

FIRST SESSION OF TRACK 1

The Role of Electronic Distribution and Publishing in the Development of International Standards

T M R Ellis

Educational Technology Resources Centre

University of Oxford

Oxford

e-mail: Miles.Ellis@etrc.ox.ac.uk

ABSTRACT

The production of International Standards traditionally involves the distribution of substantial numbers of documents to individuals in many countries for use between and during meetings. The advent of easy electronic communication has opened up the possibility of dramatically reducing the cost of document distribution, while simultaneously enabling information to be distributed considerably more quickly. On the other hand the use of electronic distribution carries several dangers, notably those concerned with copyright and equality of treatment for participants in all countries. A further matter of concern is how the electronic version used for distribution is related to the printed versions distributed for international balloting and eventual publication as International standards.

Although many of the initiatives have come from the Working Group level, the whole question of electronic distribution has become a major topic of interest and concern at the highest levels of ISO. This has led to considerable tension between those at the two extremes of the process, with regard both to policy and to the technical approaches to be used, as regards both the format of documents and the distribution media.

This paper is based, primarily, on the author's experience as Convenor of ISO/IEC/JTC1/SC22/WG5 - the Fortran programming language working group - as the group has moved over a two-year period from a paper-based document distribution system, with electronic mail as an informal communication medium, to its current system of all-electronic information distribution and publication utilising email, ftp and the World Wide Web in almost equal proportions. In common with all other programming standardisation committees, WG5 has rejected the use of diskettes for distribution, despite this being ISO's preferred medium!

Initially, documents distributed via an official file server were stored in WordPerfect, Rich Text Format (rtf), and PostScript in accordance with ISO recommendations. However, it very quickly became apparent that this was unsatisfactory and that only the PostScript version stood any chance of being reproduced by recipients so as to provide them with a more-or-less identical copy of the original document. This was particularly true of documents containing graphical images, but was also the case with plain text documents. Further experiments resulted in the current pattern of PostScript, Acrobat (pdf) and, in most cases, ASCII - the latter primarily for the benefit of those who

wish to produce their own documents incorporating extracts from other documents.

Another approach which is finding favour in some areas of International Standards community is the use of HTML for document distribution. This obviously has the advantage that documents can be published on the Web in the same way they are distributed by other means. On the other hand it does provide considerably less detailed control over the layout than is possible with PostScript or pdf, and is not so readily converted into the final document for publication.

The paper discusses the relative advantages of (ASCII) text, PostScript, pdf and HTML for widespread document distribution/publication methods. The process of moving from a working (electronic) document to a final published (paper) standard is also discussed as part of this comparison. Finally consideration is given to the social, legal and economic effects of the move from paper to electronic distribution, with particular reference to copyright issues and the effect on the future involvement in standardisation of individuals from less electronically and/or financially well-endowed nations.

1. Document distribution and the international standardization process

The International Organisation for Standardization (ISO) is responsible for the production and maintenance of tens of thousands of International Standards, covering almost every aspect of twentieth-century life. Eighty-five countries are currently full members of ISO, with a further 26 correspondent and nine subscriber member countries. Although the detailed structure of the Organisation varies slightly in different areas, the production of an International Standard always involves three main stages:

- The development of an initial [working] draft by the appropriate Working Group.
- The voting on the acceptability of that Working Draft by those countries who are participating in the process has traditionally involved huge amounts of paper being shifted around the globe every day at each of the three levels, and is obviously something which could, potentially, benefit enormously from the use of electronic means of distribution, both in terms of reducing the cost of document reproduction and distribution and of speeding up that distribution process.
- The voting on the acceptability of the Committee Draft by those countries who are participating in the broad area of standardisation which is the responsibility of the Technical Committee to whom the Sub-Committee reports. If the draft is approved then it becomes a Draft International Standard, and is forwarded for publication; otherwise it is returned to the working group for further development.

This process has traditionally involved huge amounts of paper being shifted around the globe every day at each of the three levels, and is obviously something which could potentially, benefit enormously from the use of electronic means of distribution and speeding up that distribution process.

However, there are also a number of complications which have meant that the holy grail of rapid economical electronic distribution has been very much more difficult to achieve than might appear, at first sight, to be possible.

The first problem was that the level of electronic sophistication varies dramatically around the world. Whereas the idea of sending documents across the Internet causes relatively little concern to a Working Group concerned with the standardisation of a computer programming language, and whose members are almost exclusively based in computer-sophisticated organisations in the developed West, such a concept is far more difficult to accept by those who have little or nothing to do with computers in their everyday life, or who live in countries which do not have well-developed, and economical, communications systems. Since it is imperative that any country which is a member of the relevant Technical Committee or Sub-Committee is treated equally, this puts a severe constraint on the unlimited use of electronic distribution.

