations group is programming in Java. Programming efforts within the Operations group is always focussed on operational problems. The operations group generates most of the statistics related to data usage.

IRIS DMC Functions

Data archiving

One of the primary functions of the IRIS DMC is the archiving of data from a multitude of sources. The DMC archives four copies of all data it receives. Two copies are placed in a time sorted order in the tape based mass storage system, two copies of the waveform files are stored in the same mass storage system in a station sort order and a fifth copy of the time sorted data is placed on a DLT tape library. As DLT tapes fill, data are sent to UCAR, the University Consortium for Atmospheric Research, in Boulder, Colorado, for deep archival and offsite storage.

During 2001 we received data from 28 different permanent networks and 16 temporary networks. The DMC has been very successful at acquiring data from multiple networks and temporary experiments. We have data from 48 different permanent networks, from 24 different data centers. A total of 1216 permanent stations have their data available at the DMC. Additionally, as of the end of 2000, we have data from 73 portable experiments (mostly PASSCAL) with 1948 stations available in SEED format.

We are currently archiving data at a rate of about 5.5 terabytes per year, with PASSCAL data being the largest contributor. PASSCAL data is arriving at a rate of about 2.1 terabytes per year, GSN data at about 1.4 terabytes per year, FDSN 1.0 terabytes per year and regional network data at about another 0.8 terabyte per year.

At the present time the DMC has 22 terabytes of dual sort order data (time and station). The DMC has a combined volume of 7.5 terabytes of GSN data, 2.2 terabytes of FDSN data, 1.3 terabytes of data from networks operating in the Former Soviet Union, 2.7 terabytes of data from US regional networks, and 8.0 terabytes of data from PASSCAL. All of this data is available with identical access tools and in a unified format (SEED).

The DMC has the largest collection of seismic data of this type in the world. The earliest data we have are from 1970 and the most recent data we have are data recorded now.

Data Distribution

The DMC has developed easy to use and powerful methods of accessing data. We have three primary sources from which users access data;1) collections of assembled data in SEGY format, 2) on-line event based holdings in the FARM and SPYDER® area and 3) the large archive of continuous data.

Assembled data are infrequently requested from the DMC and the request tools are limited to fairly simple WWW or email based request mechanisms. The result of this is for tape images to be copied to magnetic tapes of the user's choice and sent to the user.

Data from the FARM and SPYDER® can be accessed with only a few minutes delay. Tools to access the FARM and SPYDER® are WEED, WILBER, ftp or CROP. WILBER and CROP allow subsetting of a data volume so that users can reduce the volume size and receive only the data of interest.

We believe one of our great successes is the number of data requests that we service every year. From an original estimate of 200 requests per year, our request load has risen to a projected number of more than 60,000 in 2001. We serviced about 27,000 requests from the primary archive and 35,000 requests for data from the FARM/SPYDER® on-line event volumes.

Software Development

We are currently developing the following new software

- Data Handling software for end users in conjunction with University of South Carolina
- New FARM building software
- Sticky data request tools
- Real time data handling software for ingestion at the DMC (Interface software to Antelope, Earthworm, LISS, NRTS, etc)
- Java versions of
  - Steim compression software
  - SEED reader (jsr)
  - POD (SEED writer)
  - Evalresp (contracted to ISTI)
  - Java Quick Look in conjunction with PASSCAL

Software Maintenance

- All DMC data archiving software (hundreds of programs)
- All DMC request processing applications (hundreds of programs/scripts)
- miniSEED data handling software (packing, editing, unpacking, etc)
- PDCC Toolkit
- NetDC
- EVALRESP maintenance is now being done by contract to ISTI
- Rdseed -
- Oracle - support for experiment information (netpeople) and support for timing corrections
- SeismiQuery, continued development

Publications

- We have scanned all PASSCAL Data Reports, IRIS DMS Shortcourse, and other documentation to place on line in PDF format.
- FDSN Station Book
  - Originally in paper
  - Electronic PDF and HTML versions are available
  - On CDROM
- FDSN SEED manual
  - Originally in paper
  - Electronic PDF and HTML versions are available
  - On CDROM
- Production of DMS Electronic Newsletter
  http://www.iris.washington.edu/newsletter/about.htm

VII. DMS Subawardee Work Statements

The following request for information was sent to all DMS subawardees on March 8, 2000. A similar request was repeated for the 2001 subawards in fall 2000. The DMS subawardee responses were varied in detail but generally provided the necessary background information. We have included the response to the request for 2000 subawards; the 2001 subaward requests are included in an Appendix.

To: IRIS Data Management System Subawardees

Subject: Summary of Tasks and Associated Costs for Your DMS Activity

At the recent DMSSC meeting, the Standing Committee felt that it could more effectively meet its oversight responsibility by receiving additional information from each node of the IRIS DMS that receives a subaward. For this reason I am asking you to prepare a 1 to 2-page summary that contains the following information:

- A Statement of Work with each task you perform for the DMS identified
- Identify how each task fits into the larger DMS mission.
- For each task, indicate the approximate cost and number of FTEs required.
- Indicate how you feel this function might be improved. Identify new or additional tasks that you think your facility could perform and provide an estimate of how much that might cost in terms of FTEs and level of financial support. Indicate what tasks you are currently performing that could require reduced effort, could be transferred to DMC personnel, or could be phased out entirely.

I do not think that it will take you long to prepare these summaries. If possible please provide this summary to me by March 23, 2000, please send me an email (tim@iris.washington.edu) indicating when you think you can provide it.

In the future, I would like to have this 1-2 page summary every year before the end of January. The DMC staff will remind you of this requirement, but please be proactive and plan to provide this information each January. This will help the DMSSC and DMS personnel in our planning process and eliminate the timing problem related to when your proposals are submitted and when the DMSSC budget meeting occurs each spring.

Thank you for your assistance

Sincerely,

Tim Ahern, IRIS DMS Program Manager

NOTE: The following are the 2000-2001 Budget Requests.

1) University of California San Diego: IRIS/IDA Data Collection Center: (Pete Davis)

Subaward size: $636,053

Summary of Tasks

The IRIS/IDA Data Collection Center is a facility funded through the IRIS Data Management System to process data from the IDA portion of the IRIS Global Seismic Network and make them available to the research community at large. To this end, IDA DCC personnel routinely perform tasks that include, but are not limited to, maintaining contact with, and reviewing instructions for, station operators; documenting and reporting data problems; maintaining records of instrument response and equipment modifications; receiving, logging, and unpacking data; performing quality control and reviewing of that data to assess station performance; and reformatting and distributing data in a timely manner via the IRIS Data Management System.

The personnel to support this facility may be distributed among the following categories:

DCC Core (2.67 FTE) perform routine processing of all data including quality assurance testing and review to assess station performance; report all problems discovered to IDA engineering for correction and document same for public reference; reformat all data to SEED, and ship data to the IRIS DMC; and write or modify software to accomplish the above tasks.

Facilities (0.67 FTE) create computer environment required by DCC Core; perform computer systems administration and network analysis of the IDA LAN and WAN; monitor telecommunications over the IDA WAN, and interface with communications firms as necessary to correct circuit interruptions.

Administration (0.25 FTE) provide administrative support, including budget preparation, purchasing, billing, travel and shipping; process all personnel documentation required by IRIS and UCSD; and maintain financial records required for accounting purposes.

Management (0.5 FTE) oversee scientific, staff, management, and financial issues; provide scientific direction in areas of telecommunications, computing facilities, data acquisition, processing and dissemination; and maintain working relations with research-sponsoring agencies and the university research community.

The personnel required above totals 4.1 FTEs with salary and benefits of $307k. This figure is subject to UCSD overhead at the rate of 51.5%. The following non-personnel costs are associated with the facility:

Computing ($66K, subject to overhead) covers computer hardware replacement, hardware maintenance contracts, networking devices and connectivity, data archive, and software licenses. The costs are proportioned among four IGPP PIs based upon personnel and usage on a recharge basis.

Telecommunications ($45k, not subject to overhead) costs represent charges levied by the phone company to dial internationally ($33k) and a consultant to provide software maintenance support for our telecommunications software ($12k). The IRIS GSN program covers the costs of telemetry over dedicated circuits. Of the FTEs mentioned under Facilities above, the DCC uses 0.3 FTE to monitor telemetry circuits and station performance remotely.

In the near term, there is very little that can be done to reduce the level of effort required to accomplish the above tasks. In the long term, if the IRIS GSN can improve connectivity so that all data can be telemetered to the DCC, then the positions of the students who "spin tapes" and run initial processing software (0.67 FTE) could be eliminated. The capital costs to the GSN to accomplish this would be substantial but having a uniform communications plan for the GSN would reduce telecommunications costs to IRIS overall.

2) U.S. Geological Survey: USGS/ASL Data Collection Center: (Harold Bolton)

Subaward size: $59,700

The Data Collection Center of the USGS Albuquerque Seismological Laboratory is a data gathering facility with diverse responsibilities. A primary obligation of the ASL/DCC is a cooperative agreement between IRIS and the USGS to provide high quality data from the Global Seismic Network to the IRIS/DMC. IRIS, through the DMC, purchases and maintains most of the computer systems used in the DCC functions at ASL. The USGS covers the majority of the costs such as salaries, travel and other items.

The primary processes needed to provide data to the IRIS/DMC involve quality control, data volume production and delivery, meta-data maintenance, software development, computing facilities, and management. These functions as implemented at the ASL/DCC are summarized below.

