

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

CERN – SL DIVISION

SL-Note-99-015 DI

**POST MORTEM OF
THE ELECTRONIC PUBLICATION OF
THE EPAC'98 PROCEEDINGS**

**L. Liljeby, Manne Siegbahn Laboratory, Stockholm, Sweden
J. Poole, CERN, Geneva, Switzerland**

Abstract

The proceedings of the Sixth European Particle Accelerator Conference were the second in the series to be published electronically. This report describes the preparations before the conference, the activities at the conference and the work afterwards which was required to produce the CD-ROM and HTML versions of the proceedings. The whole process was a great success, with the proceedings available on WWW and for the publisher in record time. There were some areas where procedures could be improved and these are described in this report and statistics concerning the various aspects of the editorial activities are included. Details of the procedures and programs are given in the annexes to the report.

Geneva, Switzerland

February 26, 1999

Contents

1	Introduction	1
2	Abstract Submission and Publication	1
3	Instructions for Authors and Website	1
4	Templates	1
5	Electronic submission	1
6	Resources for the Proceedings Office	2
6.1	At the Conference	2
7	Processing the Files	2
7.1	Activities at the Conference	2
7.2	Post Conference Activities	3
8	Statistics	4
8.1	Manpower	4
8.2	Computer Platforms	4
8.3	Software used by Authors	4
8.4	Failure Rates	4
8.5	Fault Analysis	5
9	Publication	5
9.1	Preparing the CD-ROM	5
9.2	Preparing the Website	5
10	Problems Encountered in 1998	5
11	Improvements for the Future	6
11.1	Coloured Stickers	6
11.2	Keyboards	6
11.3	Information to processing staff	6
11.4	Database	6
11.5	Postscript fonts for L ^A T _E X	6
11.6	Scheduling Paper Processing	6
11.7	Paper Submission	6
11.8	Conference Identifier in PDF Files	6
12	Acknowledgements	6
ANNEXES		8
1	Acrobat Parameters	8
1.1	Distiller	8
1.2	Exchange	8
1.3	Catalog	8
2	Cropping	8
3	HTML Author Index	8
4	HTML Table of Contents	10
5	Page Numbering	12
6	Visual Basic Routine to Modify PDF Files	14

7	Exporting Data from FileMaker on the Macintosh	16
8	Creating PDFMARKS	16

1 Introduction

The proceedings of the Sixth European Particle Accelerator Conference were the second in the series to be published electronically. This report describes the preparations before the conference, the activities at the conference and the work afterwards which was required to produce the CD-ROM and HTML versions of the proceedings. The paper volumes were produced from the electronic files.

Since the previous conference there has been a lot of progress in all areas concerning electronic publishing: there is new software, the authors are more experienced and new procedures have been developed. However, the total number of man-months required to produce the proceedings was roughly the same as for the previous EPAC conference allowing for there being $\sim 10\%$ fewer papers in 1998.

The whole process was a great success, with the proceedings available on WWW and for the publisher in record time. There were some areas where procedures could be improved and these are detailed in this report together with statistics concerning the various aspects of the editorial activities. The publication of the proceedings of EPAC'96 was described and analysed in detail in [1].

2 Abstract Submission and Publication

For the first time abstract submission used a Web form which was connected to a script which separated the replies into appropriate fields and sent E-mails to the conference secretary and the submitting author. This scheme worked very well and allowed a certain amount of checking of the information to be done automatically. The information was entered in the Filemaker database which was used at every stage of the conference organisation.

The form only allowed the use of ASCII characters so that greek letters, super- and sub-scripts had to be indicated separately.

The abstract brochure was produced in the same way as in 1996, constructing individual word documents by importing the abstract and author information and then the individual documents were strung together into one large document by a program.

In 1996 authors were asked to submit abstracts in postscript for the first time to train them in the new techniques which would be required later for the papers. This made a lot of extra work for the editorial team at the time but the methods used for EPAC'98 reduced the man-hours required for abstract processing by about a factor of five to ~ 1 man-month which is comparable to the traditional method.

The PDF¹ files for publication on the web were made using automated procedures. The table of contents and author index were produced from the information in the database, the author index requiring a separate program

(perl script - see Annex section 3) in which the unique identifiers for papers were concatenated against common author names and then the HTML file was written.

3 Instructions for Authors and Website

The EPAC website was an updated version of the Sitges pages and was therefore much faster to build. Changes in templates and experience gained at EPAC'96 and PAC'97 were reflected in the instructions to authors and in the help on electronic publication. The Website and templates were ready for authors at the beginning of April 1998.

Before the templates are made publicly available the network of support personnel in laboratories around the world were asked to test and review the templates and procedures. This revealed several small bugs and inconsistencies which were fixed before authors required the templates and information.

4 Templates

It was decided that templates would be provided for the two most recent versions of WORD for both Macintosh and PC and for $\text{\LaTeX} 2_{\epsilon}$. Following the idea proposed by Leif Liljebj the templates were prepared so that the text would be centred on a PDF page A4 paper wide by US letter paper high: this means that the PDF files will print with the text centred on both types of paper. Unfortunately this means that templates for both US letter paper and A4 paper are required, making a total of 10 templates which have to be prepared and tested.

Templates were installed on the EPAC'98 FTP server which was located at the Manne Siegbahn Laboratory and retrieval was possible by anonymous FTP or from the EPAC'98 Web pages.

5 Electronic submission

EPAC'98 encouraged submission of papers ahead of the conference by FTP. A web form was designed to collect the information normally carried on the submission sheet. The script behind the form checks that the fields have been filled in, sends the E-mails to the editors and a confirmation to the author via the web browser and in E-mail.

The FTP server was an NT system located at the Manne Siegbahn Laboratory. Several Gbytes of storage were made available and ~ 2.3 were used by the end of the processing (after compression). The server was available from the time that submissions opened (early April) until well after the conference. After the conference it was used for authors to make their re-submission and to make files available for processing at CERN.

