

NATIONAL RADIO ASTRONOMY OBSERVATORY
Green Bank, West Virginia

December 5, 1991

Requirements for Data Analysis Software for the Green Bank Telescope

*Roger Foster, Martha Haynes, Mark Heyer, Phil Jewell,
 Ronald J. Maddalena, Henry Matthews, Wolfgang Reich, and Chris Salter*

I. Introduction:

The following is the requirements document for data analysis software for the Green Bank Telescope (GBT). The document is meant not only as a working guide for those who will be designing software for the GBT, but, more importantly, we hope it will serve as a starting point for a discussion by other potential users of the GBT. That is, we anticipate that the formal requirements for data analysis software for the GBT will be dynamic over the next year or so as others generate more ideas.

II. The Requirements Committee:

The requirements were researched and collected in the following way:

One of the authors (RJM) was placed in charge, by NRAO management, of all aspects of the data analysis needs of the GBT. He asked various observers outside the NRAO to be members of a committee overseeing the drafting of a data requirement document for the GBT; he also acted as a chairman for the committee. The observers were chosen because of their expertise in the wide range of types of data analysis which are anticipated to be important for the GBT. Two members of the committee were chosen from inside of the NRAO to assist the chairman in aspects of data analysis in which he felt he was weak or in which he felt he needed a second opinion. The committee members, the authors of this document, could gather information by whatever means they desired. Their expertise and responsibilities lie in the following areas, although each could give opinions about any topic:

Roger Foster (NRL):	Pulsar
Martha Haynes (Cornell):	Extragalactic H I
Mark Heyer (FCRAO):	Multi-feed and large scale spectral-line mapping
Phil Jewell (NRAO):	Monitoring telescope performance, pointing, system-support software
Ronald J. Maddalena (NRAO):	Chairman; large scale spectral-line mapping; system-support software
Henry Matthews (JCMT):	Spectroscopy
Wolfgang Reich (MPiR):	Continuum
Chris Salter (NRAO):	Continuum

A charge to the committee, copies of which are available from RJM, was drafted by the chairman and distributed to all members. The committee members were given about a month to submit their requirements to the chairman who then distributed their reports to all members for comments; some questions and requests for clarification were satisfied. The chairman collated the various opinions

into an all encompassing requirements document, which you now have in your hand. Although each author has contributed to the document, it was the chairman's responsibility to put together an adequate document and to see that all common opinions of committee members were represented. The chairman is responsible for most of the failures in this document.

The committee members were then given the chance to examine, critique, and alter the final document until all were satisfied with it. In addition to being a part of the GBT memo series, the document will be used as part of the requirements document for the AIPS++ programming effort, though the requirements presented here are only a small subset of what we would desire to see available in AIPS++. That is, we have outlined the anticipated on-line data analysis needs (see definition below) of GBT observers but have not come close to describing the full data analysis needs of single-dish radio astronomy.

III. The Presentation of Requirements:

The document is arranged as follows:

1. We present some ideas which we hope will be the underlying philosophy behind the GBT data analysis software.
2. We discuss requirements for on-line data access. Some requirements are the responsibility of the group writing monitor and control software, while others are the responsibility of the group writing analysis software. Some fall in the area in between and we suggest that both groups as soon as possible begin discussing these issues.
3. We give a preliminary list of system support software that we deem will be needed for the GBT. The list, which is far from being complete, will be expanded as more is known about the problems the GBT may have that can be analyzed or solved through analysis software.
4. We discuss some general policies and practices concerning data analysis which we would like to see adopted or expanded upon by the NRAO.
5. We discuss general issues concerning the analysis system for the GBT that we would like to see supported by the NRAO. We present requirements for user-interfaces, algorithms, and data presentation for the analysis system which we arbitrarily will call XYZ.

Our requirements, given in sections V through VIII, are rather lengthy and, although we could have given each item a relative priority, we deemed that doing so would be highly controversial and open to misinterpretation. In the sections that follow we will state explicitly whenever one of our requirements is of a low priority. If we do not give a priority to a request, then one should assume that we strongly suggest that that requirement be fulfilled.