Quality Control: Currently QC at the ASL/DCC is mostly occupied in checking timing, sensitivities and the general quality of tape-based data. These mostly interactive processes will take on a different character as more QC functions are converted to automatic analysis from the nearly live LISS data. There will always be significant interaction between DCC QC personnel and the GSN field engineers to maintain efficient problem solving. (2.5 ftes)

Data Volume Production and Delivery: Recently the DCC has begun to ship all of the GSN data directly to the IRIS/DMC via ftp over the frame relay circuit. While it is no longer necessary to have daily packages overnighted to the DMC, there are still significant daily production duties in loading station tapes, running the routine staging programs, archiving and participating in various ongoing transcription processes. (1 fte)

Meta-data Maintenance: The database of site data and instrument responses is a continually evolving resource. It is updated for newly installed stations, for new station configurations and by QC recommendation. (0.25 fte)

Software Development: Any modern data facility is dependent on the on-site software developed. Changing technology and hardware requires regular updating of all software processes. More resources are needed here. (2.25 ftes)

Computing Facilities: The DCC maintains the networks and hardware necessary for data production. Given the current amount of incoming data and a five-year replacement policy and apart from the mass store, 60K per year basically covers our DMC related hardware requirements. (0.5 fte)

Management: The DCC management provides administrative and scientific leadership. It is responsible for all aspects of data production and interaction with the scientific community. (0.75 fte)

We do foresee some reduction in actual physical work as we phase in near-live data production. This long-term process will most probably be offset by an increased load of incoming data. In fact, I feel it will be necessary add to our human and computing resources if anticipated data increases become a fact.

3) University of Washington: Hosting of IRIS DMC: (Steve Malone, Ken Creager)

Subaward size: $164,907 (Alan, should this be reduced to the new level of ~85k)

Task breakdown for 2000-2001 fiscal year

The UW contract for hosting the IRIS DMC supports 2.2 FTEs. The total budget is $206,522

The major tasks performed under this contract are the following:

1) SPYDER Maintenance (0.5 FTE): Proposed routine maintenance is totally based on passed experience. It includes the routine day to day checking of the operation of the SPYDER system. It is designed to be mostly automatic needing little direct intervention (when everything works right). Because of the complexity of the system, both at the data acquisition end and the data distribution end there are many things that can and do go wrong from time to time. Changes and additions to the station access list must be taken care of both at the main DMC hub and also at the remote SPYDER nodes. Problems with individual waveform files arise from time to time which can break more aspects of the system than just the one file. Tracking these down and fixing the individual problem as well as changing code to reduce the chances of that particular problem again takes much of the maintenance time. Other aspects include upgrading software locally and remotely including SPYDER code, telemetry code, and Operating Systems (in a few cases).

2) SPYDER Development (0.6 FTE): Major changes or additions to the SPYDER system are covered under the "development" task. Predicting exactly how much effort will be required for this is not possible but the following is based on passed experience and the probable types of near term development and personnel to work on this task.

Two major developments are planned over the next year. The first is to change the main data format used by SPYDER from SAC to mini-seed (mseed). SAC has been the primary data format used by SPYDER for data acquisition and local storage from the beginning. Since most dataloggers can now provide data in mseed format and this format is much more appropriate for use in the products of SPYDER, we will be converting our procedures to use this format. This conversion will be done in coordination with the likely change at the DMC of changing the main storage format for the FARM/SPYDER database files accessed via WILBER. The second major development is to incorporate data arriving at the IRIS DMC via real-time links into the SPYDER system. This also will take close coordination with the DMC staff.

3) SEED writing of Regional Network Data (0.2 FTE). Triggered waveform data from the PNSN regional network continues to be archived at the IRIS DMC on a quarterly basis. As PNSN recording capabilities change, the generation of these data products changes too. New efforts in this task include getting continuous waveform data from PNSN high quality broadband stations to the IRIS DMC in a smoother manner. We also plan to develop efficient ways of including strong-motion (accelerometer) data for larger earthquakes in our data products to the DMC . We have acted in the past as consultants to other regional networks on problems of converting their data to SEED for DMC submission but only in a few cases when asked. A new task we would propose to do is become more pro-active in helping such networks.

4) Consulting for DMC on data or design issues (0.2 FTE). As in the past we propose to continue consulting with DMC staff on issues of software design, seismological interpretation, and other common issues. This includes design considerations for specific projects, doing independent testing, making suggestions on improvements, bug reporting and diagnosis, and general review of procedures or policies. During this coming year the FARM is being redesigned to include data in addition to the GSN. We will be working closely with the DMC staff in the design as well as parts of the implementation of the new FARM .

5) Waveform picking (0.4 FTE). During the past year the level of effort for travel time picking has been at the 0.4 FTE level. This includes the coding, picking, checking, and comparison of methods. For example, we picked all the P times in one year of FARM data using the raw broadband records and using records filtered to a standard short-period response. We found no significant difference. The system for picking data and checking them is now fairly streamlined so we anticipate more picks this year for the same 0.4 FTE level of effort. We propose to continue picking first arrivals from the FARM data. As the FARM grows to include PASSCAL data and other networks we will pick those as well as time permits. The efficiency can be improved by incorporating an auto picker that is re viewed by an analyst. We propose to evaluate autopickers during the next year.

6) Quality Control/FARM record sections (0.2 FTE). In support of the FARM/SPYDER system we maintain code to make record sections for each event and post them on the web where they can be viewed or downloaded using WILBER. When focal mechanisms are available we also make record sections of just the long- period P and SH waveforms correcting for the predicted travel time, source radiation, geometric spreading, and instrument response so each seismogram should be the same in time, shape, amplitude and polarity. These plots are used routinely for QC purposes. The maintenance and monitoring of this process takes roughly 0.2 FTE per year depending on the problems encountered.

7) Administration (0.1 FTE)

4) Harvard University: Waveform Quality Control Center: (Goran Ekstrom)

Subaward size: $47,500

Tim:

In your March 8, 2000 email you requested information about the activities of the Harvard WQC. This is my belated response to your request. Part of the delay in getting back to you is due to some uncertainty on my part regarding how we might best provide QC on IRIS data, given our other activities. I believe we are proving useful QC under the current arrangement, but I think there are some new functions that we might perform with greater impact than some of the things we have been doing until now.

Responses to your specific requests:

1. *Work Statement.* Our annual renewal proposal (which I still owe you for 2000-2001) lists the specific tasks that we perform. In summary, we attempt to identify data problems in our routine analysis of data from the GSN. We report these problems by filing IRIS Data Problem Reports. We produce a quarterly report that gives statistics of GSN data usage and GSN station utility for the Harvard CMT project, and presents figures of discovered data problems.

2. *Utility of our efforts.* Our WQC analysis provides a regular check on the quantity and quality of GSN data. The Quarterly Reports provide a summary of the recent performance of the GSN. Most of our recent DPRs relate to errors in the SEED header information (erroneous gain or transfer functions, reversed polarities, etc.). These errors clearly affect all other GSN data users and identifying these problems should be a high priority for DMS/GSN. We are becoming less successful at being the first group finding other problems (dead channels, timing problems), primarily due to the good QC provided at the DCCs.

- **Comment:** My own assessment is that some other QC activities that we have been involved in recently have been more useful for IRIS DMS and GSN. For example (1) identification of errors in the dataless SEED volumes for older data; (2) recovery from the Harvard archive of GSN data that were lost at ASL and never archived at the DMC; (3) working directly with the SPYDER personnel to correct problems with SPYDER data; (4) identification of format errors in IRIS SEED writer for combined GSN and PASSCAL data; (5) identification of response description errors in IRIS SEED writer for PASSCAL data; (6) working directly with ASL and IDA personnel to operationally test the LISS and NRTS software at Harvard, and to provide feed-back on problems (bugs?) and advice on how to improve the real time data distribution software; (7) Determination of misalignment of GSN horizontal component seismometers using polarization analysis.

3. *Approximate cost of effort.* At present, our WQC work reflects the following funded effort: 50% of a technician, 25% of a post-doc and 1 month of my time. The IRIS WQC award is what motivates us to continue to analyze all available GSN data (sometimes more than 100 stations) in the calculations of Harvard CMTs -- this, of course, being a prerequisite for being able to evaluate how the whole GSN is performing.

4. *Possible improvements.* I believe it would be useful for DMS/GSN if the activities of Harvard's WQC were re-focused on the following three quality aspects:

1. **Analysis of real-time data and real-time data distribution software.** Over the last month, we have developed the software necessary to robustly retrieve and put into a database all continuous long-period GSN data available in real time via LISS, NRTS and the WWW. We now routinely have more than 50 stations available for the Quick CMTs, excluding a few additional SPYDER stations. By becoming an end-user of the real-time data, we are in a position to help improve these new data distribution mechanisms. Though both NRTS and LISS work quite well, there are several aspects of both systems that need improvement, and we are providing advice as we monitor the performance of the software. We believe that our direct waveform QC should focus on these data, since we might then be able to spot data problems early.
2. **Noise characteristics of GSN stations over time.** We have recently developed some software to calculate signal (noise) levels for the GSN with respect to Peterson's NLNM and display these as a function of time. Data and figures from this analysis could be used to document the performance of, and identify specific problems of, the GSN.
3. **Identification of problems in response data (dataless SEED volumes).** We now have nearly all GSN LP data since 1972 on line and are therefore well set up to systematically review the dataless SEED information from ASL and IDA. In particular, we are in a position to work with ASL and IDA to identify and correct gross (>5 degree) misorientations of sensors (of which there are many). I also believe we would be able to identify many gain errors in a (time-) longitudinal noise study.