The early submissions were processed at Stockholm and the author informed if there was a problem. The number of pages, the type of submission and the keywords were

¹Adobe Acrobat Portable Document Format

entered into the database. This was done manually but could be automated for the future. About 20% of the papers were submitted ahead of the conference by FTP.

Authors also had the possibility to submit the postscript file for distillation. A watched directory was set up (distiller option) and the postscript was distilled almost immediately so that authors could retrieve their PDF file and see the result. This process does not guarantee that the editors would accept the final postscript but does eliminate cases where there are obvious problems (like, it wont distill or the fonts are very bad).

All papers submitted by FTP received stickers at the conference and authors were instructed to bring a paper copy to the proceedings office.

6 Resources for the Proceedings Office

6.1 At the Conference

The factors which determine the resources required by the proceedings office are:

- the number of papers to be processed
- the need to process the postscript on an appropriate platform - Macintosh files should be distilled on a Macintosh and PC or UNIX generated postscript on a PC and therefore the relative split MAC/PC+UNIX is important.
- all files should be accessible on all platforms (some form of networked file system)
- adequate software installations

The hardware requirements for 1998 were based on EPAC'96 experience where 31% of the papers were prepared on Macintosh and therefore 6 PC's and 3 Mac's were requested for electronic processing. At the conference there were 6 PC's and 4 Mac's, plus the FTP server. The files were shared through Windows95/NT for the PC's and with Dave for the Macintosh's.

All of the machines were equipped with CD-ROM readers and floppy disk drives. Some PC's had zip drives and a networked printer was available for all machines. The computers were connected via an ISDN line to Manne Siegbahn Laboratory.

A very important addition to the conventional processing was the availability of Asian computers (PC and Macintosh) where Asian font problems could be solved. With the aid of our colleagues from KEK it was possible to resolve most of these problems at the conference.

The software inventory was as follows:

PC

Netscape
WORD97
Excel97
L^AT_EX 2_ε

Acrobat Reader, Distiller and Exchange
WinZip
emacs

Mac

WORD6
WORD98
Netscape
Acrobat Reader, Distiller and Exchange

In addition several of the PC's had **Pitstop**²: new software which allows one to edit PDF files. It is extremely powerful, allowing one to select parts of a document and change their attributes, scale them and so on. Typical applications for Pitstop were correcting fonts, resizing objects which flowed into margins and correcting text.

On the Mac's there was also **Tailor**³, a GUI postscript editor which was quite good at fixing the KaleidaGraph graphics problem - in many cases it was sufficient to open the offending postscript and save it again.

It had been suggested to the EPAC Organising Committee that an effort was made to develop a group of people who could perform the various tasks involved in the proceedings production with punctual support from students for each conference. This led to invitations to PAC97, PAC99, PAC2001, APAC98 and EPAC2000 asking them to nominate people to help in the proceedings office who would then stay on for a further week to help with the problem papers. We asked for sufficient people to man the computers i.e. 9 people but in fact more than this were available. There was strong support from the other accelerator conferences (PAC and APAC) and also the linac conference. Also a number of people volunteered to come and help in order to gain experience for their own conferences. In addition to some members of the EPAC editorial board⁴ there were about 12 staff available for document processing during the conference and eight of them were available for part or all of the week following the conference.

7 Processing the Files

7.1 Activities at the Conference

The main aim of the activity at the conference was author feedback. Once again the sticker boards were used but there were 2 this time: one close to the proceedings office and the other at the entrance to the main auditorium. One person worked full time on keeping the boards up-to-date.

The contributions are handed in at the proceedings office by the authors and some initial checks are made:

- checking completeness of the submission
- cross-checking the information in the database (title, co-authors)

²Enfocus Pitstop - <http://www.enfocus.com>

³Enfocus Tailor - <http://www.enfocus.com>

⁴L. Liljebj, C. Petit-Jean-Genaz, J. Poole

- inserting the keywords and number of pages in the database.

The submitted documents and accompanying sheets were placed in a transparent plastic folder so that the the submission sheet faced out at the back and the processing sheet was added at the front. The folders were split into two piles, for MAC and PC, and passed on for electronic processing. The processing sheet was used to record who had worked on the paper in the proceedings office, the acceptability of the contribution, record any problems and keep track of the status (processed OK, bad - see author, author seen and actions pending etc.).

Processing the files involved distilling the postscript and then verifying:

- the fonts
- the margins
- overall quality

In Stockholm an additional step was performed on the PDF files - they were cropped to A4 width by US Letter high (or smaller) with the text centred. A directory had been set up for each paper, containing title and authors as two file names, so that we could check the programme code against the title and authors as they appeared on the paper copy. The person doing the processing copied the postscript from the diskette into the directory and then 'dragged' it from the file explorer window and 'dropped' it on the distiller icon. This method had the advantage that the PDF was automatically written in the correct directory. Successfully distilled papers were opened in Acrobat Exchange, checked and cropped. In case of problems, the remaining files were retrieved from the diskette and copied to the directory for repairs.

Once the files had been processed the folders were passed back to the proceedings office where the database was updated to indicate the success or failure of the processing. The folders were then taken for the stickers to be placed and the boards and finally the folders were filed in the processing office.

7.2 Post Conference Activities

The post conference activities are principally fixing the problem files, careful checking, page numbering and making the indices, table of contents and other pages for the wrapper (photos, copyright, prize winners, conference organisation, list of participants, titles and separators etc.). In the problem cases, authors were contacted by E-mail and reminded that they had two weeks after the conference in which to resubmit.

The files for the CD-ROM are prepared first and then the paper copy is produced and finally the Web version is prepared. The Author Index and Table of Contents are produced with scripts (see the Annex, section 3).

As many of the procedures as possible have been automated and these all require input data from the database. **It is therefore of fundamental importance that the information in the database is complete and consistent.**

Once all of the files have been processed and deemed 'OK' they have to be scrutinised very carefully. The following steps are required:

- Open the file with Acrobat Reader or Exchange, check for error messages on all pages.
- Print the file and check that the margins and fonts are correct.
- Compare the printed version with the author's original.
- Cross check the title, number of pages and list of authors with the database.