An appendix describes some a few terms, which may be new to some or whose meanings are unclear or ambiguous.

IV. General Philosophy:

We have a few general, underlying ideas concerning data analysis with the GBT. Our list of requirements were constructed with these points in mind.

1. We are describing the on-line data analysis needs and not the needs of off-line data analysis. [We will leave the determination of off-line analysis needs to a separate forum.] That is, we are considering what must be available while someone is observing with the GBT. The analysis one needs to perform while observing includes not only checks on the integrity of the data but also having enough analysis capability so that the observer can make decisions during the observation about how to proceed with the observing. In many cases, complicated analysis steps must be performed not only for making decisions but also for checking data integrity.
2. It is imperative that the system used by the astronomer must be familiar to the observer or quick to learn and intuitive in nature. Spending time learning an analysis system while observing, instead of worrying about the quality of the observations, is detrimental to science.
3. We hope that the NRAO will want to be proud of the GBT and will not skimp on data analysis software; we expect most of our requirements will be fulfilled. If resources are limited, however, we would have a serious time deciding what requirements should be relaxed.

We expect to be kept informed of decisions regarding GBT data-analysis; we want to monitor whether or not our requirements are being fulfilled or what constraints are preventing our requirements from being fulfilled.

4. The nature of single-dish telescopes, and especially the GBT, promotes flexible observing styles, techniques, and equipment. The software that must be available should be as flexible as the telescope. It should be easy to modify the software to match new equipment or new ideas.
5. Single-dish data analysis of data has been done, and should continue to be done, in a very interactive fashion. While certain large-scale projects may require batch-like processing of data, most single-dish astronomy will benefit from a highly interactive, fast-to-set-up program with extremely quick feedback.

V. On-line Data Access:

In the following, we are assuming that the control system is taking care of the on-line data storage problems.

1. All the data provided by the backends must be accessible by the analysis system. For example: If a backend has multiple phases, then all phases of data must be accessible. For an Autocorrelation spectrometer, both the ACF and the post FFT data must be accessible.
2. The analysis system must have access to each subscan's data as soon as possible after the completion of every subscan. Because of limitations on disk space, the number of accessible on-line subscans will also be limited. When disk space limitations are exceeded, the oldest subscans can be overwritten by the newest.

3. All subscans are to be archived on some yet-to-be-determined medium. We don't deem a permanent archive a high priority item.
 - a) The archive medium should change as new technologies develop.
 - b) Archives must be copied from older media to newer whenever the physical medium of the archive starts to deteriorate (e.g., if 9 track tapes, all tapes older than 5 years must be copied to newer tapes to ensure the archive will survive.)
 - c) A suitable physical environment must be provided for the long-term storage of the archive.
 - d) A data base must be maintained so that someone wanting to access old data can find where they are located.
 - e) The personnel and software needed to maintain the archive and satisfy requests for archived data must be provided.
4. The observer should, on request, have access not only to data- and analysis-related header parameters but also information about the status of the telescope. For example, the settings of each surface actuator, the voltage drops across various drive motors. This information may prove important to some astronomers in diagnosing possible problems with their data. This request is of moderate priority.
5. In addition to providing subscans, the analysis system must have the ability to access scans (i.e., averaged or concatenated subscans). It is deemed that scans are to include a certain set of header parameters that is a subset of those contained in subscans. Because of limitations of disk space, the number of accessible on-line scans will also be limited. When disk space limitations are exceeded, the oldest scans can be overwritten by the newest.
6. Data access must be extremely fast. This may mean building special purpose data-base managers and I/O facilities.

For example, in the case of spectroscopy, data access rates exceeding a few hundred kilobytes per second (or a few dozen scans per second) are imperative. Since data access will be most frequent for scans and not subscans, the access speed for subscans can be slower than for scans [a hundred kilobytes per second (a dozen or so subscans per second) or slightly less].

