THE DATA LABORATORY

by

Weichert, Anglin, Hayman, Lewis, Mair, Leblanc

Seismological Service of Canada

INTERNAL REPORT 75-5

Seismology Division

Earth Physics Branch

Department of Energy, Mines & Resources

1975

THE DATA LABORATORY

1. Introduction

2. Constraints & General Policy

3. Configuration

   a) The old DDP-124 and peripherals
   b) The Eastern Canadian telemetered network-dedicated PDP11-15
   c) The PDP11-40 system
   d) The CANSAM on-line system
   e) The Western Canadian telemetered network system
   f) Physical reorganization of the Data Lab

4. Divisional Requirements

   a) Special reformatting
   b) Analog playback
   c) Digitization of field tapes and digital tape creation
   d) Digital tape editing, reformatting for CYBER, stacking
   e) Digital/analog playback and display
   f) Interactive Programs
   g) On-line event files, real-time uses

5. Staffing and Duties

6. Conclusions

1. INTRODUCTION

Data acquisition, recording practices and modes of analysis in seismology have changed dramatically over the years. Starting with one playback deck for analog field tapes, the Division has over the years acquired several small computers, which are used for as simple a process as digitizing and formatting of digital tape to the relatively complicated tasks of real-time beam forming of array data, interactive processing of multichannel recordings, and displaying results on a CRT screen.

The original justification for setting up the in-house digital special-purpose processing facility, outside the jurisdiction of the Departmental Computer Science Centre (CSC), was the massive computing time requirement for a pseudo-on-line detection program for the Yellowknife array. Moreover, the needs of the Division for digital data acquisition were not well catered to by CSC, and the in-house facility was consequently expanded to serve these requirements. At most stages of this development, it appeared that the available hardware was lagging the immediate Divisional user demand. This situation changed over the last few years, precipitated by the dramatic decrease of the computer cost/performance ratio and the policy of the former Division Chief to keep abreast of new technological developments for their application to seismology, leading to the hardware currently available or still on order.

This study group has been formed to examine the current situation in the light of new management, the recent re-organization of the Division and changes in project priorities, and to recommend and coordinate plans for the operation and further development of the Datalab.

2. CONSTRAINTS & GENERAL POLICY

During the various phases of hardware expansion of the current data laboratory, some opposition was frequently encountered from the CSC. Now Departmental computing policy appears quite clearly defined: although small computers not under the jurisdiction of CSC have multiplied, their use is to be limited to special-purpose operations such as on-line, real-time application, interactive editing or preliminary reduction of massive data flows, or for guaranteed 24-hour, 7-day accessibility. It was agreed that we must not attempt to create a small, general-purpose computing facility. The CSC has acquired a proven record of fast turnaround and reliability where batch jobs are concerned. Priority for time on our in-house equipment must be given to "non-standard" projects for which use of CSC facilities is not practical. In case of real doubt, a program should be developed to run at CSC. Notable exceptions are standby programs needed on an emergency basis, such as an epicenter determination program which might well be required early on a Sunday morning.

Within this constraint on its development, the group discussed different approaches to the Datalab, eventually rejecting the total systems approach and agreeing that the individual scientists will have to be closely involved in programming their specific applications.

Scientists may now accept FFT programs or other complicated packages and use them after thorough testing, but do not, in general, yet accept programs as black boxes similar to amplifiers or other products of a proven technology. To understand the results and to gain confidence in their interpretation, it is still considered necessary that the scientist handle data at a very basic level, if only in a supervisory capacity. This means he should and wants to take a leading part in constructing his data-handling programs.

More practical reasons for this attitude are found in manpower considerations and in experience with program packages. Firstly, it is not realistic to let the individual scientist-user wait until a package is designed for his particular need; secondly, most or all scientists have experience with trying to understand and use other peoples' programs: often they end up wishing they had written their own. From the point of view of providing independent checks on methods and procedures, this would be preferable in many cases.