A further problem relates to the control of access to documents at all levels of the process. As long as documents were distributed on paper through official channels it was relatively easy to ensure that they were only available to those authorised to receive them, and that the relevant copyright restrictions were adhered to. However, once documents start being made available by electronic means, especially over the Internet, then it becomes increasingly difficult to control who has them. A related difficulty concerns the integrity of such documents, for it is potentially much easier for modified documents to be mistaken for the genuine ones when the changes are made electronically than when only paper copies are available.

These various concerns have differing levels of importance at different stages of the process, of course, but a further difficulty that arises with electronic documents concerns the very movement of documents between these stages. During the discussions within a Working Group of particular aspects of its draft standard it is usually perfectly acceptable to use conventional e-mail as a rapid means of distribution and debate. During this process the text under discussion will usually appear in a relatively simple form, without any of the presentational details which will appear in the final printed version. When the final document is ready for balloting, however, the responsible editor will need to ensure that the document is properly laid out and presented in accordance with the style and other directives approved by ISO. Such a document is often not easily distributed by e-mail and, moreover, will need to be distributed to a quite different set of people from those involved at the Working Group stage. Once the final Draft International Standard has been approved then the document must be made available for publication in a form that will ensure that it is typeset and printed correctly. The smooth transfer between these differing electronic versions without any changes is another matter of concern.

Nevertheless, discussions have been carried out at various levels within ISO for several years, and during the last 2-3 years a policy has begun to emerge. Unfortunately, this policy, at least in its early stages, appeared to have been developed by people with little or no practical experience of how the electronic world really operates, and flew in the face of the methods, techniques and policies developed, mainly at the Working Group level, by those actually using electronic methods of distribution and publication.

The remainder of this paper discusses these evolving procedures from the perspective of the author's experience as Convenor of one of the more sophisticated Working Groups, namely that

responsible for the development of the Fortran programming language – ISO/IEC JTC1/SC22/WG5.

2. The Fortran experience

The Fortran programming language is the oldest programming language still in regular use – the first Fortran program having been processed on an IBM 704 computer in 1956 – and was the first such language to be standardised in 1966. The first Fortran standards were purely American ones, and it was not until 1980 that the first International Standard for a programming language was published – with Fortran, once again, setting an example that other languages were to follow. Although the ISO Working Group has the responsibility for the production of each successive revision of the International Standard, together with various other complementary Standards and Reports, most of the detailed technical work has, for many years, been delegated to the American Fortran Standards Committee, X3J3.

That committee first established an e-mail discussion list in 1986, at a time when the technology was at a very primitive stage, and its use of electronic communications has evolved over the succeeding years as the technology became more all-pervasive. By the time that the author of this paper was appointed as Convenor of WG5 in early 1995, both committees had been carrying out extensive discussions between meetings by electronic mail for some years, while X3J3 had established a computer archive of its papers which was accessible by ftp from two sites in the United States. Nevertheless, although X3J3 did accept a limited number of email messages as valid papers for their meetings, the vast majority of their papers, and all of WG5's, still began life as paper documents and were distributed to the entire membership in that form.

In April 1995, however, WG5 decided to move towards a completely electronic distribution, although initially documents were also to be distributed on paper to anyone requesting them in that form. This move more or less coincided with a first draft proposals from ISO/IEC JTC1 concerning the electronic distribution of documents, and it seemed sensible to follow the recommendations incorporated therein. [Joint Technical Committee 1 (of ISO and IEC – the International Electrotechnical Commission) oversees all international standardisation work in the field of information technology.]

The initial JTC1 proposals were totally directed towards distribution on diskettes, and went into great detail about such matters as the suffixes that should be given to files and the format of the names given to the diskettes. However, the meat of the proposals, as far as WG5 and other similar bodies were concerned, was contained in the recommended formats for distribution. Here, experience took a back seat while inexperience reigned. The preferred format was Rich Text Format (rtf), while Word 2.0 or later and WordPerfect 5.1 or later were also acceptable. The most obvious mistake here was the specification of the earliest acceptable version of the two specified commercial word-processing packages, and not the latest, since the approach specified would very soon result in documents being legally distributed in a format that many of their recipients would not be able to read without purchasing, possibly expensive, upgrades. However, a far more serious objection was the specification of a word-processing package at all.

