THE DESIGN GOALS FOR A NEW GLOBAL SEISMOGRAPHIC NETWORK

DRAFT

Prepared by the

INSTRUMENTATION and DATA COLLECTION SUBCOMMITTEES

of the

STANDING COMMITTEE FOR THE GLOBAL SEISMOGRAPHIC NETWORK

INCORPORATED RESEARCH INSTITUTIONS FOR SEISMOLOGY, INC.
IRIS, Inc., 2000 Florida Avenue, N.W., Washington, D.C. 20009

March 13, 1985

Table of Contents

FOREWORD

The purpose of this document is to translate the scientific goals and objectives of the INCORPORATED RESEARCH INSTITUTIONS FOR SEISMOLOGY (IRIS), as presented in the *Science Plan for a New Global Seismographic Network* (hereinafter referred to as the *Science Plan*), into realizable technical goals and objectives for a 100-station global telemetered network with support facilities. In this conceptual stage it is important to identify those many elements of design and planning that must be considered, survey available options, and identify those areas of uncertainty where trade-off studies, design studies, and development are required. The specifications and operational plans that result from this initial study will serve as the master planning document for design, deployment, and operation of the new network. It is planned to distribute this document widely to obtain the comments from interested parties before the overall design goals are translated into a definite system design.

Stated in its simplest terms, the goal of this new generation global seismographic network is to produce broadband, wide dynamic-range digital data from a global network of at least 100 stations and provide for the timely collection and distribution of these data to a wide variety of users. In order to achieve that goal, technical specifications must be developed which respond directly to current and anticipated future scientific needs discussed in the *Science Plan* and which take full advantage of current technology while providing the flexibility to incorporate future developments. It is not our purpose herein to present the details of the the network design, but rather, to present the general design goals which will lead to the development of detailed technical specifications of the system.

Figure 1.1 illustrates schematically the relationship of the various elements of the overall GLOBAL SEISMOGRAPHIC NETWORK (GSN) project. In section 1, the design of the seismic stations is considered. This includes the sensors, the analog to digital conversion system, the on-site data processing, the local recording system, and auxiliary systems that may be located at or

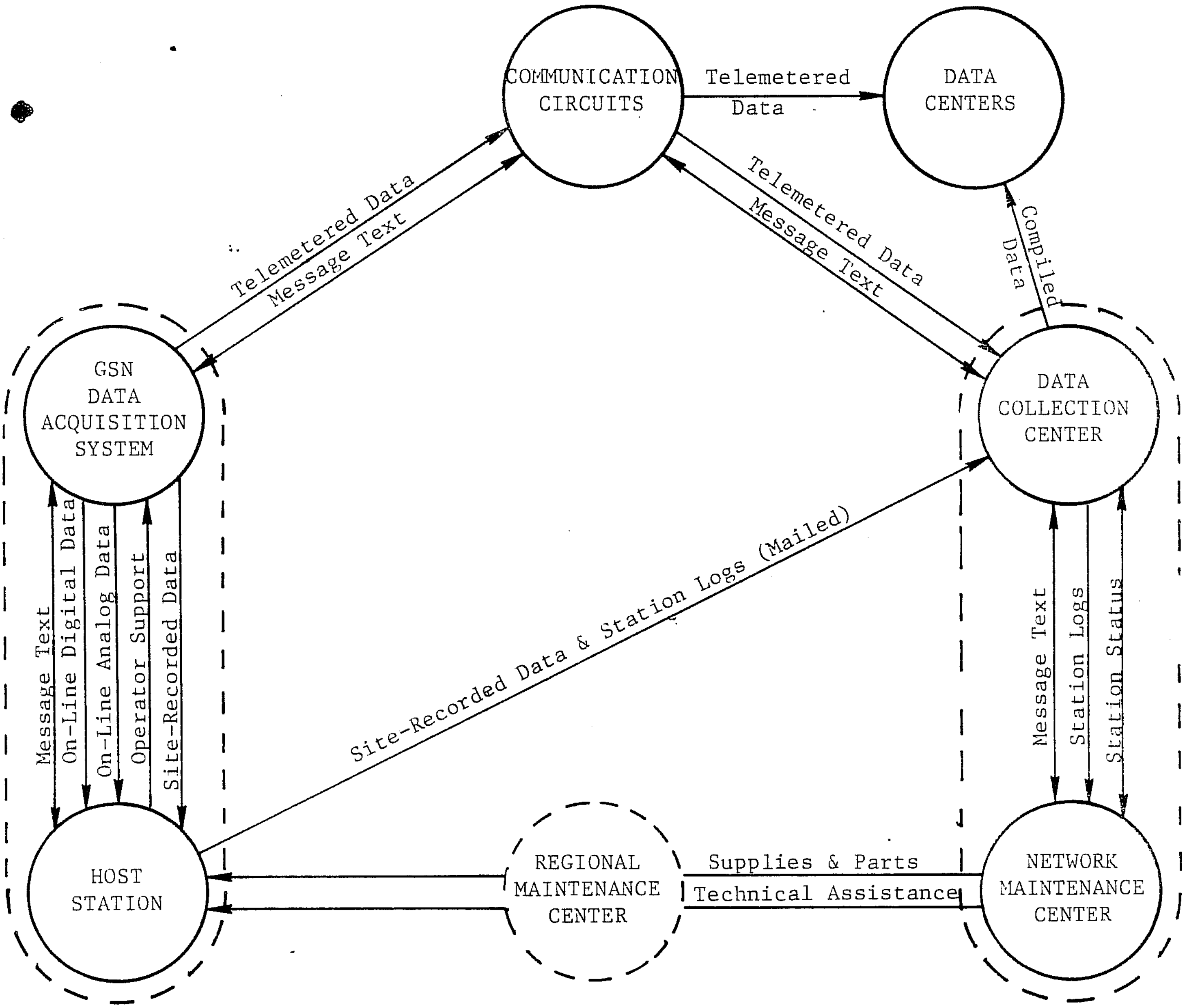

Figure 1.1. Elements of the proposed Global Seismographic Network

near the station for the use of the host organization. Section 2 discusses the requirements for the data collection system which functions as a central facility for data gathering (taken in the broad sense) and monitoring the operation of the network as a whole. Later documents will describe the telecommunications system which connects the GSN stations together into a network and provides the means to collect data and communicate with the remote stations.

This document has been prepared by the IRIS STANDING COMMITTEE FOR THE GLOBAL SEISMOGRAPHIC NETWORK and its several subcommittees. It represents the work of many individuals on those committees and is meant, at this stage, to be a preliminary document subject to review and revision by other interested parties.

1. DATA ACQUISITION SYSTEM

1.1. Introduction

The data acquisition system in the context of this document refers to the equipment installed at the seismic stations and, in some cases, at other sites of the host organization. Functionally, the system gathers seismic data, prepares it for transmission over the telemetry system, records it locally, and performs whatever control and monitoring tasks are necessary. Figure 1.1 illustrates the relation of this system to the overall GSN system.

The key technical requirements of the new network identified in the *Science Plan* that relate to the data acquisition design are:

- Digital data acquisition with real-time or near real-time data telemetry;
- Bandwidth sufficient to record the entire spectrum of teleseismic signals;
- Dynamic range sufficient to resolve ground noise and to record the largest teleseismic signals;
- Low noise instrumentation and environment;
- Linearity
- Standardization of system modules.

The purpose of this section is to translate these general scientific goals into design guidelines that may, in turn, lead to a set of technical specifications. The decisions on the type and characteristics of the instruments, the mode of recording, transmission and collection of the data, the procedures of archiving and dissemination will be preceded by a review of design goals developed in this document and by some experiments and further design studies.

