Document Interchange Possibilities at CERN

Version 1.0
1993

Maria Dimou-Zacharova
Mik Ferran
Silvano de Gennaro
Per Hagen
Rainer Többicke

1 Introduction

1.1 Purpose of this document

- establish an inventory of document interchange compatibilities for the main platforms and document producing applications at CERN,
- sketch out the main characteristics of different document types: revisable, print-only, *etc.*,
- summarise the available document transfer techniques,
- provide recommendations on how to further handle documents received.

1.2 Organisation of this document

The possible options characterising document interchange could be drawn as a 3-dimensional grid representing type of document, sender's and receiver's platforms. Instead, in order to overcome representation difficulties, this document starts by taking the point of view of the sender of a document, enumerates document types the sender is likely to produce and then considers how a possible receiver could further handle the document.

The order of presentation is

1. **Document types** - page 2 - describes document types which are likely to be produced on more than a single platform at CERN.
2. **Document transfer methods** - page 5 - describes the methods available at CERN for transferring documents from one computer to another.
3. **Platform-specific considerations** - this section deals with points not covered by the above, in particular applications which are available only on one platform or information which applies only to a particular platform,
   (a) page 7 - from the sender's point of view,
   (b) page 10 - from the receiver's point of view.
4. **Examples** - Section 6 on page 12 offers rapid access to the documents' conclusions for the occasional reader.

1.3 The future of this document / User Support

This document is meant to be kept up-to-date closely following the evolution of document processing and communications software. Observations arising by actually applying the document's recommendations to real-world interchange problems are hoped to inspire the authors to further updates. Comments are therefore appreciated and should be addressed to `docint@dxmint.cern.ch`. This Email address should **not** be mistaken as a help desk for document processsing or general computer handling support! Hints on where to ask for help are given within the various sections of the document and summarised in the appendix on page 16.

2 Document types

The sender of a document decides on the format in which the document is sent out. The following considerations apply:

- dissemination of information such as printable documents sent out for reference can be in non revisable form such as PostScript,
- if the document is going to be worked upon, it has to be revisable and the document format must retain structural information; TEX, LATEX, SGML and RTF are typical formats for text documents; IGES and GIF are examples for graphics formats;
- pictures and drawings which do not have to be revisable may be sent out as PostScript. However, due to the size of the resulting document it may be more convenient to choose a revisable format.

2.1 PostScript files

PostScript programs are descriptions of printable pages. A PostScript document contains text, drawings or formulae formatted for a given page format, a given set of fonts and colours. As a final representation (before the actual paper) of a document's contents, a PostScript document does not contain semantical information about drawings or formulae and is (practically) **not revisable**.

PostScript programs can be printed on the majority of CERN's printers. Additionally, users with Unix workstations can display PostScript documents on their screen (with e.g. ghostview).

Most document processing applications are capable of producing PostScript. Transferring PostScript documents is mostly unproblematic[1] since it consists of US-ASCII characters.

Usually application-generated PostScript programs are recognisable by the characters '%!' in the first two columns of the first line.

PostScript fonts PostScript documents can only be reproduced faithfully when the printer or the PostScript previewer has access to all the fonts used in the document. Normally, PostScript printers/previewers provide a certain number of fonts by default.

The (most basic) Apple LaserWriter built in fonts are: Times, Helvetica, Courier and Symbol. It is advisable to stick to this minimal font set whenever the final printer is not known. The Apple LaserWriter II NTX built in fonts are: Times, Helvetica, Helvetica Narrow, Courier, Symbol, Palatino, ITC Avant Garde Gothic, ITC Bookman, New Century Schoolbook, ITC Zapf Chancery, ITC Zapf Dingbats.

When other fonts are used, they should be supplied as outline type 1 fonts inside the document, unless it is known that the target printer will have them installed. The size of the resulting PostScript files will increase (e.g. 25-50 kB/font).

Support: Help for basic PostScript problems is given by Michel Goossens, `goossens@cernvm.cern.ch`.

2.2 TEX, LATEX

TEX and LATEX describe documents in a high level document description language. The documents are revisable and contain information concerning document structure (e.g. chapters, lists) and formulae. TEX and LATEX (or LATEX 's cousins such as AMSTeX) are almost universally available and very popular in the High Energy Physics world.

Document transfer is unproblematic (all files are normal text files), as long as standard macros/styles and no version-specific features are used.

LATEX documents can be formatted for a character-only terminal or into a text file (which can then be sent by electronic mail (Email))[2]. However, during the process most positioning information, font information and special characters get lost or are simulated in a very crude fashion, resulting in poor-looking documents most of the time.

[1] problematic: Macintosh-created text containing national (e.g. French accented, German Umlaute) characters requiring "binary" transfer

[2] by including \documentstyle[screen]{...} into the document and processing the .dvi file with dvi2tty

2.3 Plain text

US-ASCII text Probably the only (!) document format presenting neither conversion nor transfer problems. Worth mentioning here is that even extending the type to UK-English is not unproblematic[3]

Text containing national characters The encoding for the so-called *national* characters, e.g. French (è, é, œ, ç, *etc.*), German *Umlaute* (ä, ö, ß, *etc.*), or others differs from system to system. A minority of systems offer the use of several *code pages*. In practice however text containing national characters created on one system has a small chance of being readable on a different system, even when only systems from the same vendor are considered.