Word-processing packages are designed to allow a particular user to create documents quickly and easily, and typically incorporate many features to enable the user to customise the application to

his/her own specific style of writing and presentation. When a word-processed file is transferred to another computer it is the exception for it to appear, and print, in an identical fashion to that on the original computer – even if both computers are owned or used by the same person. The likelihood of the document appearing the same when transferred to another computer system in another part of the world is fairly remote – a disastrous situation when comments and suggestions for change are usually specified by reference to page and line number! A further disadvantage of word-processed documents concerns the incorporation of non-textual images, where even more problems can occur than with text-only files, especially when moving between different platforms.

Nevertheless, in accordance with the JTC1 recommendations, the initial set of documents placed on the new WG5 document server were provided in three formats – WordPerfect 5.1 and rtf, together with PostScript versions since the majority of WG5 members requested this. Almost immediately it became apparent that rtf is far from being the universal format that it was claimed to be, and that an rtf document stood possibly even less chance of looking the same on two machines than did the original Word or WordPerfect document. At the same time there was a substantial demand from some members for an ASCII version to be made available, despite the fact that such a document contained none of the formatting information provided in all the other versions. There were three main reasons for this – the fact that everyone can read ASCII files, the ease with which ASCII files can be searched, and the ease with which extracts from an ASCII file can be incorporated (using cut and paste) into proposals for changes to documents.

After the first six months, WG5 members indicated that, as anticipated, the WordPerfect version of the documents was of little interest. However, at about the same time, Adobe started to make their Acrobat Reader software widely available via the World Wide Web, and an increasing number of suppliers began to use Acrobat as a means of distributing their documentation electronically. Since the beginning of 1996, therefore, WG5 has made its documents available on its server in Acrobat (pdf), PostScript and, where possible, ASCII.

One problem with the distribution of documents in any form other than plain ASCII is the creation of the documents in a consistent fashion, and it is here that the greatest difficulties have arisen. PostScript files, in particular, come in a great many variations, and by no means all of these are printable on any particular printer. Indeed, one major problem with PostScript is that, although it is usually referred to as a page description language it is also a printer control language, and instructions such as “set paper tray 2” will usually cause immediate and catastrophic failure on a printer without a second paper tray. Indeed, over the twelve months to March 1997, the author has received more than one hundred papers in PostScript format for distribution, of which almost one-third have failed when downloaded to an Apple Laserwriter printer. Often it is relatively simple to edit the source file to remove the offending command(s), but this is a somewhat hit-and-miss approach.

A far more satisfactory solution to the problem of non-generic PostScript files is provided, completely unintentionally, by the Adobe Distiller program. This program converts a PostScript file into an Acrobat (pdf) file, and will successfully handle most PostScript files that cause errors with a printer – possibly because it is designed to ignore any printer control commands. It is then possible to “print” the resulting Acrobat document to a file, thus creating a PostScript file that will print successfully on any (PostScript) printer. Unfortunately, although Acrobat readers are readily, and freely, available, the tools for creating Acrobat files are not generally available as part of regular word-processing packages in the way that PostScript writers are. It has not, therefore, proved

realistic to require members of WG5 to submit documents for distribution in Acrobat format. Instead, it is mandatory for them to be submitted in PostScript format, for conversion to Acrobat before placing on the server, with ASCII versions being a desirable, but optional, extra.

By the end of 1996, therefore, WG5 had a well-established, three-level means of distributing documents electronically:

- Discussion documents and drafts of formal documents are distributed widely by e-mail.
- Formal (numbered) documents are placed on a server in Acrobat, PostScript and (usually) ASCII.
- Descriptive and informational material is placed on the WG5 World Wide Web server.

At present, the World Wide Web is not used as the primary means of document distribution for two reasons, one technical and the other procedural.

The technical reason is related to the hardware on which the WG5 ftp and Web servers run. The ftp server facilities are generously provided by NAG Ltd on their own ftp server, which provides all the facilities that might be expected. Mirrors of the main server are provided in the United States by the National Centre for Supercomputing Applications at the University of Illinois, and by NASA, both of whom also provide similar facilities for the American Fortran Committee, X3J3. The WG5 Web server, on the other hand, is a small Macintosh computer in the Convenor's department at the University of Oxford, which is (barely) able to provide an acceptable Web service, but would be quite unable to support an ftp archive of the size required to support the work of WG5.