1.2. Data Requirements

In designing the seismograph system for the new network, the foremost issue is the definition of the data requirements in terms of resolution, bandwith, and dynamic range. Both the amplitude and frequency range of interest in seismology cover many orders of magnitude (nanometers to centimeters; thousands of seconds to tens of Hertz). In the past it has not been technically possible to cover this entire spectrum of signals with a single system. The application of modern techniques in electronics and control theory, however, now offers major improvements in seismometer design not fully utilized in existing networks.

There are three prime aspects of the system to consider in defining the requirements of the instrumentation:

- BANDWIDTH; The network is intended primarily for global and regional seismic studies and hence the bandwidth should be adequate to record the full spectrum of seismic signals that propagate over distances large compared with the station separation, on the order of 20° for a 100 station network.
- RESOLUTION AND FULL-SCALE; Over the specified bandwidth, the system should be capable of resolving signals at the level of minimum ambient ground noise and of recording signals from the largest expected teleseismic event. Neither the full scale clipping nor the system noise should obscure any signal of interest.
- LINEARITY; The system output should be a linear time-invariant function of the ground motion. Further, the linearity should be such that signals near the ground noise minimum can be resolved in the presence of ground noise at other frequencies near the expected ground noise maximum.

It should be stated clearly at the outset that the network is designed primarily as a global seismological tool. From a technical viewpoint, this means that the design is predicated by the nature of the seismic signals which propagate over distances large compared with the station separation, on the order of 20 degrees. Thus, both the bandwidth and dynamic range will be limited to that necessary to acquire such data. Specifically, the system is not designed to monitor local seismicity or strong ground motion. Occasionally, a single station may experience saturation of its data acquisition system, but a significant fraction of the network should never be saturated.

1.2.1. Bandwidth

Over what frequency range should the system be specified to record signals? The response at long periods must allow us to record the gravest modes of free oscillations of the Earth for sufficiently large events. Further, the earth tides with a principal period of about 12 hours provide a very useful calibration signal and help to monitor the absolute calibration of the instrument.

A high frequency limit cannot so easily be set. This limit determines the sampling rate of the network and so controls the data volume which must be handled. The observations summarized in the *Science Plan* suggest that a sampling rate of of about **20 samples/second** would be adequate for data acquired continuously. In this document, channels sampled at this rate will be referred to a "broadband." Occasionally, when large events occur, or at some particularly quiet sites, a higher sampling rate will be required to capture the high frequencies excited.

1.2.2. Ground Noise and Resolution

The seemingly random and ever present ambient seismic noise forms the fundamental limit of the resolution of seismic signals. The design should specify that the system noise be less than the minimum ambient ground noise over the operating bandwidth. The spectral density of power per unit bandwidth of ambient acceleration has been well determined over the band of interest for global seismology. Figure 1.2 illustrates the range of ground noise reported at seismic installations. In this figure, the power spectral densities are converted to 1.25 times the rms amplitude over a 1/3 octave bandwidth. This corresponds roughly to the mental filtering one applies when estimating average peak amplitudes at differing frequencies from a time domain plot. The maximum noise levels at periods of about 3 to 6 seconds are generated by marine microseisms. At shorter periods, wind and cultural activities are the principal noise sources. At longer periods,

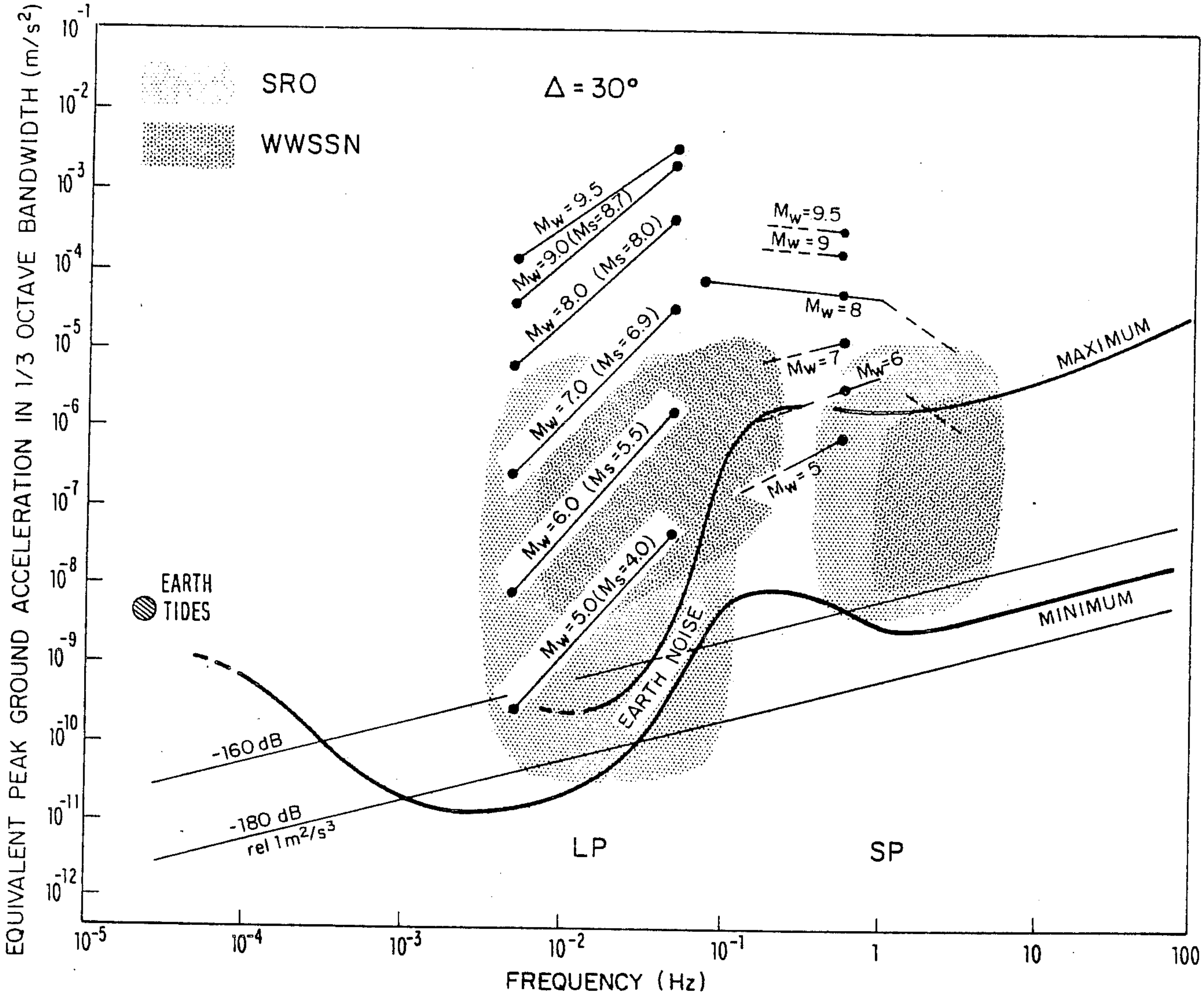

Figure 1.2. Acceleration amplitude as a function of frequency taken as 1.25 times the rms values over a 1/3 octave bandwidth. The heavy lines indicate estimates of the minimum and maximum ambient ground noise. The lines marked −160 dB and −180 dB indicate levels of frequency independent power spectral density relative to $1(\mathrm{ms}^{-2})^2/\mathrm{Hz}$. The accelerations expected at 30° for a range of magnitude earthquakes are shown. For periods from 200 to 20 s, the estimates are time domain amplitudes of surface waves. For 20 to 1 s, the estimates are for time domain amplitudes of P and S waves. For shorter periods, the estimates are derived from P wave spectra and are subject to large uncertainties. The shaded areas indicate approximately the principal operating ranges of the WWSSN and SRO systems.

While many sites achieve the lowest values of ambient noise at the longer periods, few sites will have the minimum noise levels at periods short compared with the microseisms.