Before page numbering can be done one has to be sure that there will be no further modifications to the files and entries for the separator pages (category titles) between sections of the book have to be inserted in the database. The latter are given program codes starting AAA and have the appropriate classification so that when the database is sorted in proceedings order they appear in the correct place. The numbering is done so that these separator pages always start on a right hand page and are followed by a blank; this of course means that in some cases an additional blank page has to be inserted ahead of the separators.

Once the page numbers have been attributed (perl script) they are imported to the database and the table of contents can be generated. From the file of programme code and page numbers the pages of the PDF files are numbered (see section 5 in the Annexes) by editing a postscript file made from the PDF and then re-distilling the processed postscript. This process is not 100% efficient and the numbered pages have to be printed and checked - sometimes the number does not fall in the correct place or is the wrong colour.

Once the pages have been numbered and the whole lot printed and checked again for consistency, the PDF files can be cropped, thumbnails generated, default parameters set for file opening and finally the hidden fields can be filled. This requires an input file generated from the database which contains the program code, title, author, category (in the subject field) and keywords. This file is put into EXCEL and a visual basic program uses the data to open the files in Acrobat exchange and make the modifications (see the Appendix, section 6). After making the modifications the files are saved using the 'Save As' option with the optimise option checked.

For the PDF files one needs a file with a field containing all of the authors (or submitting author, et al. if the list is more than 256 characters long); for the authors index one needs a field containing all of the papers for which the person was an author. This puts further constraints on the data in the database, like the name of the author must be written in the same way for each of the papers - something which is often not the case (there was at least one author who spelled his name incorrectly on the paper in EPAC'98).

In 1996 one of the most arduous tasks was inserting all of the links in the authors index for the CD-ROM. Using Acrobat Exchange, each link had to be drawn as a box

around the text and then the destination typed in. It took nearly one man-week to do both the table of contents and the author index and was very tiring. For 1998 Leif Liljebj created a program which used pdfmarks to automatically create the links with the correct names and at approximately the correct place on the page. All that had to be done was to move the active with Acrobat Exchange so that it covered the correct text on the page i.e. the corresponding page number. This new program reduced the manpower required by almost an order of magnitude. For details of how the program works see the Annex, section 8.

The post conference work was divided between teams working in Stockholm and CERN. The basic repairs to the papers and checking were done at Manne Siegbahn Laboratory and the database work at CERN. Once all of the files were deemed to be ready and the numbering done, two people came from Stockholm to complete the work at CERN. One week was set aside for this and most of the work was completed during this week - the paper printed, the author's index prepared and the table of contents made.

8 Statistics

8.1 Manpower

Substantial reductions in the manpower requirements were achieved in the pre-conference period in 1998 with respect to 1996 as Table 1 below shows.

Table 1: Manpower Resources for EPAC Proceedings in Man-weeks

	1996	1998
R & D	12	2
Planning	8	2
Build/maintain WWW pages	12	4
Author documentation	4	1
FTP services	4	2
Abstract Processing	24	5
Author education	8	0
Conference infrastructure	4	2
Processing at the Conference	8	10
Post Conference Processing	24	20
	<u>108</u>	<u>48</u>

8.2 Computer Platforms

There was a shift in the distribution away from Macintosh (31% in '96) and UNIX (18% in '96) towards PC in 1998. This trend was also present in PAC'97 compared to PAC'95.

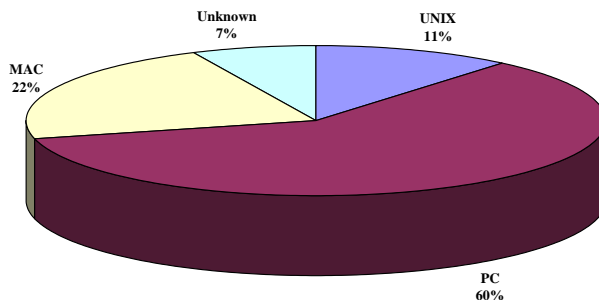


Figure 1: Computer platforms used for the preparation of EPAC'98 papers

8.3 Software used by Authors

The distribution of software packages used by authors remains dominated by Microsoft Word and \LaTeX . In 1998 the percentage of \LaTeX users remained essentially the same as in 1996 and WORD usage increased slightly as authors moved away from the more obscure word processing packages. WORD dominates \LaTeX by two to one and FrameMaker was the only other package which was used in 1998, as Figure 2 shows.

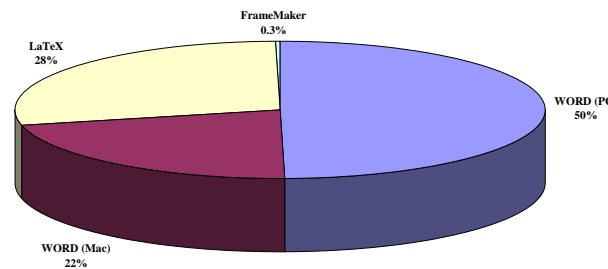


Figure 2: Software used for the preparation of EPAC'98 papers

8.4 Failure Rates

71% of the postscript files submitted were processed without problems. In total only four figures were scanned - no paper was scanned. Many equations had to be re-written because of font problems and often this was because people had used a plug-in (MathType) in WORD which uses special fonts. If the WORD source is opened on a system which does not have the plug-in, the only option is to re-type the equation. In one case a paper had many formulae which contained bad fonts and the whole thing was re-typed.

8.5 Fault Analysis

Fault	Percentage
Format problems, source edited	10.0
Bad or no postscript file	5.1
Fonts substituted	3.9
Figure problems	3.5
Unreadable disks or missing files	2.2
Too long, shortened by author	1.3
Equation problems	1.2
Changes requested by author	1.2
Cursor, page numbers or empty pages	0.6

This analysis is based on the comments which were written on the processing sheets and are therefore probably slightly optimistic - in some cases where it was trivial to fix a problem it was not recorded. The figures therefore represent the papers which took up the time at and after the conference.