A related technical reason is that, although it is quite straightforward to arrange for the mirroring of ftp sites, this is not a realistic option for a Web server. Although the provision of a single Web site (on a powerful server) with links to several ftp servers in different geographical locations is increasingly prevalent in the computer software business, it requires a support structure which is quite unrealistic for an organisation dependent entirely on volunteer effort from people with full-time jobs which do not, usually, allow for any significant time to be spent on a regular basis on standards infrastructure work.

The procedural reason is, however, probably more important, for some of the documents available from the ftp server are either very large, or protected by copyright, or both. In particular, draft standards that have reached the Draft International Standard level become the copyright of ISO, since they and the various national standardisation bodies obtain a not insignificant income from the sale of standards.

3. The evolving ISO and JTC1 electronic distribution procedures

During the second half of 1995 and throughout 1996, considerable discussions took place at various levels concerning the implementation of the JTC1 recommendations regarding the move towards electronic document distribution, and the simultaneous revision of the initial recommendations.

Within SC22 (the programming languages sub-committee) there was a universal dislike of diskettes as a transmission medium, despite the fact that the recommendations saw this as a first step to be achieved before moving forward to the use of direct electronic communication. Indeed, the SC22 Secretariat was instructed not to distribute documents in this way, but instead to use either email or a controlled ftp server. In the event, because of problems relating to copyright and access control, the Secretariat only uses e-mail, with documents unsuitable for distribution in this way still being distributed on paper.

JTC1 has also moved towards this position, and has also revised its list of acceptable distribution formats. By December 1995, the original list of acceptable alternatives to rtf had been extended by the addition of DOS text files (ISO 646:IRV, also known as 7-bit ASCII), generic PostScript, Acrobat Portable Document Format (pdf) and machine-processable Table and Tree Combined Notation (TTCN). Four months later, generic PostScript was removed from the list of acceptable alternatives. Twelve months later, in January 1997, HTML was added to the list of acceptable distribution formats, even for documents that are not being distributed in the form of Web pages. In this context, HTML is seen as something between ASCII and Acrobat, in that it is easily viewed and edited, but does not retain the formatting of the original document – an important factor in certain situations as mentioned earlier. At the same time, rtf was removed from its position of preference and made one of seven acceptable formats, with pdf being identified as the format of choice for documents containing complex embedded graphics.

The move towards all-electronic distribution is not purely a one-way process, however, for an integral part of the standardisation process is the international balloting on documents at defined stages of their development. It follows, therefore, that if the document being balloted is to be distributed in electronic form, then the national votes and accompanying comments should be returned to the relevant secretariat in the same way. Once again, this can cause potential and actual problems with some of the poorer and/or less technically sophisticated countries who, nevertheless, wish to have an influence in the development of International Standards which will impinge upon their national economy. Nevertheless, at least within JTC1, an aggressive timetable has been agreed for the move towards all-electronic distribution of documents and, at the same time, of submission of approved Draft International Standards for publication.

4. The publishing of International Standards

The above discussion has concentrated on the distribution of working documents at the various stages of the standardisation process. The final stage still involves the production of a paper document – the International Standard itself – and this poses quite different problems from distribution.

As has been indicated, the use of word-processor files for document distribution is generally not to be recommended, whereas page description languages, such as Acrobat and, to a somewhat lesser extent, PostScript meet the requirements quite well. When it comes to publication, however, the issues are more complicated. The arguments against word-processed documents are equally valid, for, at this stage, even more than at the earlier ones, it is essential that the exact page layout is retained. On the other hand, it will frequently be necessary for additional (standard) pages to be added at the beginning or the end, and this is clearly not really possible if the document is fully determined by a pdf file, or a PostScript one. The best solution would, undoubtedly, be to use a powerful markup language, such as SGML (which is, incidentally, the only internationally standardised method of defining a formatted document). Unfortunately, creating a document in SGML is not easy, and, in particular, is not normally an option when the basic document has been defined in a word-processor such as Word or WordPerfect. A number of word-processors do allow the creation of an HTML file, but, despite the fact that HTML is an extended subset of SGML, this is not a sufficiently precise method of defining the layout for most standards purposes.

At present, therefore, the only solution is either for the final document to be provided in camera-ready form, so that ISO can simply add their own foreword and/or other introductory material in paper form or for the final electronic document to consist of two or more documents, one describing the additional initial material, one being the body of the Standard itself, and a third, possibly, describing any extra material at the end. A possible variation of the latter approach might be for the document to be supplied as one or more pdf files and for the additional material also to be produced in Acrobat form, thus permitting the use of the capability of Acrobat Exchange to add or remove complete pages to insert the additional material in the correct place. However, this is a relatively new feature of Acrobat and it seems unlikely that it is a realistic option at present.