Ground tilt from wind turbulence is a major source of noise on horizontal component instruments in this band and may be expected at most surface sites. Some locations may be equipped with borehole seismometers in order to achieve the required performance.

A conservative design specifies that the total system noise should be less than the lowest ambient noise. Over much of the band of interest this level is a frequency independant power spectral noise density shown in Figure 1.2 as a line that varies approximately as $f^{-1/2}$.

1.2.3. Full Scale Range

Keeping noise levels low is an important factor in site location but the real interest in seismology is in recording transients associated with seismic "events." In particular, we must primarily be concerned about recording large transient signals. It is unlikely that a single seismograph will record the full range of ground motion experienced in the near field of an earthquake, so a decision must be made as to what should be recorded on scale with high fidelity. As the station separation for a 100 station network is approximately 20°, it is reasonable to use this distance as the limit beyond which all stations should remain on scale.

Figure 1.2 shows the peak accelerations in a 1/3 octave bandwidth expected at regional distances from a range of magnitude earthquakes. An earthquake with a moment exceeding 10^{28} dyne-cm occurs every one or two years; earthquakes approaching 10^{30} dyne-cm occur every few decades. It would therefore be desirable to require the system to be capable of producing low distortion, on scale records of an earthquake of moment $\sim 10^{30}$ dyne-cm ($M_w \approx 9.3$) at a distance of 20°. Over most of the seismic frequency band, accelerations will rarely exceed 10^{-3} ms^{-2}. Only for very large events which occur about once a decade will the accelerations reach 4×10^{-3} ms^{-2}. These peak accelerations occur for surface waves with a period of about 20 seconds. The system should be designed to record these levels of acceleration while resolving ground noise at quiet sites.

1.2.4. Analog-to-Digital Encoder Resolution and Format

The required resolution, full-scale, and response of the system lead to the specification of the number of bits in the analog-to-digital encoder. The resolution (or precision) of analog-to-digital converters has until recently ranged from 72 dB (peak to peak) for a 12-bit encoder to 96 dB for a 16-bit encoder. In order to take advantage of the analog dynamic range of the seismometers, the operating range of encoders has been increased by a gain-ranging technique in which the analog signal is attenuated in a series of steps as the encoder output approaches full scale. This technique permits large signals to be recorded, but the resolution of the encoder, referred to input voltage or equivalent earth motion, decreases in proportion to the attenuation.

Recently, 24-bit encoders have been commercially developed. The increased resolution of these encoders is achieved by bit enhancement (a process involving oversampling and averaging) and the actual resolution is frequency-dependent. However, they are expected to have a resolution of at least 120 dB at 0.1 seconds and higher resolution at longer periods. At the same time, until the 24-bit digitizers pass field tests, considerations should be given to the development of gain-ranged digitizers with less resolution, as they are readily available, involve lower-level technology, and are likely to be less expensive. If this technique is employed, in order to achieve the desired performance, the gain should not change due to the ambient noise conditions, but only when large signals are present.

1.2.5. Very Short-period Channels

At some sites that combine low ambient high frequency noise with a high apparent Q, it will be desirable to record signals beyond the 10 Hz bandwidth of the normal system. In these cases. a module should be added to the Data Acquisition System which samples at 200 sps, triggered whenever significant signals are detected in the 1 Hz to 10 Hz band that indicate the presence of significant higher frequency energy. Little is known about the seismic signals propagating over large distances at these frequencies, but it is felt that the addition of this capability will not significantly add to the system complication or cost.

1.2.6. Very Long-period Channels

For some purposes it is desirable to record seismic channels with a very low sample rate; 1 sps or less. These channels could quite easily be derived from the 20 sps data stream and would represent an insignificant addition to the data volume.

1.2.7. Low-gain Channels

It may be that existing single sensor systems can achieve the dynamic range discussed in section 1.2.3. More likely, however, a set of low-gain data channels designed specifically to record the largest accelerations will be required. These could be triggered in the same way suggested for the very short period channels. With such a system, it would be feasible to record even the very large accelerations experienced from occasional nearby events.

Figure 1.3 illustrates schematically the ranges of a possible two level data acquisition system. Clearly, the primary broadband data channels should be designed to have the highest dynamic range technically feasible. If the sensor system has sufficient dynamic range, but the ADC does not, these low-gain channels can simply be obtained by a second set of encoders digitizing the outputs of low-gain amplifiers. If, on the other hand, a single sensor system cannot cover the entire required range, then a less sensitive sensor system will be needed.

A relatively inexpensive set of low-gain data channels could be implemented that would have a full-scale of 1 ms^{-2} (.1g), a dynamic range of 90 dB, and a bandwidth of from .01 to a least 100 Hz. With the addition of such modules, the overall GSN system would have an operating range of 200 dB and would be capable of recording signals from the largest credible earthquake even if within a few degrees of the epicenter.

1.2.8. Supplemental Data Channels

In addition to the multiple components of seismic data, there is likely a requirement to record several channels of supplemental data such as:

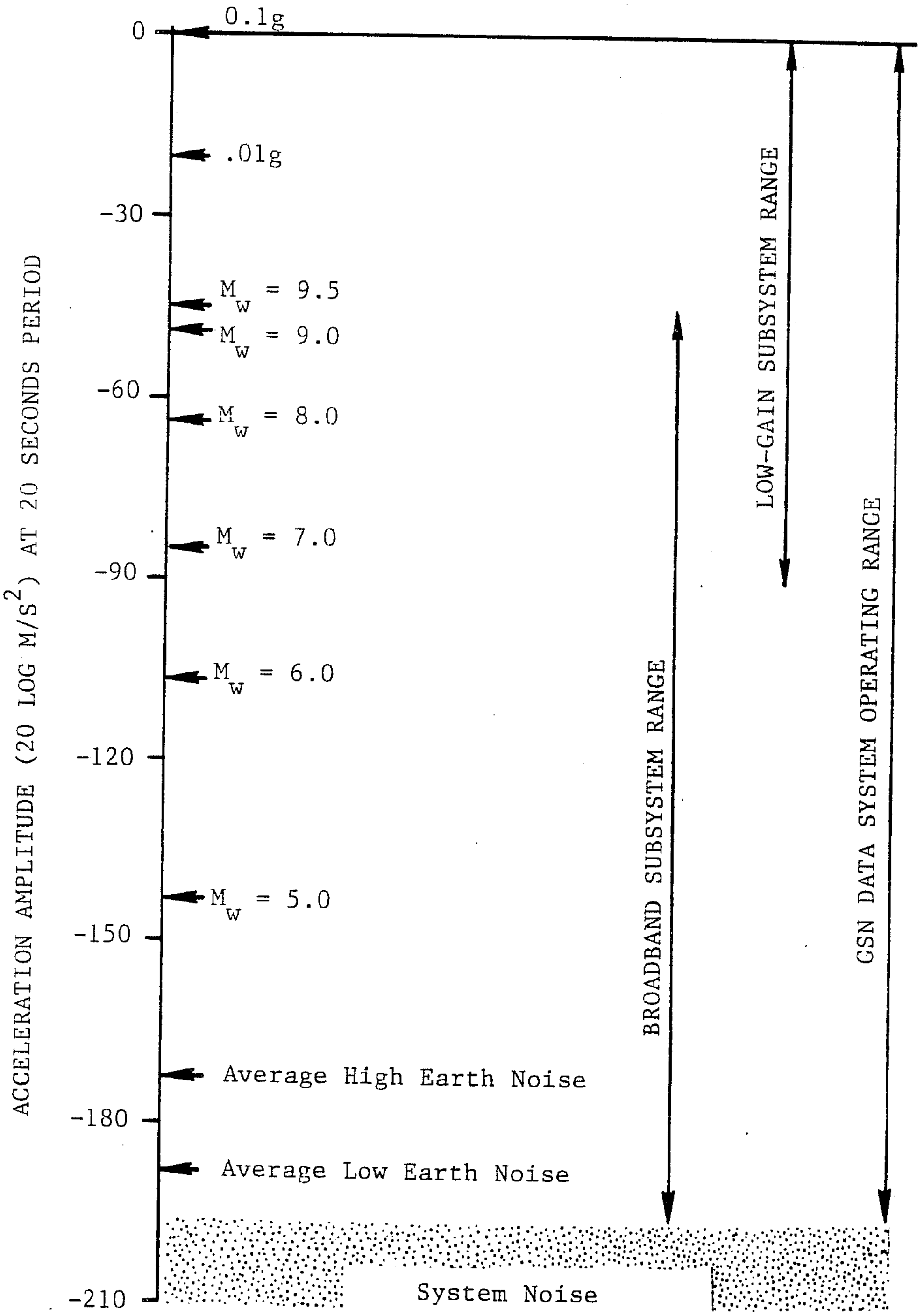

Figure 1.3. Operating range at 20 seconds of a possible two level data acquisition system.