Thus, the desired approach to applications programming in the Datalab is still similar to that of a few years ago. The user requirements should be taken care of with a few Fortran-accessible simple service routines, and the scientist-user be allowed to build around these his own logic as complicated or simple as he wishes.

3. CONFIGURATION

A descriptive outline of the current Datalab hardware and software was deemed desirable as a basis for discussion for those members of the Study Group not intimately familiar with the developments of the last few years. The list is included in this report to make it more self-contained.

a) The old DDP-124 and peripherals

i) DDP-124 computer, 16 K - 24-bit memory CPU;
input/output:
paper tape reader/punch, typewriter,
2 7-track digital magnetic tape decks
30 channels analog/digital input (22 simultaneous-sample-and-hold)
10 channels digital/analog output
12 discrete input and output lines

ii) analog playback:
2 7-channel, ½-inch Sanborn analog playback decks
1 24-channel, 1-inch EMI deck
1 14-channel, 1-inch Honeywell deck
24-channel analog filters with several short-period pass bands
several channels of Khronhite filters

iii) visual display:
2 8-channel hot-wire trace recorders
1 CRT storage display

The system is supported by an adequate software system, including Fortran, a mnemonic machine language assembler and several general-utility routines. Compared with modern disk-oriented operating systems and associated software, the DDP system is outdated and program creation is awkward. New program development is wasteful in time and should be avoided except for cases where an equivalent hardware configuration is not yet available in the new system.

Moreover, the DDP system is now no longer supported by the manufacturer; simple malfunctions have been repaired by our staff, but major breakdowns may result in quick decommissioning of this system. We have called the system obsolescent for several years; now "obsolete" is probably fair. Over a relatively short time range the newer PDP11/40 system should therefore be configured in such a way that the DDP tasks can be transferred on short notice. It is not, however, recommended to discard the DDP mainframe and peripherals as long as they run, as their space requirement in a physical reorganization of the Datalab is not considered objectionable.

b) The Eastern Canadian telemetered network (ECTN)-dedicated PDP/11/15

This is currently a 16-K - 16-bit configuration, dedicated on a 24-hour basis to receiving and processing 4 digitally telemetered seismic channels. It has 4 D/A channels connected via band pass filters to helicorder monitors; teletype I/O and punch are available and can be attached. A 1.2 M word disk and the hardware for a data channel to the PDP/11/40 (cf. c) has been installed. The ECTN program runs currently under RSX11-A, a real-time multi-programming operating system, which should be updated to RSX11-M for easier communications control, unless this requires a second RK05 disk drive. ECTN receives 4 digital seismic channels over telephone lines, reformats and outputs them to helicorders, and operates an envelope trigger on each channel in a certain pass-band. Disk circulation and transfer of data to magnetic tape via the 11/40 has not yet been implemented. It is recommended that Hayman implement the data circulation over the disk, i.e. the arrangement of disk buffers and delays, and the transfer to the 11/40 via the DR11 data channel. Anglin will then take responsibility for further development of the ECTN data handling and reduction.

Still existing and future problems in design concepts, development and operation of this system will be discussed and resolved by a standing Datalab Committee as recommended in Section 5.

c) The PDP/11/40 system

This system has currently 24-K - 16-bit memory, a 1.2M word cartridge disk, one 9-track digital magnetic tape deck, a DECwriter (kind of typewriter) I/O, card reader and lineprinter and high-speed paper tape reader and punch. At the present time a disk-oriented operating system (DOS) is available, which allows very flexible and fast program development on this machine.

The main reasons for acquiring this system are the need for program development for the on-line CANSAM in Yellowknife, for the telemetered arrays in eastern and western Canada and as a replacement for the obsolete

DDP system. For this last purpose, we need analog-to-digital input, a second digital tape deck, sufficient digital/analog output channels and a few discrete output lines for automatic (and semi-automatic) control of tape deck functions. Analog/digital inputs and one 16-bit output register are designated a high priority and should be ordered for delivery early in the 1975/76 financial year. A DEC interactive GT44-type display unit with lightpen input is on order for delivery in December. Associated software allowing Fortran-access to this CRT is not yet available on the market, but is promised under RT11 (cf. below).