Some figures were very slow to display in PDF and these were fixed by saving them as bitmaps using Adobe Illustrator. This technique was a new development for Stockholm and replaced scanning as a solution for this type of problem.

9 Publication

9.1 Preparing the CD-ROM

The files on the CD-ROM include all of the papers, the table of contents, the author index, the Acrobat index and the various other parts of the wrapper. The Acrobat index is generated by Acrobat Catalog and in the internal files it records filenames in upper case. In the past this would have caused problems on case sensitive platforms, like UNIX, where references will only work if the case matches. As a consequence all directory names and filenames on the CD were forced to upper case. Since publication it has been discovered by Yong Ho Chin of KEK (APAC98 proceedings) that the latest versions of the Acrobat Reader and Search engines (version 3.02 and later) are no longer case sensitive.

The structure of the CD-ROM was cleaned up a little for 1998, with only three files (Table of Contents, Author Index and Acrobat Index) in the root directory together with directories for Acrobat software, the wrapper, Acrobat index, the papers themselves and the whole of EPAC'96.

Information has to be gathered from a number of sources for the wrapper - the ISBN numbers from the publisher, the preface from the chairman of the local organising committee and the photographs from wherever. For the CD-ROM a suitably artistic display has to be created and the files made into PDF.

The cover graphics and instruction booklet have also to be prepared. For 1998 the previous booklet was updated with the new structure, references to the Joint Accelerator Conference Website (JACoW), new examples for boolean searching and updated documentation for the installation

instructions.

9.2 Preparing the Website

Preparing the Website was a relatively simple task after the rest of the work. The authors index and a full table of contents are written by scripts as was done for the abstracts (see the Annex, section 3 and 4. The table of contents then has to be split into its constituent parts but this was achieved with relatively little editing.

10 Problems Encountered in 1998

The FTP server had to be moved from Manne Siegbahn Laboratory to the conference centre. When it was installed at the centre it took much longer than expected to set it up again. Although the time at which this was done was well after the deadline for submission by FTP there were several complaints from authors. These complaints were received with little sympathy since we had clearly stated the deadlines. Those who complained that they could not get hold of templates got even less sympathy since it was only two days before the conference.

Some strange effects were observed when re-working papers, most noticeably that colour was lost from figures sometimes. Installing colour printers and then printing to file did not always solve the problem and the reason remains a mystery - perhaps it was related to the postscript level in the eps file.

Some authors who used WORD had paragraphs indented with TABs which are not necessarily defined in the same way on different platforms and this led to problems when reworking from the original.

The Malyshev postscript computer modern fonts were not available on the L^AT_EX installation in Stockholm. In order to circumvent this it was necessary to transfer the files to CERN and rework them over the internet.

It was found that Acrobat Exchange under Windows 95 will change the case of filenames when using 'Save As' with the same name as it proposes. In order to maintain the case it is necessary to type the name in full again. Also when creating links by hand, Exchange will make the file extension in lower case by default, even if the file already exists and has 'PDF' as the extension. These problems were fixed by editing the PDF files which contained the links on UNIX and making global replacements - in general this worked without problem. There can be difficulties if the file is edited and the size of the file changes - when it is opened with the reader there is an error message which says that the file is damaged. This can be repaired by Exchange and it is only necessary to open the file and save it again.

During the week when the team reassembled at CERN there were quite a number of papers (~50) which required further work because there were still font problems with the PDF files. It is believed that this probably went un-

detected earlier because non Base 14 fonts⁵ were installed on the computers/printers in Stockholm but not on those at CERN.

11 Improvements for the Future

11.1 Coloured Stickers

It was pointed out after the conference that the colours used (red and green) for the stickers on the author feedback boards are the two colours that are most commonly indistinguishable to colour-blind people. It is therefore proposed to use shaped stickers (e.g. stars and circles) to avoid this problem in the future.

11.2 Keyboards

The keyboards on the computers in the proceedings office were Swedish - they had a different layout and more characters than a standard US or UK type. This created some problems when one was trying to find 'standard' special characters and made life a little less easy for everyone. The software installation was more complicated and trying to use the keyboards as if they were standard keyboards was not easy.

11.3 Information to processing staff

The volunteers were not given clear enough instructions about what they should do. The mechanics of the process are fairly clear but it takes a little experience to know what is acceptable and what is not. The instructions should contain clearer guidelines and it would be better if a short training course (one or two hours) could be arranged. The volunteers could then be presented with a selection of problem cases and their solutions.

11.4 Database

Some time was lost in the CD-ROM preparation and some errors were introduced because the data in FileMaker was not coherent. This underlines the important role of the database and the requirement for rigorous controls.

11.5 Postscript fonts for L^AT_EX

The LaTeX installation for paper repairs should have postscript versions of the computer modern fonts installed. It should also be possible to take a L^AT_EX document which has been prepared without Times fonts and use the Malyshév fonts when making the postscript (thus taking a *dvi* file and using the *-Pbkm* switch on *dvips* will produce an acceptable postscript file.

11.6 Scheduling Paper Processing

There are more than enough helpers available for the paper processing now that more conferences are using Acrobat

software. At times there were too many people in the office in Stockholm and at other times there were not enough. It would be more efficient if there was a rota established, establishing continuous and uniform coverage.

11.7 Paper Submission

In an effort to avoid having several versions of names and addresses for the same institute it is proposed that authors are asked to select the information from a list of values on the web form. The information from the submission form should also be loaded automatically into a database.

11.8 Conference Identifier in PDF Files

When a paper is retrieved/printed from JACoW⁶ there is no way to tell at which conference it was presented, apart from the rather obscure URL. PDF files of papers presented at future conferences should carry the conference name on each page. This will be done at the same time as the insertion of page numbers and the system was tested for the Chamonix Performance Workshop in 1999.

12 Acknowledgements

The EPAC editorial board are greatly indebted to the following group of people who worked very hard on the electronic processing of the papers in Stockholm.