- environmental data (temperature, pressure, wind speed, etc.),
- state-of-health data (timing errors, tape parity errors, mass position, etc.),
- parametric data derived from the seismic channels (event counts, noise levels, etc.).

For design purposes, it is assumed that these channels will be sampled at a rate low compared with the 20 sps rate so the additional data stream will not represent a significant increase in the overall data volume.

1.3. Data Acquisition System Modules

Figure 1.4 shows a block diagram illustrating the principal features of the Data Acquisition System. The modules discussed in this section are:

- Sensor Subsystem
- Station Processor
- Power System
- Host Subsystem
- Analog Monitor

The design of the data acquisition system will allow for unattended operation for extended periods of time but it is hoped that most stations will be located at sites where local organizations will supply operational personnel. At these sites, the "host subsystem" will provide the station personnel with the capability to control the data acquisition system operation and provide access to the data.

The design goals of the GSN call for data acquisition systems capable of real-time telemetry of the entire data stream. The plan is to equip the stations with transceivers for data telemetry and message communication via satellite circuits to the data collection centers. However, it is realized that some stations may have to function without the telemetry links capable of real-time data transmission at the required rate, and design should permit this.

There are certain characteristics common to each of the hardware modules that the design should achieve. Principally they are:

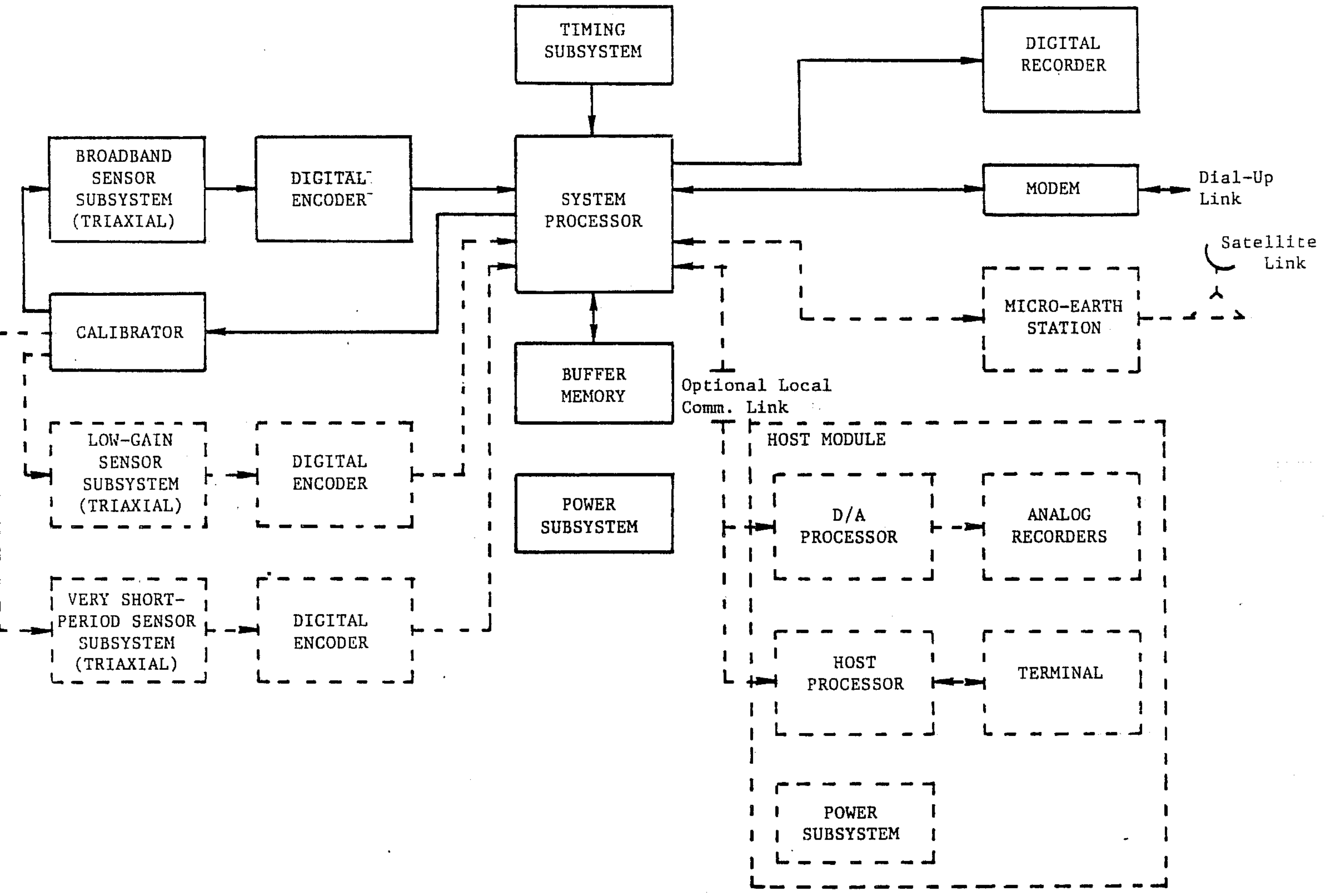

Figure 1.4. Block diagram illustrating the principal components of the Data Acquisition System.

- **Standardization.** From a practical viewpoint, this characteristic is highly desirable. Past experience with operating seismic networks, regional or global, has graphically demonstrated the disadvantages of constructing a network from a diversity of individual stations with differing characteristics. It is cumbersome and costly from a viewpoint of network operations, and increases dramatically the complication of data collection and analysis, often diminishing the reliability of the final scientific results. Of course the network must maintain the flexibility to encourage participation with stations of differing characteristics as long as they meet certain standards.
- **Modularity.** Modular system design is generally considered desirable as it supports flexibility in operational configuration and eases maintenance. Such a system is easily modified and hence can be updated as new technology is developed and still maintain the necessary degree of standardization. Thus, the proposed network, if carefully configured should provide the science with its observational tools for at least twenty five years after deployment.
- **Reliability and Maintainability.** A practical design for the overall system reliability is about 90 %. This figure is based on past experience with global network operations where it is found that, no matter how reliable the system hardware is, other factors result in approximately 10 % of the stations being non-operational at any one time. A higher uptime goal, while technically achievable, can be costly because of the requirement for better-than-commercial grade components and the need for a much higher level of station spares.
- **Exportability.** Export licenses will be required for the station equipment for most destinations. Licenses cannot be obtained until system design is complete and destinations are known, but it may be possible to obtain a preliminary judgement on the exportability of critical components while design is underway.