I hope this information and my comments are helpful for your review of the Harvard subaward. I believe that, because of the focus of our research program on analysis of large volumes of GSN data, we can offer unique and valuable evaluations of the performance of the GSN. I would welcome a mechanism by which the specific form of this QC was regularly reviewed, so that our efforts could be assured to have the greatest impact. For example, I think it would be very useful to have an annual meeting at the DMC involving the DCCs (including PASSCAL) and the WQCs. I think you have a meeting with the data providers each year, so maybe adding the WQCs and half a day of QC discussion would be a possibility. I believe this would be a useful forum for presenting, reviewing and obtaining feedback on our activities.

5) University of South Carolina and 2AB, Inc: IRIS DATA HANDLING INFRASTRUCTURE DEVELOPMENT (T.J. Owens)

Subaward size: $169,354
Project Tasks, 2000-2001: **Overview**

The IRIS Data Handling Infrastructure (DHI) project is designed as an effort to improve the delivery of DMC products and services to its end-user clientele. These products include customized data requests, SOD requests, FARM data, and real-time data. The DHI project is using state-of-the-art internet-based exchange protocols (CORBA) and object-oriented design to build a distributed system to efficiently deliver data to DMC users.

Project participants include Philip Crotwell at USC and Mitchel Sanders of 2AB, Inc. The mode of operation thus far has revolved around quarterly design sessions in which operation goals are defined. Sanders designs and writes IDL (interface definition language) that forms the framework of the system. Crotwell then implements specific elements of the system (the "business model") in Java. Crotwell and a graduate student programmer then implement code to exercise the system and provide feedback for the next design session. The DHI project benefits greatly from a related project at USC, the South Carolina Earth Physics Project (SCEPP), which is utilizing the DHI framework in a small-scale real-time seismic network and data center. This allows the DHI project to benefit from the services of a full-time programmer for only 0.25FTE.

DHI System Design

The DHI framework is built as a tiered system that allows separation of various system components without duplication of effort. These tiers are primarily conceptual and design entities. They are not rigid in that a level 4 tier can speak directly to a level 1 tier under appropriate circumstances.

The first tier is the "*DataCenter*". These are designed to provide an interface between a data provider to a data manager. A *DataCenter's* sole purpose is to deliver data from a provider to a manager. It provides “read only” access to a data source. Thus, the *DataCenter* will be the point of access to the FARM. Specialized *DataCenters* such as an *EventDataCenter* or a *NetworkDataCenter* will provide similar services to managers needing to access information in the DMS Oracle database. *DataCenters* will have the capability to control, manager, and prioritize their external request loads.

The second tier is the "*Manager*". *Managers* inherit from *DataCenters* in an object-oriented sense and add the functionality of being able to store data. Thus, *Managers* may serve as both a provider and a recipient. In other words, they have both “read” and “write” capabilities. They may receive data from a *DataCenter*, pass it upstream to a *Service*, or store it in a locally useful format. Work to date has been on the *SeismogramManager*, which knows how to retrieve data from a local or remote repository or establish a feed from a real-time data source and store or forward the data to a *Service*.

The third tier is the "*Service*" tier. *Services* implement specific access points for end users (clients). For instance, a user on the lookout for specific types of seismic events would start a *SODServer* locally and configure it for the events of interest and the intended destination of those events (local storage, an automated picker, etc). The *SODServer* would connect and listen to an *EventManager* to learn when events occur, then correlate the event parameters with any pending orders. Events which meet an order's criteria would trigger the appropriate actions, such as connecting to the appropriate *DataCenter* to request seismograms.

A fourth tier of the system exists, which is specific user applications. The DHI Project is not charged with developing any specific research applications. However, some data access applications will be developed to allow users to make data requests. For instance, a general purpose event request GUI allowing users to access DMC data would be developed and made available to the community. Eventually, conversion of current DMC WWW-based utilities to access the DMC archives through the new DHI would be a logical step as well. The data viewer under development to demonstrate the capabilities of the system would continue to evolve as well. These examples, hopefully, will entice users to try the DHI approach. The source code will provide simple examples of accessing data through DHI interfaces. In addition, sets of templates, tutorials, and other documentation to encourage users to tap into DHI Services will become increasingly important as seismologists come to grips with the volume of data that they need to process.

Tasks for 2000-2001: The 1999-2000 funding will result in:
- Design and implementation of the seismogram DataCenter and the Seismogram Manager as well as a simple data viewer suitable for SCEPP/IRIS E&O applications.

In 2000-2001 funding will result in:
- Design and implementation of other *DataCenter* tier elements, specifically the *FARMDataCenter*, the *EventDataCenter* and the *NetworkDataCenter*.
- Design and implementation of the appropriate *Managers* to interact with the DataCenters that are implemented.
- Design and implementation of two Services. By the end of the 2000-2001 funding cycle, we would like to have prototypes of the *FARMService* and the *SODService* that could be distributed to the user community for testing and feedback.

Budget Increases, Budget Cuts and DHI Development Efforts

Continued growth of the DMC archive and the shadow of USArray on the horizon suggest that more efficient and automated data delivery services will be needed. The main limiting factor in DHI development is designer time and programmer time. Budget increases will allow the project to proceed more rapidly. Budget cuts will delay the end products. In the 1999-2000 budget cycle, the funds available to 2AB were only enough to get a fraction of Sander's time over 6-7 months. Sander's involvement is critical since he is an expert in distributed information management with CORBA and now has an understanding of seismological data handling issues. Maintaining his involvement is key to rapid progress. USC support is minimal compared to the return and central to the development of useful products. If more rapid development is desired, increasing Sander's time and adding either another graduate assistant or a full-time programmer to the USC budget would be the best way to increase the rate of development in the coming year. The DHI design allows for the existence of multiple DataCenters and multiple destination services, so there is nothing to prevent more complex Services from being developed to deliver USArray data from the DMC and/or mirror sites to end users.

6) University of California San Diego: Summary of Tasks and Associated Costs for DMS support of KNET (Frank Vernon)

Subaward size: $40,000

- A Statement of Work with each task you perform for the DMS identified
- Identify how each task fits into the larger DMS mission.
- For each task, indicate the approximate cost and number of FTEs required.
- Indicate how you feel this function might be improved. Identify new or additional tasks that you think your facility could perform and provide an estimate of how much that might cost in terms of FTEs and level of financial support. Indicate what tasks you are currently performing that could require reduced effort, could be transferred to DMC personnel, or could be phased out entirely.

Data Acquisition

Task 1: Kyrgyzstan Network Data Preparation

Subtask 1.1: Maintaining Real-Time Data feeds from KNET to the DMC

- Real-time data access is made through an Antelope data source in Kyrgyzstan. Real-time data are copied to IGPP and retransmitted directly to the DMC over the Internet. Data are placed directly into the DMC archive on a daily basis.

- Provides access by the IRIS community to the data from the 10 broadband stations in KNET.

- .25 FTE required

Subtask 1.2: Bulletin preparation of local, regional, and teleseismic events recorded by the Kyrgyzstan network.

- Phase arrival times from the Kyrgyz network are be determined as part of normal network operations. A magnitude scale calibrated to US standards has been developed, and is applied to all local events recorded by the network. Events will be compared with those listed in the PDE to verify detection limitations of this array. Unassociated events will be further processed and categorized into local and regional events. Local event locations will be determined using conventional earthquake location algorithms and regional events locations will be determined from single array location methods. A Kyrgyz network bulletin will be produced listing location estimates for events located in these ways.

- This task provides the quality control assurance for the KNET data, verifying timing quality and data system quality.

- .25 FTE required

Subtask 1.3: Preparation of event parameter database products.

- Phase arrivals and waveforms corresponding to the Kyrgyzstan network bulletin are be made available as part of a comprehensive data set that will contain all necessary information for doing research tasks with the Kyrgyzstan network data. The data will be sent to the IRIS DMC in SEED format for general distribution to the IRIS community.

- Provides access by the IRIS community to the parametric data from the 10 broadband stations in KNET.
- .1 FTE required
- Indicate how you feel this function might be improved. Identify new or additional tasks that you think your facility could perform and provide an estimate of how much that might cost in terms of FTEs and level of financial support. Indicate what tasks you are currently performing that could require reduced effort, could be transferred to DMC personnel, or could be phased out entirely.

There is need for modification or change in this project as long as the real-time data from KNET are being archived in the DMC. There is little prospect for additional stations to be added to KNET and the data collection and quality control have been automated to minimize the operational costs. The only additional tasks which my group could offer is to perform the same quality control steps on other data sets. The cost of this in FTEs is dependent on several factors including number of stations and level of processing and whether there is real-time data. None of these tasks currently being performed should be reduced in effort, transferred to the DMC or phased out.