"binary" This document uses the term *binary* in some places to mean text not only consisting of characters from the US-ASCII character set. What happens is that the application interprets the contents[4] of the document entirely in its own fashion.

2.4 Microsoft Word

Word is the recommended word processor for Macintosh and PCs at CERN. It is not used on other systems. Word stores documents in an internal format which preserves font, character size, page layout and other information and is only understood by Word itself. Word internal documents are only usable between PCs and Macintoshes. As of Word version 5 for Macintosh and Word for Windows version 2 for PC the formats are at least largely compatible. The current version number can be inquired by pulling down the *Help/About* menus on a PC or the 'apple' menu on a Mac, but is also normally displayed when the application starts. The compatibility between Macintosh and PC versions is subject to restrictions which are documented in a file on the Word distribution diskettes.

RTF Format Word can store documents in RTF format designed by Microsoft specifically for interchange. RTF files use US-ASCII text only and therefore do not require special encoding or binary transfer. They can be interchanged between PC and Macintosh versions. Interchange of graphics is not supported and the same restrictions apply as in the case of interchange of Word internal files. The RTF format is of interest since a published specification exists. Based on this specification, converters to other word processors start to emerge[5].

Text version Word can create a special text-only version (cf. Section 4.1) of the document, which has a significant part of formatting information stripped off or crudely simulated. Still, the document is readable on a character terminal and, while not as beautiful as the one printed on paper, can be sent to other systems via Email.

2.5 SGML

SGML marked up documents describe the document contents and its structure on a high level, usually without specific layout/formatting information. Documents are not printable as such but must be processed by an SGML parser and a formatter.

SGML is currently not widely used at CERN: few SGML parsers/formatters exist yet for computer platforms used at CERN. However, this is likely to change as a growing number of publishers accept SGML documents and the number of available applications, including WYSIWYG text processing applications increases.

2.6 Application-Internal formats

Many applications store a revisable representation of the document contents in some internal format, likely to be only usable by a similar application, often on a similar platform. For example AutoCAD stores drawings naturally in 'DXF' format, Excel uses 'XLS' although it can produce other formats (e.g. SYLK, the Multiplan format, WKS for Lotus 1-2-3, *etc.* with sometimes loss of functionality).

Usually application internal formats contain binary information and/or system-specific structural information which require appropriate methods for transfer, e.g. *binary* FTP, *binhex* (for Macintosh files requiring transmission of both *resource* and *data* forks), or *uuencode*ing for electronic mail transfer.

[3] e.g. $ ↔ £

[4] naturally 8 bit

[5] There are now Unix-based public domain converters to TEX or LATEX. Unfortunately they still work only for the most trivial cases.

2.7 Graphics formats

We distinguish two types of graphics formats:

1. structured graphics which describe a picture in terms of graphics objects, such as circles, lines, their colour(s) and position relative to each other; a typical example is revisable CAD/CAM data;
2. bitmap graphics which describe a picture in terms of the individual pixels at a given resolution, their "brightness" or their colour; a typical example is the output of a scanner.

2.7.1 IGES

While some graphics files are standard for a platform e.g. for the MacIntosh, some other graphics files are application-dependent. There is clearly a need for an independent graphics standard: PHIGS and GKS have tried to provide this.

The ANSI standard IGES (Initial Graphics Exchange Standard) is used to exchange CAD/CAM data, mainly for mechanical engineering, and allows the user to transmit some semantic information along with the drawing. Elements which make up the picture, circles, straight lines, B-splines, *etc.*, are identified as such and so may be manipulated by the receiver. Other information such as annotations, dimensions and parts lists are sent in a standard way and they too may be understood and manipulated.

However the standard leaves too many details to the implementor and allows him[6] to choose between alternate ways of expressing himself. The result is that an IGES file written by a typical CAD system and containing a considerable amount of information will have some parts which are illegible by other CAD programs.

In order to minimise such transfer problems, an on-line service at CERN, CIFAS (CERN IGES File Adaptation System) has been set up to "flavour" IGES files. CIFAS changes the IGES file "dialect" written by one CAD system, say, Euclid, to the IGES file "dialect" of the receiving system, say AutoCAD.

A similar service, CADDFAS, is available via electronic mail requests for those who have no access to the CERN VXENG cluster. For more details of this send a Email to `caddfas@cadd.cern.ch` with subject: `info`.

The writeup `cad_exch_guide.ps` which is in the CADD FTP account, directory `/doc`, on CADD.CERN.CH tells the whole story. Otherwise contact Nils Hoimyr, `hoimyr@vxeng.cern.ch`.

Support: For help on interchanging graphics files from engineering applications contact Mik Ferran, `PMFDJ@CERNVM.cern.ch`

2.7.2 Bitmap graphics

GIF, TIFF, JPEG, X-windows These are popular bitmap graphics formats for which viewing programs and converters from one format into another exist.

Scanner output Most scanner support software (e.g. Adobe Photoshop on a Macintosh) [7] can save files in EPS (encapsulated PostScript), GIF or TIFF. EPS files can readily be included into LaTeX documents when these are to be printed on PostScript printers. These files are normal (big) text files for which no special transfer precautions are necessary.

GIF or TIFF files can be displayed on Macintosh, PC and Unix workstations.