1.3.1. Sensor Subsystems

The selection of a broadband seismometer (or seismometers) will be a most important design decision. It is clearly desirable, if possible, to have a single sensor that covers the entire band with sufficient dynamic range and linearity to meet the data requirements. Preliminary investigation, however, indicates that this is unlikely to be realizable with today's seismometer technology. Current concepts are for three separate sensors:

- broadband to 10 Hz.
- very short period from 10 Hz. to 100 Hz.
- low gain to record accelerations up to 1 ms^{-2}

The design of the sensor system should allow for several modes of installation depending upon local conditions. Based on previous experience deep (~ 100 m) is preferable but the more conventional vault installation will be needed at some sites.

Sensors are available that come close to meeting the broadband data channel requirements at least for vault installation. Research and development are needed for borehole systems. Very short period sensors are also available that will likely be able to meet the requirements. Accelerometers of the type used in strong-motion programs, but with higher sensitivity, are a likely choice for the low-gain sensors.

1.3.2. Station Processor

The station processor controls acquisition, processing, transmission, and recording of the data as well as system calibration and "house keeping" functions. The processing is likely to include:

- data digitization and timing
- sensor system control (leveling, mass resets, etc.)
- monitoring system state-of-health
- data compression and event detection
- filtering and decimation (to produce long-period channels)
- maintenance of 24 hr. data buffers
- recording of all data on site (~ two weeks capacity)
- system calibration

- duplex communication (data and message transmission, etc.)
- selective replay of data (within data buffers)

While the system should be designed for unattended operation, many stations will have operators and all will have personnel occasionally visiting them. The station processor should clearly have a operators console for testing, trouble shooting, maintenance, and parameter adjustment. A keyboard and hard-copy terminal will be the minimum requirements for these purposes. The standard printout might include a log of operator actions, state-of-health messages from the system, automatic signal detections, and messages to and from the data collections center via the telemetry link. In addition, it would be very useful to have the capability to display waveforms of selected data stored in the systems buffers.

Consideration of existing satellite and other communication channel capabilities and tariffs plus knowledge of the spectrum of seismic signals to be acquired leads to a specification of

- **1200 bit per second (bps)**

as the overall aggregate data rate to be either transmitted or recorded locally. To achieve this rate, modest data compression will be necessary to acquire three channels of 20 sps data plus the supplemental data and overhead associated with formatting the data.

A station clock is needed to synchronize data sampling and to generate time codes for digital transmission and recoding. Automatic and nearly continuous synchronization to Universal Time to an accuracy of 10 milliseconds should be combined with a local timing source of short term stability better than 10^{-8}.

Although modern force-balance seismometers are more stable than their forerunners, calibration is still essential in a seismograph system. It is needed during the installation to set the parameters and periodically during operation to confirm sensitivity and other transfer function characteristics. Calibration of each individual channel is extremely important if full scientific benefit is to be made of the data. Periodic application of calibration signals such as 1/2 full-scale steps or "random" binary sequences of steps precisely timed should be automatically applied in order to maintain an overall system calibration to 1% in amplitude and 1° in phase.

Maintaining a 24 hr data buffer at the specified data rate amounts to about 10 megabytes (MB, one byte = 8 bits) and hence, with todays technology, it is likely to be most practical to store this data on disc. It is not unlikely, however, given the trends in memory size and price, that it will soon be practical to have this entire data buffer memory resident.

Digital recording on magnetic tape or such medium will be essential at those station that are not linked by telemetry circuits and it may be required at all stations for backup and local data access. A two-week minimum storage capacity (140 MB) is desirable and a one-week (70 MB) should be considered as minimum.

Finally, the station processor should be designed to be tolerant of the results of malfunctioning data acquisition equipment. In many cases (eg. bad recording format, clock errors, polarity reversals, and component mis-identification), the quality of the data is not compromised if the fault can unambiguously be determined. The system should be designed to eliminate known historical problems and/or to provide auxiliary information with which such problems may be resolved. Suggested candidates include:

- a completely self identifying data stream,
- data block sequencing information independent of the system clock,
- self identifying components (eg. performing component calibrations at different times),
- a means of testing component polarity,
- an audit trail of all operating parameter changes included in the data stream,
- an automated tape changeover system (to eliminate tapes improperly terminated by the operator).

1.3.3. Power System

According to past experience, a significant percentage of system failures are related to power, partly because power reliability is poor in many areas of the world, and partly, perhaps, because system power is often the last item of design and given insufficient attention. The power system should provide for "conditioning" of the local mains power and the capacity to provide the data acquisition system with 2 to 4 hours of operation in the event of local power outage. Clearly some elements of the data acquisition system can be shut down with little impact on the overall system performance. Other items are critical. In the event that local power is off for longer than this, the power system should be capable of "shedding" its load gracefully.

1.3.4. Host Subsystem

The purpose of the host system is to provide local personnel with some modest seismic data processing capability so that they may access and analyze the data being collected by the station or stations they are running. Typical functions this system should be capable of performing are:

- duplex communication with the station processor;
- selective retrieval of data within the station processors 24 hour buffer;
- graphics display of selected waveforms
- data and message exchange with the data centers;
- monitor the real-time telemetry stream.

It is likely that these capabilities can be provided by a system not unlike today's top-of-the-line personal computer, configured with an appropriate set of peripherals.

1.3.5. Analog Monitor

Analog recording of selected filtered channels should be available as an option to those stations that request it. The analog channels could be designed to emulate the WWSSN-type short-period and long-period responses, and the recording method could be chosen to eliminate the need for expensive photographic recording at stations that operate both WWSSN and GSN stations.

2. DATA COLLECTION SYSTEM

2.1. Introduction

This document specifically addresses the design goals of the data collection system being planned as part of the IRIS initiative. However, the IRIS network, which itself may be composed of sub-networks, is considered in the context of a cooperative venture in which the IRIS contribution to the global seismic network will combine with other networks, both national and international in scope to become part of this larger network. In the following, the data collection problem has been addressed in a general way so that concepts presented could be applied to each network which is a part of the global seismograph network. By clearly defining the role of each element in the data handling system and their relationships to one another, it is hoped that the timeliness of data collection and dissemination from the entire network will be facilitated.

When the Global Seismograph Network is fully operational it will produce an aggregate digital data stream which could approach 8×10^{12} bits/year. While this is a monumental data volume by the standards of research seismology, it is several orders of magnitude smaller than the amount of data acquired each year in space physics. Obviously, the global seismograph network data collection effort can be managed, but it will require careful planning to ensure that design goals are met throughout the life of the network.

During the course of two decades of experience in global seismograph network operations, an organizational structure has evolved. Organizational units of interest here (shown in Figure 1.1) are network maintenance, data collection, and data distribution. The data collection operation entails processing a full day's output of a network each and every day, a task best done in a stable production environment, whereas network maintenance and data distribution are primarily event driven. Thus, this three part organizational structure serves to isolate data collection from events which would upset the stability of its environment. This observation has been used to determine what functions are best performed by the data collection operation.

By its very nature, data collection implies centralization. Thus, the following discussion assumes that there will be data collection centers (DCC's). Network maintenance and data distribution also imply centers which will be referred to as maintenance centers (MC's) and data management centers (DMC's). Despite their functional separation, the operations of data collection and network maintenance are necessarily closely related since maintenance problems will usually be discovered during data collection and since messages to stations may be transmitted over telemetry links required to collect data. If a sub-network is defined as an ensemble of stations which are maintained by one maintenance center, then all data from each sub-network, and no data from any other, must be collected by one data collection center which should be co-located with its associated maintenance center. In contrast, the data management centers are simply the destination of the finished product of the data collection centers. The relationship of sub-networks, data collection centers, maintenance centers, and data management centers is shown schematically in Figure 2.1. The number of networks, sub-networks, and data management centers is variable. The IRIS network may consist of one sub-network or several and other networks may be divided into several sub-networks. Notice that all IRIS data collection centers will send their data directly to all IRIS data management centers. Data exchange between cooperating networks will be performed by the respective data management centers.

