Report of the IAU WGAS Sub-group on Electronic Distribution

G. H. Kaplan
U.S. Naval Observatory
Washington, DC 20392, USA

1 Task of Sub-group

The task of the Sub-Group on Electronic Distribution (SGED) of the Working Group on Astronomical Standards (WGAS) is to make recommendations for direct worldwide computer access to astronomical standards of various kinds. Specifically, part of the charter of the WGAS was "to prepare a draft report on possible electronic access to these units, constants, quantities, and procedures..."

Like the other sub-groups, the SGED operated as a "committee of the whole", the whole being the entire Working Group on Astronomical Standards. The scope of interest includes numeric data and software that has some sanction by the IAU or has been validated (in some way) as implementing some IAU-adopted model, set of constants, catalog, ephemeris, or algorithm. Thus, the controversial matters of what defines an IAU standard, whether such standards are still useful, what subject areas are appropriate, what specific standards are to be adopted, how software is to be validated, etc., are addressed by the other sub-groups.

2 Survey

In early 1992, I drafted a list of survey questions on the use of electronic means of communication and remote data access by people in the positional astronomy community. An advanced draft was distributed for general comment in WGAS Circular No. 46 of 24 March 1992 (46.3.1). Once in final form, the survey questionnaire was distributed in WGAS Circular No. 48 of 16 April 1992 (48.3.1). The mailing list for the Working Group circulars is comprised of about 200 astronomers and other scientists worldwide, working in the fields of positional astronomy, Earth rotation, geodesy, geophysics, and celestial mechanics.

Over the next six months, I received 30 replies to the questionnaire, a response rate of 15arrived by E-mail, from J. Ray, N. Kawai, E. M. Standish, M. Tokumaru, M. Davies, T. Sasaki, T. Huang, M. Yoshikawa, C. Pavlis, S. Klioner, G. Nina, T. Fukushima, J. Wenjing, J. Lieske, H. Kinoshita, H. Schwan, B. Schupler, G. Petit, and E. Proverbio. The other eleven arrived by paper mail or FAX, from M. Bursa, D. McCarthy, G. Kaplan, R. Martin, I. Mueller, C. Tscherning, S. Zhu, C. Hohenkerk, L. Jones, S. Deguchi, and R. Langley.

The geographic distribution of the responses probably reflects reasonably well the distribution of members of the IAU Commissions involved in the WGAS. However, the total number of responses was small compared to the total size of the community potentially affected by the work of the WGAS. There were no responses from South America, Africa, south or west Asia, or Australia. Of course, the WGAS mailing list is largely E-mail based, so the distribution mechanism for the questionnaire tends to pre-select respondents that already have advanced communications facilities. Thus, we must extrapolate – with caution – from a small data set which is biased by selection effects.

3 Interpretation of Survey Results

The questionnaire and the tally of responses for each question was presented in WGAS Circular No. 58 of 16 November 1992 (58.3.2). In general, I would describe the questionnaire responses as "conservative". FORTRAN is still the overwhelming choice for a computer programming language, and simple ASCII file formats are favored for most data exchange, even for documentation.

I. Appenzeller (ed.), Highlights of Astronomy, Vol. 10, 190–192.

Question 6 asked about the preferred means of receiving astronomical data and software. Of the 30 respondents to this question, 21 marked "electronic file transfer (use of computer network)" as their first choice. However, the responses make it clear that even within the community that is electronically well-connected, there is still a need for hard-copy distribution of data and software. And, several of the general remarks sent in appealed to us to remember those who are not well-connected. It seems unlikely that electronic file transfer can serve as the sole means of data and software distribution at this time. Certainly we must be cautious, since 30 people is a very small sample of IAU.

The responses to question 12 indicate that most respondents think it is better for there to be one institution that collects standard software and distributes it (with possible re- distribution nodes), rather than each institution that develops a piece of software distributing it itself (with some centrally-maintained listing of available packages).

Finally, a comment on the second part of question 12 – what institution should take on the responsibility for collecting and distributing the standards. A number of institutions were suggested; all were "write-ins" since no institution was named in the questionnaire. However, few people nominated their own institution, and nobody was in a position to make an institutional commitment. The problematic aspect is the one of software validation, which can be quite labor-intensive if it is done right. Clearly software validation would be a major service to the international astronomical community, but also a major task. Many institutions are under severe staffing constraints, and it is not clear that a software validation service is a realistic goal. Distribution of existing routines – with documentation and test cases supplied by the author – may be as much as we can expect.