Sarah Amandusson (MSL)
Yong Ho Chin (KEK)
Cathy Eyberger (ANL)
Lisa Freyhult (MSL)
Andreas Jansson (MSL)
Romain Lhéritier (CERN)
Leif Liljeby (MSL)
Peter Lucas (FNAL)
Winfried Mitaroff (EPAC 2000)
Kazuyuki Nigorikawa (KEK)
John Smith (BNL)

The processing staff could not have completed their task without the support of the other volunteers in the other offices:

M. Campbell (BNL)
P. Carlé (MSI)
L. Freyhult
J. Karlson-Forestier (CERN)
M. Nadalin
P. Possanza (INFN)

References

- [1] J. Poole, "Post Mortem of the Electronic Publication of the EPAC'96 Proceedings", CERN-SL-Note 96-68 DI, November 1996.

⁵Courier, Helvetica, Times, Symbol and ZapfDingbats; the first three in roman, bold, oblique and bold-oblique making the 14 in total. These fonts are usually installed on printers and are the most commonly available fonts and they do not need to be embedded in postscript or PDF files

⁶Joint Accelerator Conference Website at <http://www.cern.ch/accelconf/>

ANNEXES

1 Acrobat Parameters

1.1 Distiller

It was noticed in Stockholm that one has to set each preference tab and click on 'OK' before moving to another tab.

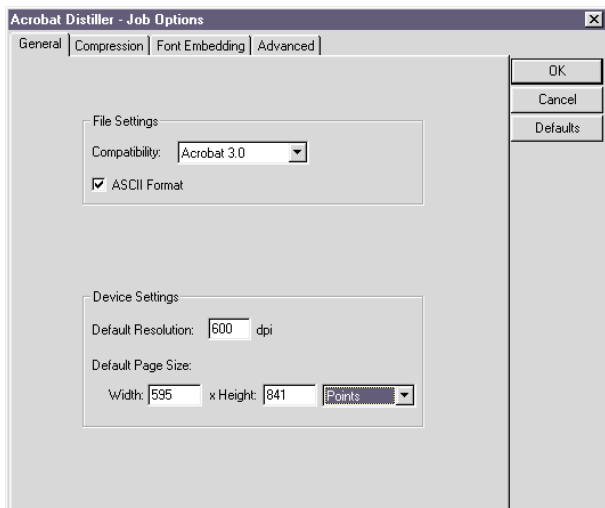


Figure 3: Job Options - General

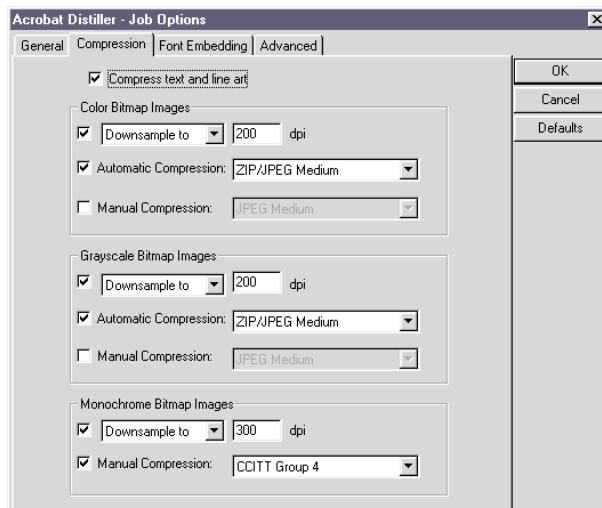


Figure 4: Job Options - Compression

The default values were left for the Font Embedding and Advanced Job Options and for the Preferences.

1.2 Exchange

Always save using the 'Save As' option with the 'Optimise' option checked.

1.3 Catalog

Select the 'Optimise for CD-ROM' option.

2 Cropping

The following table gives the cropping necessary to accommodate the PDF page layout resulting from the combinations of word processor and printer page setups. The standard text should fall within a rectangle 241mm high by 170mm wide, so for cases where the text was within these limits but the margins did not fit any of the expected values, the pages were cropped until the text was centred. Care was necessary to see that the cropping required was the same for all pages.

Table 2: Cropping dimensions for the expected PDF page layouts

PDF Format	Crop
A4 on A4	18mm top
US letter on US letter	6mm right
US letter on A4	18mm bottom
A4 on US letter	55mm top

3 HTML Author Index

There were two versions of this script, the first which is listed below was for the author's index which was made at the abstract stage and the second one for the final version.

```

#!/usr/local/bin/perl -w
# Script to generate data for Author Index for EPAC98
# J. Poole, 19th March 1998

# Setup some variables

$refno = 0;
$startingnew = 0;
$countr = 0;
$lastname = "";
$lastinit = "";
$newname = "";
$navigline = "<BR><BR><A HREF=\"#A\"> A </A><A HREF=\"#B\"> B </A><A HREF=\"#C\"> C
</A><A HREF=\"#D\"> D </A><A HREF=\"#E\"> E </A><A HREF=\"#F\"> F </A><A
HREF=\"#G\"> G </A><A HREF=\"#H\"> H </A><A HREF=\"#I\"> I </A><A HREF=\"#J\"> J
</A><A HREF=\"#K\"> K </A><A HREF=\"#L\"> L </A><A HREF=\"#M\"> M </A><A
HREF=\"#N\"> N </A><A HREF=\"#O\"> O </A><A HREF=\"#P\"> P </A><A HREF=\"#Q\"> Q
</A><A HREF=\"#R\"> R </A><A HREF=\"#S\"> S </A><A HREF=\"#T\"> T </A><A
HREF=\"#U\"> U </A><A HREF=\"#V\"> V </A><A HREF=\"#W\"> W </A><A HREF=\"#X\"> X
</A><A HREF=\"#Y\"> Y </A><A HREF=\"#Z\"> Z </A><BR>";

# Input file
$coauthr_file = "/home/poole/epac98/authmon.txt";

# Output file
$htmlfile = "/home/poole/www/epac98authr.html";
#
# Open the file with the coauthors in it
# - which was created with by Christine
#
open (COAUTHRS,$coauthr_file) || mydie("Cannot open file $coauthr_file: $!");
open (WWW,">$htmlfile") || mydie("Cannot open file $htmlfile: $!");

print WWW "<HTML>\n<HEAD>\n<TITLE>EPAC'98 Author Index</TITLE>\n";
print WWW "<BODY BGCOLOR=#FFFFFF>\n";
print WWW "<H1>Author Index </H1>\n";
print WWW $navigline;
print WWW "<H2><A NAME=A>A</A></H2>";
#
# Index the page numbers
#
while (<COAUTHRS>) {
    chomp($_);
    ($newname,$newinit,$newfile) = split(/\t/, $_,3);    ## split into Name, Initials
    if ($countr == 0) {                                    ## and filename
        $lastname = $newname;
        $lastinit = $newinit;
        $lastletter = "A";                                ## keep track for navigation bar
    }
    elsif ($lastname eq $newname && $lastinit eq $newinit) {
        ## need to add to list of references
        @pdfs = (@pdfs, $newfile);
    }
    else {
        ## print
        $prname = $lastname . ", " . $lastinit;           ## Name, Initials, ...
        $line = "<BR><B>$prname</B>";
        foreach $fref (@pdfs) {