Data access rates needed for some, but not all, pulsar work could be many orders of magnitude higher than the above given rate; processing of such data may have to be done off-line.
7. Data access must be available across the network to workstations located anywhere in Green Bank as well as elsewhere on the Internet or comparable network system.
8. Multiple (~ a dozen) analysis sessions, either on a single machine or on multiple machines, must be able to access data simultaneously.
9. We would like to see a uniform format used throughout the system (sub-scan, scan, and archive), though that may be impractical. We strongly suggest that all formats should be self-documenting; the Single-Dish FITS binary-table format, devised in Green Bank in October, 1989, can be used as a model for such a format.

10. The disk space for subscan storage should be 10 GBytes or more since we anticipate data rates exceeding, for some types of observing, a few GBytes per day. Likewise, the disk space for scan storage should be 10 GBytes or more. [For some types of pulsar searches, where data rates could exceed a few hundred Gbytes a day, data probably could only be storable on the archive medium and all data analysis would have to be performed off-line.]
11. The access routines that provide the data to the analysis system must be well documented and of a very general nature. This will allow multiple analysis packages the ability, with a minimum amount of work, to access the telescope's on-line data (see below).
12. The observer should have the power not only to get at each individual sub-scan but also the more complicated collection of scans we are calling observations (see the appendix). This might require a sufficiently powerful data base manager which will allow observations to be constructed and accessed by such user-specified quantities as: range of positions on the sky; feed number; range of frequencies; range of LST or UT; range of system temperature; etc. The list of parameters by which data can be selected must be at least those provided by other single-dish analysis systems (e.g., the UNIPOPS SELECT and the CLASS SET commands) and probably more. Some trade offs may need to be made between performance and number of quantities by which data can be accessed.
13. Security measures must be taken so as to ensure that only the observer or members of the observing team who took the data has the privilege of accessing that data. A model of such a system is that used at the 12-m or 140-ft telescopes. Data security may be more of a problem as more and more data reduction will be done remotely.

VI. System Support Software:

System support software is most often used by staff members to check on the performance of the telescope and to increase the efficiency of observing with the telescope. Observers will sometimes use the software to help them characterize their data. Such software is a significant part of the analysis software needed for the GBT (for example, system support software makes up about 25% of the total amount of analysis software for the 140-ft). The need for this software, some would argue, is more important than the need for other analysis software -- that is, without the results from system support software, the telescope is virtually useless to the astronomer. Most of these requirements may be fulfilled cheaply by a commercial data analysis system of the class of IDL.

1. Software must be provided to reduce holographic data. The software may need to be specific to the GBT's planned holographic experiment but we recommend that as much software as possible should be shared for analyzing holographic data among all the NRAO sites.

The software must be able to provide at least the following functions:

- a) Access holographic observations.
- b) Access and reduce the necessary calibration observations.
- c) Apply calibration factors to the holographic data, regrid the data into two-dimensional arrays.
- d) Window the data so as to minimize leakage problems.
- e) Inverse Fourier transform the data.
- f) Apply feed- and position-related corrections to the data.

- g) Display the arrays as contour, half tone, and raster diagrams.
- h) Rotate and scale arrays so that they can be averaged together or differenced.
- i) Reduce phase errors into panel setting errors. Provide a table consisting of errors at the locations of each surface actuator.

Other functions will be required as the data analysis group learns more of the needs of the group performing the holographic experiment. The thesis of Charles E. Mayer is an excellent beginner's guide to the rudimentary analysis stages for analyzing holographic data, but the GBT will require software not completely described by that thesis.

2. Software must be provided to assess telescope performance from calibration measurements. Since the data sets will be rather large, and such data will be frequently taken, the data reduction must be as automated as possible. For example, bad measurements should be automatically thrown away by the software. Software will be needed to perform at least the following functions:
 - a) Access data that meet a certain set of criteria (e.g., frequency, date, time of day, zenith distance range, etc.).
 - b) Reduce a few hundred flux measurements of continuum sources -- data reduction may include baseline removal, optimal filtering, Gaussian fitting, interference excision, correcting for atmospheric attenuation and pointing offset.
 - c) Arrange the data so as to produce flux estimates as a function of elevation.
 - d) Fit arbitrary curves to the reduced data.
 - e) Plot the results of the data reduction.
 - f) Keep a catalog of the results for use by other users.