An RT11 foreground-background operating system is on order and will complement the present DOS. This system will, hopefully, facilitate inter-computer communication, etc., and simultaneously allow new program development, or other utility operations to occur in the background.

d) The CANSAM on-line system

Although not explicitly included in the terms of reference for this Study Group, it is considered to obtain a better overall view. This is a PDP/11/45-centered system with one digital magnetic tape, a 1/4M-word, fixed-head disk, a teletype and a Datapoint 3300 display unit, all located in Yellowknife. The disk is used as delay buffer for the short-period array data. The magnetic tape deck produces an edited digital data tape; on the typewriter a detection list and a paper tape copy is produced. The CRT is used for operator intervention. A suggestion has been made to repatriate the CRT and substitute a second teletype which would also serve as a backup for the bulletin device, but this is not planned for the near future. Currently, this system also runs under DOS, with a limited foreground-background capability achieved by running the main beam-forming and detection program under ODT, the on-line debugging utility. The RT11-operating system should be installed there also, if the fixed-head disk will allow this. This should ease the problems of computer communication with Ottawa and would also give the YKA operator a true foreground-background capability to allow for minor program development.

e) The Western Canadian telemetered network (WCTN) system

This will basically be a copy of the ETCN (cf. b). WCTN is at present of only tangential interest to this group, but the possibility of program and data exchanges and a later direct computer link must be kept in mind. Note that WCTN will not be a dual processor system and the basic software supplied (ECTN 11/15 Software) will need to be modified by Victoria Staff to support TM11 tape in place of DA11 printer processor link.

f) Physical reorganization of Datalab

Present layout of facilities is historical and haphazard, and almost any change would be an improvement. The group has not considered a new layout, apart from general recommendations on noise abatement measures, moving one key punch and the CSC intercom terminal upstairs and provision of quieter office space in the north-west corner of the lab. The recommended Datalab discussion group will plan this reorganization.

4. DIVISIONAL REQUIREMENTS

Present and envisaged processing requirements listed here are based on, and include, all suggestions initially submitted by Section representatives. It is not implied that Datalab would ideally serve each particular need; on the contrary, it is specifically recommended that certain requirements are better served by CSC. However, the list under (f) shows the latent interest in in-house interactive computing, which would take up any slack in operation.

a) Special reformatting

In the past, data on paper tapes have been converted to digital magnetic tapes for the Geothermal Section. This required about 1 hour every month. Since it is to continue indefinitely, the fast paper tape capability must be maintained. Similar conversions for "backpack" cassette tapes to standard magnetic tapes will soon be required on a larger scale.

b) Analog playback

One important role of Datalab is to provide the necessary facility, hardware and assistance for the playback of analog magnetic tapes after field surveys. Digital playback via an intermediate computer is not needed for picking arrival times of phases and comparing relative amplitudes.

i) The facilities for ½-inch high-speed tape exist and should be preserved; the one-to-one ratio is ideal.

ii) The facilities for ½-inch L.S. tape are not judged to be ideal but acceptable as the playback/recording ratio is 10, thus creating a reproduction problem at high frequency (150 Hz max).

Max freq. of LF

iii) For the 14-track tape (1") used with the portable array, the analog playback facilities are not adequate, as the playback/recording ratio is 20. Although a multichannel (e.g. 14) recorder with about 1KHz response may be desirable, it was agreed for the present time that the available facilities can be used, but that the problem will require further attention by the Datalab Committee.

All analog playback facilities suffer to varying degrees from a common problem: lack of a time code makes positioning on the tape difficult and time consuming, and often prevents obtaining conclusive evidence on the absence of small signals in the records. It is, therefore, recommended that, within the prearranged priorities of the Instrument Section, a suitably coded time identification be added to the present time track. For the array this should occur before the next planned field deployment, while for the other systems implementation is less urgent. This time code could then be interpreted by the computer, and tape positioning and event search can be partially automated. Requirement b(iii) and c(ii) may then also strongly interrelate and eventually be eliminated by developments under e. below.