2.2. Functions of a Data Collection Center

Data collection center functions include:

- acquisition of all data from one sub-network,
- communication with station operators and station processors,
- monitoring station supply inventories,
- monitoring data completeness and quality,
- reporting data problems to the sub-network maintenance center,
- maintaining calibration information,
- creating merged data volumes,
- initial data distribution, and

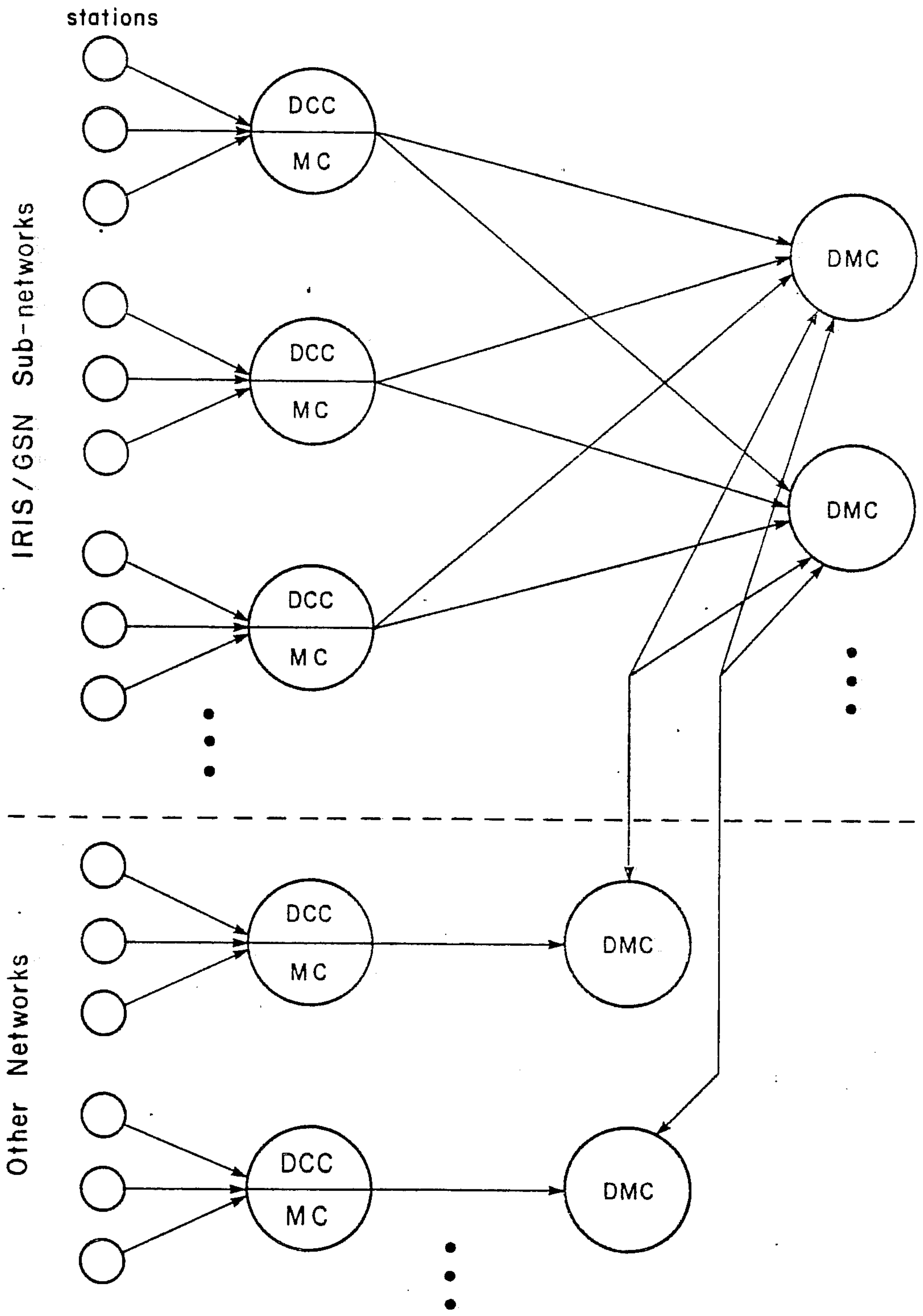

Figure 2.1. Data flow, showing multiple sub-networks, data collection — maintenance centers, and data management centers.

- maintaining a data archive.

Figure 2.2 shows these functions in the context of data flow.

2.2.1. Data Acquisition

Almost all data acquired by a data collection center is contained in digital data streams from the stations. The contents of this stream are described in detail in section 1. To summarize, the bulk of the data is compressed, three component, continuously transmitted, broadband seismometer output. In addition, other special purpose seismic channels may appear along with supplemental channels including environmental, state-of-health, and parametric data. All of these data types must be separated into contiguous time series. In addition, operator logs will be available from manned stations. Relevant portions of the operator logs must be attached to the appropriate time series.

Some of the stations will transmit their data to the data collection center in real time via satellite. The data rate may be variable and, under exceptional circumstances may significantly lag real time. This data must be acquired at the data collection center as reliably as possible. For the forseeable future, all but exceptionally remote stations will record the data stream on-site. Most telemetered stations will thus have completely redundant on-site recording. On-site recording media will contain about two weeks of data and should arrive at the data collection center (by mail) within 60 days of real time. All on-site recordings must be entered into data collection center on-line storage. Redundant data streams must be merged into a single complete and reliable data stream.

2.2.2. Message Text

Routine communications with station operators or directly with station processors is a data collection center function. In the former case, the data collection center provides a convenient single point of contact, being the receiver of on-site recordings and operator logs and the supplier of recording media. In the latter case, messages will often be transmitted over the same telemetry links by which data is received at the data collection center (though for some stations, messages may be transmitted via a station processor dial-up port). This arrangement allows a data