7) Lamont-Doherty Earth Observatory of Columbia University: KZNET: Integration of Kazakstan Broadband Seismographic Network and Seismic Arrays Operated by LDEO/NNC into IRIS/DMS Program: (Won-Young Kim)

Subaward size: $40,000

SUMMARY

We propose to integrate the seismic facilities established in Kazakstan since the summer 1994 under the auspices of various programs of the IRIS Consortium into the IRIS/DMS. The essence of the integration plan is to maintain the acquisition of high-quality seismic data at quiet sites in Kazakstan and make them available for IRIS and global seismological communities. The integration will promote earthquake studies and efforts to mitigate earthquake hazards in the former Soviet Republics in the Central Asia, and to contribute for the verification of the Comprehensive Test Ban Treaty (CTBT).

The key features of the integration during FY00/01 are:

- data acquisition at five stations of the Kazakstan broadband seismographic network (KZnet) - AKT, CHK, VOS, ZRN and TLG;

- near-real time data acquisition and transfer from IRIS/GSN stations in Kazakstan - BRVK, KURK and MAKZ to IRIS-DMC;

- development of the IRIS/LDEO/NNC Joint Seismic Data Center (JSDC) in Almaty, Kazakstan for a stable, long-term data acquisition and direct network-day SEED volume submission by JSDC to IRIS-DMC.

STATEMENT OF WORK

We will continue to acquire continuous, seismic waveform data from the Kazakstan Broadband Seismographic Network stations AKT, CHK, VOS, ZRN and TLG during July 2000 through June, 2001.

We will quality control the waveform data, generate network-day SEED volume and submit them to IRIS-DMC in timely fashion.

We will coordinate efforts with UCSD/IDA, USGS/ASL and JSDC/NNC under the IRIS/GSN program to implement near real-time data acquisition and transfer to IRIS-DMC from the IRIS/GSN stations - BRVK, KURK and MAKZ.

We will promote NNC, Kazakstan to establish a stable, long-term data acquisition and submission to IRIS-DMC through development of the IRIS/LDEO/NNC Joint Seismic Data Center in Almaty, Kazakstan.

In the following, we will elaborate further on the tasks and associated costs, as well as on a long-term plan for the future of the KZnet under IRIS/DMS program.

1) Development of the IRIS/LDEO/NNC Joint Seismic Data Center (JSDC): In the summer of 1999, IRIS/LDEO/NNC Joint Seismic Data Center was formally established at its new office in Almaty, Kazakstan under a decree by the Ministry of Science & Technology - Academy of Sciences, RK. Currently, JSDC is led by Dr. Natalia Mikhailova and has a staff of five. The JSDC is also functioning as KNDC (Kazakstan National Data Center) for IMS system. It has two SUN workstations, two hard disk drives with a total of 30GB capacity and two DAT tape drives. These were furnished by supports from ISTC and IRIS/JSP program.
The JSDC is working as data collection center of the KZnet and it will be the hub of the satellite telemetry for GSN stations in Kazakstan. Hence, healthy functioning of the JSDC is important to IRIS/DMS and GSN programs in ensuring a stable data collection mechanism for KZnet.

LDEO plans to furnish JSDC with a data collection and distribution system consisting of a SUN workstation (Ultra 10), network access server (router), a hard disk drive (50GB) and UPS during the spring of 2000. This is a one time costs and it will be furnished by LDEO's FY99/00 IRIS/DMS budget. No further cost will be requested to DMS.

2) Data acquisition & submission to DMC: this is the main task and has been performed by staff at LDEO in collaboration with staff of the IGR, NNC since July 1994 under the auspices of IRIS JSP program. Since the summer of 1998, when this activity was transferred to the DMS, LDEO had been aggressively promoting local host institution (NNC, Kazakstan) to acquire the data, perform quality control, generate network-day SEED volumes and submit them to IRIS-DMC.

During the first half of FY00/01, LDEO must work with JSDC staff to establish a new procedure to generate SEED volumes, because four stations in the Borovoye region has been upgraded to the GSN quality instrument (Q680 with STS-2) in the spring of 1999. It will require training of JSDC staff at LDEO as well as a minor programming work.
Currently, JSDC staff is working on generating the network-day SEED volumes for 1999. The first such network-day SEED volume can be submitted to the DMC during the FY99/00.

For FY00/01, LDEO requests a budget of $74K, which includes a total of 3.5 months of salary, support for LDEO staff, one round trip to Kazakstan, subcontract to JSDC/NNC ($25K) and tape and other material ($5.5K). LDEO plans to transfer all of its activity under IRIS/DMS by June 2001 to JSDC/NNC. Hence, starting from July 2001, the local host institution (JSDC/NNC) may need some funding support from DMS. It is anticipated that the reasonable costs may be around $25K/year (without ICI) for DMS program to support the JSDC, since it will require at least two full time JSDC staff to work continuously.

3) Near real-time data transfer from KZnet to USA: Since the spring of 1997, under the ISTC support, LDEO and NNC have been working hard to make the Kazakstan Broadband Seismographic Network stations into permanent GSN quality stations with real time data acquisition & transfer capability. In addition to three GSN stations in Kazakstan, three stations in northern Kazakstan as well as Kurchatov seismic array are being upgraded to GSN quality stations. At the moment (March 20, 2000), satellite telemetry between these sites (including all three GSN stations) and JSDC are under construction and it will be functional at around July 1, 2000. The plan is for the JSDC to start transferring near real-time GSN and other data stream via Internet to the recipients in the USA. For this facility to become functional and robust system, personnel at LDEO, UCSD/IDA, USGS/ASL and JSDC/NNC under the IRIS GSN program must cooperate closely. At the moment, the best scenario is that the facility becomes functional during FY00/01.

4) Future plan: Under a favorable condition, we hope that LDEO's involvement in the IRIS/DMS program for KZnet will be substantially reduced or completely eliminated and that JSDC/NNC should be able to work directly with staff at IRIS/DMC as a data collection center starting from July 1, 2001.

Milestones

May 1, 2000 start submission of KZnet network-day SEED volume for 1999 by JSDC/NNC to IRIS-DMC via LDEO

Jul 1, 2000 near real-time data transfer from JSDC to UCSD/IDA, USGS/ASL and LDEO for IRIS/GSN stations

Jan 1, 2001 direct submission of KZnet network-day SEED volume for 1999 and 2000 by JSDC/NNC to IRIS-DMC

8) Synapse Science Center, Ltd., Institute for Problems of Mechanics (Moscow Data Center)

Subaward size: $25,000

VIII. **Questionnaire results**

The ultimate goal of IRIS is to provide seismograms to seismologists, as a result the usefulness to the research community of DMS request tools and DMS software, station and experiment history information, and help in dealing with data related problems are important measures of consumer satisfaction, and can provide indications of directions for DMS development efforts. A questionnaire designed for the DMS user community was posted on the web in the summer of 2000, accompanied with a Bulkmail request to respond to the questionnaire. Prior to release the questionnaire was tested and tuned with a small group of respondents chosen from the IRIS committees.

Initial response to the questionnaire was mediocre; 45 people responded to our Bulkmail requests. We next targeted a group of 107 users whose opinions we wanted. Thirty-eight people responded to the targeted request, increasing total responses to 83. (For comparison the DMS had data requests from ~1500 unique users in 2000, PASSCAL fielded ~50 experiments in 2000). During discussion of the survey results at the DMSSC meeting in February 2001 we decided to target two other groups who are poorly represented in the first two: graduate students, and primarily PASSCAL users (20 respondents total). The responses to the questionnaire, including short suggestions, are included as Appendix 1. In the following we summarize the results of the questionnaire.

Of the ~95 respondents, 78 are employed by academia, 12 by government, and the 3 by industry. Seventy-five are faculty or professional research staff, 7 are graduate students. The primary data interest of 48 respondents is now GSN data, 13 PASSCAL data, and 14 regional data. For future research 71 respondents intend to use GSN/FDSN data, 62 PASSCAL data, 63 regional data, and 59 USArray data. Fifty-seven respondents have used IRIS data in 10 or fewer publications, 11 have used it in more than 10 publications.

Data requests are made daily by 5 respondents, weekly by 12, monthly by 22, and quarterly by 25, with most respondents (69) making their own data requests. Most respondents (67) had requested data in the last year. The DMS request tools are all well used, with SeismiQuery, Breq_fast, Weed, and Wilbur used by the largest number of scientists. Each respondent gave the range of data they requested. The average min-max varied between 8.3 and 30 Mbytes, and 287 to 472 traces.

Half of the respondents felt that near-real time data would be of value to them, the other half didn't. Common uses cited for near real time data are earthquake and aftershock source parameter estimation (5), quality control (2), Education and Outreach (6), and emergency management (2). Two respondents indicated that they are already using near-real time data. Thirty-two respondents felt that their lab could handle an order of magnitude increase in the data it ingests, 19 felt they couldn't, 18 felt that they could in the near future.

Much of the IRIS written software is used widely as indicated in Table 8.1, particularly rdseed, evalresp, the PASSCAL database, and Weed.

Table 8.1: IRIS Software Use

evalresp	29
relish	1
rdseed	71
verseed	9
pdcc	5
pod	9
passcal database	25
weed	25
xretrieve	17
NA	7

Use of various non-IRIS tools for data handling and analysis is summarized in Table 8.2. The most commonly used data handling/analysis software is home written code (46/54), SAC2000 (36/46), and SAC (29/27). Matlab (16/43), AH (16/10), Antelope (14/9), and the PASSCAL database (14/3) are also well used. Home written software is most frequently the favored software for analysis. SAC2000 is the most favored software for data handling, and is highly favored for analysis.