---

3 Document Transfer

3.1 "binary" Transfer

Certain document types have to be transferred without character-set translation, which is normally applied when transferring between system using different character sets. Additionally, some transfer methods (e.g.

[6] or **her**; perform grammatical corrections where required

[7] announced for PC, no experience yet

Email) require the document to consist only of characters out of a specific character set, which is generally not the case for e.g. application internal documents or compressed documents.

Using FTP (see below), the appropriate method is "binary" transfer, specified at the beginning of the FTP session by the `binary` command.

Using Email, the appropriate method is to convert the document into a stream of Email compatible characters, using a procedure called *uuencode*. The receiver will be able to reconstruct the original binary contents of the document be applying a *uudecode* procedure. Some Email implementations (see Section 3.3) apply this procedure automatically under certain conditions. Section 6.6 contains an example how to use uuencode.

3.2 FTP

Transferring files via FTP[8] requires access rights on both the source and the target machine. This limits the possibility of actively 'sending' the document to somebody. The most commonly used technique is therefore for the sender to store the document in some place accessible to the receiver and then inform the receiver via some other method (e.g. electronic mail), specifying the location (e.g. TCP/IP address, subdirectory) of the file and requesting him to copy the document for reading or further processing.

Anonymous FTP This assumes that sender and receiver both have access to the place where the document is stored. While this is unlikely to cause a problem on mainframe-like systems like VM or VMS where often sender and receiver both have access, other systems can set up an *anonymous FTP* archive which allows anybody(!)[9] to access a controlled set of files without requiring the requester to have an account on the machine where the archive is stored.

Although less convenient, FTP servers exist equally for Macintosh (Versaterm Pro) and PCs (FTPSRV from FTP Software, Inc.) which can run continuously in the background (control panel resource on Mac, DOS Box on PC/Windows) and allow access control through userid/password combinations.

Alternatively, in many cases it is sufficient to store the document on some central file server, which could be an AFS volume, a folder in some AppleShare fileserver or a Novell directory depending on whether the intended receiver has access to the server.

Archives and Compression A frequently used technique on Unix systems is to group a set of files together in a single file as a *tape archive*, recognisable by the file name's **.tar** extension. Such a file is designed to be 'exploded' by the Unix `tar` command.

In addition, commonly files or even tape archives are compressed in order to save disk space and network bandwidth required for transfer. Such files are typically recognised by the **.Z** extension. They can be expanded using the Unix *uncompress* command. Other compression techniques have been invented, currently a popular one is 'ZIP' which combines compression and archiving into a single format and utility. Compressing or 'ZIPping' files can lead to spectacular space savings. ZIP is popular in both Unix an PC environments.

Transfer modes Text document types such as TEX source files, PostScript programs[10] or English text should be transferred in *text* mode.

Documents in internal text processing format (e.g. MS Word), compressed archives or bitmaps should be transferred in *binary*.

3.3 Electronic Mail

In general a file transfer protocol should be used instead of Email for the following reasons:

Document size The formated documents especially if they include images and moreover if they are PostScript files are very large. In most cases Email messages have to cross a number of relays or even gateways (hidden by the uniformity of the address format user@domains) before they reach the destination. Therefore, body truncation, fragmentation or even rejection might occur as there is no common size limit across Email protocols.

[8] TCP/IP File Transfer Protocol

[9] at CERN the ftp daemon on some Sun systems has been modified to grant access to groups of users based on criteria such as IP addresses

[10] except Macintosh, see above

Content type The most commonly used Simple Mail Transfer Protocol (SMTP) can handle 7-bit ASCII data only. Thus, accented characters, alphabets other than US-ASCII and binary data (naturally 8-bit) get truncated [11] when crossing a SMTP relay/gateway unless :

- either the sender explicitly runs the 'uuencode' program before sending the message and the recipient (recognising the uuencoded data by the 'begin' statement at the beginning of the message) runs 'uudecode' hoping to restore the original data format (not always possible between platforms),
- or a more sophisticated mail reader encodes 8-bit message contents into 7-bit and a program of similar functionality exists on the recipient end. More specifically:
  - MSmail enclosures between PC's can be exchanged without problem. When a MSmail enclosure crosses the SMTP gateway it gets automatically uuencoded. MSmail messages should include bodyparts with accented characters as enclosures (instead of the sender simply typing the text while composing the mail) to make sure that they will be correctly deployed at the recipient end.
  - QuickMail enclosures between Macintoshes can be exchanged without problem. When a QuickMail enclosure crosses the SMTP gateway it gets automatically uuencoded. QuickMail messages should include bodyparts with accented characters as enclosures (instead of the sender simply typing the text while composing the mail) to make sure that they will be correctly deployed at the recipient end.
  - Enclosures can be exchanged between MSmail and QuickMail with a use of a CERN-developed program provided the Mac and PC are registered at the level of the MSmail gateway. For CERN internal mail no registration is necessary. Outside CERN users should send a request to `admin@smtpgw.msm.cern.ch` including the domain name of the external QuickMail community. If messages are autoforwarded from other systems (VM, VMS, Unix) to MSmail or QuickMail then the above inter-ex-changeability is not available; therefore PC and Mac users should register in EMDIR their PC or Macintosh preferred Email address, if they want to enjoy the additional functionality of the relevant Email systems.
  - Mail reading programs claiming to be conformant to the Multipart Internet Mail Extensions (MIME) standard (mostly public domain, a few commercial ones for the moment) running on almost any Unix platform provide similar functionality between Unix systems. Currently MIME requires the use of a 'Content-Type' header where the ISO-8859-x (for the character set used) is explicitly mentioned. Thus the mix of different character sets within the same text is not convenient. If QuickMail and MSmail become MIME conformant too, then all these content types will be exchanged between Mac, Unix and PC mail readers.
  - Mail from/to VM and VMS mail is US-ASCII only.
  - Security consideration! Beware of receiving binary files (check the file type[12] where possible) that might hide harmful programs.