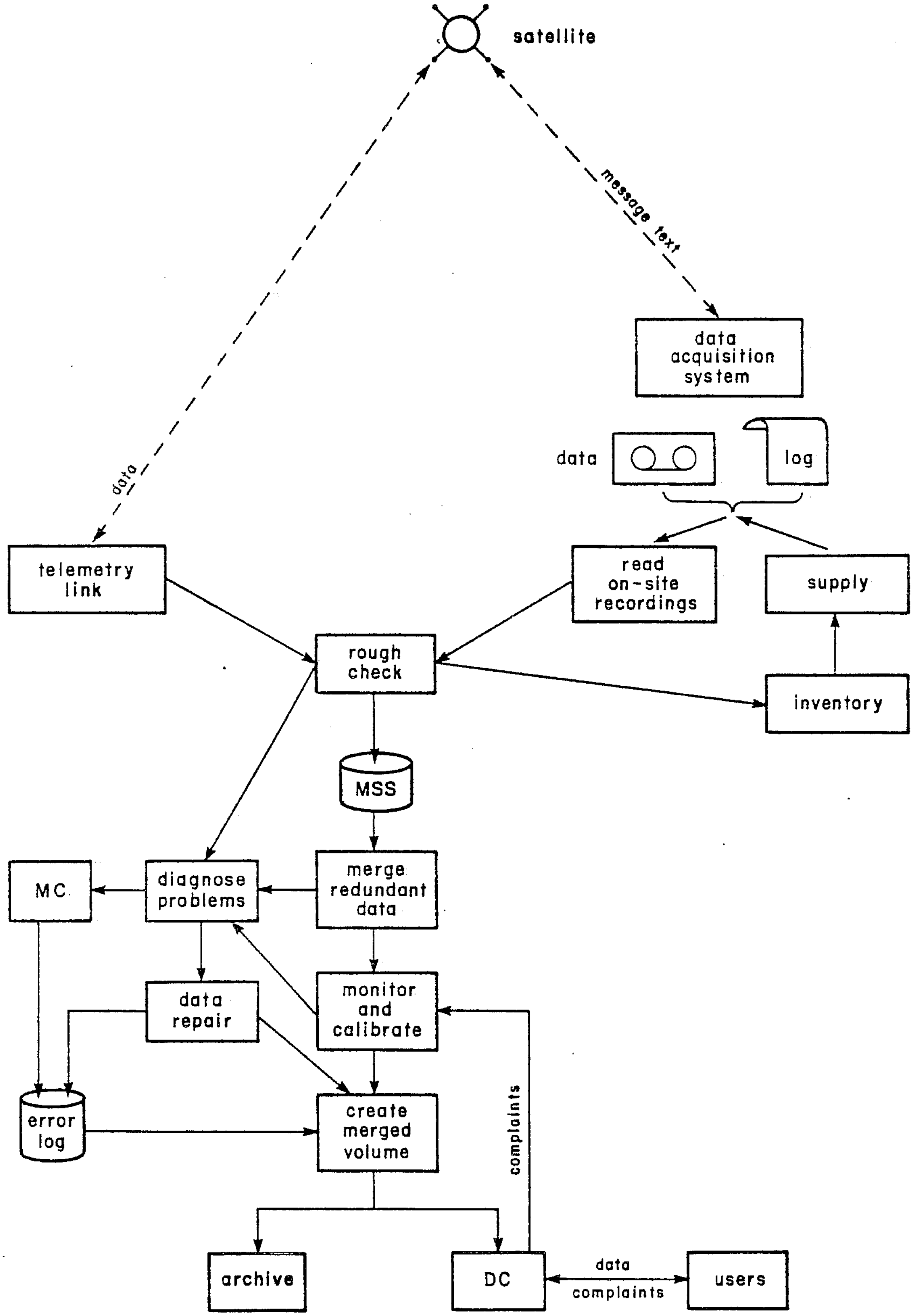

Figure 2.2. Data flow within the IRIS Global Seismographic Network.

collection center to monitor changes in station operating parameters. Messages may be initiated by maintenance center personnel for purposes of remote diagnosis.

2.2.3. Inventory

As the data collection center receives on-site recordings, it will monitor station inventories of recording media in order to ensure that an adequate supply is always available. In some cases, on-site recordings may be returned to the originating station on the original or some other medium as part of an agreement with station operators.

2.2.4. Monitoring

Monitoring data completeness and quality is the single most important and demanding of all data collection center functions. The absolute time, continuity, and identity of each time series must be verified. All data must be examined as completely as the production schedule permits for problems in the sensors, analog-to-digital encoder, station processor, data logger, on-site recorder, and telemetry link. Minor data problems (eg., clock errors, formatting errors, etc.) must be fixed if this can be done unambiguously. A summary of unresolved data problems must be attached to the relevant time series.

2.2.5. Maintenance

All data problems must be reported promptly to the sub-network maintenance center for evaluation. The results of all maintenance visits and field upgrades must be reported promptly to the sub-network data collection center.

2.2.6. Calibration

Complete calibration information for each station-channel will be available from the associated maintenance center. Relative changes in gain will be monitored by the data collection center through information periodically placed into the data stream by station processors. Calibration information must be attached to the appropriate time series.

2.2.7. Merging

All redundant data streams must be combined to form one complete and reliable data stream. Merged, verified, and corrected time series, with all of their attached information, from the entire sub-network for one time period, must be combined into a "merged data volume." By making this merged data volume complete within itself, requests for event data can conveniently be satisfied by the data management centers. Supplemental data channels should be included as separate time series and the volume should include information from operator logs, data quality monitoring, and calibration.

2.2.8. Initial Distribution

A data collection center will distribute merged data volumes to associated data management centers only. Copies of on-site recordings from stations with no real time telemetry also may be distributed to data management centers in advance of merged data volumes. Note that timely distribution to users could also be accomplished by the data collection centers sending data to the data management centers as soon as it is verified. Merged data volumes would then be compiled independently by each data management center. This approach involves significant duplication of effort to satisfy a small percentage of the data requests.

2.2.9. Archiving

Experience indicates that it is worthwhile to archive merged data volumes and, if possible, all raw data streams as well at the data collection centers. In addition, facsimiles (perhaps on microfische) of operator logs and output from data quality monitoring and calibration processing should be saved to help resolve historical problems discovered long after initial distribution.

2.2.10. On-line Analysis

It is clearly in the interest of the seismological community to make real time data available to facilitate the rapid determination of seismic event parameters (eg., the National Earthquake Information Service). While direct support of such work is not a data collection center function, data collection centers will benefit by close cooperation with analysis organizations. Analysts performing near real time parameter estimation may be the first seismologists to examine data quality. Conversely, analysts will require data collection center information on known, real time data problems.

2.2.11. Exclusions

The acquisition and organization of real time telemetered data for on-line analysis is not a data collection center function. The acquisition and organization of real time telemetered data and event data which may be available via station processor dial-up ports (for non-telemetered stations) and the organization of on-site recordings distributed in advance of merged data volumes will be done by data management centers with the understanding that all such data will be superceded by the merged data volumes. Also, data management centers will acquire and organize data from other data collection centers handling associated sub-networks and from other data management centers associated with other networks.

2.3. Implementation of a Data Collection Center

Despite the substantial uncertainties at this early stage of global seismograph network planning, some guidelines may be established for data collection center hardware, operations, personnel, facilities, and security. These guidelines are based on preliminary estimates of a variety of network parameters and on what computer technology has to offer today and in the near future.

2.3.1. Network Parameters

The global seismograph network will have about 100 stations, most with real time telemetry and (probably) most with on-site recording as well. Approximately 5 to 10 years will be required to completely deploy the network currently planned. Telemetry rates will be no higher than 1200 bps/station (bits/second/station), but may average as little as 600 bps/station. This represents a worst case of 12,000 sps (samples/second) or an aggregate bandwidth of .12 MHz (in serial form) for the 100 station network. Table 2.1 shows the daily aggregate data volume in GB (giga-10^9-(8-bit)-bytes) for both average and maximum data rates for number of stations, N, in the global seismograph network versus number of telemetered stations, M. On-site recorded data volumes will contain about two weeks of data, so on the average 7 volumes/day will be received at data collection centers from a full global seismograph network. All on-site recordings should be

Table 2.1
Data rates in GB/day for N stations with M telemetered
(minimum/maximum)

M	0	10	25	50	75	100
N						
10	.07/.13	.13/.26				
25	.16/.32	.23/.45	.32/.65			
50	.32/.65	.39/.78	.49/.97	.65/1.3		
75	.49/.97	.55/1.1	.65/1.3	.81/1.6	.97/1.9	
100	.65/1.3	.71/1.4	.81/1.6	.97/1.9	1.1/2.3	1.3/2.6

Table 2.2
Active storage in GB for N stations with M telemetered
(minimum/maximum)

M	0	10	25	50	75	100
N						
10	3/5	6/12				
25	6/13	10/19	15/29			
50	13/26	16/32	21/42	29/58		
75	19/39	23/45	27/55	36/71	44/88	
100	26/52	29/58	34/68	42/84	50/100	58/117

received by data collection centers within 60 days, at which time merged data volumes will be created. As redundant telemetered data may be discarded once merged with the on-site recording, about 50 days of telemetered data and about 40 days of on-site recorded data (allowing for some overlap) must be saved at any one time pending the creation of a merged data volume. Table 2.2 shows this aggregate buffered data volume in giga-bytes for the same range of conditions as in Table 2.1.

2.3.2. Technology

Fast micro-computers and mini-computers have bus bandwidths ranging from 1 to 30 MHz (in parallel) and processor speeds of .3 to 4.5 MIPS (million instructions per second). Although much faster computers are available, their cost-performance ratio for this application is not attractive. Local area networks for connecting a number of micro- and mini-computers are available with serial bandwidths in the range of 2 to 10 MHz. Dual processor and clustered mini-computers are available where the interconnect is actually a shared computer bus.