Table 8.2: Data Handling and Analysis Software

	Data Handing	Most Important	Data Analysis	Most Important
AH	16	1	10	0
Antelope/ORB	14	7	9	4
Datascope	12	4	10	3
Earthworm	11	4	3	0
Geotool	4	1	7	2
Matlab	16	4	43	5
Matseis	1	0	3	0
Omega	1	0	1	0
Passcal DB	14	2	3	0
PDCC	4	1	0	0
Promax	3	1	5	1
SAC	29	5	27	5
SAC2000	36	25	46	22
SPW	1	0	1	0
Home written	46	14	54	30
Other	3	0	3	3
NA	4	14	4	8

The data handling and analysis question indicates that a rather broad spectrum of software is in use by our community. The questionnaire continued by asking what the community would like in customized software by way of data handling/data conditioning capability (Question 15) and what software IRIS could develop (Question 16). The summary of Q15 is given in Table 8.3. More than half of the respondents would like to possess integrated software having basic data handling, trace and event sorting, plotting and visualization, travel time picking, event location, time series filtering, and 3-component rotation capabilities. Individual comments in response to Q16 are given in the appendix, and ranged from 1) no development to 2) large scale development to 3) supporting other groups' development efforts. One comment of interest was that IRIS' software development prevented graduate students from learning programming skills.

Table 8.3: Question 15: If you could order a customized software package to aid you in data handling and data functionality would it include:

	Handling	Analysis
a. Seismic event/trace management	62	37
b. Event/trace sorting	52	40
c. spherical geometry	17	37
d. plotting	39	51
e. travel time picking	32	55
f. band pass filtering	29	60
g. 3-C trace rotation	26	53
h. visualization	31	52
i. Database capability	44	26
j. standard reflection trace processing	8	14
k. standard reflection multichannel processing	5	13
l. interpretation software for reflection data	3	13
m. event location	23	49
n. seismic processing history	23	20
o. integrating other data with seismic data	18	24
p. other	4	9

The respondents program in Fortran (76), C (51), and Matlab (40), with a small number programming in JAVA (17), C++ (12), and other languages (17). The most commonly used operating systems are Unix/Linux (59/32), the Windows systems (59), and Mac (27). The low number of graduate student responses may influence the results towards Fortran and C and away from more modern languages. Given the often-asserted faculty statement that students no longer want to code it is unclear how these numbers will change with a better representation from the graduate student population. We are targeting a group of graduate students to find out.

Question 18 asked about dataset size in terms of number of events, traces/event, samples/trace, and sample rate. The minimum and maximum number of events/dataset, traces/event, and samples/trace are plotted in Figures 8.1, 8.2, and 8.3. Figure 8.1 shows that half of the respondents work with fewer than ~300 events/dataset, and half with more than ~300. Figure 8.2 shows that half of the respondents work with ~150-200 traces/event, and half with more. Figure 8.2 shows that half of the respondents work with ~150-200 traces/event, and half with more. Figure 8.3 shows that half of the respondents work with ~50000 samples/traces, and half with more. (This appears to violate Fowler's law that all seismograms are 2000 samples in length. In fairness to Jim, his rule was developed for active source experiments and that group is poorly represented by this survey, and does tend to work with traces that long).

Using a 4-byte data word this gives an average dataset size of 10 Gbyte. An order of magnitude increase in data would be 100 Gbyte, serious processing and manipulation of which would conceivably tax some seismologists computer systems. The optimism expressed in the answers to Question 11 may be unrealistic. I emailed a few of the more active labs using GSN data and found that their total disk/robotic tape holdings range from 500 to 5000 Gbyte. Asked individually if they could manage and analyze the increase in data expected from USArray, they waffled.

Figure 8.1: Question 18A: Minimum and Maximum Events/Dataset

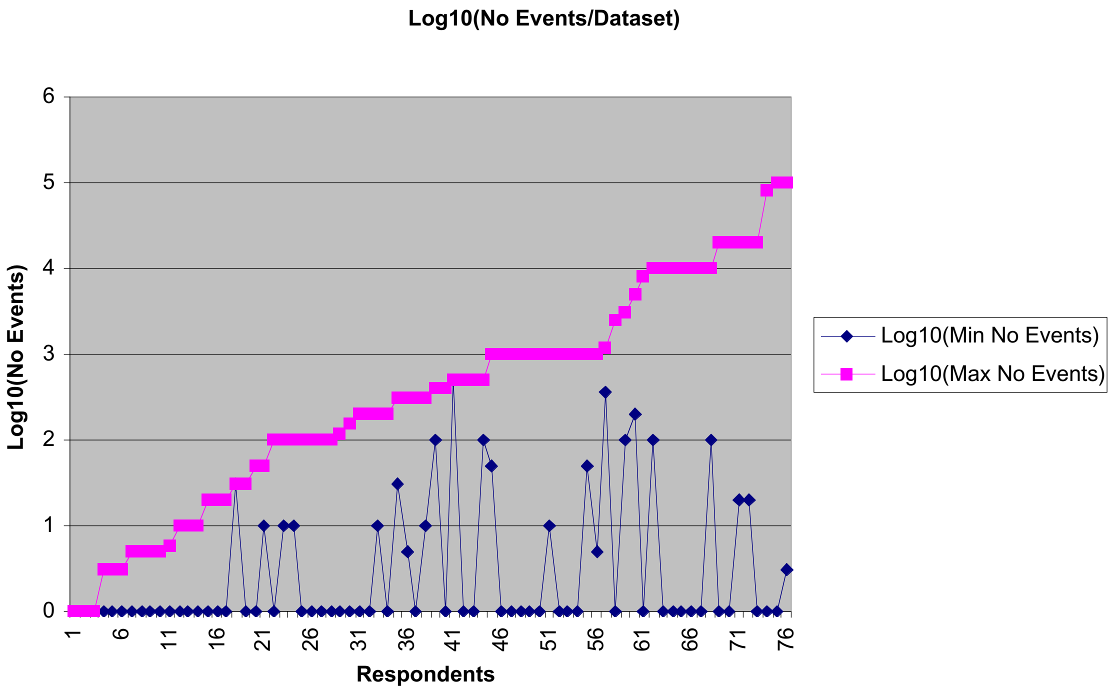


Figure 8.2: Question 18B: Minimum and Maximum Traces/Event

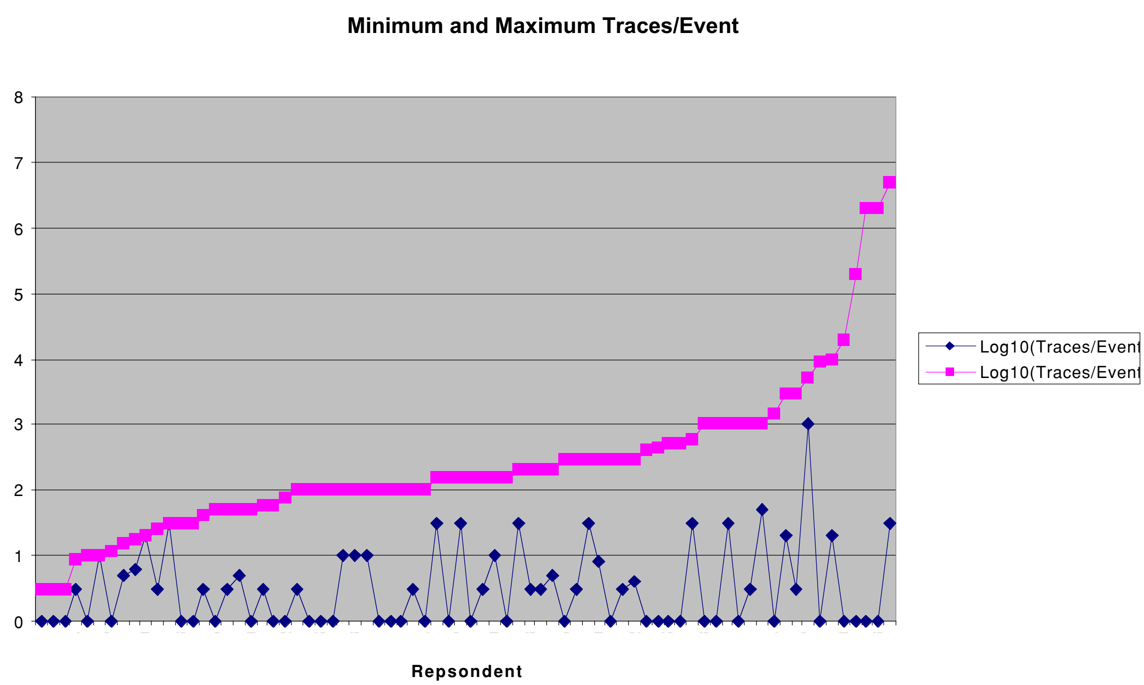
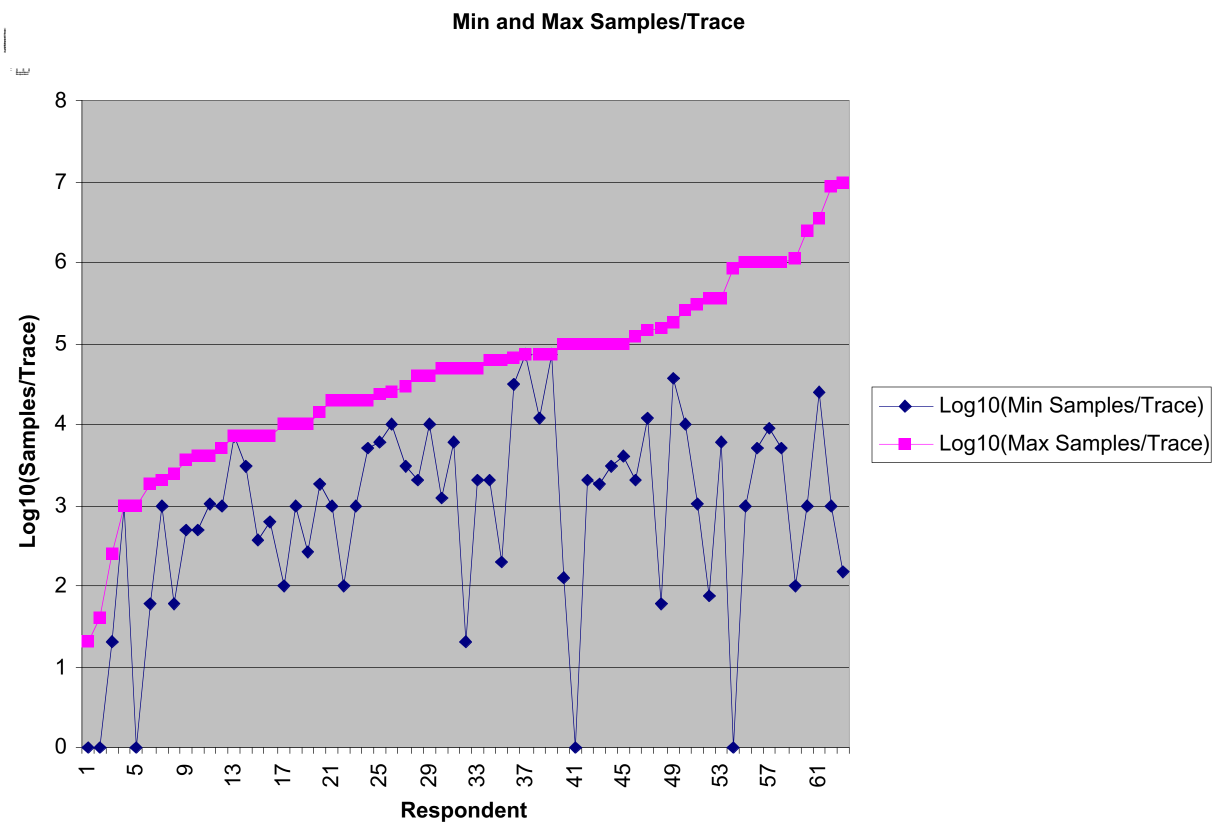