(5)

It is recommended that Hayman investigate the possibility of installing electronics for a single channel (time) for all speeds available on the Honeywell tape deck.

(6)

c) Digitization of field tapes and digital tape creation

i) Various groups have in the past used the Datalab predominantly as a sophisticated digitizer. High speed and low speed ½-inch FM modulated tapes were played back, prefiltered and channelled into the DDP124 through a plugboard. The required sampling speed of 1000 Hz on 7 channels (7 KHz) is within the DDP capabilities, both for the digitizer (7.33 KHz) and the magnetic tape drive (about 15 K seismic samples/s, depending on block length). This capability can be maintained on the PDP-11/40 when the currently ordered A/D channels are installed. Mair will take the lead in program conversion.

ii) The Seismicity Group will have the requirement to digitize the 1-inch, 14-channel FM tapes from the portable array. Digitization facilities of these tapes are currently inadequate, but as a consequence of group discussions were given high priority as indicated earlier. On the DDP system digitization may currently be done with multiple passes, using only a subset of all channels; on the 11-40 the limiting speed will be the tape drive at 15K samples/s for continuous digitization or the 34 KHz A/D speed for shorter record section that fit the disk (its speed is about 50K samples/s); this would allow, e.g., 14 channels at 50 Hz, in 20 x real-time speed, for a Nyquist frequency of 25 Hz, continuously. Record section of customary lengths could be stored on disk and later transferred to digital tape, thus, in effect, more than doubling the capacity to over 100 Hz (50 Hz Nyquist). The lead role for implementation of this facility goes to the S.&.S.H. Section, with assistance, as required, from Weichert and the Instrumentation Section.

iii) YKA-LP tapes. The 1-inch, 14 channel tapes have occasionally been digitized on the DDP and used for beamforming; no problems are foreseen.

iv) YKA-SP tapes. The edited digital CANSAM tapes will not obviate the necessity of digitizing the 1-inch, 24-channel EMI tapes. The number of A/D channels now being ordered will serve this requirement. Responsibility for leading the conversion will fall on the scientist planning to use these data. Assistance will be provided by Weichert.

d) Digital tape editing, reformatting for CYBER, stacking

In the past it was found indispensible to have the digital tape-handling capability for the above purpose, which requires considerable direct user interaction. This was afforded by having 2 tape decks on the DDP. It is recommended to maintain this capability by adding a second 9-track tape deck to the present 11-40 configuration.

e) Digital/analog playback and display

Many of the above applications require a visual monitor output. The available Sanborn recorders will do if the required number of D/A channels are installed. However, the presently available 4 channels may suffice, especially if good use can be made of the GT44 display and the hardcopy device planned as a possible addition after the CRT is operational.

GT44 - deliver?

Programming for optimum use of the GT44 will be one of the most urgent projects in the short range. Suggested is a Fortran-accessible plotting package compatible with CSC plotters, and similar to the one written by Anglin for the DDP124. As soon as the GT44 is installed and support documentation delivered, Anglin, Hayman and Weichert will cooperate in setting up specs and deciding how to utilize this display in a most useful way.

f) Interactive programs

A number of programs previously running interactively on the DDP, plus several other programs, have been suggested for running in the Datalab:

i) EPDET, a teleseismic epicenter determination program using P phases only, (Weichert).
ii) HYPO 71, or CANSESS, epicenter determination using other phases as well, (Anglin).
iii) ARA, a versatile multi-option, short-period, pseudo-real-time array data processing program, (Weichert).
iv) ARAFT & ARACRT, non-real-time interactive programs with options for analysis, e.g., filtering, least squared $dT/d\Delta$ and azimuth solutions, in both the frequency and time domains, Fourier spectral analysis.
v) a number of other utility programs for data acquisition, editing and processing, (Weichert, Anglin).

low priority

vi) a stacking program.
vii) seismic ray tracing.
viii) adaptive beam-forming à la King and Mereu.
ix) f-k spectra forming.
x) power estimation using maximum entropy or maximum likelihood.