Support: Support for Email problems is can be obtained by sending an Email to `postmaster@cern.ch`. In case this leaves you in a catch 22 situation, the person normally answering the requests is Maria Dimou-Zacharova, phone 3356.

3.4 WWW

Sharing documents using WWW (World Wide Web) is conceptually similar to the *anonymous FTP* technique described above (Section 3.2): documents are made accessible to a WWW *server*[13], and the receiver accesses the document through a WWW *client* program. As in the anonymous FTP case the transfer of the document is initiated by the receiver, who therefore may have to be informed of the existence of the document by some other method.

The advantages of WWW are (among others) guided access to documents through hypertext links, which on some systems (e.g. Unix) allows mouse point-and-click access, facilities to search for documents, and relieving the user of sometimes complicated interaction with a file transfer program.

Sharing documents with WWW requires some setup, and is therefore suitable for a group of users sharing many documents.

[11] some places support 8-bit non-standard SMTP

[12] e.g. `.EXE`, `.COM`, or `.BAT` on PC, executable files on Unix, *etc.*

[13] in a variety of ways: local files, FTP, output of commands, *etc.*

3.5 Other document transfer methods

In addition to the general methods described above of transferring files from one computer to another, several methods exist, mostly involving only a subset of the computer platforms considered. The following (incomplete) list illustrates this:

Interlink DECnet-VM gateway for transfer of files from VMS or Ultrix to VM.

EtherShare or a public domain counterpart **aufs** (Columbia AppleTalk Package) emulates an AppleShare server on a Unix machine. Macintosh text files however are not directly usable by Unix applications owing to a different encoding of the *newline* character (decimal 10 on Unix vs. 13 on Macintosh). Conversion is trivial but not applied automatically.

Novell NetWare dedicated file server for PCs and Macintosh. Macintosh/PC text files are not directly interchangeable due to differences in *newline* character(s): decimal 13 on Macintosh vs. the sequence 13+10 for PC. Some applications can cope with this.

With the additional NetWare/NFS product Unix systems can access files on Novell servers. Conversion of text files is even more complicated.

AFS (Andrew File System) is a distributed file system for Unix machines, however files residing in AFS can be made accessible from VMS and PCs via NFS, from Macintosh via **aufs** (see above).

---

4 From the Sender's Point of View

4.1 Sender: Macintosh or PC

Considered file formats:

- TEX, LATEX
- Microsoft Word
- PostScript
- Application internal formats
- Scanned images

TEX, LATEX Several implementations of TEX exist for Macintosh and PC, but there is only very little support available for them at CERN. For PCs the usual product is PCTEX, for the Macintosh TEXtures.

MS Word Sending out MS Word documents in Word's internal, revisable format only makes sense if the receiver is a Mac or PC with Word installed. Compatibility considerations between Macintosh-PC versions apply as mentioned earlier.

For all other receivers the document should be sent as a printable PostScript file; alternatively, sending a **text-only** version should be considered:

- choose "File" menu "Save As"
- from the "Save File As Type":
  - choose "Text only with line breaks" or "Text with Layout"; the latter will try to keep text from tables aligned
  - choose a suitable file name and click "OK"

Internal Word files contain binary information (i.e. not consisting only of ASCII characters) so that appropriate care has to be taken for transfer to another machine. Difficulties can be avoided of the document is saved in RTF format, another option on the "Save As" menu, before being sent.

Support: Microsoft Word is supported by the responsible for the Desktop Publishing Centre, Michele Jouhet.

PostScript On a Macintosh a PostScript program can be created by selecting any available LaserWriter in the *Chooser*, causing the application to print and then clicking on the *PostScript file* button in the *Print* dialog box. The resulting file can be printed on standard PostScript printers subject to font availability and perhaps memory.

Documents containing accented characters will result in PostScript programs which are not 7-bit clean, therefore the file should be transferred in *binary* or precautions have to be taken for transfer via electronic mail.

On a PC (under Windows), the printer port can be set to `FILE`. When the application prints, a dialog box appears prompting for the name of the PostScript file to be created.

Application internal formats Application internal formats (e.g. "raw" Excel files, PowerPoint and the like) usually require binary transfer. For the Macintosh these formats also contain additional structural information, the so called *resource fork*. In order to preserve this information, the file is best wrapped into a special transmission format which can be easily transferred across mail gateways and nevertheless preserves all binary and structural information.

For the PC an appropriate encoding method is *uuencode*, for the Macintosh *binhex*[14].

Macintosh - PC compatibility Several applications exist in versions for Macintosh and the PC. However, experience shows that the PC and Macintosh versions of many applications are not 100% compatible. The user is advised to consult the documentation for the application.