Models of what is desired are found at the 12-m and 140-ft telescopes.

3. Similar software must be provided for reducing pointing measurements so as to check the pointing of the telescope and to update pointing coefficients. Again, the data sets will be rather large and the observations will be made frequently so having the analysis as automatic as possible is highly desirable. The software must provide at least the following functionalities:
 - a) Access data that meet a certain set of criteria (e.g., frequency, date, time of day, zenith distance range, weather conditions, etc.).
 - b) Reduce a few hundred pointing measurements of continuum sources -- data reduction may include baseline removal, optimal filtering, Gaussian fitting, interference excision.
 - c) Arrange the data so as to produce pointing errors as a function of elevation and azimuth.
 - d) Fit arbitrary curves to the reduced data.
 - e) Plot the results of the data reduction.
 - f) Keep a catalog of the results for use by other users or by the on-line control system.

Models of what is desired are found at the 12-m and 140-ft telescopes.

4. The GBT will also require software for ascertaining the best parameters for the telescope's radial and lateral focus position and for estimating atmospheric attenuation.
5. As the NRAO learns more about the nature of the GBT, additional support software will be needed. For the first few years of telescope operation, staffing must be such that the as-yet-unknown problems of the GBT that can be solved or analyzed with software will be solved.

6. It is unclear whether data-analysis or 'monitor and control' programming groups are providing the software that engineers will use to analyze their data. (For example, the plotting of drive voltages as a function of azimuth for a set of epochs.) If these capabilities are to be provided by the data-analysis group, then the group will need to acquire information from the engineers about their needs and then to design and provide the necessary software. Present staffing may need to be increased as these desires become known.

VII. Observer's Analysis Software -- General issues:

1. We deem it necessary that an ANALYZ, CLASS, COMB, SPA, UNIPOPS, ... class single-dish analysis package be provided by the NRAO as its supported analysis system at the GBT. From now on, we arbitrarily will call this default, NRAO-supported system XYZ. XYZ could be one of the above systems, an expansion of one of these systems, or a new system, though we shy away from a new system because of the costs and risks involved. The choice will depend upon how the cost of a new system that fulfills our requirements compares to the cost of modifying and maintaining an existing system.
2. We also deem it necessary that observers with experience with packages other than XYZ can easily alter their software system to access on-line data. One way to accomplish this is to keep the I/O interfaces to the on-line data well documented and simple to use so that anyone with familiarity in modifying code in, for example, SPA could add calls to I/O routines that will provide SPA with data. We recommend using the current data access schemes used at the 140-ft and recently installed at the 12-m as models of such an I/O interface.

We request a fast-to-read, efficient disk data format and suggest that NRAO examine how the performance of the single-dish binary FITS, or some other kind of efficient FITS format, compare to special purpose data formats.
3. We do not want the NRAO to support all single-dish packages -- support will be for the XYZ system and for the I/O interfaces. Support for other systems must come from outside of the NRAO. The NRAO must make clear to users of the non-XYZ package that support for their chosen system will be extremely limited. These other systems should be handled in the same way the NRAO now provides software systems like TeX, Mongo, Emacs, etc. -- that is, NRAO should place ANALYZ, SPA, etc. on the GBT system but will not be responsible for fixing internal bugs.
4. Disk space and necessary hardware and software utilities for running the non-XYZ packages must be provided unless the demands of the non-XYZ system are too unreasonable for NRAO to satisfy.
5. At least 10 GBytes of disk space must be provided so as to satisfy these software needs as well as providing sufficient space for observers to store reduced data. Other hardware requirements are: 1) Exabyte and DAT tape drives; 2) 1600/6250 BPI 9 track tape drives; 3) CD-ROM readers; 4) fast laser-graphics (postscript) printers. At least two of each kind of device must be provided; the actual number and location of the devices should be such as to minimize walking distance between where the astronomer analyzes data and the devices. Spare computers, disks, and other hardware must be available on a moments notice -- it would be a shame if an astronomer couldn't observe because a monitor's power supply had died.