General agreement was reached that programs of the type vi to x should have very low priority in the Datalab. Firstly, they violate the non-general-purpose-computing-center restraint rather radically; secondly, many already run on bigger machines as batch jobs and it would be difficult to fit them to 24-K - 16-bit memory. Only in case of urgent demand for scientist interaction should such programs be moved into Datalab.

Programs of type v have been considered already: they are naturals for Datalab. Program types iii and iv are not basically different from types vi to x, except for their options requiring analog input or operator interaction. It is not recommended to convert these programs from DDP to PDP until an actual requirement arises. It should only then be decided by the Datalab Committee, whether program options should be moved into Datalab or to CSC. New program development of a similar type for CANSAM and ECTN data are planned within the constraints.

Programs i and ii fall under the 24-hour access demand and should be transferred soon, since they are urgently requested from various sides (earthquake contingency planning, public relations, seismicity). Responsibility for this is indicated under i & ii.

g) On-line event files, real-time uses.

We are committed to making the YKA detection file accessible to outside users on a near real-time basis. The most probable way of accomplishing this is via automatic-dialling modem attached, initially, to the 11-40 for development purposes. The YKA detection file would be called in daily, edited and placed on a CSC file accessible to outside callers (US, UK). The facility may later be transferred to the 11-15, but the detection file will be kept for outside access on the CSC computer. Technical details are left to the discretion of the staff directly involved in this project. (Hayman, Weichert).

Pressure from outside agencies (USGS) to modernize and expedite our transmission of phase reports raises the question of moving the data flow and files of the standard Canadian network into or through the Datalab. Discussions on basic policy (c.f., constraints) and utility have been held and will continue for a while. The demand for quicker CSN data transmission to international agencies, similar to YKA data, will be met by writing the CSN daily phase data directly, after decoding, on a disk file at CSC instead of punching this file onto cards and sending it to Colorado via mail. USGS can then call CSC and read out their file. Responsibility for implementation rests with Shannon. Billing considerations are beyond the terms of reference of this Group.

In the longer term the current difference between the YKA detection file and CSN data may disappear. YKA may develop into a prototype digital Canadian standard station and the data flow unified under the same management. As above, decisions on the precise way of doing this belong mainly in the technical domain with some policy input from above, and are not within the terms of reference of this Group.

5. STAFFING & DUTIES

Work must be done in three areas: firstly, system programming and building is not completed, e.g. data exchange between the 11-15 and 11-40 require DOS driver and its counterpart RSX driver, telephone linkage to Yellowknife is not implemented. New versions of the operating systems should be obtained from DEC (as they appear) and installed as long as reasonable returns, i.e., improvements, for our operation can be expected; this should not exceed the order of 2 weeks per annum. Secondly, development of programs to ensure the various user requirements is still in an early stage and should be accelerated. Thirdly, the user himself must enter the scene more forcefully, thus providing some feedback and more stimulation to further development.

Under the constraints of available staff, the following shifts in duties have been considered by the Group and have also been discussed with the people involved. No guarantee can be given, but a reasonable hope exists that these recommendations are workable.

J. Edwards should be encouraged to upgrade her skills, possibly by attending a course yet to be identified; she should then look after system updating with nominal supervision and also be able to take greater part in

generating and installing, from time to time, special system programs. Thirdly, she will play a stronger part in development of user programs, preparing and debugging programs in Macro and Fortran from rough program manuscripts provided by the scientists involved. In the past she has performed similar duties under more or less close supervision.

W. Tyrlik will assume a considerable part of J. Edwards' former duties. In addition to the digital B.C. long period data, which he handles currently, he will receive and catalog the day-to-day data flow through the Data lab. At the present time this includes receiving and cataloging CANSAM detection lists and edited event tapes as well as analog long period and short period tapes, and analog tapes from field experiments. In the future this responsibility will be expanded to tapes from ECTN, the SRO (seismic research observatory), back-pack cassette tapes, and possible others. He will assist in providing playback of events from the tapes as requested by scientists, using analog facilities or computer programs provided, and digitize record sections as required. In addition, his long experience as a field officer must be utilized by involving him substantially in consultation with regard to, and organization of T.F.S.S. field equipment requisitions.