```

```

        $line = $line . " , <A HREF=ABSTRACTS/$fref\".\".pdf>$fref</A>";
    }
    $thisinit = substr($prname,0,1);
    if ($thisinit ne $lastletter){
        print WWW "<BR><BR>$navigline\n";
        print WWW "<H2><A NAME=$thisinit>$thisinit</A></H2>";
        $lastletter = $thisinit;
    }
    print WWW "$line\n";
    $lastname = $newname;
    $lastinit = $newinit;
    @pdfs = ($newfile);
}
$countr++;
}
##
## Do the last one
##
    $line = "<BR><B>$lastname</B>";
    foreach $fref (@pdfs) {
        $line = $line . " , <A HREF=$fref>$fref</A>";
    }
    print WWW "$line\n";

print WWW "</BODY>\n</HTML>\n";
close WWW;

exit 0;
#-----
sub mydie {
    my $Msg=shift;
    print "Error executing the script\n";
    die "$Msg\n";
}
#-----
sub by_surname {
    return (($nameorder{$a} cmp $nameorder{$b}) || ($a cmp $b));
}
#-----

```

4 HTML Table of Contents

The logic for the table of contents is simpler because it contains one record per paper. The input data already has the special characters in HTML format.

```

#!/usr/local/bin/perl -w
# Script by J. Poole to generate co_author lists
# Modified from abstract Author Index WWW generator
# 30th July 1998

# Setup some variables
$countr = 0;
$newcode = "";
$lastcode = "";
$sauthbits = "";
# Input file
$sauthr_file = "Authors.html";

```

```

$toc_file = "tocwww.data";
# Output file
$html_file = "contents98.html";
#
# Open the file with the authors in it
# Was created with by Christine
#
open (AUTHRS,"$authr_file") || mydie("Cannot open file $authr_file: $!");
while (<AUTHRS>) {
    chomp($_);
    s/\r//g;
    $name = "";
    $newcode = "";

    ($newcode, $totname) = split(/\t/, $_,2);
    ($name2,$name1) = split(/, /,$totname,2);
    $name = $name1 . "&nbsp;" . $name2 ;

    if ($countr == 0) {
        $authbits = $name;
        $lastcode = $newcode;
    }
    elsif ($lastcode eq $newcode) {
        chomp($name);
        $authbits = "$authbits, $name";           ## need to add to list
    }
    else {                                       ## print
        $authors{$lastcode} = $authbits;
        $authbits = $name;
        $lastcode = $newcode;
    }
    $countr++;
}

## Do the last one
    $authors{$lastcode} = $authbits;
close AUTHRS;
##
## Fix the missing authors
##
$authors{"MOZ02A"} = "D.L.&nbsp;Burke";
$authors{"WEI02B"} = "M.&nbsp;Leenen";
$authors{"FRY01A"} = "A.&nbsp;B&aacute;r&aacute;ny";
$authors{"WEI04B"} = "B.R.&nbsp;Nielsen";
$authors{"FRY02A"} = "C.&nbsp;Rubbia";
$authors{"MOX01A"} = "B.&nbsp;Svedberg";
$authors{"TUOC01A"} = "V.N.&nbsp;Litvinenko, I.V.&nbsp;Pinayev,
    S.H.&nbsp;Park, Y.&nbsp;Wu";
$authors{"THOA03A"} = "R.&nbsp;Ryne et al.";
$authors{"THP03C"} = "S.I.&nbsp;Kukarnikov, V.K.&nbsp;Makoveev,
    A.Yu.&nbsp;Molodozhentsev";

open (PAGENUMS, "/user/poole/tex/epac98/pagenos/numbered.data") ||
    mydie("Cannot open page numbers file: $!");

#
# Index the page numbers

```

```

#
while (<PAGENUMS>) {
    chomp($_);
    ($ppcode,$ppnum) = split(/\t/, $_,2);
    $pagendx{$ppcode} = $ppnum;
}

$rowcoa = 1;
open (WWWDAT,"$toc_file") || mydie("Cannot open file $toc_file: $!");
## Print the header
open (TOC,">$html_file") || mydie("Cannot open file $html_file: $!");
print TOC "<HTML>\n<HEAD>\n<TITLE>EPAC'98 Table of Contents</TITLE>\n</HEAD>\n";
print TOC "<BODY>";
while (<WWWDAT>) {
    s/\r//g;
    chomp($_);
    ($pcode,$acode,$title) = split(/\t/, $_,3);
    if (substr($pcode,0,3) eq "ZZZ") {
        $pcode = substr($pcode,3);
        #
        # Got a NOT RECEIVED
        #
        print TOC "<BR><B> $title </B> <BR>\n";
        print TOC "$authors{$pcode}\n";
        if (substr($acode,0,3) ne "XXX") {
            print TOC " (<A HREF=\"ABSTRACTS/$acode.pdf\">Abstract</A>)\n";
        }
        print TOC "<BR><B>Contribution not received</B><BR>\n";
    }
    else {
        print TOC "<BR><B> <A HREF=\"PAPERS/$pcode.PDF\"> $title </A></B><BR>\n";
        print TOC "$authors{$pcode}\n";
        print TOC " (<A HREF=\"ABSTRACTS/$acode.pdf\">Abstract</A>)\n";
        print TOC "<DIV ALIGN=RIGHT> <B>$pagendx{$pcode}</B></DIV>\n";
    }
}
## Finish off the HTML
print TOC "</BODY>\n</HTML>\n";
exit 0;
#-----
sub mydie {
    my $Msg=shift;
    print "Error executing the script\n";
    die "$Msg\n";
}