6. Most analysis problems will be satisfied by a Sparc 2 class computer, although we recommend purchasing at least one significantly higher-class machine for unusually compute-intensive projects. At least three (and possibly five) workstations, plus the bigger machine, should be purchased solely for use by observers. Staff members who develop or support analysis software should be provided with suitable software and hardware tools (e.g., workstations, color monitors, disk space, etc.). In addition, IBM-PC-class and Macintosh II-class computers should be available for use by the observer. All hardware purchases should be postponed as close to 1995 as possible (but without compromising the ability for programmers to develop code) so that the best, most up-to-date hardware can be purchased.
7. As much as possible, the software and hardware purchases and development should adhere to an open computing environment.
8. The physical environment in Green Bank in which the observer reduces his or her data must be spacious and comfortable (i.e., lots of desk space, book shelves, good lighting, isolated from staff members and other observers, close to reference library, no noisy electronics.... The NRAO should provide at least three such environments, one of which is located in the telescope control room close to but not adjacent to the operator's console. Given the choice, however, we will sacrifice comfort for better software.
9. Remote observers must have all the powers that an observer in Green Bank has except for restrictions imposed by the hardware the observer is using (e.g., limited support of dumb terminals but full support for Sun Workstations) and the state of the link between observer and telescope (i.e., limitations in data access speeds due to effective baud rates between Green Bank and the observer). That is, an observer should be able to use any analysis system on their own computer and access data that has just been taken in Green Bank.
10. Remote observers should be given the option of performing analysis on their home CPUs or remotely logging into a Green Bank computer and using its CPU.

Remote data access will require at least a T1 connection. Two types of uses of the network connection to the outside world should be provided: 1) For low data-rate projects, users will benefit by running the analysis system on their home CPUs and using the network to supply the data. 2) For high data-rate projects, users would benefit by logging into a Green Bank computer to analyze data and using the network for sending commands and graphics. The availability of both methods will reduce the number of CPU's needed in Green Bank and reduce traffic on a possibly heavily burdened network.

11. The NRAO should provide an adequate pool of support personnel for helping observers, both those who come to Green Bank and those who observe off site. Support personnel should be available every hour of the day and have close by the tools needed to assist the observer. Operators may need to be taught more about data reduction so that they can effectively check the integrity of the data in the absence of an observer. We do not care how the NRAO achieves full support but, as a guideline of what may be needed, we have come up with the following possible strategy:

The NRAO may want to minimize the number of support personnel and so will not have someone on duty 24 hours a day; all support personnel are at home at night. They could be given hardware and software for home use so that they can help with a problem that has occurred during nighttime observing.

The NRAO should ensure that staffing is not so thin that support becomes unavailable if a person were, for example, away on a trip or on vacation.

In addition to these support people, system administrators must be available for keeping operating system up to date, adding new packages to the system, performing system backups, fielding general questions from users, setting up user's accounts, fixing system problems, managing e-mail and other system resources, and so on.

12. A whole suite of on-line catalogs should be available for perusal by the observers. Catalogs may be either on-site or available across the network from, for example, NASA archives.
13. Besides analysis software packages, observers will need general purpose software systems. Here is a very short list of the types of things that will be needed: editors (EMACS, EDT, ...), compilers (F77, C, C++, ...), plotting packages (PGPLOT, MONGO, ...), scientific subroutine libraries (IMSL, Numerical Recipes, Mathematica, ...), astronomical utility programs (Floppy almanac, satellite positions, general astrophysical formulae, ...). An accurate planetary ephemeris will be required for pulsar timing observations.
14. The NRAO should also ensure that the computing environment for staff astronomers and support personnel in Green Bank is of a very high caliber so as to ensure that NRAO will be able to entice excellent people to Green Bank (and to keep them there). We believe that the higher the caliber of staff in Green Bank, the more likely the GBT will succeed and become an excellent research instrument.