The ECTN has been developed to a point where a more systematic evaluation by the user-scientist becomes highly desirable. F. Anglin has accepted the leading responsibility for developing the seismological application of ECTN, as indicated under 3b. He has already started to familiarize himself with the 11-40 system, so as to understand and influence the flow and format of edited digital data to the point where he will participate or lead in setting up routine display, editing and dumping programs for the data flowing from the 11-15 disk to magnetic tape via the 11-40. This will also aid and accelerate considerably his experimenting with the real-time programs.

The load of application programming, i.e., of providing a compatible (self-consistent) set of Fortran-accessible routines to serve the requirements set out earlier, should be carried more or less equally by the research scientists most closely involved with the project, with guidance by Weichert as necessary. Anglin and Weichert have agreed to look after implementing epicenter programs, with Weichert rewriting the present EPDET and Anglin rewriting CANSESS to fit the 11-40. When the programs are running, a number of staff, as recommended by the Study Group on Earthquake Contingency Planning, must learn to operate the 11-40, access these programs and learn to use them.

Weichert will write the equivalent of the DDP utility routines for digitizing and digital tape formatting, as the need arises. C. Wright might conclude his plans for $dT/d\Delta$ measurements on the DDP124, but if his work is to be transferred to the 11-40, or new programs are to be used, he will participate extensively in programming.

Shannon has expressed interest and initiative to become involved in the Datalab. His long range interest can generally be described as covering the automation of the dataflow, quality control and archivation of the CSN data, for which he has the support of his Section. It is, therefore, recommended that Shannon's responsibilities be slowly shifted into the Datalab, as personnel become available to relieve him of part of his present responsibilities.

To avoid problems that have arisen in the past it is recommended to setup a standing datalab Committee. This will consist of an informal discussion group centered around Anglin, Hayman and Weichert and meeting about monthly, or as necessary. Other staff will join in the discussion according to their interest and need. This Committee will ensure that the seismologists are involved in future system design from the beginning, and routes of useful feedback and control remain open; it will coordinate projects, adjust their priorities and discuss and recommend equipment purchases.

After the system has stabilized, it may become useful to appoint a Datalab manager to direct the day-to-day operation.

6. CONCLUSION

Discussions of this Study Group were lively and often controversial. Differing viewpoints resulting from the history of development of the Datalab were smoothed out and combined into a course of action as laid out in detail in this report. The general principles arrived at can usefully be summarized as follows.

The aims of development and operation of the Datalab are twofold:

i) the processing, reduction, interactive editing and reformatting of on-line real-time or near-real-time dataflows;
ii) the reformatting & some interactive editing of all special recordings of limited duration.