Figure 8.3: Question 18C: Minimum and Maximum Number of Samples/Trace



Summary of Written Response Questions

Question 9. What information/data requests tools do you use and how well do they work for you? We welcome comments on any of the products above, especially if you are not at all satisfied.

A wide variety of DMS request tools are in use.

Appendix II: Questionnaire Responses

Question 9. What information/data requests tools do you use and how well do they work for you? We welcome comments on any of the products above, especially if you are not at all satisfied.

Would like to see NetDC requests coming out of the Weed program.
Let me first say that DMC has been invaluable for my work! However, I do feel that there have been too many changes in the methods for data request. Unfortunately, not all changes are for the better. In particular, it is important for requests to be downward-compatible, so time is not wasted for the user to reinvent the wheel several times, each time using a different tool
Some of VBB data at CDSN have problems of wrong format and are not usable.
I have always found BREQ_FAST to be an efficient and easy means of requesting data
What is the seismic monitor? How to do event searches?
I was very pleased with the generous and efficient service of the DMC people, in particular Rick Benson.
Need better SEGY conversion tools (e.g. from SAC, SEED or whatever)
Wilber pictures are nice, but should contain a key--what do all the numbers mean? Which phases are which?
I continue to have problems obtaining PASSCAL data subsets using weed and rdseed. I don't know yet if it is a problem with rdseed or with the archiving process using the results of pdbtools on the PASSCAL side of things. Hopefully the DMC and PASSCAL will continue to communicate, as they have been to make the requests less problematic.
most or all are helpful; I find breq_fast most useful, xretrieve helps me find out
I needed to request multiple months of continuous data for one of our regional networks. Your request tools are not setup to do this. You might also make a DLT tape an option for media type when choosing the method of data transfer. I was able to retrieve the data needed by contacting you directly
seems perfectly adequate for what I need.
I have not used the IRIS DMC in a number of years, simply because I have been busy in other research areas and I no longer have support for students for such work. For this reason, I have left many of the items on this survey blank. In the future, I hope to be contributing data to the IRIS DMC, but I need to find the time to get it formatted correctly.
it will be nice to see an option for "standing requests" in weed and/or breq_fast
instrument responses are sometimes wrong, DMC staffs have been helpful, but better mechanism of solving the problem is desirable.
I would find it very useful to have a means to request data from all stations within a Geographical region (i.e. Southern California, California, Western US, US, North America) for a specified event or time window. The use of wildcards in station names in breq_fast has helped this, but it is sometimes quite difficult to put a coherent station list together that covers GSN, FDSN, regional networks (e.g. Terrascope), and PASSCAL networks.
I'll use Weed as soon as it runs under Linux
I've downloaded events data from the web in SAC format. I contacted IRIS about plotting this data, as I wanted to develop some instructions for teachers. The suggestion was to use Alan Jones' Seismic Waves or a program (unnamed) from PEPP. When I tried Seismic Waves, the events would not plot because some of the header information was missing! This required running SAC to add the missing information, a task that one could not ask a teacher to perform. It would be great if IRIS would develop a Java-based seismogram viewer program with an intuitive interface that a teacher could use.
The variety of tools can be confusing - it is not at all clear the best way to retrieve data at first.
I'm sorry I cannot give any comments on most of the request tools because I almost always using the WebRequest tool, which I find very good. What I specially like is that the data are in SEED format, with the correct instrument response. I'm sometimes using the Event Search as well, for making event lists. I would prefer to obtain a list which contains the single events only once as it

was before, but this is not such a big problem, which I can solve with a simple script
SeismiQuery; station information is not always correct FARM/SPYDER; it would be handy if both data sets would be combined at a certain stage Event Search; there are multiple entries for the same event, sometimes that is annoying The ideal situation would be if all the data (including national/regional networks and PASSCAL) would be combined and accessible though IRIS. Response information is not easily obtained for the SPYDER data.
Need better organization of station response information, and better notification fixes to station response errors.
There seems to be no way in which seismology networks can be wildcarded in waveform requests such that both all permanent and all temporary networks that operated in that period are included.
My students use wilber and it is only out of being lazy that I have not started using it
I will be using netDC once it's working. is it already?? and yep, Tim, i will try to use SeismiQuery more often. I STILL HATE THE SLOW WEB THOUGH and the IRIS web page is becoming more complicated to find things! FDSN station book; sometimes a little out of date.
I am very satisfied with everything, but you should continue to expand your service
interfaces on things like weed are still incredibly brittle and frequently can generate unintentional messes without really doing anything wrong. Interface design is not so great—it can be a true pain to get event data trimmed in a good manner. Wilbur had a slick interface but I ended up having to make the request by hacking up some stuff from Weed and manually stuffing it into BREQ_FAST. Seems like too many tools--should basically be a way to get data in continuous chunks or by event (either from a list or with some flexible search parameters)--each of these tools might have a web and an email version (and maybe a paper version for occasional out-of-way field experiments wanting data back)--could then have, say, weed_web, weed-email, weed-paper and users would know that the functionality is the same even if the mode of use is different. Right now I keep wandering in and out if all these things trying to figure how it works
I find the public ftp site useful and easy to use. It is normally quite fast.

Question 11. Would near-real-time data be useful to you?

socket connection to run on our own detection system.
T wave data -- see recommendation of LA2000 workshop
(1) During aftershock studies.(2) As outreach / education tool
I would use it to monitor the performance of dataflow within the IRIS DMS
Education data viewer for SCEPP, screensaver???
I see important research and emergency management potential from such data, for both high resolution real-time locations as well as real-time rupture models
in education -- current spyder does a pretty good job of this
Designing real time applications for improved event identification and association. Automate waveform segmentation, preparing for off-line data processing for each experiment. Set up real-time displays for outreach. Monitoring seismicity and data quality from stations of interest.
Check data quality.
I already use it with Wilber, mainly to show classes recent data but sometimes for research
Monitor the Earth's ground motion to see something.
Enhancement of Alaskan regional network monitoring in any way possible
event location, source parameter estimation
Event detection online from local/regional arrays
model source parameters for events in our region
We would incorporate it into our real-time data feed for better initial locations.
We have 5 Gb of data/day coming in as it is, so our problems focus on data reduction and analysis rather than pulling in new data from the DMC or other arrays. I responded "No" only because We have or will soon have direct sharing relationships with neighboring arrays.