Scanned images Files should be saved in EPS (Encapsulated PostScript) when they are to be included into LaTeX documents destined for PostScript printers. EPS files can be rather big.

Images may also be saved in GIF or TIFF format. These files can be viewed on Unix workstations or translated into EPS later. They contain binary data and therefore have to be handled like binary application internal formats.

TIFF files should be stored in the *byte order* of the targeted receiver: *Photoshop* (on the Macintosh) prompts whether to store the file in PC format ("little endian", which applies also to VAXstations or DECstations) or Macintosh format ("big endian", which is also what nearly the rest of the world uses).

4.2 Sender: VMS

Considered file formats:

- TeX, LaTeX
- `DOCUMENT` (CDA)
- PostScript

TeX, LaTeX The VMS version of TeX maintained on VXCRNA is compatible with the versions on VM/CMS and Unix. TeX documents can be formatted for PostScript printers.

DOCUMENT ...is a VMS utility that lets you create structured documents for a variety of output devices. It also includes a graphics editor. In particular DOCUMENT can create PostScript files. DOCUMENT input is a marked up in a way vaguely resembling SGML mark-up. The input file consists of text with national characters encoded according to Digital's own code page, requiring *binary* transfer when national characters are used.

Support: Support for DOCUMENT on VMS systems is given by Julian Bunn, `julian@vxcern.cern.ch`.

PostScript Created by TeX (i.e. *dvips*), or DOCUMENT. File may contain non US-ASCII characters if the original file (e.g. TeX document, DOCUMENT input file) contained other than US-ASCII characters, e.g. accented characters typed in from a French keyboard.

[14] Macintosh: *stuffit* allows grouping of files and compression, *UULite* performs uuencode/uudecode

4.3 Sender: VM

Considered file formats:

- TEX, LATEX
- SGML/Bookmaster
- PostScript

TEX, LATEX The versions on VM are centrally maintained. TEX documents can be formatted for PostScript or IBM 3812 printers.

SGML Currently available on VM are two SGML-style text formatters: Waterloo Script[15] (the `SGML` command) and IBM Bookmaster (the `BOOKIE` command). Documents written for both formatters (i.e. marked up using SGML) can in principle be transferred to anywhere, provided the DTD (*Document Type Definition*) is available to the receiver. In practice however this is rarely used: 1. there are no other systems at CERN which understand SGML and 2. many publishers today accept TEX or LATEX style mark-up which is still more commonly used.

Both formatters (Bookmaster and Waterloo Script) can produce output for IBM 3812 printers as well PostScript. Output for IBM 3812 printers contains binary information which can only be printed on those printers.

Bookmaster PostScript output Bookmaster produces PostScript in ASCII! The resulting file is a valid PostScript file using a character encoding which is completely unnatural to VM. Such files cannot be transferred by electronic mail. They can be printed using the `LWPRINT` command which recognises their special encoding, or transferred to an ASCII (e.g. VMS, Unix, PC) machine **bypassing translation** (i.e. `binary`).

In addition, Bookmaster can produce a 80-column formatted character-only file (by specifying the `EMAIL` option on the `BOOKIE` command), suitable for viewing on a terminal screen or sending via Email. This option is useful only if the document consists mainly of English text. Some effort is made to preserve national characters (e.g. German Umlaute 'ä', 'ö', *etc.* replaced by 'ae', 'oe', *etc.* and French accents dropped), but most special symbols just appear as their SGML identifiers in the text.

Waterloo Script PostScript Waterloo Script produces PostScript (in the for VM natural EBCDIC encoding this time). However, the first column of every line starts with a blank, which does not invalidate the PostScript program, but causes several printers to misinterpret the program as normal text. They will then print the program but not interpret it – the resulting output will be unrecognisable.

The VM `LWPRINT` command recognises this and takes corrective action, care should be taken only when such files are sent to users on other machines. In that case it is easiest to edit the PostScript file and to delete the leading blank on the first line(or add a `%!` in front): the resulting file will look like a standard PostScript program.

4.4 Sender: Unix

Considered file formats:

- TEX, LATEX
- PostScript
- X-Window graphics (e.g. screen/window dumps)
- FrameMaker

TEX, LATEX A centrally maintained version of TEX is available from the ASIS server. The level is normally equivalent to the one maintained on VM and VMS, but Unix workstations which do not use the ASIS software itself via AFS or NFS may have a local copy of a different level installed. TEX documents can be formatted for PostScript printers.

[15] No longer recommended as of CNL 198, March 1990, but still used

PostScript almost all word processing/drawing packages on Unix are capable of producing PostScript. Sometimes special versions of some common utilities have to be used, e.g. GNU `groff` instead of the standard `nroff`.

X-Window graphics Most commonly window dumps produced with `xwd`. The result is a bitmap in X11 internal format. In order to be usable on non-Unix machines such bitmaps should be converted to more popular formats, such as GIF or PostScript. A typical window dump producing a GIF file could be initiated through `xwd | xwdtopnm | ppmtogif > /tmp/X.gif`.

FrameMaker FrameMaker is a word processor available for Unix workstations as well as Macintosh and PCs. It stores files in internal format which can be interchanged between Unix and the Macintosh and PC versions. Additionally, it can produce PostScript output.