4 Other Considerations

4.1 Network Accessibility

The number of nodes on the Internet has been growing exponentially, with an annual growth rate of about 50exceeds five million. Access to the Internet (directly or indirectly through other networks) has greatly facilitated the exchange of scientific information worldwide over the last five years, and has transformed the nature of scientific communication in a number of rapidly-developing fields. Indeed, much of the work of the WGAS has been made possible because of Internet connectivity.

However, we must remember that access to the Internet is not uniformly convenient, or even possible, worldwide. Accessibility is greatest in North America, Europe, and east Asia. In many parts of the world, access is difficult or impossible. Even within the well-covered areas, not all institutions are able to establish a connection. And even with an established connection, billing mechanisms or other inconveniences may exist that place a high premium on network use. Thus, although the Internet is clearly a primary means of communication within the IAU, it cannot be the sole means of communication. Any standards established by the WGAS must be made available in several forms, one of which should be the Internet. Some standards (e.g., planetary ephemerides or large star catalogs) may be too large to economically or reliably transmit over the Internet.

"Anonymous FTP" is an obvious, lowest-common-denominator choice for a file-retrieval mechanism. It is widely used across the Internet. A World Wide Web connection and a MOSAIC home page are now de rigueur for astronomical information on the network. The distribution institution(s) should also maintain an awareness of evolving methods of data access, such as virtual work environments, and should be prepared to implement the schemes that gain widespread acceptance in the astronomical community.

4.2 Security Concerns

Some system managers have security concerns about anonymous methods of file retrieval, since it provides remote users with some interactive capabilities. Thus it would probably be advisable to use a small dedicated workstation, whose sole function is that of a file server, for the distribution of the standards.

An alternative method of file distribution, which is more secure, is available. An E-mail daemon can be established that handles file requests sent by E-mail. Remote users simply send an E-mail file to a specified address with a particular subject line, or with specific keywords in the message. The daemon – a software "robot" – receives E-mail, interprets the subject line or message, and mails back to the sender the files requested. This method is more secure than anonymous FTP, it minimizes network time, and it allows for the transmission of the files during off-peak hours. To partially compensate for the lack of an interactive file "browsing" capability, a file directory and set of instructions is sent to any user who requests it, or who sends an undecipherable request.

4.3 Computer Language

FORTRAN and C should be the primary computer languages for implementing WGAS models and procedures. This recognizes the large volume of astronomical code that already exists in FORTRAN and the preferences of the WGAS as indicated on the SGED questionnaire responses. The C language has a growing number of devotees in the scientific community, and a significant amount of new astronomical code is being written in C. It should be recognized, however, that C can be trickier to port from one system another, especially for inexperienced users.

The computer trade journals would have us believe that object-oriented programming, implemented in languages such as C++ and Smalltalk, is the wave of the future. However, the difficulty in acquiring skill in these new methods has slowed their acceptance in the astronomical community. Given the small number of current users, it would be inadvisable at this time to base standard astronomical software on these languages. However, developments in this area (for example, the NRAO AIPS++ project in the US) should be closely watched.

Several correspondents have pointed out that, in any event, numerical test cases must be supplied with all code. The reproduction of test cases on the local system is the only way to verify that an acquired routine has been correctly implemented.

4.4 RECOMMENDATIONS

The recommendations for electronic access to standard constants, models, and procedures are:

1. A single institution should be identified to take on the task of collecting, validating, and preparing for electronic distribution the standard units, constants, quantities, and procedures, along with any documentation needed for their use. The ready-for-distribution electronic form of these standards will be referred to simply as "the standard files".

2. To the greatest extent possible, all of the standard files (even documentation) should be simple ASCII files. Any use of TeX, PostScript, or wordprocessor-specific formats for documentation should be in addition to, rather than as a substitute for, ordinary ASCII files. (This does not rule out the use of compression schemes, as long as the compression- decompression software is available to all.)

3. Software source code should be written for the FORTRAN 77 or 90 standard or for ANSI standard C. All code should be self-contained. Numerical test cases should be supplied with all code.

4. Several institutions (for example, one on each continent) should act as distribution centers for the standard files. The Astronomical Data Centers would be obvious choices for this task. Each of the distribution centers should be accessible via the Internet.

5. Except for very large files, the primary means of distributing the standard files should be through the Internet, using TCP/IP protocols such as "anonymous FTP". Regardless of the technical details of the file distribution mechanism, a World Wide Web connection and a MOSAIC home page should be established for the standard files. At least one of the distribution centers should agree to handle requests for distribution on MS-DOS floppy diskettes or 9-track magnetic tape. (Other forms, e.g., 8mm tape or CD-ROM, may be supported at the institution's discretion.)