Teaching -- show a seismogram to class -- SPYDER stuff
to compute seismic source parameters
Teaching, Outreach, Global Integration of data from monitoring earthquakes and possible nuclear events
preliminary locations, magnitude and event type will be useful
distribute for hazard analyses, qc for stations, etc
Rapid moment tensor estimation
analyze special events of interest for hazard analysis, educational purpose, or explosion discrimination
I have a near real-time Moment Tensor code running for teleseismic events
in some cases quicker access to data for individual events would be helpful
Near-real-time analysis of events of interest
But I am currently working on getting BFO data out in near real time.
quick source parameter determination
Education and outreach.
an autodrm would be ideal, but i do plan on starting to use NetDC

Question 16. Is there a software package you that you believe IRIS should develop?

Real time data handshakes for all formats
Better integration of PASSCAL field acquisition programs into DMC-type applications
Real Time Data handling software
Yes, Fissures, and I'm working on it as fast as I can. ;)
Database that allows storage of lots of additional information (e.g. travel time picks). Maybe PASSCAL database can already do this I have not spent much time with it. Possibility of link to MATLAB and home-grown FORTRAN or C routines would be extremely useful.
NO - I've found the most useful software is developed outside the normal IRIS programs. Some IRIS support for such activities might be useful...
A standard package for generating receiver functions and inverting for V. Could be provided as ProMax tool to allow power of ProMAX to address passive recordings.
improvements on the PASSCAL database software- this is very difficult to use to apply small time corrections and write SEED volumes
They should stick to getting data transferred and make sure interfaces with other companies products work smoothly
Integrate Antelope and Earthworm better
greatly expand Ken Creager's Coral to make it more versatile, then we can take advantage of Matlab
I think IRIS should support the continued development of Antelope/Datascope, and distribution to member institutions
Geotool, SeismicHandler (see ORFEUS), SNAP (see ORFEUS)
Polarization filtering package
hard to say. I write my own in-house stuff.
single and multichannel processing, which is very expensive for individual researchers to buy, will be useful
One of the problems is that there are so many software packages out there, each with different pros and cons. It makes life very confusing! It would be nice to have one system rather than these many different ones, though that is probably too much to hope for.
No.
Instrument response correction.
Scope of existing software is OK, but could it be more extensively tested and debugged before release.

At the recent SRS in New Orleans I was told that the future of SAC at LLNL is uncertain. Is it possible for IRIS to take this over?
visualization, removing instrumental effects properly before data analysis (because I'm afraid that many people ignore this important step)
seismic handler
The next generation of SAC would be good.

Question 20. What additional data products would you like IRIS to offer?

Metadata (earthquake related)
More strong motion
Hawaii Network Data, Japanese Network Data, Australia, California Short-period (possibly all with a NetDC request)
Special event products, FARM fro Nuclear tests, FARM from other networks
Near-real-time data
Any interest in historic regional-network data? I have a box of Shumagin Net tapes in my office (no $$ to convert to seed though)......;-)
At this point I would like the DMC to debug and streamline requests for subsets of PASSCAL passive source data more than anything else.
Array data from GERESS, access, TXAR, LASA, etc.
AutoDRM A method to produce a yearly report that lets produces a clear graphical display and text output. Maybe it should allow the user to tell it to show all gaps greater than "X". I have used a tool similar to this on your site that has a calendar display but it only showed complete days that were missing data.
Longer near real-time SPYDER broad-band traces
Near-real time seismicity maps longer time series' in spyder
barometric pressure at stations
Data from other seismological data centers, national and regional networks.
Just keep the data flowing.
References to publications on the stations and networks of which IRIS distributes the data.
What would most help me is the addition of other national data and regional data that can be accessed with breq_fast for example.
Not sure
None specified.
Comment to 19; assembled data are either too short or have too many channels I'm not interested in. so i use customized requests 99% of the time. the only time i use assembled data is immediately after an interesting earthquake. A very important product I'm using is the FDSN station book! to 20)IMS data! IMS data! IMS data! Berkeley data if not available through netdc (same true for GEOSCOPE, GEOFON and MEDNET!!)
A tool which allows me to ask for strain meter data or microbarometer data etc. without me having to know the seed channel codename, station name and network code.NB; II and IU use different channel names for microbarometer data!
AutoDRM requests in GSE2.0 or IMS1.0 format.
The next generation of SAC would be nice. (with better vespagram analysis). Alternatively, could go to seismic-handler (written by Klaus Stammler). Could include shear-wave splitting analysis.

Question 21. What kind of data - including non-seismic Earth Sciences Data - could the DMS provide? Please b e as specific as possible.

Photos (satellite, other eq information), barometric pressure
Engdahl et. al, hypocenters topography, geoid + gravity, heat flow, bathmetry
Gravity, topography, magnetic,
GPS data other data (GPS, geology, etc.) database for data obtained in conjunction with PASSCAL seismic field experiments
I look forward to seeing NOBSIP OBS data in the DMC
local or regional gravity.
data for CTBT technologies (infrasound, hydroacoustic)
Digitized fault traces for given areas.
(this may exist); Geologic/site specific metadata that may affect seismic records, as a table and perhaps as a hyperlink to original info. Consider a digital DEM map of the (say) 10 km around the station; rock type; local velocity model if the station is in an established array; borehole velocity vs. depth if known; measured local velocities.
Gravity data Fine scale topography and bathymetry GPS(could have pointers to where this data can be found)
IRIS should promote the idea of "Seamless Archives" of Earth Science data. That is, to a variety of data in different physical archives distributed throughout the community. Mass storage is becoming enormously massive and less expensive. Storage densities are increased with a doubling time of 10 months, which exceeds the doubling time of processors (18 months). In this environment, there will be a growing number of archives, but the connectivity will almost certainly decrease absent a concerted effort on the part of the geophysical community. For example, there is already a viable "seamless archive" of GPS data within the geodetic community.
Hydrophone records.
coverage maps, preferably in ARC format, of gravity, high resolution topography and surface geology would be useful.
topography
Some selected industry-type datasets may be of interest and may be available (such as the Texas Bureau of Economic Geology Stratton 3D dataset).
Global gravity Global topography Global heat flow Global volcanism Global geology
not sure
Hydroacoustic and/or OBS data.
additional non-seismic data products; weather data, e.g. wave height. locations of thunderstorms. locations of storms. IRIS should NOT become "the weather channel" but weather data are extremely useful to have to study microseisms.
Infrasound data (this is of interest in connection with CTBTO)

22. How do you usually report problems with data requests?

Email Tim and Debbie
Email to Debbie
I send e-mail to Debbie.
Fire off an e-mail to Rick Benson or Debbie
Email
Email
E-mail to data technicians and sometimes to Tim Ahern (sorry, Tim.)
To DMC staff
Email someone at the DMC
Email
Have never had any significant problems
Email to Tim or Rick
have not reported any problems
email
email + call DMC
I complain
using email
By phone or email.
Contact person in breq_fast request response
email the person who sent the email telling me the request started
email
I contact Rick Benson or Chris Laughbon.
I send email to Rick
pestering email to Rick Benson
e-mail
email
Send mail to "webmaster" link at the bottom of the page or to Rick or Debbie.
email
I haven't had to.
Email to IRIS personnel
yes -- I complain and people respond
to the people listed at the end of the e-mail I receive using breq_fast
email one of the technical people.
get in touch with IRIS staff
talk directly with Rick.
e-mail
trouble report by email, telephone call to DMC staff,
Depends on the problem -- through the technician handling the request for simple request problems, someone like Rick Benson for more serious data problems
mail them to Debbie
Communicate by email with IRIS employees.
email
By sending emails
Haven't had any problems.

Email to DMC staffer that I believe would be able to help with problem.
to the managers at DMC
Politely via e-mail
E-mail to listed contact
email
email Rick Benson
Contact DMC personnel directly. Used problems@iris.washington.edu just last week.
tim@iris/rick@iris
I used to knock on Pete Davis door. At times I sent e-mail to Rick at irisdmc.
email
By e-mail to DMC.
email
e-mail the appropriate person

24. We welcome your suggestions for ways to improve the IRIS website.

Frameless wilber
Debbie does a great job - I like the web site
Streamline the instrument response format. Presently, some responses are given in M, some in M/S; some with convention A for poles or zeroes, some with convention B. I suggest that all instrument info. be given the SAME format, for example everything in M And format A. It would greatly simplify the extraction of instrument responses2. try to regroup all stations in a single category -- eliminate the need to specify IU, II, GE, etc. Often times, it takes a considerable effort to figure out which network belongs to.
Need a clear, nested site map and in some instances, duplication of relevant information (after all, it's just another pointer on a web page). For instance, if I need a GSN station map/list, I should be able to get it from both the GSN page and the DMC page.
It is so extensive it is sometimes hard to find things. SeismiQuery is a very valuable tool.
The "current newsletter" on the publications page is Fall/Winter 1998, a bit out of date.
It's difficult to find the button sequence to get to the waveform menus from the IRIS homepage. A few of the labels leading to this page are not informative enough for some people to determine exactly where the pages are.
Generally it is a great web site and I am always amazed at the efficiency with which IRIS gets the data out, so just two detailed comments.1. WebRequest is cumbersome to use. Would it be possible to allow copying of whole lines, so that for example if I want to request the same time frame for a number of stations I don't have to enter the time constraints again and again.2. It would be good if there was a way to cancel faulty requests. This feature would be extremely useful, if the initial email confirming receipt of a data request would also contain an estimate (does not need to be accurate) of the expected size of this delivery. This make a number of blunders obvious (once I requested one year and 10 minutes of data instead of 10 min because of a typo in the BREQ_FAST request - luckily someone at IRIS caught my mistake before time was wasted on this request.
Spend more time with testing web pages and underlying software by people who did not design the specific applications. Spend more time during design and development phase with the end users who will use each specific application.
Cut down on frame usage, and hosting windows in frames. Better to open up another browser window because you get more width.
I appreciate the constant tinkering with/improvement with software, etc., but I worry a little about the rapid pace of innovations. I have become attached to breq_fast, for example, and have built it into my own programs. If it were to be superceded by a program that is somewhat or marginally better I would be disappointed. Can you really maintain all these different access tools? I think interactive, web-based access methods are a step back from batch methods. They take so much more time to use.
See above comments on collecting data from a variety of networks. Reliance on network codes is understandable, but it'd b nice if there was a more simple way to find all stations in a specified region.
Too much Java (means pages are heavy sometimes). Text oriented pages (as an option) would be useful for foreign users.
Think quick loading...
Web site is OK.
SeismiQuery has evolved beautifully! What a great tool!
de-clutter! I personally don't really like "Frame" websites because i cannot add a bookmark, if i found "my" page. The IRIS page is getting complicated enough so that i usually don't remember how the hell i got to the page i wanted. i do realize that IRIS got an award for their pages once. How about de-coupling the IRIS page from the DMC page? Bad idea, hu? I still think that important notices on changes in evalresp and rdseed (or whatever) are not really well announced. perhaps the letters announcing changes need to be redder and bigger?? at least for me.
Back buttons and error buttons too frequently dump you all the way back to start pages--not very forgiving. Too often links to external content end up in a frame (sometimes a really small frame)--should either give an option to leave IRIS site or open a new window, etc.