Support: There is currently no CERN-wide support for FrameMaker!

---

5 From the Receiver's Point of View

An obvious problem for the receiver of a document is to recognise in which format the document has been sent/stored. Unfortunately this is not always easy, but the following general guidelines may help:

Filename extensions Often the file name of the document, mostly the *extension* (e.g. **.tex**, **.tar** in Unix, VMS, and PC, or the file type in VM) may hint at the method the document has been encoded with:

- an extension of **.tex** obviously means a TEX or LATEX document, only looking into the document itself can give further information on what TEX macros or LATEX document type and style files are required to process the document,
- a name ending in **.ps** is probably a PostScript file,
- the extension **.tar** probably means a *tape archive* file created under Unix; the Unix *tar* command is used to restore the individual files contained in the archive,
- the extensions **.Z** and **.zip** mean that the document has been compressed using the Unix `compress` command or the `zip` command. Often such an extension is appended to an existing extension such as .ps.Z which means a compressed PostScript document. A frequently encountered combination is **.tar.Z**, i.e. a compressed tape archive to be processed first by `uncompress` and then by `tar`.

uuencode Files containing non-ASCII characters may be transferred in uuencode format. This format can be recognised by the word 'begin' on the first line (some implementations of uudecode allow for varying amounts of text to precede the 'begin'), followed by a Unix mode specification consisting of 3 digits in the range 0-7, followed by a Unix file name. Example: `begin 644 docint.dvi`. An optional block beginning with a 'table' line may actually precede the 'begin' line, this block is part of the uuencode mechanism.

The content of the file can be reconstructed (bitwise) using the *uudecode* command under Unix. Whether the result then makes any sense on the receiving system depends on the application that created the file.

Section 6.9 contains more hints on how to guess the document type by looking at the first few lines of the document.

5.1 Receiver: Macintosh or PC

5.1.1 PostScript files

Macintosh The *sendps* or *LaserWriter* utility can be used to print PostScript files directly on Apple LaserWriters.

PC The `LPR` command (part of the TCP/IP command suite) can be used to print on printers served by the `springer` print server[16]. PCs connected to a Novell server may print on the server's PostScript queues.

[16] for registration: send mail to `printsp@dxcern.cern.ch`

5.1.2 Graphics, bitmaps

GIF and TIFF files can be displayed using applications such as *Photoshop* on the Macintosh or on a PC.

Additionally, PC programs `wingif` and `winjpeg` available on NICE can be used to display GIF/JPEG graphics files. These programs also support conversion between major bitmap formats. For Macintosh XlateGraf, JPEGView and QuickGIF are available in the public domain.

5.2 Receiver: VM

PostScript PostScript files can be printed using the `XPRINT` command on the following printers:

- PostScript printers managed by springer; use `xfind printers` on VM or `www` to obtain a list of possible printers,
- IBM 3812, the IBM PostScript interpreter is automatically run in the user's virtual machine if the file's file type is `PS`[17],
- Xerox 4050, by passing the PostScript file to the `PSXEROX` command.

In all cases files should start with the PostScript signature '`%!`' in the first two columns. It is generally safe to add a line containing just '`%!`' before the first character of the file.

5.3 Receiver: VMS

PostScript PostScript files can be printed on the following printers using the `XPRINT` command:

- PostScript printers managed by springer; use `www` to search the Computer Centre Documentation with the keyword *printer*; the variant `XPRINT/LW` should be used where applicable,
- IBM 3812 printers, the IBM PostScript interpreter is run in a VM batch job.

PostScript files can also be displayed on VMS workstations using the CDA PostScript interpreter.

5.4 Receiver: Unix

File formats considered:

1. PostScript
2. X11 bitmaps
3. GIF, TIFF and other bitmaps

PostScript PostScript files can be printed on PostScript printers served by `springer`. The printer has to be defined on the Unix system: refer to the appropriate CERN User's guide for your Unix system. PostScript files can also be displayed on the screen on Unix workstations with an X-windows display[18] using GNU's *ghostscript* program, e.g. the `ghostview` command available from ASIS.

X11 bitmaps Typically X11 screen/window dumps, produced by `xwd`. The utility to display such bitmaps on the screen is `xwud`. The `xpr` utility can be used to convert to PostScript or HP LaserJet format for printing and is even one of the few Unix utilities that can produce IBM3812 printer format.

GIF, TIFF and other Bitmapped Graphics The public domain *xv* program available on ASIS displays GIF/TIFF images on X11 screens. It can also be used to edit the picture's colours and to convert the file into different formats: in particular it converts GIF to/from TIFF, JPEG, X11 bitmap, PBM, and can produce PostScript files (with limited resolution).

Other programs for converting bitmap formats into each other exist in the public domain, e.g. the PBM (Portable Bit Map) suite of converters. They, too, are available on ASIS.

[17] careful! The resulting bitmap files are **big!**
[18] or X-terminals

6 Examples - by application

6.1 FTP Transfer to Macintosh

At CERN the standard TCP/IP software is NCSA Telnet, where the FTP protocol is handled within a Telnet session as FTP server. Therefore, in order to transfer a file to a Macintosh via FTP is usually initiated on the Mac using the following procedure is the

1. double click on the NCSA Telnet icon to start it,
2. in the *File* menu, select "FTP Enable" and "Set Transfer Directory" to specify the Mac folder in which the transferred file will be put,
3. open a connection and log in to the system from which the document is to be transferred,
4. on the system, invoke "FTP" followed by the IP number of your Mac; the latter can also be sent automatically by selecting "Send IP Number" in the CONNECTION menu,
5. use standard FTP subcommands to find transfer the document.