VIII. The NRAO Supported Analysis System:

As stated above, we desire that observers be given the ability to choose the analysis system they want to use. Nevertheless, it is imperative that the NRAO provides its own well-supported analysis system. The users of the NRAO system will be a) those who prefer it over its competitors; b) those who have never used any of the currently available systems; and c) staff members, telescope operators, and engineers (for monitoring the quality of backends, presence of interference, etc.). As stated above, this NRAO system that we have been calling XYZ could be an existing single-dish analysis system, a modified system, or a new system -- we are more interested in what will fulfill our requirements in as effective a manner as possible more than in what that system is called.

1. We do not deem it wise for the NRAO to develop another visualization system in the class of IRAF, AIPS, IDL, etc. for use as the on-line data analysis system at the GBT. We think that a UNIPOPS, CLASS, SPA, ANALYZ, etc. class program should be adopted, developed, and supported by the NRAO as the GBT on-line data analysis system.
2. Observers should be able to use commercial software packages to analyze on-line data although we do not recommend that a commercial system should be the NRAO supported system for the following reasons:
 - a) Astronomers will want to use the system at home and the cost of a commercial software system may be prohibitive for most university-based observers.
 - b) Even if the initial cost can be met, upgrades in operating systems and in the software system will increase the amount of money outside users will need to spend.

- c) The NRAO will still have to tailor the commercial system to the needs of the problems encountered by single-dish astronomers. That is, the cost benefits from going with a commercial system may not be that great.
3. The most frequent single-dish observers on an NRAO telescope usually have a few days of observing per year (if they are lucky). Many observers will never have used the GBT before and will have to learn not only how to take data but also how to analyze it. Many observers will never have done radio astronomy before and they will have the added burden of learning a new technology. Many observers who use the NRAO telescopes have a very weak background in using computers or are familiar with a type of computer the NRAO does not possess. Many must do all their analysis with an NRAO computer since their resources at home are very limited. Because the NRAO must conserve personnel resources, it can't afford to teach each new observer how to analyze data.

We deem it imperative to the success of the GBT that the XYZ system be very quick to learn (e.g., an hour or two with a cookbook-like manual should be sufficient for an average user to learn most of the essential capabilities of the system) and should not be aimed at black-belt computer users. This does not imply that the system will be insufficiently powerful for challenging experiments. That is, the power provided by the analysis system should automatically match the problem at hand and the ability of the user.

4. We believe that the XYZ system should be portable to a small set of other machines and architectures than those Green Bank may own. The set of possible machines and architectures should represent what is most prevalent within the outside community that may use the program. If a type of computer is infrequently found at universities then the analysis system should not be ported by the NRAO to that machine. With limited resources, NRAO has to be judicious in assigning priorities as to what machines are to be supported. Nor should the NRAO support rare or outdated kinds of hardware unless the hardware is still very common.

Some users may want to run the analysis program on their PC-class computers. Some research should go into seeing whether this can be achieved in the XYZ system without compromising the quality of the system.

5. The XYZ system should be distributable to interested observers. Support for outside users must be excellent. In order of decreasing priority, support should entail: rapid (~ 1 day) replies to bugs found, frequent updates, ease of installing updates, frequent documentation revision, 24 hour phone support (for remote observers), quick response to desires of observers for a copy of the system, distribution of a newsletter, establishing forums where users can discuss problems and ideas, etc. Secretaries and clerks may be needed to handle some of these support functions. The level of support could be reduced if severe constraints on NRAO funding dictates it.
6. The XYZ system should have full support. This includes maintaining the program, fixing bugs, adding new algorithms, and providing good documentation. The system should be exceptionally responsive to users' needs. For example, if an observer needs to use an algorithm and discovers it has a bug, then all support must be aimed at fixing that bug immediately. We deem it irresponsible of the NRAO if observers cannot ascertain whether their data is good or bad because of broken or inadequate software.