```

5 Page Numbering

The following is the script used to insert the page numbers in the postscript files. It is specific to the version of Acrobat 3.0 running on a UNIX station.

```

#!/usr/local/bin/perl

# Script by J. Poole for putting page numbers into PostScript
# FEB 1998
#
# Set up to run from /user/poole/chamonix98/pagenos

```

```

#
# Called with file name followed by first page number
#
# e.g. add_page.pl /voll/epac/pnostest/johnps/WEP017G.ps 41
#
# Modified July 1998, for EPAC98
#     uses US letter paper, no need for edits
#     place page number in different place
#
#     Pitstop can move the origin - checks BoundingBox
#
#     Color set to black before writing page number (August 98)
#

MAIN:
{
    $nxtfile = $ARGV[0];
    $pzero = $ARGV[1];
    @fbits = split(/\/\/,$nxtfile);
    $nextfile = $fbits[$fbits-1];
    print "Starting $nextfile\n";
    $IFILE = "psin/" . $nxtfile;    # Input PS Filename
    $OFILE = "psout/" . $nextfile;  # Processed PS Filename

    open(PSFILE,$IFILE) || die "cannot open $IFILE\n";    # Open the input file
    open(PAGED,">$OFILE") || die "cannot open $OFILE\n"; # Open the output file

    $abort = 0;
    $imem = 0;
    $gotone = 0;
    $yzero = 25 ;                # Standard offset
    $yoff = $yzero;

    while (<PSFILE>)            # Step through the file
    {
        if (/BoundingBox/) {    # Look at the values
            $bbline = $_;        # to check for offsets
            ($bbtext,$bbox) = split(/: /,$bbline,2);
            ($llx,$lly,$urx,$ury) = split(/ /,$bbox,4);
            if ($lly ne "" && $lly > 0) {
                $yoff = $yoff + $lly ;
            }
        }
        if ($imem < 4) {
            $imem++;
        }
        $psline[$imem] = $_;
    }
    else {
        $psline[0] = $psline[1];    # My own circular buffer
        $psline[1] = $psline[2];
        $psline[2] = $psline[3];
        $psline[3] = $psline[4];
        $psline[4] = $_;
        if ($_ =~ "%PageTrailer") {

```



```

        $gotone = 1;
    }
    if ($gotone == 1) {
        if ($psline[0] =~ /^PDFVars/) {
            print PAGED "0 0 0 rg\n";
            print PAGED "/Times-Roman findfont 1 scalefont setfont\n";
# A4      print PAGED "0 Tc 10 0 0 10 293 25 Tm [(" . $pzero . ")]TJ\n";
            print PAGED "0 Tc 10 0 0 10 293 $yoff Tm [(" . $pzero . ")]TJ\n";
                #
                # Set the colour black (August 98)
                # Set the font for the page number
                # Stick in the page number
                # 0 Tc sets the char spacing
                # Tm is translate/moveto
                # [ ] TJ prints it.
                # $yoff takes care of offset (PITSTOP)
                #
                $pzero++ ;
                $gotone = 0;
            }
        }
        print PAGED "$psline[0]";
    }
}
print PAGED "$psline[1]";
print PAGED "$psline[2]";
print PAGED "$psline[3]";
print PAGED "$psline[4]";
close PAGED;
close PSFILE;
$watched = "/nfs/srv2-home/div_sl/drop_box/Poole/in/" . $nextfile;

    `cp $OFFILE $watched`;                # Copy file to watched directory
                                           # for distilling

    `rm /tmp/AcroAA*`;                    # Clean up after Acrobat Reader
exit;
}

```

6 Visual Basic Routine to Modify PDF Files

The EXCEL workbook contains the data necessary to drive the visual basic routine. It works by having Acrobat Exchange open on the desktop and clicking a start button on the EXCEL worksheet which activates the visual basic programme. The following is a sample of the input data (lines have been wrapped for convenience).

TUX01B	A 08 Linear Accelerators	A Review of Linac and Beam Transport Systems for Transmutation	linac, proton, rf-structure, superconducting - rf	Lagniel, J.M.
TUX03B	A 09 Pulsed Power Accelerators	A Review of Spallation Neutron Source Accelerators	storage ring, injection, H minus	Gardner, I.S.K.
THZ03A	A 11 Advanced Concepts	Accelerator Schemes based on Lasers and Plasmas	laser, plasma, acceleration, beat-wave, wakefield	Mine, P.
THZ04A	A 11 Advanced Concepts	Muon Collider Progress	acceleration, lepton	Noble, R.
WEI01B	A 12 Accelerators and Storage Rings, Other	Technology and Commercial Supply of Components for the LHC Project	collider, superconducting magnets, hadron, cryogenics	Faugeras, P.E.
MOX03A	A 12 Accelerators and Storage Rings, Other	HERA Upgrade Plans	collider, luminosity, polarization, proton, electron	Gianfelice-Wendt, E.
WEY03B	A 12 Accelerators and Storage Rings, Other	Present Status of KEKB Project	collider, factory	Kurokawa, S.I.