Support: Help for Macintosh problems is given by: Patrick Herzog, `patrick_herzog@macmail.cern.ch`

6.2 Word Document on Macintosh or PC

If the document is to be **revisable** and the receiver is a **PC or Macintosh**, send out the native Word file using *binary* transfer (see Section 3.1). Using QuickMail or MSmail, encoding/decoding is performed automatically when the internal Word file is transferred as an *enclosure* (refer to Section 3.3). Alternatively, save the document in RTF format which can be sent out as a standard text file.

If **printing quality** is **not required**, the document does not contain **graphics** and is for **reference only**, create a text version of the document (cf. Section 4.1) which can be sent by Email.

Otherwise create a PostScript file which has to be transferred binary (cf. Section 3.1) if the document contains non US-ASCII characters such as French accented characters.

Support: Questions about Microsoft Word should be addressed to the Desktop Publishing Centre, Michele Jouhet, or to Tony Shave, `tony_shave@macmail.cern.ch`.

6.3 SGML on VM

In order to transfer a document of **typeset quality**, the only practical way as of today is to create a PostScript file (by specifying the `BOOKIE` options NOPRINT and printer type LASERWRITER) and to transfer the file using a binary transfer method (cf. Section 3.1).

Alternatively, if **no typeset quality** is required and the document consists mainly of **English text**, use the `BOOKIE` command's `EMAIL` option to create an 80-column formatted text file (cf. Section 4.3) which does not require any precautions for transfer.

Support: BOOKIE on VM is supported by M.Goossens, `goossens@cernvm`.

6.4 TEX, LATEX

The commands mentioned in this section apply to Unix systems, VM and VMS at CERN. On VMS (VXCRNA), the `cerntex` command has to be executed to set up the required symbols. TEX implementations on Macintosh or PC require different handling.

A **revisable** document must be sent out in TEX or LATEX **source form**. The receiver must have TEX or LATEX installed as well as any non-standard LATEX style sheets, TEX macros or auxiliary files (e.g. a logo) that the document requires. Since the document consists of US-ASCII text, no specific transfer problems exist.

A document for **reference only** should be formatted using the `tex` or `latex` command and the resulting .dvi file converted to PostScript using `dvips`. LATEX documents printed on PostScript printers normally look better when typeset using PostScript fonts instead of standard TEX fonts: include

```
\documentstyle[times]{...}
```

at the beginning of your document[19].

Support: TEX and LATEX on the central systems/servers are supported by M.Goossens, `goossens@cernvm`.

6.4.1 Character-only version of LATEX document

This option produces acceptable results for documents consisting solely of English text, using a single font and no graphics, formulae or special page layout (cf. Section 2.2). Specify

```
\documentstyle[screen]{...}
```

in your LATEX document, then reformat the document using the `latex` and `dvi2tty` commands.

6.5 Scanner on Macintosh or PC

After scanning the picture, decide in which format to save it in: if the picture is to be integrated into a PostScript document (or e.g. a LATEX document to be formatted for a PostScript printer), it is easiest to save it in EPS (encapsulated PostScript) format (if your scanner software supports it).

If the picture is to be sent around mainly for on-screen viewing, or if you want to modify it, save it in GIF format. Such a file has to be transferred in *binary* (cf. Section 3.1). The GIF file is usually much smaller than the corresponding EPS file, and can be converted to PostScript later, e.g. with *Photoshop*.

Support: `patrick_herzog@macmail.cern.ch` may help in case of problems with Macintosh and PC attached scanners.

6.6 Encoding documents with uuencode

The `uuencode` command on Unix systems, PC and VM is used to convert arbitrary input data into a stream of characters out of the US-ASCII character set. On VM the utility is located on the Archive-Tools disk, available using the command `GIME ARCHIVE-TOOLS`. For Macintosh use the UULite program available in the public domain. On VMS no such utility exists on the central system (although there is a `uudecode` utility).

The `uuencode` command normally takes two arguments: the name of the file to be uuencoded (or standard input on Unix systems if only one argument is given) and the name under which the file should be restored on the destination system; if there is no reason to change the file name, the file's name will actually appear twice although this may look silly, e.g. (on Unix):

```
uuencode picture.gif picture.gif | mail rtb@cernvm.cern.ch
```

6.7 Combinations of `tar`, `compress` and `uuencode`

Sometimes documents are processed by several packaging/compressing/encoding methods in a row before the actual transfer takes place. A typical example is a set of documents sent together as one file, which may have been processed in the following manner:

1. the set of documents is processed by the `tar` command resulting in a single tape archive (**.tar**) file,
2. the result is compressed using `compress` producing a file with a **.Z** extension,
3. the compressed file is then processed by `uuencode` to be able to survive Email gateways intact.

Once the file has been sent out, the receiver has to reverse the operations in the correct order:

1. `uudecode` the file received,
2. `uncompress` the **.Z** file resulting from the previous step,
3. split the tape archive into individual files using `tar`.