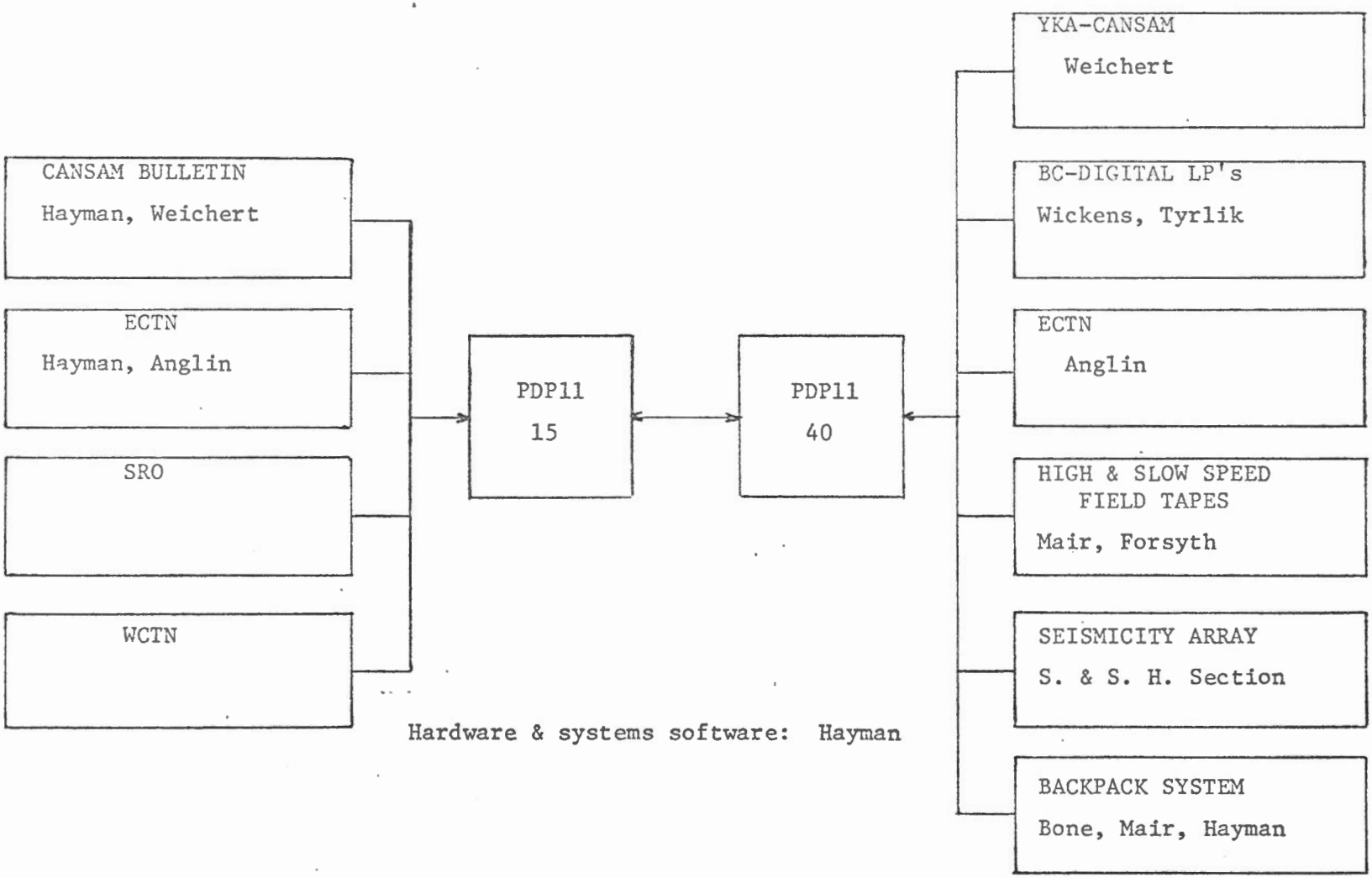
All other computing will be done at CSC, especially of jobs that can be efficiently and simply entered as batch jobs.

Routine data files for access by outside agencies will also be kept at CSC as required.

All hardware development maintenance and system software is done by the Instrument Section under Hayman. He is also responsible for data acquisition and program coordination for the 11/15 system. Application programming for the various projects on the 11/40 will be the responsibility of the scientists involved. A schematic of projects presently planned, the systems used and personnel involved is attached. The initial burden of some general application programming will be carried by Weichert and Anglin, who will later also be available for consultation and assistance to other seismologists.

A Datalab Committee in the form of a general meeting of interested parties about once a month will be instituted to discuss progress of projects currently underway, problems that have arisen or are forseen, equipment purchases, and to arrange job priorities.

DATALAB SYSTEMS, PROJECTS & ASSIGNMENTS OF PRIMARY RESPONSIBILITIES



Hardware & systems software: Hayman

Routine 24-hour processing;
limited editing

Routine near-real time processing &
editing;
Reformatting of special recordings of
limited duration