5. Conclusions

The past two years have seen a major shift from paper to electronic distribution within the international standardisation world, or at least that part that is concerned with information technology. The over-riding lesson has been that the process is far more difficult than might be expected and that no fully satisfactory system yet exists.

As far as the distribution of small documents is concerned, e-mail is usually sufficient. For larger documents, this is not a satisfactory method, especially where, as in many cases, the recipient has either limited file store availability, or is using a form of access to the Internet which involves time-based charges. (The cost of Internet access is one of the few areas where, at present, academia, both in the UK and in many other countries, has an advantage over the commercial and industrial world.)

In theory, larger documents can be sent more economically as attachments to an e-mail message, but this is far from being 100% reliable, and a substantial proportion of such attached files never arrive at their destination – frequently without the recipient being informed of the loss. A test by the Director of the Computing Laboratory at the University of Liverpool in 1996 on a randomly selected group of colleagues around Britain produced a 100% failure rate! Anonymous ftp is currently the most complete form of document distribution, in the sense that it can generally deal with documents in many different formats, both binary and text-based, without any problem. The major disadvantage of such servers as a *distribution* medium is that nothing is actually distributed;

it is the recipient's responsibility to initiate the downloading of any required documents from the server, which in turn implies that the availability of documents on a server must be announced to all potential recipients – presumably by e-mail.

The World Wide Web is already playing an increasingly important role in this process, but mainly as a means of access to one of the two primary forms of distribution already discussed. As a general rule, the information available directly from the Web sites operated by standards groups is intended to be available to anyone accessing the site, whereas many of the documents on file servers are (or should be!) protected from unauthorised access by some form of security check – typically based on a password which is changed at intervals and communicated directly to those authorised to access the files, usually by e-mail.

Experience in various standards committees and working groups has confirmed that word-processing files are inappropriate for general document distribution, where the preservation of the format of the resulting document is of any significant concern, and also that there are too many potential problems with PostScript files for them to be used as a widespread distribution medium. Acrobat files have many advantages for document distribution where the format is important, while ASCII is unsurpassed where the format is not important.

Despite its many advantages, Acrobat also has several flaws – mainly in situations where the document contains non-standard fonts. The intention of Acrobat is that it is not necessary for the recipient to have the fonts used in the document, as substitute fonts will be used or created (using Adobe's Multiple Master Font technology) in those cases where the original font is not available on the recipient's computer. Unfortunately, this only works well where the original font is a Type 1 PostScript font. In particular, if the original document has been created using a text-processing system such as TeX then the resulting Acrobat document is frequently almost unreadable – although it will usually print perfectly successfully. A further problem with Acrobat, which makes it impossible to use as a universal distribution format, is that cutting and pasting from an Acrobat document into, for example, a word-processed document is virtually impossible because the copied text loses all formatting information, including line breaks. Moreover, it is only possible to copy from one page at a time. Taken together, these mean that, despite Acrobat's powerful searching capabilities, a pdf file is far less useful than an ASCII file for those who wish to create proposals for changes to a document, or who wish, for whatever reason, to include extracts from a distributed document in a document of their own.

The use of electronic methods to define the final document for publication is still at a rather earlier stage, and it is unlikely that any major steps will be taken to provide an electronic "master" for publication, following the final approval ballot, until further improvements are made in the software. In particular, it is important that the electronic format used both preserves, absolutely, the formatting, page numbering, etc, of the original document, while permitting the easy addition of extra initial and/or final pages.

Finally, the importance of cheap and easy access to the recommended software and communication systems is essential if the international standardisation committees are ever to endorse fully the move to electronic distribution and publishing. The basis of international standardization is that any countries that wish to participate may do so, as long as they have the technical expertise to contribute in a meaningful manner. It is essential that the move towards electronic distribution does

not prevent the participation of member countries from outside the developed Western and Pacific Rim worlds.

Miles Ellis is Director of the Educational Technology Resources Centre at the University of Oxford, and is also Deputy Chairman of the Educational Television & Media Association. Before moving into educational television he spent nearly 30 years in computing, including 12 years in the Computing Services at the University of Sheffield and 8 years as Director of the Computing Teaching Centre at the University of Oxford. During the latter period he became heavily involved in the International Standardization of Programming Languages, especially Fortran, and has been the Convenor of the ISO Fortran Working Group since 1995. His paper is largely based on his experiences in the latter role.