[19] this uses 'Times Roman', 'Helvetica' and 'Courier' fonts; consult the 'TEX at CERN' manual for details

6.8 MIME compliant Multi-media Electronic Mail

The following shows a typical header of a MIME compliant multi-media electronic mail message. Such a message is recognisable by the `Mime-Version:` phrase in the header. The `Content-type:` phrases will indicate the different parts of the message. Multi-media messages may contain e.g. sound and pictures. To correctly handle such a message the receiver needs a MIME capable E-mail reading program such as *elm* or *pine* on Unix[20]. MIME-capable E-mail readers are currently not available for VM or VMS. If such a message is received on VM and VMS with `Content-type:` indicating types other than text, parts of the message will be unreadable. In this case the message should be forwarded to a MIME-capable system. Refer to the document 'Functional Requirements for Multimedia Electronic Mail', M.Dimou-Zacharova, Nov. 1992, for more information.

```
From: nsb@sunmed.cern.ch (Nathaniel Borenstein)
Message-Id: <9109191641.AA12840@greenbush.bellcore.com>
To: abel@sunmed.cern.ch, bianchi@sunmed.cern.ch, braun@sunmed.cern.ch,

   . . .

Subject: Star Trek Party!
Mime-Version: 1.0
Content-Type: multipart/mixed; boundary=Outermost_Trek
Sender: dimou@sunmed.cern.ch

   . . .
```

6.9 Recognising the Document Type from the first few Lines

The following samples documents whose types can easily be recognised by looking at the first few lines. Shown are the first few lines of US-ASCII text files only. Care should be taken when displaying documents that contain other than non US-ASCII characters: on Unix, a naive 'cat' may lock up your screen if parts of the document are misinterpreted as terminal control characters; rather use an appropriate editor. *Binary* documents are normally recognisable by their file name extensions (cf. Section 5).

PostScript

```
%!PS-Adobe-2.0
%%Creator: dvips 5.515 Copyright 1986, 1993 Radical Eye Software
%%Title: docint.dvi
%%CreationDate: Thu May 13 16:59:56 1993
%%Pages: 13
%%PageOrder: Ascend
%%BoundingBox: 0 0 596 842
%%DocumentFonts: Palatino-Bold Palatino-Roman Helvetica-Bold
%%+ Palatino-Italic Courier Courier-Bold

   . . .
```

uuencode

```
begin 644 docint.ps
M)2%04RU!9&]B92TR+C`*)25#<F5A=&]R.B!D=FEP<R`U+C4Q-2!#;W!Y<FEG
M:'0@,3DX-BP@,3DY,R!2861I8V%L($5Y92!3;V9T=V%R90HE)51I=&QE.B!D
M;V-I;G0N9'9I"B4E0W)E871I;VY$871E.B!4:'4@36%Y(#$S(#$V.C4Y.C4V
M(#$Y.3,*)25086=E<SH@,3,*)25086=E3W)D97(Z($%S8V5N9`HE)4)O=6YD

   . . .
```

[20] extensions to QuickMail (for Macintosh) and cc:Mail (for PC) have been announced

LaTeX

```
\documentstyle[palatino,a4wide,draftcopy,rtb]{article}
\title{Document Interchange Possibilities at CERN}
\author {Maria Dimou-Zacharova\\Mik Ferran\\
Silvano de Gennaro\\Per Hagen\\Rainer T\"obbicke}
```

. . .

RTF

```
{\rtf1\mac\deff2 {\fonttbl{\f0\fswiss Chicago;}{\f2\froman New York;}
{\f3\fswiss Geneva;}{\f4\fmodern Monaco;}{\f13\fnil Zapf Dingbats;}
{\f14\fnil Bookman;}{\f15\fnil N Helvetica Narrow;}
{\f16\fnil Palatino;}{\f18\fnil Zapf Chancery;}{\f20\froman Times;}
{\f2040\fnil Oxford;}{\f2046\fnil Swing;}
```

. . .

---

7 Appendix - Summary of User Support Contacts

The following addresses may be useful to sort out problems with document transfer: except for urgent matters, contact via electronic mail is to be preferred since less likely to lead to frustration arising from busy phone lines, or people repeatedly not in the office because of meetings or other matters. Also, in many cases electronic mailboxes are checked for urgent matters by backup personnel during the main contact's holidays.

BOOKIE on VM Michel Goossens, goossens@cernvm.cern.ch

DOCUMENT (VMS) Julian Bunn, julian@vxcern.cern.ch

electronic mail Maria Dimou-Zacharova, postmaster@cern.ch

graphics from engineering applications Mik Ferran, PMFDJ@CERNVM.cern.ch

Macintosh + PC applications Patrick Herzog, patrick_herzog@macmail.cern.ch

PostScript Michel Goossens, goossens@cernvm.cern.ch

TeX, LaTeX Michel Goossens, goossens@cernvm.cern.ch

Unix utilities Philippe Defert, defert@dxcern.cern.ch

Microsoft Word Tony Shave, tony_shave@macmail.cern.ch; questions can also be addressed to the person responsible for support in the Desktop Publishing Centre, Michele Jouhet

WWW (World Wide Web) Robert Cailliau, CAILLIAU@cernnext.cern.ch

other matters User Consultancy Office, UCO@cernvm, phone number 4952