25. Do you feel there is a need for a better station history and data problem reporting system? If yes, can you describe what information and features it should have?

Visual time line of problems or instrument changes
Station maps in operation by year could be very helpful for planning requests(ok the Info exists but I usually have to get and plot it myself then go back to make a request)
Perhaps a file, which would give station name and first day of operation. This would help greatly in identifying adequate station/earthquake pairs when trying to solve a seismological problem. Such a file could be downloaded and read directly into search routines.
It is very important that the return of requested data has the correct station information (i.e., the information appropriate for the particular time window.) In addition, a pointer for station history will be very useful as well, because sometimes The history can give us hints as to what may have really gone wrong (reports have errors too.)
Based upon what I am hearing I would agree with this
The current DPR system seems to be designed for the QC centers, but there d doesn't seem to be a good and easy way for the average seismologist that notices a data problem to report it. At the least some more information on the DPR page to direct seismologists as to whom (at IRIS or elsewhere) to send data quality questions. Even if it is just a statement like "the DMC just takes what it gets, go complain to XXX" would be useful.
Needs to more explicitly include times/dates of no clock locking, plus time windows where sensor are poor or incorrect.
You could well have the features I require in place. Basically I need to be alerted to any problems with the data I have received, such as timing problems (I am assuming response or component labeling problems have been resolved already).
Inclusion of the appropriate information either in the rdseed file or in the email IRIS sends out to notify users of the availability of the data. This way one would see immediately potential problems with data or a data product. The email could also contain a clear message explaining how to report problems one might find in the data, which were previously unknown. It would be truly great if you could keep track of all requests made within some extended period of time, say for one year, and then send out an email to users if problems are discovered with data after they have been sent out. This sounds like quite a logistic feat to match changes to the problem report base to past user requests so I don't know whether this is feasible.
I find the old noise curves very useful. They help me figure out where to expect to find good/bad data, and help me understand signals I see. However about half the existent stations came on line after the noise survey was run. What would it take to update those plots, at least for Iris-GSN?
There should be a succinct summary of all significant maintenance work at a station.
Develop a detailed reporting history for PASSCAL passive source stations even more detailed than the dataless seed volumes. This would involve cooperation of PASSCAL experiment PI's.
it should be possible to append a sort of flag to data request, so that the flag/message always goes with the data and does not get separated from it
I haven't looked at what currently exists... but when a data request is made for a station that has a problem some type of flag should alert the user to check a "problem log". The log should have operator provided comments about the type of error and the status (i.e. it was fixed on 5/15/2000).
Don't know.
Actually, I'm not sure what the current system is capable of.
Station responses updated more frequently.
The use of dprs needs to be better advertised to the user community (in original email that notifies us of data requests being ready).
complete history should be described. in particular, any possible visit and service of the station should be described. This is because even when the service person thought he was

aware anything thing done to the instrument, he may not be aware that he adversely changed something else (such as a gain change as an accident).
Wouldn't it be wonderful if RDSEED sets a flag in the SAC data header that warns if there is a problem reported for this particular station on this particular day?
I am not aware of where to report data-quality problems
Information on how instruments are calibrated. Information on possible timing problems. Etc.
Details of any problems with station orientation, more information on external sources of noise caused by near station
timing, orientation,
No opinion
IRIS DMC needs to have a mechanism in place to notify users when station response information has been changed or modified. The only I way I keep up with changes to response information is to occasionally go back and re-request all response data for stations of interest.
Only when data have a problem but are kept on-line anyway.
Would be ideal to be able to see at once everything known of the accuracy (or uncertainties) of timing and response correction for each station/channel
Station history, descriptions of problems, 'features', and time range where found in waveform archive, researchers who may have noted problems, when, and what was done to correct or prevent them.
timing problems, orientation problems, calibration errors, missing waveforms (sometimes, there's just nothing on the records---sensor dead, gain too low?? — I could provide a list for some GSN stations) the DPRs are ok but not very attractive to use. Is there a way to put these into an interactive web-tool? (like the FDSN station book) Actually, it would be great if the DPRs were in the station book, for each particular station--because that's where I usually look first for station info!
Current PASSCAL feed of data tends to put quality control after data archiving; I suspect there are a number of problems not in the DMC databases because of this. So some better system would find a way of making it very natural for these discoveries to be added in to the system--this is more a people than software problem.

26. Would you be interested in an on-line IRIS bulletin board/chat room that could serve as an avenue of exchange for problem solving, etc. (though it would not be verified for validity of content)? Please comment.

Maybe not a live "chat room" but something closer to an information/e-mail question and answer archive site (does this counts as a "chat room"?). The ability to easily search the archive would be extremely important and useful.
From my experience with the GMT software, this type of format can be useful for a while then gets out of hand very quickly. The volume of traffic is so large that I no longer read any of the mail regarding GMT.
The mailing list and/or net news model seems to work well in the rest of the computer world. Seismologists don't seem to use it though. As a simple start, a FAQ would be a nice place to put dumb questions that are asked frequently.
This could possibly be useful. I do not have time to keep up-to-date with bulletin boards. An email newsgroup could be useful if traffic was sufficiently low. As an overall comment I think that the IRIS DMC service has always been excellent, and greatly facilitates research in earthquakes around the globe.
This might be useful but generally I would consider it much more useful if IRIS staff condense any tips or comments put forward by IRIS users into help pages/FAQ etc. I know this means more work for you guys but even a little time spent on this would be adding more quality I think than an unmoderated chat room.
no time....
that will be very useful. particularly for people in the field running PASSCAL type experiments.
I would probably chew up too much time manipulating the chat room before getting to issues
PIs with same problem (in my case instrument gain error and data format error) can share experience and save other’s time tremendously!
Bulletin boards quickly get cluttered. I don't think I'd look at it that much. I'd be much more interested in effort going into a system to automatically check a "problem log" for a specified station/channel/time-window for each request, with an email to the requester being automatically sent if a problem is found.
General comment on this form; I am myself of course only an occasional user, most of the direct data retrieval and handling is done by my students. The responses I gave represent my own experience, mixed in with some impressions gained when looking over student's shoulders.
I doubt I would use it much but some people like to chat. General comment, IRIS is going an amazing job distributing data please keep up the good work.
Being POC for IRIS data requests from a group of ~20 researchers at LLNL, the replies herein represent the whole group and not just me.
I used to look into the bulletin board back in the old days. I’ve never used a chat room. No time. I still think that the DMC should have an email list of users, if a severe change has been made to the data format (e.g. that dreadful location code), RDSEED or evalresp (i.e. found bugs). It's probably not really a good idea to just put this on the Web as announcement (because I request data A LOT more often that going onto the web page). We only realized a few days ago that we have used the wrong instrument responses for some of the IU stations (luckily only those with that dreadful location code--why do GSN stations need this???? ---why not use UVW and FGHas originally done?), because I didn't know about the consequences of using the one but last version of RDSEED. Things like these might be resolved faster if I'd get an email from IRIS saying that I need to get an updated version of whatever. IS NETDC available yet?? If anything I said is offending anybody; I don't say things to attack anybody, I just thought you'd want to know the "users' opinion". And sometimes, I'm a frustrated user
Searchable web-based system with threads and an archive could work well
It could be useful but I don't see having the time for it.

I have been particularly frustrated using the PASSCAL database software and think that many people could benefit from problems and solutions that have been encountered. Year 2000 PASSCAL instrument problems have also been plentiful and future experimenters can benefit from learning about specific problems encountered.
Mailing system seems to work fine.
I would probably be an infrequent contributor and until I see it and use it I won't know how valuable it will be to me, so I wouldn't push this chat room idea to strongly. Still, if you started one, I'd check it out.
Maybe with a searchable FAQ area. Also possibly useful, a contrib area for people to share tricky scripts or data requests that could serve as a pattern or example set.