High performance magnetic discs are readily available in a variety of sizes up to 1.2 GB. The maximum aggregate disc storage which may be connected to a large mini-computer ranges from 4 to 40 GB. Off-line storage on magnetic tape is limited to .18 GB/volume (2400 foot tape at 6250 bpi). Off-line storage on optical digital disks range from .3 to 4 GB/volume and should be generally available in about one year. A number of mass storage devices are now or soon will be available. For example, automatic tape libraries using standard 6250 bpi tapes are available in capacities of 100 to 1000 GB. Magnetic tape cartridge or spool mass storage devices are available with on-line capacities up to 470 GB. Optical digital disk carousel or "jukebox" systems with capacities in the range of 100 GB should also be available in about one year (such devices are already available overseas). Note that a mass storage device requires considerable software to become an mass storage system and that an existing mass storage system may require considerable effort to interface with a particular computer. Generally mass storage devices consist of a large number of data volumes which are mounted in a reader/writer by a robot mechanism. These data volumes may be selectively taken off- and on-line (removed from or restored to the device) as desired.

2.3.3. Hardware

Real time telemetered data must be buffered on-line until preliminary data monitoring can be done and then saved until either the redundant on-site recording is received or until the merged data volume is created. Likewise, on-site recordings must be read as soon as they are received, checked for quality, merged with redundant real time data (if available), and saved until the merged data volume is created. In both cases, it is vital that the data be examined immediately to detect field malfunctions. Because of the labor often involved, field recordings should be read only once. Maintaining all active data on-line is technologically feasible and will simplify development, speed up operations, eliminate labor intensive bottlenecks, and facilitate a high degree of automation. A data collection center handling a small number of stations may use conventional magnetic discs for on-line storage and conventional magnetic tape for initial distribution and archiving. A data collection center handling a large number of stations may require a mass storage device for on-line storage and use the mass storage data volumes for initial distribution and archiving. The dedicated, and relatively simple storage requirements of a data collection center may allow direct use of a mass storage device without the necessity of the generality and complexity of a true mass storage system. Also, a substantial amount of magnetic disc "scratch pad" storage must accompany any mass storage device (enough to hold 2 to 5 days of the aggregate output of the sub-network). Thus, software for data collection centers need manipulate only scratch disc storage. Only the mass storage hardware need depend on the size of the sub-network handled.

In order to eliminate interference that could result in data loss, real time data should be acquired by a dedicated computer. The data rates indicate that a fast micro- or moderately fast mini-computer could easily handle the worst case load. Non-real-time processing will probably require fast mini-computers since, for the worst case aggregate global seismograph network output, only 83 microseconds/sample will be available even if processing is done 24 hours/day and 7 days/week. These considerations dictate that several computers must interface with the same mass storage at each data collection center. Some sort of computer network will be required. This strategy has many advantages including:

- data collection centers handling different size sub-networks can share the same architecture and software,
- the time available to examine each sample from one sub-network can be increased by splitting the data stream between several non-real-time processors,
- processing power can grow with each sub-network,
- unanticipated processing loads can be accomodated without redesigning the entire system, and
- upgrades can more easily be made, delaying system obsolescence (the life cycle of a computer system is 3 to 5 years, a short time compared with the design life of the global seismograph network).

Although most of the data must flow smoothly through the system to ensure proper data collection center operation, problems requiring special equipment will inevitably arise. For example, real time visual monitoring of telemetered data with known or suspected problems would be very useful. Other problems, such as reading incorrectly formatted tapes, will require specially equipped computer work stations. In fact, experience indicates that reading on-site recordings may take as much as several man/machine hours each. The visual examination of stored data will require very high speed interactive graphics terminals and highly automatic, high resolution hardcopy devices such as laser printers.

2.3.4. Operations

Given a computer network and a mass storage device, all known bottlenecks in data acquisition, organization, initial distribution, and archiving can be eliminated. The task which will dominate both hardware and operational considerations is data monitoring. Monitoring operations applied to every sample must be very simple and highly, if not completely, automated. More sophisticated monitoring, including visual examination, can only be applied to a small portion of the data. Minor repairs (e.g., time errors, continuity, and polarity reversals) to the data should be attempted when they can be done unambiguously in a reasonable amount of time. Continuous input from the user community (via the data management centers) will be needed to determine what data are not of sufficient quality to distribute.

Similarly, the algorithm used to merge redundant data streams must be simple and highly automatic. If both telemetry and recording mechanisms are functioning properly, current technology indicates that the recording will probably be significantly more reliable. However, arrival time information inherent in telemetry can be used to check acquisition system clocks. Situations in which both the telemetry and recording are suspect will require sample-by-sample decisions to recover a correct and complete data stream.

A scenario for data monitoring operations is:

- perform sample-by-sample monitoring on all data streams as soon as they are acquired,
- merge redundant data as soon as the on-site recording is received,
- perform sophisticated monitoring on the merged data stream.

Experience indicates that the time available for data monitoring is never sufficient to detect all data problems. Thus, data monitoring procedures and algorithms should evolve in response to data user problems reported to the data management centers. Ideally, imaginative user developed monitoring techniques (eg. using tidal analysis, synthetic seismograms, or new spectral techniques) would be incorporated into data collection center operations. This one aspect of data collection center-user interaction should be encouraged as much as possible, perhaps by means of visiting scientist programs.

2.3.5. Personnel

Clearly, there are important trade-offs between levels of computer automation and human intervention and between staffing levels and operating schedules. Obviously, real time data acquisition must be completely automatic and can never stop. Likewise, sample-by-sample data monitoring must operate at all times, as automatically as possible. Thus, a data collection center must be staffed by custodial personnel at all times. Further, hardware and software maintenance personnel should be on call at all times. However, there seems to be no reason why many tasks such as calibration monitoring, creating merged data volumes, initial distribution, archiving, software development, and inventory control should not be single shift tasks. More demanding tasks such as reading incorrectly formatted on-site recordings, sophisticated data monitoring, and data repair will require a number of people and perhaps several shifts.

2.3.6. Facilities

Each data collection center will require adequate space for computer equipment, offices, and supply and archival storage. Existing facilities should be utilized as much as possible. The computer facility will require raised flooring and adequate conditioned power and air conditioning. Special ventilated space will have to be provided for uninteruptible power supply equipment.

2.3.7. Security

The primary enemy of smooth data collection center operations is system failure. The risk of a catastrophe is probably not sufficient to justify extreme measures such as complete, duplicate, backup facilities. Given the data value, cost versus risk does favor other security measures including:

- regular backups of operating software,
- storage of some backups at a remote location,
- fire, water, heat, humidity, and power alarms,
- a non-destructive fire extinguishing system (eg. haylon),
- well documented emergency procedures,
- spares of high maintenance or critical equipment,
- complete redundancy of real time data acquisition hardware,
- power conditioning,
- 10—15 minutes of uninteruptible power supply (UPS) for the entire computer facility,
- 6—12 hours of uninteruptible power supply for real time data acquisition equipment,
- perhaps a diesel generator for extended power interuptions.

In spite of precautions, there will be system failures. Reliable local maintenance must be available for all equipment. Maintenance for critical equipment should have a guaranteed response time no longer than a few hours. In addition, in order to recover from extended down time (and to handle exceptional seismic events), there should be 25—50% excess capacity, relative to the nominal load, in both processing and mass storage equipment.

Mass storage hardware is a special security problem. When it fails, data it contains is unavailable and most data collection center activities will stop. Thus, redundancy for critical mass storage storage device components is a necessity. In addition, provision should be made for conventional off-line storage (magnetic tape) sufficient to prevent data loss and support at least minimal data collection center operations during extended mass storage hardware down time.