The support staff must also be flexible enough that it can quickly respond to observers' needs. For example, if an observer knows that he or she needs a certain not-yet-written algorithm for his or her next observing run, and the NRAO is informed of that need sufficiently prior to the

observation, then the NRAO must have the resources available to provide the observer with the necessary algorithm. Again, we deem it irresponsible of the NRAO if good science is prevented from being done because of lack of software support.

7. Data must be portable between the XYZ system and commercial, public domain, and other observatory analysis systems. This is essential since many of the desired analysis steps that observers will need are already performed within these packages and we deem it wasteful for the NRAO to reproduce what already exists.
8. One should be able to easily set up the on-line data-analysis system so that the completion of an observation or scan by the control system will automatically trigger some user-defined data analysis algorithms. This will prove useful for monitoring, on-the-fly, the integrity of data or for reducing large quantities of data efficiently.

On a related note, the programmers working on data analysis should work closely with telescope operators and the monitor-and-control programming group to make sure that their needs for data monitoring will be met by the on-line analysis system.

9. The XYZ analysis program should be supported at all NRAO sites. Collaboration with the 12m and ascertaining the needs of their observers is essential.
10. The NRAO should include within the powers of the analysis system the ability for each user to add application software. This can either be done by having a powerful macro building capability, or a means whereby users can add Fortran or C subroutines to the system, or both.
11. Common programming standards should be adhered to so as to increase portability and reduce maintenance costs. For example, the list of the supported systems' window environments should include X-windows and other commonly found environments (e.g., SunView). A programmer's manual is deemed a necessity. We encourage the GBT data analysis group to adopt an object-oriented programming language (like C++) as the standard language for future code development.
12. The XYZ system should run not only in a windowing environment but also from a dumb terminal with, of course, loss of graphics capability and other such powers.
13. In addition to highly interactive processing of data, batch processing of data must be available and simple to implement. For batch processing, intermediate results should be stored automatically so as to provide the user a means to go back in the analysis and redo certain steps.
14. We are divided as to whether or not publication-quality plots and tables should be provided by the XYZ system. At least, the NRAO should provide the means whereby data from XYZ are gotten into a system that can produce such quality output.
15. We are also divided in our opinions about the need for a graphical user interface (GUI). We see supporting a common GUI on different architectures to be a possible maintenance problem but a GUI may be an important tool for a user to learn the analysis system quickly. All of us agree that a command-line-interpretor must be the default means whereby users enter commands and that a GUI should be an optional interface each user can decide to use or ignore. In all cases, the user interface should be simple to learn, intuitive, and powerful.

16. Pulsar observers, who have traditionally done all their own programming, see a growing need for some analysis steps to be performed in a NRAO-supported system. The GBT data analysis group is strongly encouraged to continue corresponding with pulsar observers to learn more about their needs in this opening field of on-line, observatory-supported analysis software.
17. The XYZ system must be well documented, both on-line and in the form of manuals. Various on-line help facilities should be available.
18. The analysis algorithms we desire are those that can be encountered in most other single-dish analysis systems. At a minimum, we need the algorithms in the UNIPOPS, ANALYZ/GALPAC, SPA, CLASS, NOD2, and SPECX systems. The designers of the GBT data analysis software should look at these packages as guides to what will be needed. We also strongly suggest that, as much as possible, they incorporate existing algorithms and code, from whatever source, into the XYZ system.

The algorithms must be easy to use and fast to set up (i.e., they should not require a massive effort in setting up input parameters). We need extremely strong 1-dimensional analysis algorithms and strong algorithms for analyzing two- and three-dimensional images and data sets. Judgements occasionally will need to be made as to what algorithms should be included in the on-line analysis system and which are almost always used for off-line data analysis.