The Visual Basic Script is as follows:

```

Sub TransferOne()
Start:
    FilePath$ = "H:\UNIX_BOX\EPAC98\To_do_again\"           'Set path for PDF files
Restart:
    If ActiveCell.Column <> 1 Then GoTo Screwup           'If no column A then Exit Sub
    If ActiveCell.Value = "END" Then GoTo EndSub         'If cell contains "END" then Exit Sub
    FileName$ = ActiveCell.Value                         'Get File name
    ActiveCell.Offset(0, 1).Range("A1").Select          '1 right
    Subjct$ = ActiveCell.Value                          'Get subject (session name)
    ActiveCell.Offset(0, 1).Range("A1").Select          '1 right
    Title$ = ActiveCell.Value                          'Get Title
    ActiveCell.Offset(0, 1).Range("A1").Select          '1 right
    Keywords$ = ActiveCell.Value                       'Get Keywords
    ActiveCell.Offset(0, 1).Range("A1").Select          '1 right
    Author$ = ActiveCell.Value                         'Get Authors
    ActiveCell.Offset(0, -4).Range("A1").Select         'Back column A
On Error GoTo Abort
    CheckFile = Dir(FilePath$ & FileName$ & ".pdf")     'Check if file exists
    If CheckFile = "" Then GoTo NoFile                  'If not then Message + exit
    AppActivate "Acrobat Exchange", True               'Activate Acrobat Exchange
    SendKeys "% R", True                               'Restore Exchange
    SendKeys "^o", True                               'Ctrl-o to Open a file
    SendKeys FilePath$ & FileName$ & ".pdf{ENTER}", True 'Open FileName
    SendKeys "{ENTER}", True                          'Clear any error message
    SendKeys "%D", True                               'Document Menu
    SendKeys "C", True                                'Cropping
    SendKeys "3", True                                'Take 3mm from left
    SendKeys "{TAB 2}", True                          'Skip crop top
    SendKeys "3", True                                'Take 3mm from right
    SendKeys "{TAB 2}", True                          'Skip bottom
    For J = 1 To 6000000                               'Loop for a while
    Next J                                             'End of loop
    SendKeys "A", True                                'Crop All pages
    SendKeys "{ENTER}", True                          'Close menu
    SendKeys "{ENTER}", True                          'Confirm it
    SendKeys "%D", True                               'Document Menu again
    SendKeys "T", True                                'Generate Thumbnails
    SendKeys "^d", True                               'Document Information
    SendKeys "t", True                                'Go to Title field
    SendKeys Title$, True                             'Type Title
    SendKeys "{TAB}", True                            'Go to next field
    SendKeys Subjct$, True                            'Type subject
    SendKeys "{TAB}", True                            'Go to next field
    SendKeys Author$, True                            'Type Author

```

```

SendKeys "{TAB}", True           'Go to next field
SendKeys Keywords$, True        'Type Keywords
SendKeys "{ENTER}", True        'Press Enter
SendKeys "%F", True             'Open the file menu
SendKeys "i", True              'Select Document information
SendKeys "o", True              'Select Open settings
SendKeys "t", True              'Set for Thumbnails and page
SendKeys "{TAB 2}", True        'Go to magnification settings
SendKeys "{UP}", True           'Choose fit visible
SendKeys "{ENTER}", True        'Press Enter
SendKeys "+^S", True            'SaveAs
SendKeys "{ENTER}", True        'Save file
SendKeys "y", True              'Save file with same name
For J = 1 To 3000000             'Loop for a while
Next J                           'End of loop
SendKeys "^w", True             'Close file
SendKeys "^w", True             'Close file
SendKeys "{ENTER}", True        'Dont know why ...
SendKeys "% ", True            'Get the Window menu
SendKeys "N"                    'Minimise Exchange window
AppActivate "Microsoft Excel - pdf98.xls" 'Back to Excel
Selection.Font.ColorIndex = 4    'Change text colour to Green
ActiveCell.Offset(1, 0).Range("A1").Select 'Down 1
GoTo ReStart                     'Do it again

NoFile:
ActiveCell.Offset(0, 0).Range("A1").Select 'Up 0 mod 28/1/97 from (-1,0)
Selection.Font.ColorIndex = 3      'Change text colour to Red
Dummy = MsgBox(FileName$ & ".pdf does not exist!", 0, "Error")
Exit Sub

Screwup:
Dummy = MsgBox("Wrong Column", 0, "Error")
Exit Sub

Abort:
Dummy = MsgBox("Aborting Task", 0, "Error")
Exit Sub

EndSub:
End Sub

Sub ResetColour()
' Selection.Font.ColorIndex = 4      'Reset selected cells col to Auto
Selection.Font.ColorIndex = xlAutomatic 'Reset selected cells col to Auto
End Sub

```

7 Exporting Data from FileMaker on the Macintosh

Files exported from the database contain Macintosh special characters (accents, greek letters etc.) which do not translate on other platforms or work in HTML. It is necessary to make more than one version of the file containing author names and programme codes which is used to make the author index (PDF and HTML) and the input file for the EXCEL program. For the HTML version, all of the special characters are translated to their HTML equivalent using global edits. For the version which is used to make the PDF authors index, the raw data is processed and re-imported on the Macintosh. For EXCEL all special characters have to be removed since only ASCII characters are put in the hidden fields of the PDF files.

The table of contents is only exported from the database for the HTML version and therefore the special characters are translated into HTML on the Macintosh. The PDF version of the table of contents is generated from a print-merge into WORD on the Macintosh and then into postscript for distilling.

8 Creating PDFMARKS

PDFMARKS are a construct inside PDF files which create links and govern their behaviour; they can be inserted in the postscript file before distillation. From an export of the WORD file as 'Text only with line breaks' it is possible to compute the approximate location on the page of the page numbers which will become the active text. Care is required to make

sure that the line spacing is set at an exact amount in the WORD document and then using the text file from WORD and a text file from the database giving the relationship between page number and programme code (filename) the pdfmarks are inserted. The Pascal program reads the text file from WORD and looks for page numbers; it then calculates the approximate position on the page for each of the page numbers and generates an appropriate PDFMARK in the postscript file. For further information contact the author of the program - LILJEBY@msi.se.