For more complicated analysis algorithms, we see the XYZ system as a front-end processor for packages like IRAF, IDL, and AIPS. For example, processing is first performed in the XYZ system and, when finished, the data is passed to another system for further analysis. However, enough software must be provided in the XYZ system to satisfy most on-line data analysis needs. The other systems will be useful when an observer wants to perform very complicated or sophisticated off-line data analysis or has an on-line data analysis need that may not be available in the XYZ system. In this way, the tool (program) used matches the problem to be solved.

19. The XYZ system should be much more powerful in logging what has been done to the data than is currently provided by most existing analysis systems. A useful model to look at is the bookkeeping abilities in the ANALYZ/GALPAC system and in AIPS.
20. Graphics capabilities should include a wide range of types of one-and two-dimensional plots. We do not think that powerful visualization tools must be available for on-line data analysis but they are important for off-line analysis.

IX. Summary:

We have presented our thoughts and recommendations for the on-line data analysis system that we would like to see installed at the GBT. In many respects, our recommendations exceed what the NRAO currently supplies to its single-dish users. We believe that the first-class nature of the GBT dictates that the software system in use at the GBT should be also of a first-class nature. Fulfilling our requirements will be expensive but is justifiable because of the important science that will be accomplished by the GBT and its users. The NRAO plans to spend a great deal of money and effort to ensure that the telescope performs well -- that is, the same philosophy which impels the NRAO not to purposely miss-set surface panels should also drive it not to provide inadequate, undesired, or improper data analysis tools.

The next steps in the GBT data analysis project will be to gather together an over-sight committee from within NRAO which will monitor the work performed by the programming group working on data analysis software. Using the present requirements document as a starting point and over the next six months, the programmers in charge of GBT data-analysis software will try to see what will be the most effective way of fulfilling the requirements given above. A tentative time schedule of when the programmers hope to accomplish certain sub-phases of the data analysis project, as well as the deliverables they will provide, is available from RJM.

Appendix: Definitions:

Subscan -- The individual data records produced by the backend are called subscans. For example, if a spectrometer has a natural cycle of 10 seconds, then the data produced in each cycle is called a subscan. Similarly, if a data buffer in a continuum backend has a 1024 limit on the number of data points, then each set of 1024 points is a subscan.

Scan -- All subscans, when added together, produce a scan. For example, a spectral line scan is the $(\text{system temperature}^2) / (\text{integration time})$ weighted average of all subscans for a particular backend. A continuum scan contains the concatenated data from all subscans for a particular phase and feed.

Observation -- All scans which were observed at a certain time or can be grouped together as a single entity. For example, for continuum data taken with a beam-switched receiver, it may be the on and off position scans. For spectral line data, it may be the set of scans taken toward an object at a set of frequencies. Or it may be a set of scans which compose a map of a certain region.

On-line data -- On-line data are those that just have been taken by the telescope and which are accessible on the system disk where the telescope's computer has placed them. By definition, they are raw data.

Off-line data -- Off-line data can come from three sources: 1) Data that has been read from an archive medium back onto disk (and, thereby, is raw data); 2) data which the user has stored on the disk and which consists of processed on-line data; or 3) processed data which has been produced from a previous observing run or by another data analysis system and which has been read onto the disk. One inherent difference between on-line and off-line data is that the user of the analysis system is usually in full control of the off-line data files while the telescope's computer system is in control of the on-line data files.

Header information: Auxiliary information that describes how data were taken and the conditions of the telescope and its environment is grouped under the name of header information. Every observation, scan, and subscan will contain header information along with the data.

On-line analysis: On-line data analysis consists of the steps one needs to perform while acquiring data in order to check on data integrity, to perform preliminary data reduction, and to allow an observer to make decisions regarding how to proceed with an experiment. Without on-line analysis capabilities, the observer will be observing blindly. On-line analysis is usually performed under time pressure and can be performed as soon as possible after an observation is completed.

Off-line analysis: Off-line data analysis consists of the steps that can be performed long after the data were taken (i.e., there is no time pressure). This includes, for example, refining the data parameters derived from the on-line data analysis, exploring the physics behind the data, and producing publishable tables and plots.